Pause or No Pause?--Prosodic Phrase Boundaries Revisited

Chiu-yu Tseng and Chun-Hsiang Chang

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

Abstract

In this study, we present evidences from analyzing acoustic parameters of fluent continuous speech to show that within-paragraph prosodic phrase boundaries are related more to contrasts of neighborhood prosodic states rather than between-phrase pause durations; prosodic states receive more constraints from higher level discourse information. Revising a modular acoustic model by Tseng's Hierarchical Prosodic Phrase Grouping (HPG) framework [2, 3] and examining the much varied Prosodic Phrase (PPh) boundary B3 within speech paragraph, statistical accounts of layered contributions reveal distinct contrasts between boundary immediate duration and intensity patterns irrespective of pause duration. Contrasts of F0 contour patterns were also observed in these locations. Evidences obtained also illustrate how PPh boundary states are specified more by higher level discourse information than by lower level prosodic word construction. These results combined suggest that contrastive neighboring prosodic states are more significant cues to PPh boundaries; boundary pause duration is less significant. The results also help explain why in fluent speech between-phrase pause durations vary greatly, and can be applied to automatic speech segmentation.

Index Terms: fluent speech prosody, Hierarchical Prosody Group, prosodic state, prosodic phrase, boundary break, discourse prosody, linear regression model

1. Introduction

We have collected various types of fluent Mandarin speech data from read narratives in COSPRO [1] and designed annotations on the basis of perceived boundary breaks in relation to prosodic units. Our Hierarchical Prosodic Phrase Grouping (HPG) [2, 3, 4] specifies multiple-phrases speech paragraphs as a significant discourse prosody unit above phrases whereby COSPRO annotation [5] specifies 5 levels of within-paragraph boundary breaks, i.e., from lower levels upward Syllable (Syl) boundary B1, Prosodic Words (PW) boundary B2, Prosodic Phrase (PPh) boundary B3, change of breath (Breath Group BG [6]) boundary B4 and Prosodic-Group (PG) terminal boundary B5 where physical pause applies from B2 to B5. We have shown from quantitative analyses of speech corpora that output prosody of multiple-speech paragraphs are not at all un related phrase strings, but rather cumulative outcome of contributions from all prosodic layers specified by HPG [3,4]. Further, central to fluent speech prosody is the contribution from above-phrase higher level information related to discourse organization, in which phrases and sentences are all prosodic sub-units of speech paragraph; speech paragraphs sub units of spoken discourse. Among each and every prosodic level, prosodic boundaries in relation to discourse prosody organization are significant cues; perceived boundary breaks are therefore significant prosodic units as well.

However, in a previous study [7,] we discovered that not all boundary breaks could be accounted by pause durations. We found from consistently annotated speech data of 2 speakers at slightly different speaking rate (220, 230ms/syllable) that higher level boundaries B4 and B5 all possessed pause duration over 330 ms (m=330, 520 ms for B4, SD=162, 124 ms; m=415, 595ms for B5, SD=209, 109 ms, respectively), indicating pause durations alone can be viewed as significant cues for BG and PG boundaries. However, boundary pause of B3 varied considerably in duration (from 17-585, 21-538ms at m=224/248ms, SD=150, 207ms, respectively) from 0 to over 350 ms across speakers, indicating pause durations alone are NOT sufficient for PPh boundaries. Therefore, to develop automatic speech segmentation or recognition, pause durations are adequate cues to locate B4 and B5; speech paragraphs as discourse units could be identified. Unfortunately, the rationale would not be applied to within-paragraph prosodic phrase boundaries B3 since it could not be located by pause duration. The question then is why PPh boundary break B3 varies so much in duration across speakers and yet is still perceived consistently across transcribers?

Note nevertheless that the perception based annotation is it makes examination of signal-perception discrepancies possible, especially when perceptions are consistent across transcribers. We therefore hypothesis that there must be cues in the speech signal other than pause duration that are significant to PPh boundary, and are significant to the human ear as well. The same previous study also demonstrated by including boundary immediate prosodic state by one syllable of immediate B3 neighborhood, predictions of B3 were improved by 8.3% [7]. We therefore hypothesize now that B3 predictions can be further improved by including more neighborhood prosodic states in the prediction.

In the following sections, we will show how the previous model is revised to accommodate more boundary immediate syllable duration allocation patterns along the time domain, as well as intensity distribution patterns, and compare newly obtained predictions from the same speech materials with those from the previous model.

