Prosodic phrasing and F0 in Singapore English

Adam J. Chong[1], James Sneed German[2]

[1] University of California, Los Angeles
[2] Aix-Marseille Université, CNRS, LPL UMR 7039, Aix-en-Provence, France
ajchong@ucla.edu; james.german@lpl-aix.fr

ABSTRACT

A number of studies have investigated the phonetic properties of intonation patterns in Singapore English (SE). Additionally, it has been proposed that SE intonation is structured by a level of phonological phrasing above the word called the Accentual Phrase (AP), which encompasses a content word and preceding function words. This study investigates experimentally whether there are any durational correlates of AP boundaries and explores their possible relationship to established F0 patterns. In addition, the sensitivity of F0 movements to lexical stress within an AP is examined. Our findings show that word-final syllables at an AP boundary are marked by longer vowel durations than those in a phrase-medial word, and that this difference corresponds closely with known F0 correlates. Finally, within an AP, stress in a word-initial syllable results in an overall raising of F0 across the entire phrase, although the actual slopes of the contours do not differ.

Keywords: Singapore English, intonation, phrasing, duration, stress

1. INTRODUCTION

Singapore English (SE) intonation has often been described as having a sequence of rising tones [4]. A number of studies have explored the phonetic and acoustic properties of SE intonation [4, 11]. More recently, [3] proposed a preliminary phonological description of SE intonation within an Autosegmental-Metrical (AM) framework [8]. Specifically, [3] proposed that SE consists of three levels of phrasing above the word: an Accentual Phrase (AP), Intermediate Phrase (ip) and Intonational Phrase (IP).

The AP in SE consists of a single content word, or a single content word and preceding function words. Additionally, a function word can occur AP-finally in a number of syntactic constructions that involve ellipsis. Most commonly, the AP consists of a LH tonal pattern, and it was suggested by [3] that the AP-final H tone is a delimitative tone that marks the word or phrase edge as in Korean [5], Japanese [1, 9] and French [6].

Previous work by [3] relied primarily on qualitative evidence for the presence of different prosodic levels. This study brings empirical and quantitative data to bear on certain aspects [3]'s phonological model. Specifically, we investigate two questions related to the status and properties of the AP in SE. Firstly, we examine whether there are any durational or F0 correlates to AP boundaries. Secondly, we investigate whether F0 movement across an AP is sensitive to stress placement. Importantly, we only examine the speech produced by Singaporeans who are ethnically Chinese, since it is now well-established that members of different ethnic groups differ in terms of pronunciation including prosody [11].

2. METHODS

2.1. Participants

9 native speakers (3 M; 6 F; age 20-27) of Singapore English were recorded in Singapore. All participants were ethnically Chinese and students at Nanyang Technological University at the time of the study.

2.2. Stimuli

Two sets of stimuli were created in order to investigate the two questions addressed in this study. To investigate durational and F0 correlates of phrasing, target words consisting of either prepositions (9 items) or auxiliaries (9 items) were placed in one of three sentential positions based on the size of the prosodic boundary that was expected at each: word, AP, or IP. Since nouns and lexical verbs nearly always trigger an AP boundary at their right edge in SE, it is not trivial to create matched pairs for comparing the Word- and AP-levels. Since auxiliaries and prepositions are often phrased with the following noun or lexical verb, however, it was possible to achieve matching either through ellipsis (auxiliaries) or through stranding (prepositions) constructions. Thus, the Word-level items were followed by the associated noun or lexical verb within the same (syntactic and prosodic) phrase, while the AP-level items were instead followed directly by a temporal or locative adjunct. IP-level items were similar to AP-level items, except that the target words were sentence-final. The three conditions are illustrated in (1-3) below.

