Variability in tonal realisation in Singapore English intonation

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

[1] Queen Mary Uni. of London; [2] Aix-Marseille Université, CNRS, LPL UMR 7039
a.chong@qmul.ac.uk; james.german@lpl-aix.fr

ABSTRACT

The intonational melody of Singapore English has previously been described as consisting of a series of repeated downstepped rises over an utterance. The exact nature of this rise, however, differs by phrasal position, with significantly higher variability in contour shape and tonal alignment in non-initial Accentual Phrases than in initial ones. In this study, we focus on variability of tonal contours in utterance-medial Accentual Phrases, examining whether the two main types identified in previous studies (plateaus and rises) represent distinct categories. We then investigate the extent to which factors such as duration and f0 scaling are predictive of f0 pattern selection in utterance-medial phrases. These findings are discussed in the context of the challenges posed by the high variability, both within and across speakers, in Singapore English, and in particular, the difficulty this poses for refining an intonational model of this variety.

Keywords: intonation, prosody, Singapore English, f0 scaling, duration, downstep

1. INTRODUCTION

The intonational melody of Singapore English (SgE) has traditionally been described as a repeated regularly-timed series of rises which are downstepped across an utterance [1, 2]. The exact nature of this rise, however, has been subject to various analyses. In one line of work, it has been argued that SgE is a 'tone' or stress-accent language [3,4], with each syllable in a prosodic word specified for a tone. [4]'s model in particular posits that word-final syllables receive a high (H) tone, the first stressed syllable receives a mid (M) tone and any preceding unstressed syllable receives a low (L) tone. These accounts of the tonal patterns of SgE predict that stress should reliably be signalled by an M tone, and further does not consider how this system would interact with higher-level sentence prosody. Moreover, these studies rely primarily on impressionistic transcriptions of tone without a quantitative examination of f0 patterns.

A separate line of investigation has examined SgE intonation within an Autosegmental Metrical framework [1,5,6]. [1,5,6] argue that the domain of each rise corresponds to an Accentual Phrase (APs), a prosodic constituent that typically consists of a single content word and any preceding function words. In initial APs where the pitch range is widened [2,6,7], [5] found that when stress is on an initial syllable, f0 is raised across the entire AP, starting from the first syllable, and there is no clear tonal target on stressed syllables in these APs, with similar tonal contour shapes regardless of whether stress was initial or medial in a word.

Figure 1: L tone aligned to stressed syllable followed by small rise on final syllable in an IP-medial target.

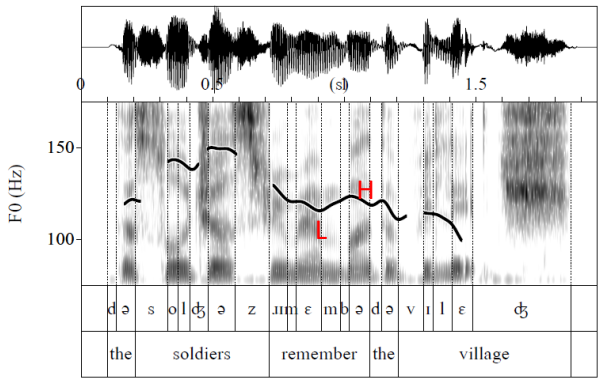
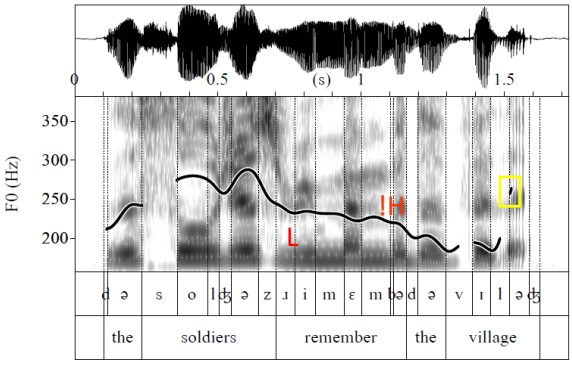


Figure 2: L tone aligned to AP-initial syllable followed by tonal plateau in an IP-medial target. (Yellow indicates pitch tracking error.)