2. Speech Data and Methodology

2.1. Speech Data

The same Mandarin Chinese speech data used for previous analysis [5, 7] were selected from Sinica COSPRO 0 [1], i.e., one male and one female speaker (F051P and M051P). Both speakers are professional radio announcers under 35 years of age at the time of recording. Each speaker read text of 26 discourse pieces in sound proof chambers. The 26 6discourse pieces ranged from 85 to 981 characters in length which amounted to a total of 11602 syllables. The corpora were first automatically labeled for segmental identities by the HTK toolkit in SAMPA-T notation [5], and then manually tagged for perceived boundary breaks by trained transcribers using the Sinica COSPRO Toolkit [8]. Annotation results were spot-checked by professional transcribers for segmental alignments as well as inter-transcriber consistency..

2.2. Methods of Analysis Speech Data

We analyzed the speech data in three steps: 1. three acoustic parameters were extracted from annotated speech data, i.e., pause, syllable duration and intensity. 2. Derived acoustic parameters were subsequently normalized. 3. Respective layered contributions specified by the HPG framework were obtained through a step-wise linear regression model. Figure 1 is a flowchart that shows the basic HPG analysis.

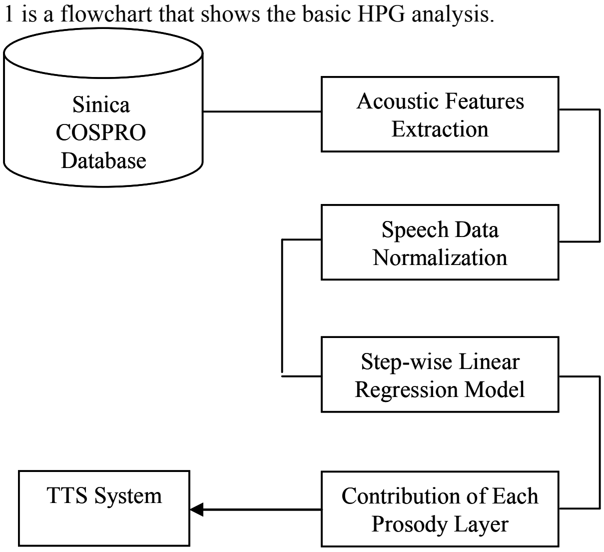

Figure 1: Flowchart of Analysis by HPG Framework.

Table 1 summarizes derived acoustic features of both speakers, where μ and σ represent the mean and standard deviation of each acoustic feature, pause, duration and intensity, respectively.

Table 1. Derived Acoustic features by speaker.

Speaker	μ_{Pause}	σ_{Pause}	$\mu_{Duration}$	$\sigma_{Duration}$	$\mu_{Intensity}$	$\sigma_{Intensity}$
F051P	37	106	200	65	3.65	0.07
M051P	45	138	190	60	3.62	0.05

2.3. Speech Data Normalization

In order to eliminate between-speaker variations, each set of data was normalized with the mean and standard deviation of the entire class. The original method of normalization [4] would easily be affected by extreme data, causing normalized data distribution of to shift, and thereby making comparisons between speakers meaningless. To rectify the situation, we modified the normalization as follows:

$$\mathrm{Ynor(i)} = (\mathrm{Y(i)} - \mu\mathrm{Y}) / \sigma\mathrm{Y}$$

$$\mathrm{Ynor} = \{\mathrm{Ynor(1)}, \mathrm{Ynor(2)}, \ldots \mathrm{Ynor(n)}\}$$

$Y_{(i)}$ and $Y_{nor(i)}$ represent each datum in Class Y and Normalized Class Y respectively. μ_Y and σ_Y represent the mean and standard deviation in Class Y. The same modification was made for the three acoustic features under consideration. Hence Y would be duration, intensity and pause in the following sections.

2.4. Revising the Duration Model

A syllable duration model corresponding to the HPG framework was constructed previously [7] to predict and locating boundary breaks B2 to B5across continuous speech rather than simply predicting pauses. The predictions thus bear discourse information in relation to prosody organization specified by HPG. Higher level BG and PG boundary breaks (B4 and B5 respectively) indicating multiple-phrase speech paragraphs across fluent continuous speech could easily be located using pause durations alone (see Section 1), whereas lower level within-paragraph boundary breaks B3 and B2 corresponding to PPh and PW respectively were predicted using both boundary break pause and durations of one immediate neighboring syllable.