(1)   He said he <u>will</u>.   (IP-final)
(2)   He said he <u>will</u> tomorrow.   (AP-final)
(3)   He said he <u>will</u> go tomorrow.   (Word)

To investigate the effects of stress placement on F0 movement, 60 test words were placed in sentence-initial position within a variety of sentence frames. Previous work has suggested that the initial word (or accentual phrase) showed the largest F0 range [3, 4]. Therefore, it is likely that F0 movement would be most salient in this position, thereby maximizing any stress-related F0 differences. Target words varied as to the number of syllables they contained (1, 2 or 3). Furthermore, underlying stress placement in bisyllabic and trisyllabic words was varied with stress falling either on the initial or second syllable. Additionally, we included multi-word targets that consisted of a mono- or bisyllabic verb preceded by one or more function words.

2.3. Procedure

Participants were seated in a sound-attenuated booth. Stimulus items were presented as written text on a computer monitor, with presentation order controlled by E-Prime 2.0 [10]. Stimulus items were pseudo-randomized into five presentation blocks. The order of presentation of blocks was randomized, as was the order of presentation of items within each block.

Speakers were instructed to read each sentence as naturally as possible. Their productions were recorded at a 48 kHz sampling rate with 16-bit encoding using a Shure SM81 microphone linked to a FocusRite Saffire PRO40 audio interface.

3. RESULTS

3.1. Duration

Duration was measured for the vowel nuclei of target words across the three phrasal positions (Fig. 2). The effect of phrasal position (Word vs. AP vs. IP) was analysed using linear mixed effect models, by comparing a base model with only random intercepts for target and speaker with a superset model which included phrasal position as a fixed factor. There was a significant effect of phrasal position ($\chi^2(2)= 424.00$, $p < 0.001$). Pairwise comparisons revealed that duration of vowel nuclei differed significantly for all three pairwise comparisons (word vs. AP: $p < 0.001$; AP vs. IP: $p < 0.001$; IP vs. word: $p < 0.001$), with vowels for targets in IP positions longer than those in AP positions, which were in turn longer than those in AP-medial words.

These results therefore support a model in which duration is used to mark a three-way distinction in

prosodic boundary strength. Since AP boundaries are expected to be marked by an H tone at their right edge, it is further predicted that the lengthening associated with AP in our study should be correlated with f0 in the same syllable. More importantly, if duration and f0 together mark a *categorical* distinction between the word and AP level as predicted by [3], then the two datasets should give rise to distinct groups in the 2-dimensional space defined by those two measures. Fig. 2 suggests that this is indeed the case. (Note that because of strong interspeaker variability in f0 range and speech rate, both measures were z-normalized before plotting).

Figure 1. Duration of target nuclei by phrasal position. Empty triangles indicate mean duration.
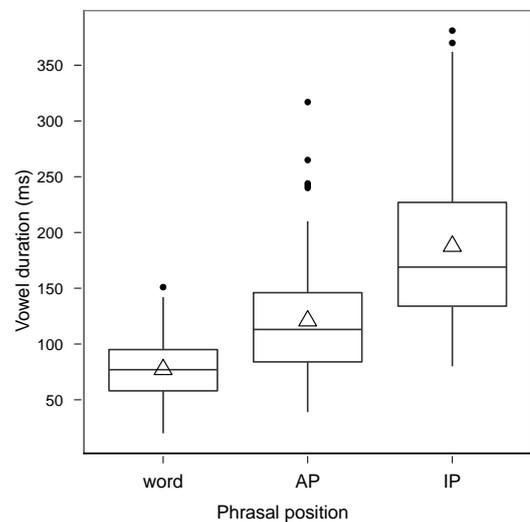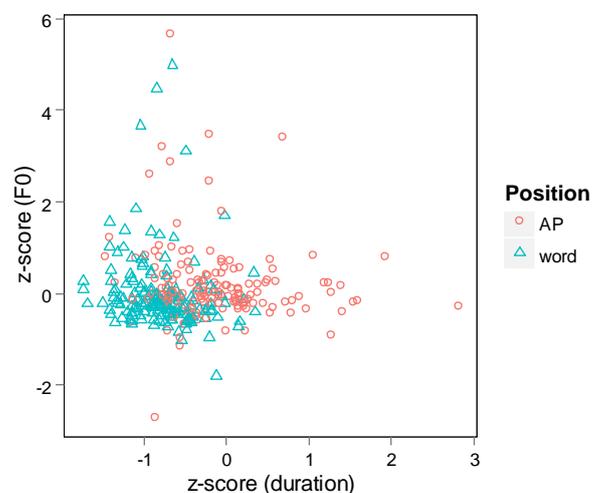