While the general shape of tonal contours in initial APs is consistently a rise regardless of speaker and stress pattern, the tonal contours in non-initial APs are much more variable [8]. [8] identified two major contour types in non-initial APs: rises and plateaus. Final APs also show an additional falling pattern where f0 falls across the entire AP. [8] also provided

preliminary evidence that the f0 valley or elbow that marks the left edge of an AP can align to the stressed syllable or earlier, a pattern that is difficult to account for under a model that assumes tone is reliably used to mark stress [4]. In fact, the same target word can display different tonal patterns across speakers (Figures 1 and 2), suggesting that this variability cannot solely be due to differences in the position of a stressed syllable.

These previous qualitative observations raise the question of whether non-initial phrases should be analyzed as phonologically distinct from initial phrases [5]. Further, from a methodological perspective, the high variability in the tonal patterns observed within and across speakers for the same task and items poses an interesting challenge for the refinement of an intonational model of this variety.

In this study, we focus on investigating one aspect of the phonetic detail of tonal realization in medial APs. In particular we address the question of whether the tonal contour types previously identified by [8] represent two distinct categories or instead exist on a more gradient continuum. We further examine whether factors such as duration and tonal scaling predict the choice of tonal pattern in this context.

2. METHODS

2.1. Participants

Eight native speakers of Singapore English (4 male; 4 female; mean age: 22) were recruited for this study as paid volunteers (same corpus as [8]). Given previous descriptions [9] of intonation patterns differing across the major ethnic groups of SgE speakers, only speakers who self-reported being of ethnic Chinese heritage were included in the present study. To further minimize possible sources of variability, participants were all university-educated, and were all undergraduate or graduate students at Nanyang Technological University at the time of recording. While this narrow sample of the Singapore population is certainly not fully representative, it permitted us to more carefully examine a specific aspect of intra-speaker variability by controlling for the even greater variability that arises from population-level sociolinguistic factors.

2.2. Materials

Target sentences involved trisyllabic words with either initial or medial stress. The target words were placed in either an IP-medial or IP-final AP. This was achieved through the inclusion or exclusion of a prepositional or noun phrase at the end of the sentence. We further manipulated the distance of the target word from the left edge of the AP by varying the inclusion of a preceding determiner or auxiliary verb. Target words were paired for overall segmental similarity and stress placement (initial or medial), with each member of the pair occurring in the same sentential contexts. (1) shows an example pair of target words, with initial and medial stress, placed in all four experimental sentence contexts. A full list of target sentences is presented in Table 1. There were 12 target words (six pairs), with a total of 48 items (12 targets X 4 contexts). The 48 experimental items were randomized in a single list along with 73 other items used for a separate experiment.

(1) Initial stress: *minerals;* Medial stress: *memorials*

a. Medial AP: They explain *minerals/memorials* to the tourists.

b. AP + Function word: They explain *the minerals/memorials* to the tourists.

c. Final AP: They explain *minerals/memorials.*

d. Final AP + Function word: They explain *the minerals/memorials.*

Table 1: Target sentences. Parentheses indicate elements that were manipulated.

Target sentences
1. They clean (the) animals (in the back).
2. They clean (the) enamels (in the back).
3. They explain (the) minerals (to the tourists).
4. They explain (the) memorials (to the tourists).
5. They compare (the) seminars (for the report).
6. They compare (the) tsunamis (for the report).
7. The officers (will) mobilise (the troops).
8. The officers (will) manoevre (the troops).
9. The players (will) dominate (their rivals).
10. The workers (will) demolish (their rivals).
11. The soldiers (will) liberate (the village).
12. The soldiers (will) remember (the village).

2.3. Procedure

Participants were presented with sentences displayed on a computer monitor one at a time. Experigen [10] was used to control presentation order, and the task was self-paced. Participants were asked to read each sentence aloud in a natural conversational style. Their productions were recorded in a sound attenuated booth at Nanyang Technological University in Singapore. Participants were asked to read the entire list twice and were given a short break between the two readings.