The goal of the present study is to revise the syllable duration model by altering both the Syllable (the bottom) layer and the PW (the immediate higher) layer of the previous regression model to better predict PPh boundary B3. Using the same step-wise regression technique [2, 3], a linear model with four layers [9, 10] was modified and developed to predict speakers' timing behavior through temporal allocation of syllable duration modification. At the syllable layer, we used 6 consonant groups and 6 vowel groups in order to decrease the difference between groups. The Revised Syllable Layer Model could be written as function (1):

$$\begin{aligned} \mathrm{Ynor} = {} & \mathrm{Const} + \mathrm{CCt} + \mathrm{CVt} + \mathrm{Tt} \\ & + \mathrm{PCt} + \mathrm{PVt} + \mathrm{PTt} + \mathrm{FCt} + \mathrm{FVt} + \mathrm{FTt} \\ & + \text{2 - way Factors of each factor above} \qquad (1) \\ & + \text{3 - way Factors of each syllable above} \\ & + \text{PW boundary constraint of each factor above} \\ & + \mathrm{Delta1} \end{aligned}$$

In function (1), we added a new condition that is constrained for each factor in PW boundary to include co-articulation effect such as tone sandhi at the PW layer. Prefix C, P and F represent current, preceding and following syllable, respectively. Ct, Vt and Tt represents consonant, vowel and tone type, respectively. Subsequently, residuals Delta1 that could not be predicted by the syllable layer are then analyzed in the immediate higher layer.

Figure 2 shows the distribution of Delta 1 (the residuals of the syllable layer) of the revised duration model from the speech data where the horizontal axis represents Breaks from B1 to B5, and the vertical axis represents the residual value from -2 to 3. Significant difference (p-value<0.001) was found with respect to the durations between the distributions of B2 and those greater than B2, as well as between speakers, respectively. The results enabled us to avoid overestimating contribution of B2 from the PW Layer, thus we decided to add a constraint condition to only calculate f(PW) in the B2 level.

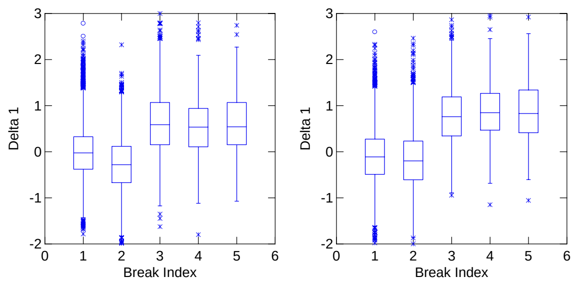

Figure 2: Distribution of Delta 1 of the revised duration model for speakers F051P and M051P.

The Revised PW Layer Model could be written as function (2):

$$\begin{aligned}\text{Delta1} &= f(\text{PW Length, PW Sequence}) \\ &+ \text{The calculation of } f(\text{PW}) \text{ is constrained in B2 level} \\ &+ \text{Delta 2}\end{aligned} \tag{2}$$

Delta2 that could not be predicted by the PW layer are assumed to be contributions from the immediate higher level and therefore are to be analyzed at the next layer upward subsequently.

The PPh and BG Layer Models are the same as our previous models, written as function (3) and (4) respectively.

$$\text{Delta2} = f(\text{PPh Length, PPh Sequence}) + \text{Delta 3} \tag{3}$$

$$\text{Delta3} = f(\text{BGIMF, PPh Length, PPh Sequence}) + \text{Delta4} \tag{4}$$

2.5. Revising the Intensity Model

Based on the revised duration model, we used the same method to analyze the characteristics of the intensity parameter. Figure 3 shows the distribution of Delta 1 of the revised intensity model for both speakers where the horizontal axis represents Breaks from B1 to B5, and the vertical axis represents the residual value for -4 to 3. Significant difference (p-value<0.001) was also found with respect to intensity patterns between the distributions of B2 and those greater than B2, as well as between speakers. Therefore, the same rationale of modification can be applied to the Revised Intensity Model as well.