Figure 2. Normalized F0 by normalized duration of vowel nuclei in Word (blue) and AP-final (red) conditions.

The presence of two categories was further confirmed by a linear discriminant analysis with normalized duration and F0 as predictors. Using the posterior probabilities of the model to predict category membership (Word or AP) resulted in an accuracy of 73.9%, which was significantly different from chance (50%) according to a binomial test (p < 0.001, 95% CI [68.7%, 78.7%]). An example showing the AP-final H tone is shown in Fig. 3. Note that the preceding content word, *said*, also bears an H tone indicating the boundary of the preceding AP.

Figure 3. Example sentence showing H tone on the auxiliary verb *will* in AP-final position.
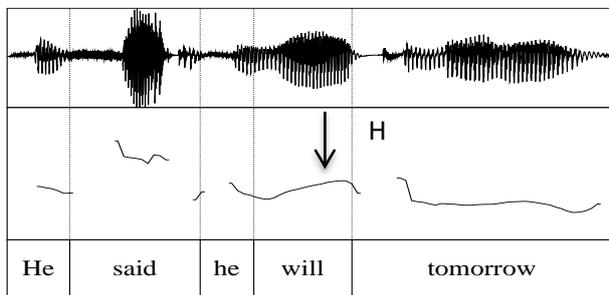
3.2. F0 across the initial content word (first AP)

In this paper, only data from targets that contained a single content word with either two or three syllables was analyzed. Target items were excluded if a speaker's production did not consist of the expected number of syllables (e.g. two syllables instead of three). A total of 9 tokens were excluded for this reason.

Sentence-initial target words were segmented and mean F0 was collected over the nucleus of each syllable using a Praat script [2]. Mean F0 (in Hz) by syllable number and condition (stress-initial or stress-medial/final) are shown in Fig. 4 and 5 for two- and three-syllable words respectively. Separate analyses were conducted for two- and three-syllable words using linear-mixed effect models. In each model, Syllable No. and Condition (Stress-Initial vs. Stress-Medial/Final) were used as fixed factors as well as their interaction. All models included random intercepts for target word and speaker. To compare models, likelihood ratio tests were conducted, comparing two models with different number of factors (in a subset relationship). Model comparisons without the variable of interest are presented as chi-squares.

With two-syllable words, there was a significant effect of Syllable Number ($\chi^2(2)$= 170.53, $p < 0.001$) with F0 being significantly higher in the second (word-final) syllable than in the first. There was also a significant effect of Condition ($\chi^2(1)$= 6.15 p =

0.01), with F0 higher overall in stress initial (Su) than stress-final (uS) target words. There was no significant interaction ($\chi^2(1)$= 0.33, $p = 0.56$).

With three-syllable words, there was a significant effect of Syllable Number ($\chi^2(2)$= 315.06, $p < 0.001$). There was also a significant effect of Condition ($\chi^2(1)$= 10.86, $p < 0.001$), with F0 higher overall in stress-initial (Suu) than stress-medial (uSu) target words. There was no significant interaction ($\chi^2(2)$= 2.49, $p = 0.29$). Pair-wise comparisons revealed that F0 in all syllables was significantly different (Syllable 3 > Syllable 2 > Syllable 1; 3 vs. 1: $p < 0.001$; 3 vs. 2: $p < 0.001$; 1 vs. 2: $p < 0.001$).