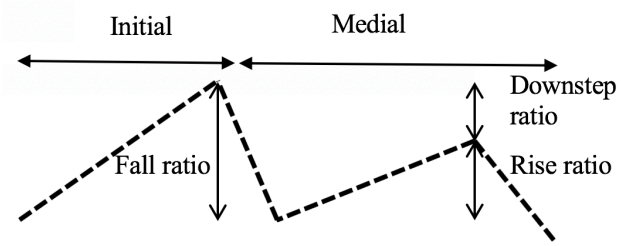
2.4. Data pre-processing and annotation

Here we analyse data from the second reading, as those productions were overall much more fluent. Productions involving pauses or hesitations within an IP were also excluded, for a total of 361 tokens. These recordings were then phonetically segmented and labelled automatically using the SPPAS force alignment tool [11]. In the present study, we confine our investigation to the tonal patterns in medial APs, and more specifically, those involving either a plateau or a rise terminating near the end of that AP. Among the 182 tokens involving medial targets, 49 involved an f0 peak on either the first or second syllable of the target word (i.e., the lexically stressed syllable); since this is not a normative pattern for this variety, those items were excluded. 23 of these represented the entire sample for a single speaker, who was therefore effectively excluded. The remaining 133 tokens were labelled using Praat [14] for the following three landmarks: (i) the final f0 peak of the initial AP (initial H), (ii) the initial f0 elbow of the medial AP (medial L), and (iii) the final f0 peak/terminus of the rise/plateau of the medial AP (medial H). F0 values were extracted automatically using Praat [12]. Peaks were labelled at the local f0 maximum using the 'Get maximum pitch' function over a manually specified window. Elbows were labelled at either the f0 minimum or following [13], at the point where change in f0 appeared to be greatest following the preceding peak. In cases where these landmarks were obscured by strong segmental perturbation or voicelessness, the label was placed at the lowest point that was outside of the perturbation. Plateau termini were labelled at the point just prior to the first sign of a decrease in f0 before the fall.

3. RESULTS

3.1 Rising vs. plateaus: a categorical distinction?

Figure 3: Schema of dependent variables.



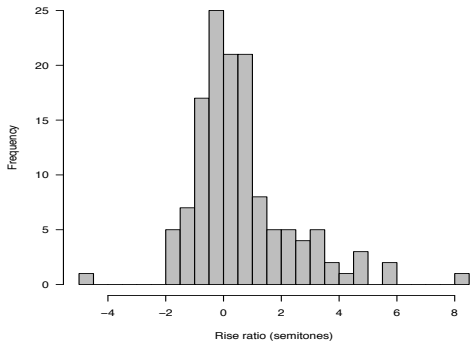
We first asked whether the global f0 patterns previously identified (rise and plateau) involve a categorical or gradient distinction. F0 excursion (here, *rise ratio* Figure 3) from the initial L to the final H of the medial AP was calculated by taking the semi-tone transformation of the ratio of those two points (2).

(2) Semi-tone formula: $12*\log_2(f0\text{max}/f0\text{min})$

If speakers select, in production, two categorical tonal patterns, we would expect a bimodal distribution for this measure. An examination of this distribution (Figure 4), however, reveals a clearly unimodal distribution (Mean: 0.65 semitones; SD: 1.72), with a positive skew (skewness: 1.25), supporting the previous informal classification of tonal patterns into plateaus (rise ratios close to 0) and rises (ratios larger than 0), but not falls.

The results of Hartigan's Dip Test for unimodality using the *dip.test()* function [14] in R [15] confirms that the distribution is unimodal with a non-significant result ($D = 0.02$, $p = 0.96$). This therefore suggests the surface patterns observed in medial APs reflect different phonetic implementations of a single phonological pattern.

Figure 4: Histogram of *rise ratios* (Medial AP H/Medial AP L).



3.2. Predicting f0 patterns in medial APs

How does one explain the variance in rise ratios observed in the current data? We hypothesize that the variability in the realization of a single phonological LH target can be explained by target undershoot due to temporal constraints [16]. Such durational effects on tonal target realizations have been shown in other languages (e.g., French [17]; Standard Chinese [18]). If this is the case, we predict that more compressed phrases (i.e. shorter duration) would result in more undershoot of either the initial L target, the final H target, or both.

In order to further examine these possibilities, we examined the effect of AP duration on the size of the f0 fall from the H of the initial (previous) AP to the initial L target of the medial AP (*fall ratio*: Figure 5L), and on the size of the f0 downstep from the final H of the initial AP to the final H of the medial AP (*downstep* ratio: Figure 5R). F0 displacement for both measures were calculated as with rise ratios. Linear

mixed-effects models with Satterthwaite approximations were fit using the *lmerTest()* function [19] in R [16]. Each model (fall ratio and downstep ratio) contained AP duration as a fixed effect and random intercepts for subject and item.

Figure 5: (L): *Fall ratio* (Final AP H/Medial AP L) by AP Duration. (R): *Downstep ratio* (Initial AP H/Medial AP H) by AP Duration.