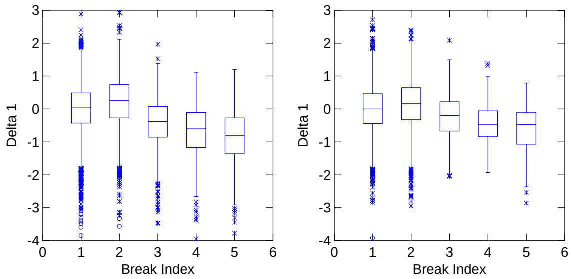

Figure 3: Distribution of Delta 1 of the revised intensity model for speakers F051P and M051P.

2.6. Revising the Pause Model

In our previous pause model [7], we calculated the contribution of pauses from B1 to B4. However, we observed that the distribution of real pauses for B1 is very narrow and therefore decided contribution of B1 could be ignored in the revised model. Figure 4 shows the distribution of pauses from B2 to B4 for speakers F051P and M051P where the horizontal axis represents normalized pause value from -0.4 to 6, and the vertical axis represents frequency of distribution.

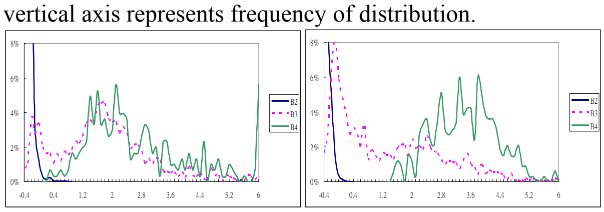

Figure 4: Distribution of Pauses as Boundary Breaks for speakers F051P and M051P.

For the new pause model, the Revised PW Layer Model can be written as function (5):

$$\begin{aligned}\text{Ynor} &= f(\text{PW Length, PW Sequence}) \\ &+ \text{The calculation of } f(\text{PW}) \text{ is constrained in B2 level} \\ &+ \text{Delta1}\end{aligned} \tag{5}$$

Delta1 that could not be predicted by the PW layer are analyzed in the immediate higher layer subsequently.

Further, we also analyzed the distribution of B3's pauses in relation to punctuation marks in the text used. Punctuations comma, period and no punctuation in text in relation to B3 occurrences in speech data were calculated for their respective distributions, as depicted in Figure 5. The results indicated that the value of the B3 pause was indeed affected by the presence of punctuation marks and that that the order of the mean values of B3 is period> comma> no mark. In other words, though both speakers did pause at where no punctuation marks appeared in text; the presence of punctuation marks did cause more B3 in the speech data.

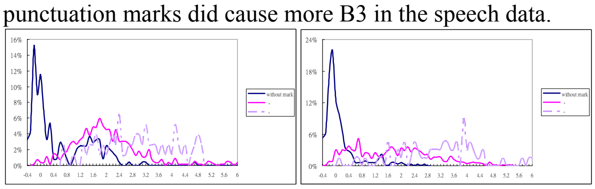

Figure 5: Distribution of Pauses as Punctuation Mark in B3 for F051P and M051P.

According the result, we categorized three groups for punctuation mark by their means, so we can use punctuation marks as a feature at the PPh Layer Model, written as function (6)

$$\text{Delta1} = f(\text{MarkGroup, PPh Length, PPh Sequence}) + \text{Delta2} \tag{6}$$

The BG Layer Model is the same as our previous model, written as function (7).

$$\text{Delta2} = f(\text{BG IMF, PPh Length, PPh Sequence}) + \text{Delta3} \tag{7}$$

3. Results

3.1. Comparison of Duration Predictions

Figure 6 shows the duration patterns of PW (1-4 syllables in length) along the temporal course by syllable number and by speaker from the previous model [7] while Figure 7 shows patterns from the current revised model. Each line represents the corresponding regression coefficient of one syllable at the specific position in a prosodic word. The horizontal axis indicates the position of each syllable and the vertical axis represents the coefficient of normalized values. From Figures 6 and 7 we can see that the PW patterns from the previous model are opposite from patterns from the revised models. Note how the previous model showed final syllable lengthening of PW by syllable number and across speakers whereas the revised model showed the reverse, namely, final syllable shortening PW by syllable number and across speakers. The results from the revised model attribute less contribution from the PW layer to total output prediction in general.