Figure 4. Mean F0 for each syllable by condition (bisyllabic words). Error bars show 95% confidence interval.
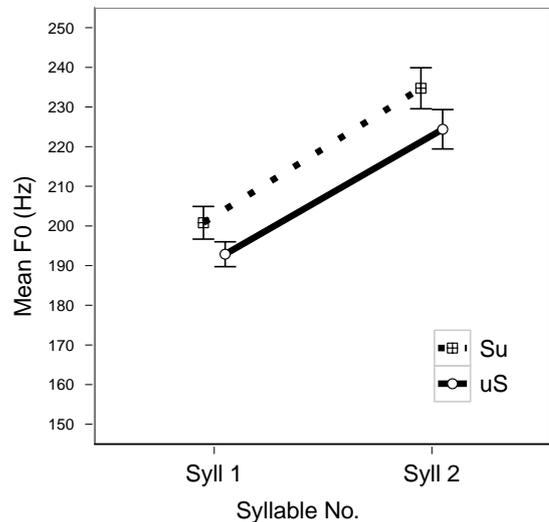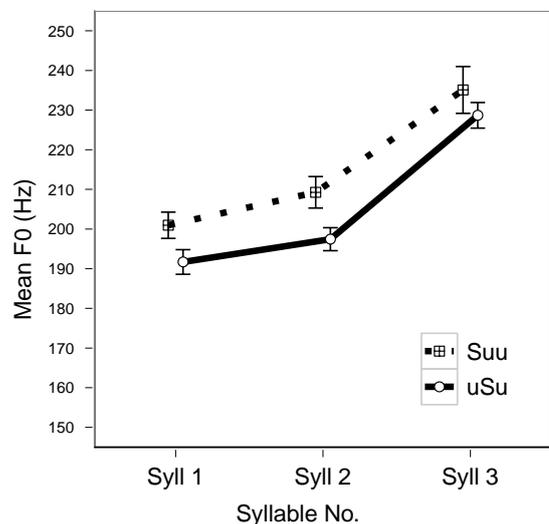

Figure 5. Mean F0 for each syllable by condition (trisyllabic words). Error bars show 95% confidence interval.

In order to explore the trajectory and height of F0 over each word in a qualitative manner, F0 contours across target words were plotted. The global F0 contour across target words from one speaker is shown in Fig. 4 for two syllable words, and Fig. 5 for three syllable words. Impressionistically, F0 is raised across the entire content word when the first syllable bears lexical stress (blue line). This pattern seems to be consistent in both two-syllable and three-syllable words. The qualitative observation therefore reinforces the quantitative results presented above: Stress on a word-initial syllable causes across-the-board raising of F0.

Figure 6. Global F0 contour of one make speaker across two syllable words in a normalized time domain. (Blue = Su; Red = uS)
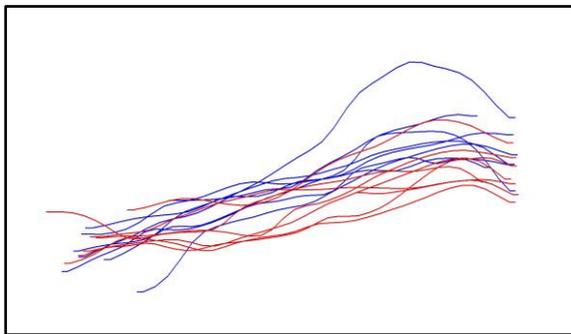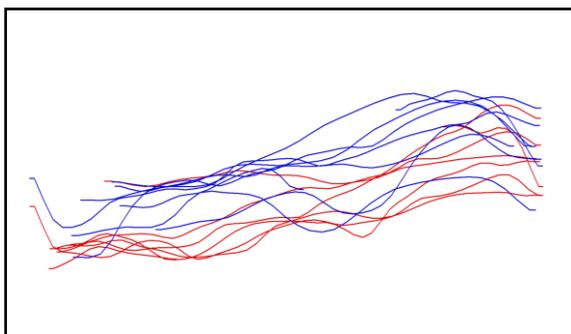