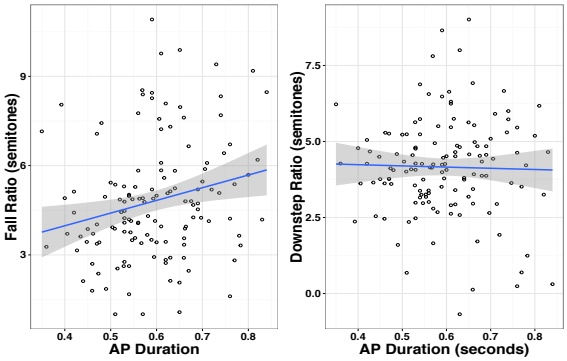
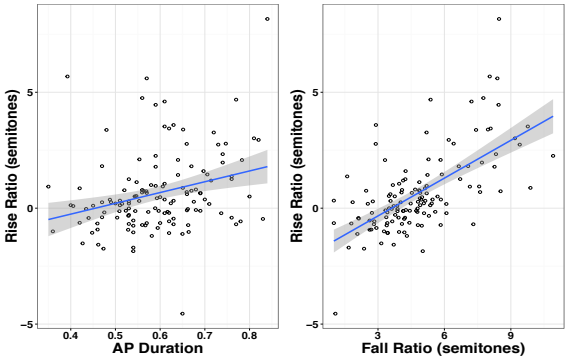


Figure 6: *Rise ratio* by AP duration (L) and *Fall ratio* (R).



AP Duration was a significant predictor of the fall ratio (β = 3.93, *SE* = 1.54, t = 2.56, p = 0.03), in that longer AP duration was associated with a larger f0 fall. Downstep ratio, however, was not significantly correlated with AP duration (β = -0.81, *SE* = 1.51, t = -0.53, p = 0.60). That the size of the fall, but not the size of the peak to peak downstep, is affected by phrase duration suggests that (a) the final H of the medial AP is a relatively fixed function of the f0 range of the initial AP, and (b) that variability in the L-to-H f0 change in medial APs (i.e., rise ratio) is therefore most likely due to constraints on the realization of the initial L of those APs, in that shorter medial APs lead to undershoot of the L targets.

Finally, we return to the issue of variability in the rise ratio by examining the extent to which it can be explained by either AP duration (Figure 6L), fall ratio (Figure 6R), or both. Another linear mixed-effects model was fit with rise ratio as the dependent variable and fall ratio and AP duration as fixed factors. Both factors had a significant effect on rise ratio with larger falls (β = 0.29, *SE* = 0.08, t = 3.48, p = 0.009) and longer APs predicting larger rises (β = 3.39, *SE* = 1.21, t = 2.80, p = 0.01).

4. DISCUSSION

In this study, we addressed two questions relating to the variability in tonal realization in medial APs in SgE. In contrast to initial APs, which are consistently realized as a rise, medial APs may be realized as rises or plateaus. We showed that these differences do not correspond to distinct categories. Instead they appear to represent gradiently different phonetic implementations of a single underlying category.

We then examined which factors best predict the surface variation in the global f0 pattern of medial APs, and found that these include both the duration of the AP and the size of the fall from the previous peak, but not the size of the peak-to-peak downstep. These findings are consistent with an account in terms of tonal undershoot of AP-initial L resulting from durational constraints, and they further suggest that the AP-final H acts as the reference pitch level for an AP [20] relative to the preceding AP.

That differences in the surface patterns of medial APs are gradient in nature suggests that this variation should not be accounted for in terms of phonologically distinct tonal melodies. Instead, both patterns may be subsumed by a single AP-level LH melody. Variation in the surface pattern has its source in phonetic implementation subject to f0 scaling and temporal constraints.

The descriptive observation of sustained level plateaus has been used to argue for an analysis of SgE intonation in terms of tonal specification over each syllable in a word (e.g., [3,4,21]). Our study, however, reveals a continuum of phonetic variation between plateaus and rises, which arises from temporally-driven compression effects. Our results therefore do not support such an analysis. While here we do not specifically address the issue of word stress and tonal alignment, this general result is in line with [8], which suggest that the AP-initial L can optionally align to either the left edge of the AP or a lexically stressed syllable, a fact which also cannot be captured by a model which involves tonal specification for each syllable [3, 4].