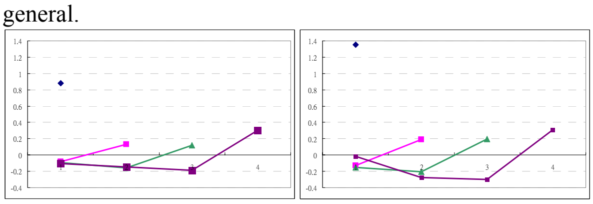

Figure 6: The PW patterns of the previous duration model for

speakers F051P and M051P.

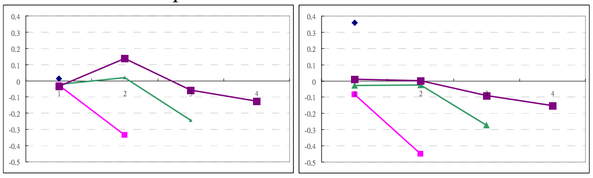

Figure 7: The PW patterns of the revised duration model for speakers F051P and M051P.

Figure 8 shows the duration patterns of PPh (6-11 syllables in length) along the temporal course by syllable number and by speaker from the previous model [7] while Figure 9 shows patterns from the current revised model. Instead of considering only one immediate neighboring syllable of annotated B3, i.e., one pre- and post-B3 syllable only, we defined immediate between-PPh neighborhood as the last 4 syllables of a preceding PPh and the first 3 syllables of the following PPh. By this definition, PPh neighborhood is defined by units that would encompass boundary immediate PW rather than single syllables, a definition that better reflected the rationale of our HPG framework. Note that the cross-boundary contrast is more distinct in the revised model than that from the previous model.

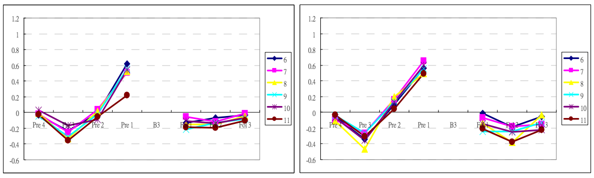

Figure 8: The PPh patterns of the previous duration model for speakers F051P and M051P.

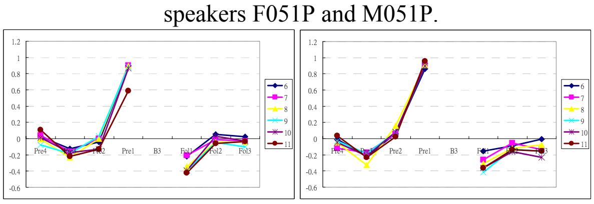

Figure 9: The PPh patterns of the revised duration model for speakers F051P and M051P.

In addition, Figures 7 and 9 combined also show how patterns derived from the revised model are more contrastive in general than patterns derived from the previous model as shown in Figures 6 and 8.

3.2. Comparison of Intensity Predictions

Figure 10 shows the intensity patterns of PW (1-4 syllables in length) along the temporal course by syllable number and by speaker from the previous model while Figure 11 shows patterns from the current revised model. Similar to results from the revised duration model, the revised intensity prediction patterns at the PW layer are also opposite from previous predictions. Figures 12 and 13 show both the intensity distribution of PPh patterns from the previous and revised models; PPh's ranged from 6 to 11 syllables. Note that the PPh patterns from the revised model decayed more drastically towards boundary, thus matching the tendency of the intensity attenuation for PPh final weakening, especially for speaker M051P. Once again the cross-boundary contrast is more pronounced from intensity predictions. Coupled with more phrase-final syllable lengthening found in Section 3.1, the prediction is closer to the physical speech data. Therefore, we believe the cross-boundary contrasts in both duration and intensity patterns are significant cues to boundary perception regardless of boundary pause duration.

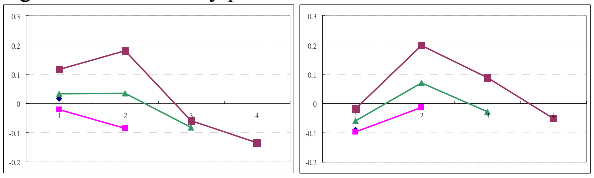

Figure 10: The PW patterns of the previous intensity model for speakers F051P and M051P.

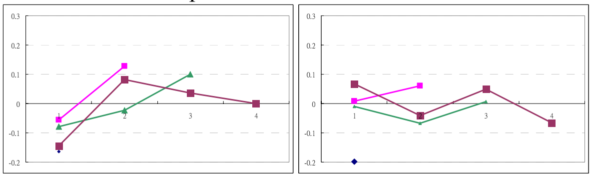

Figure 11: The PW patterns of the revised intensity model for speakers F051P and M051P.