Figure 7. Global F0 contour of one male speaker across three syllable words in a normalized time domain (Blue = Suu; Red = uSu)

4. DISCUSSION & CONCLUSION

In this paper, we investigated two questions related to prosodic structure and tonal realization in SE. Firstly, we tested for the presence of durational correlates of a level of phrasing (AP) that was previously identified on the basis of intonational cues. Our findings support the claim in [3] that SE possesses at least two levels of prosodic structure above the word: the AP and the IP. Although it is possible that the intermediate level of phrasing, what we are calling the AP, is instead an intermediate phrase (ip), this is unlikely given that the boundaries

in question are realized consistently with a H delimitative tone, and the domains that these tones typically mark consist of at most a single content word plus any preceding function words, as in other AP languages [1, 5, 6, 9]. Reinforcing this view, we found that when both F0 and duration are considered, they strongly suggest a categorical distinction between the Word and AP level.

Our second question focused on the effects of stress placement on F0 movement across an AP. Our findings show that when stress falls on the initial syllable of a content word, F0 is raised *across* the entire phrase starting on the first syllable. This effect is not predicted by previous models of tone association in SE [7, 12] which predict that only word-initial syllables should show F0 differences due to stress but following syllables should not.

Future work will extend the current analysis to multi-word phrases which include a content word and preceding function words, with the eventual goal of producing a full phonological model of SE intonation. Perception studies will also follow to assess whether SE speakers are sensitive to the stress-related F0 differences found here.

6. REFERENCES

[1] Beckman, M., Pierrehumbert, J. 1986. Intonational structure in Japanese and English. *Phonology Yearbook, 3,* 255-309.

[2] Boersma, P. and Weenink, D. 2011. *Praat: doing phonetics by computer [Computer program]*. Version 5.3.02.," Retrieved September 28, 2011 from http://www.praat.org/.

[3] Chong, A. 2013. Towards a model of Singaporean English intonational phonology. *Proc. of the Meetings on Acoustics, 19*, Montreal, Canada.

[4] Deterding, D. 1994. The intonation of Singapore English. *Journal of the International Phonetic Association, 24*, 61-72.

[5] Jun, S.-A. 1996. *The phonetics and phonology of Korean prosody*. New York: Garland.

[6] Jun, S. A., Fougeron, C. 2000. A phonological model of French intonation. In: Botinis, A. (ed.), *Intonation: Analysis, modelling and technology*. Dordrecht: Kluwer Academic Publishers, 209-242.

[7] Ng, E.-C. 2011. Reconciling stress and tone in Singaporean English. In: Zhang, L. J., Rubdy, R., Alsagoff, L. (eds.), *Asian Englishes: Changing perspectives in a globalised world*. Singapore: Pearson Longman, 48-59.

[8] Pierrehumbert, J. B. 1980. *The phonology and phonetics of English intonation.* Doctoral dissertation. Boston: MIT.

[9] Pierrehumbert, J., & Beckman, M. (1988). Japanese tone structure. *Linguistic inquiry monographs*, *15*, 1-282.

[10] Schneider, W., Eschman, A., Zuccolotto, A. 2012. *E-Prime User's Guide*. Pittsburgh: Psychology Software Tools, Inc.

[11] Tan, Y. Y. 2010. Singing the same tune? Prosodic norming in bilingual Singaporeans. In: Ferreira, M. C. (ed.), *Multilingual Norms*. Frankfurt: Peter Lang, 173-194.

[12] Wee, L.-H. 2008. More or Less English: Two Phonological Patterns in the Englishes of Singapore and Hong Kong. *World Englishes*, *27*, 480-501.