Finally, we also observed patterns much more characteristic of head-marking (e.g., in American or British English). This kind of variability (on the same stimuli and task with speakers of similar sociolinguistic backgrounds) poses challenges for deciding what does or does not count for the purposes of proposing a phonological model of this variety. Future work will examine the extent to which these productions show SgE-like features.

5. ACKNOWLEDGEMENTS

This study was made possible through support by the A*MIDEX project (n° ANR-11-IDEX-0001-02) funded by the Investissements d'Avenir French Government program, managed by the French National Research Agency (ANR), by a grant from the Singapore Ministry of Education Academic Research Fund Tier 1 (2013–T1–002–169), and by research funds from the UCLA Department of Linguistics. Thanks to Diyana Hamzah for assistance with data collection, and to Andy Xu for help with data pre-processing.

6. REFERENCES

[1] Chong, A. J. 2013. Towards a model of Singaporean English intonational phonology. *Proceedings of the Meeting on Acoustics* 19, 1–9.

[2] Deterding, D. 1994. The intonation of Singapore English. *J Int Phonet Assoc* 24, 61–72.

[3] Lim, L. 2009. Revisiting English Prosody: (Some) New Englishes as tone languages? *English World Wide* 30, 218-239.

[4] Ng, E.-C. 2011. Reconciling stress and tone in Singaporean English. In Zhang, L., Rubdy, R, Alsagoff, L. (eds.), *Asian Englishes: Changing Perspectives in a Globalised World*. Singapore: Pearson Longman, 48–59.

[5] Chong, A. J., German, J. S. 2015. Prosodic phrasing and F0 in Singapore English. *Proceedings of the 18th International Congress of Phonetic Sciences* (Glasgow, UK).

[6] Chong, A. J., German, J. S. 2017. The accentual phrase in Singapore English. *Phonetica* 74, 63-80.

[7] Low, E.-L. 2000. Is lexical stress placement different in Singapore English and British English? In: Brown, A., Deterding, D., Low, E.-L. (eds), *The English Language in Singapore: Research on Pronunciation*. Singapore: Singapore Association for Applied Linguistics, 22–34.

[8] German, J. S., Chong, A. J. 2018. Stress, tonal alignment and phrasal position in Singapore English. *Proceedings of Tonal Aspects of Language 2018* (Berlin, Germany).

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

[10] Becker, M., Levine, J. 2013. *Experigen – an online experiment platform*. Available online at: http://becker.phonologist.org/experigen.

[11] Bigi, B. 2015. SPPAS: multi-lingual approaches to the automatic annotation of speech. *Phonetician* 111–112, 54–69.

[12] Boersma, P., Weenink, D. 2015. *Praat: doing phonetics by computer*. Computer program.

[13] del Giudice, A., Shosted, R., Davidson, K., Salihie, M., Arvaniti, A. 2007. Comparing methods for locating pitch "elbows". *Proceedings of the 16th International Congress of Phonetic Sciences.* (Saarbrücken, Germany).

[14] Maechler, M. 2015. *Hartigan's Dip Test Statistic For Unimodality*. R Package version 0.75. https://cran.r-project.org/web/packages/diptest/index.html

[15] R Core Team. 2015. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

[16] Lindblom, B. 1963. Spectrographic study of vowel reduction, *J. Acoust. Soc. Am.* 35, 1773-1781.

[17] Jun, S.-A., Fougeron, C. 2000. A Phonological Model of French Intonation. In Botinis, A. (ed.), *Intonation: Analysis, Modelling and Technology*. Dordrecht: Kluwer, 209-242.

[18] Kuo, Y. C., Xu, Y., Yip, M. 2007. The phonetics and phonology of apparent cases of iterative tonal change in Standard Chinese. In Gussenhoven, C., Riad, T. (ed.), *Tones and Tunes Vol 2: Experimental Studies in Word and Sentence Prosody*. Berlin: Mouton de Gruyter, 211-237.

[19] Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. 2016. *lmerTest: Tests in Linear Mixed Effects Models*. R package version 2.0-30. https://CRAN.R-project.org/package=lmerTest.

[20] Ladd, D. R. 2008. *Intonational Phonology*. Cambridge: Cambridge University Press.

[21] Lim, L. 2004. Sounding Singaporean. In Lim, L. (ed.), *Singapore English: A Grammatical Description*. Amsterdam: John Benjamins, 19-56.