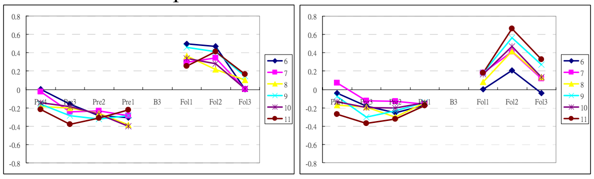

Figure 12: The PPh patterns of the previous intensity model for speakers F051P and M051P.

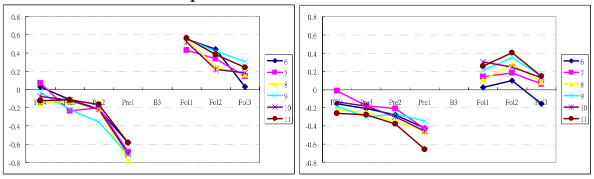

Figure 13: The PPh patterns of the revised intensity model for speakers F051P and M051P.

3.3. Comparison of Pause Predictions

Due to space limit, we will present comparison of pause prediction from one speaker only. Figure 14 shows the comparison of predictions boundary pauses from the previous and revised models for speaker M051P where the horizontal axis represents the Break index of each syllable, and the vertical axis represents pause values from 0 to 750ms. We can see that the differences of pauses between the previous and revised models for B1 and B2 are greater because the previous pause model could be mistaken for contribution from lower Break levels. In the revised boundary pause model, since the contribution of B1 is about 0.4 ms which can not be perceived by the human ear, we ignored the contribution from B1 to refine the prediction of lower Breaks.

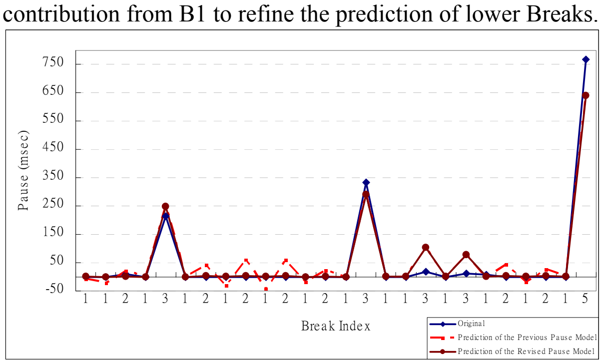

Figure 14: Comparison of the pause predictions between the previous and revised models for speaker M051P.

3.4. Prediction Error Improvement

Our analyses showed a reduction of overall T.R.E. by about 20% from the previous model to the revised model. Table 2 shows the Total Residual Error (T.R.E.) between the previous and revised models for both speakers Therefore, revising the previous model by including more boundary neighborhood state resulted in improved predictions than the previous model, indicating the current predictions deviated less from actual speech data.

Table 2. T.R.E. for speakers F051P and M051P.

F051P T.R.E.	Previous	Revised	T.R.E. Reduction
Duration	36%	32%	11%
Intensity	54%	47%	13%
Pause	32%	22%	31%
Average	41%	34%	17%

M051P T.R.E.	Previous	Revised	T.R.E. Reduction
Duration	33%	31%	6%
Intensity	48%	41%	15%
Pause	27%	13%	52%
Average	36%	28%	22%

We also noted why the T.R.E. of the intensity prediction is always higher than that of the duration prediction. Comparing the distribution of Delta 1 of the intensity model and that from patterns shown in Figures 2 and 3, one can see that the previous case has a broader distribution. It means that the variation of intensity is greater than that of duration, most notably F051P. The broader distribution of Delta 1, the greater the deviation is for the acoustic parameter. The pause prediction was increased effectively by ignoring the contribution of B1 and adding punctuation mark as a feature. Therefore, the order of prediction performance is pause> duration> intensity.

3.5. Analysis of B3 Pauses Shorter than B2 Pauses

As mentioned in Section I, the range of pauses for Breaks is very wide for B3, as plotted in Figure 4. Therefore, in addition to revising the prediction models above, we also studied B3 in more detail. We further analyzed the performances of duration and intensity predictions of between-PPh pause B3's that are shorter than B2. These are cases that contradict the annotation definition but consistently perceived by transcribers. Accordingly, we defined two conditions to analyze short B3 pauses: 1. the lengths of the preceding and following PPhs are equal to or over 6 syllables. 2. Maximum pause of B2 is used as B3's threshold.

The analysis of short B3 pauses for duration and intensity is depicted in Figure 15 and 16. In Figure 15 and 16, we chose to include the last 4 syllables of the preceding PPh at B3 and the first 3 syllables of the following PPh at B3 for analysis. Except for the first 3 syllables of the following PPh of intensity for M051P, there are significant differences between the preceding and following syllables of B3. These results also indicated that our previous model attributed more contribution from the lower PW layer to output prosody, whereas the revised model entails more contribution from the higher PPh layer instead. Since the revised model yielded better overall predictions, it is clear that more contribution from higher level information accounts for the speech data better, hence proving further the significance of higher level contribution to output prosody and how such information is perceived by the human ear..

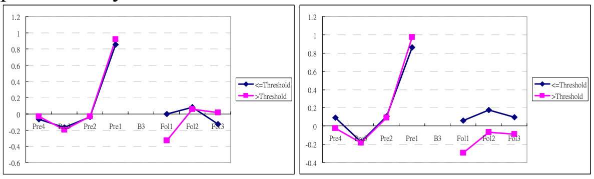

Figure 15: PPh Patterns of the Duration Model for F051P and M051P.

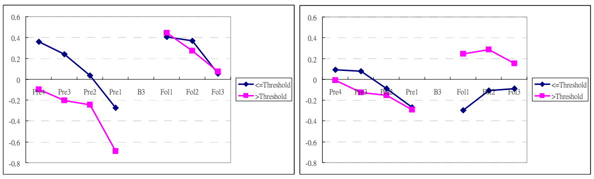

Figure 16: PPh Patterns of the Intensity Model for F051P and M051P.

3.6. Analysis of B3 without Pause

The above results also imply that boundaries between prosodic phrases within speech paragraphs could be signaled more by contrastive neighborhood prosodic state in duration and intensity patterns rather than by the physical duration of boundary pauses in the temporal domain. We also noted in the speech data that there are 7 and 18 tagged B3's without pause for speakers F051P and M051P, respectively. In other words, breaks are consistently perceived across transcribers when there is no silent pause in the speech signals. We studied these data and found the following phenomena may attribute to the perception of boundary B3: 1. there is final-lengthening before B3, as depicted in Figure 17 and as specified in our syllable duration templates [2, 7]. 2. The between-B3-boundary F0 contrast may also have an effect, as depicted in Figure 18. Note when a new PPh begins after B3, F0 reset occurred whereas no reset occurred after the B2 within the same PPh. Therefore, these cases indicated that the perception and judgment of B3 rely more on the integration of neighboring acoustic properties and from contrasts of neighboring states combined rather than on pause duration.

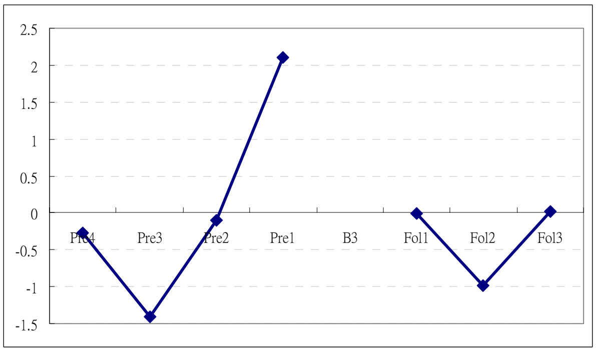

Figure 17: An Example for Final Lengthening.

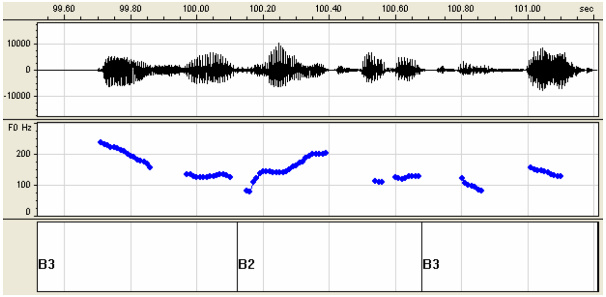

Figure 18: An example of F0 Contrast. The left most B3 occurred at a PPh end where a new PPh begins afterwards and ends by the second B3 from left.

4. Discussion

Instead of analyzing the duration and intensity patterns of one syllable before and after annotated PW and PPh boundary breaks in a previous model [5, 7], we analyzed B3's boundary immediate prosodic states in terms of duration and intensity distribution along the time domain by PW (4 syllables before and 3 syllables after), compared them with those from immediate neighboring B2's, and found different yet corresponding patterns in these two acoustic parameters. Accordingly, we included factors of duration and intensity to revise and fine-tune the linear regression model [7], and recalculated contribution predictions from the PW layer to final prosody output under the HPG framework. The Total Residual Error (T.R.E.) of duration and intensity at the PW layer is improved by 10%; overall prediction of output prosody is consequently improved by 5%. In addition, the layered predictions are now more consistent with the actual break distribution in the speech data,

Based on the above results, we believe that a detailed analysis of residual distributions of every prosodic layer (from syllable to PPh) can yield more stable and general patterns that lead to better prediction. In Figures 6 and 10, duration and intensity patterns at the PPh layer yielded clearer evidences that the coefficients of the last 4 syllables are similar irrespective of PPh lengths (from 6 to 11 syllables). Thus it became clear that to the human ear, PW boundary break B2's and PPh break B3's can be distinguished from each other not by pause duration alone, but by contrastive neighborhood prosodic states as well. Evidence of boundary neighboring F0 contour patterns also showed similar results. Our analyses also showed how contrasts were constituted more by higher level constraints from discourse information than by lower level concatenation smoothing. To the human ear, it is clear that B2 and B3 boundaries are within- rather than between-paragraph signals; their respective pause duration less relevant.

The results enable us to better predict B3 and further argue that prosodic states relate more to higher level information; fluent speech prosody is more than lower level co-articulation driven smoothing.

5. Conclusions

We believe the above results definitely offer alternative rationale for automatic segmentation of fluent speech and speech recognition of Mandarin Chinese in general, especially to the most commonly adopted approach focusing on individual syllabic tone identities and F0 contour patterns, and perhaps inadvertently disregarding boundary as well as higher level information. The improved model can also be incorporated to enhance prosody output of speech synthesis, showing where boundary breaks CAN vary greatly to yield more natural prosody. Last but not least, though we drew evidences from Mandarin Chinese, we believe boundary properties in relation to higher level discourse information are not at all language specific.

6. References

[1] Tseng, Chiu-yu, Cheng Yun-Ching and Chang Chun-Hsiang (2005). "Sinica COSPRO and Toolkit—Corpora and Platform of Mandarin Chinese Fluent Speech", *Proceedings of Oriental COCOSDA 2005*, pp. 23-28. and also http://www.myet.com/cospro

[2] Tseng, Chiu-yu and Lee Yeh-lin (2004). "Speech rate and prosody units: Evidence of interaction from Mandarin Chinese", *Proceedings of the International Conference on Speech Prosody 2004*, pp.251-254.

[3] Tseng, C., Pin, S., Lee, Y., Wang, H. and Chen, Y. "Fluent Speech Prosody: Framework and Modeling." *Speech Communication*, Vol.46, issues 3-4, (2005), Special Issue on Quantitative Prosody Modeling for Natural Speech Description and Generation, pp.284-309.

[4] Tseng, Chiu-yu (2006). "Recognizing Mandarin Chinese Fluent Speech Using Prosody Information—An Initial Investigation" *The 3rd International Conference on Speech Prosody 2006*,

[5] Tseng, C., Chou, F., 1999. A prosodic labeling system for Mandarin speech database. In: *Proceedings of ICPhS'99*, pp. 2379-238

[6] Lieberman, Philip. 1967. *Intonation, Perception and Language*, The MIT Press, Cambridge, Massachusetts.

[7] Tseng, C. and B. Fu. "Duration, Intensity and Pause Predictions in Relation to Prosody Organization" *Proceedings of Interspeech 2005*, pp.1405-1408.

[8] Sinica COSPRO and Toolkit http://reg.myet.com/cospro

[9] Keller, E., Zellner Keller, B. "A Timing model for Fast French", *York Papers in Linguistics*, 17, University of York. 53-75. (1996)

[10] Zellner, Keller B, Keller E., "Representing Speech Rhythm" Improvements in Speech Synthesis. (pp. 154-164). Chichester: John Wiley. (2001)