TITLE: The accentual phrase in Singapore English

AUTHORS: Adam J. Chong[1] and James Sneed German[2]

1. Department of Linguistics, University of California, Los Angeles, USA

2. Aix-Marseille Université, CNRS, LPL UMR 7309, 13100, Aix-en-Provence, France

ACKNOWLEDGEMENTS:

This study was funded by a grant from the Singapore Ministry of Education Academic Research Fund Tier 1 (2013-T1-002-169) and funds from the UCLA Dept. of Linguistics. This paper has greatly benefited from comments and discussion with Marc Garellek, Chad Vicenik, Sun-Ah Jun, Pat Keating, Megha Sundara, Bruce Hayes, members of the UCLA Phonetics Lab, and audiences at the 18th ICPhS in Glasgow. Additionally, we are grateful for the comments and discussion from an anonymous reviewer and an Associate Editor. We would also like to thank Audrey Kim, Eunice Liu, Sophia Sherman, Nigel Soriano and Sara Westcot for help with the processing of sound files and participant metadata, as well as Charmaine Hon & Talal Bin Amin at NTU for assistance with data collection.

ADDRESS FOR CORRESPONDENCE:

Adam J. Chong

Department of Linguistics

University of California, Los Angeles

3125 Campbell Hall

Los Angeles, CA 90095, USA

Email: ajchong@ucla.edu

Phone: +1 (310) 913 2857

28 Fax: +1 (310) 206 5743

ABSTRACT

This paper reports on a speech production experiment that explores whether the Accentual Phrase (AP) represents an abstract level of prosodic phrasing in Singapore English. Specifically, it tests whether the right edge of the AP is associated with phrase-final lengthening, the degree of which can be distinguished from lengthening associated with the Intonational Phrase (IP). Target words were produced in matched sentence contexts in three phrasal positions: AP-medial (word-final), AP-final, and IP-final. As predicted, target words in AP-final position were longer than those in AP-medial position and shorter than those in IP-final position. Analysis of target duration and f0 together shows that AP boundaries are well-discriminated from medial positions. Together, these results strongly support an AP level of phrasing for Singapore English and highlight its role in predicting timing variability.

1. Introduction

While phonological structure is by definition abstract, it can be detected indirectly through the phonetic and phonological variation that it induces. Evidence for a particular unit of abstract structure can be adduced from the finding that multiple independent phonetic measures tend to vary in unison, since this suggests a common hidden source of the variation. For Singapore English (SgE)[1], it has been proposed that variation in f0 is determined primarily by a unit of phonological structure called the Accentual Phrase (AP) (Chong, 2013). This unit groups together one or more lexical items (typically a content word plus any function words to its left) and is marked at its left and right edges by a low (L) and high (H) tone respectively. Combined with phonetic implementation rules, the analysis of a longer utterance in terms of a sequence of APs can explain the largely regular pattern of f0 rises and falls across an utterance. If a higher level unit of prosodic grouping (e.g., an intermediate phrase or intonational phrase) is assumed, such an analysis also explains the pattern of variation observed at stronger boundaries such as the end of an utterance or before a pause.

Thus far, only variation in f0 has been proposed as a phonetic correlate of the AP in Singapore English. Across languages, however, prosodic phrasing has been found to correlate with a range of phonetic measures, in particular the temporal lengthening of segments and syllables that lie at the right boundary of a phrase (Beckman & Edwards, 1990;

[1] Here we refer to a mostly standardized variety of English spoken in Singapore, roughly equivalent to the Standard Singapore English (SSE) discussed in the literature on diglossia (see for example, Gupta, 1989). All of the materials used in this study include only standard lexemes common to British, American and Singapore varieties. The selection and characteristics of our participant population are described in detail in Section 3.1.

Wightman, Shattuck-Hufnagel, Ostendorf & Price, 1992; inter alia). In this study, we explicitly test for the presence of pre-boundary lengthening as a phonetic correlate of the AP. This is important for several reasons. First, to the extent that lengthening occurs where we also observe the expected f0 correlates of the AP, this provides strong corroborating evidence for the presence of the AP as an abstract unit. Second, such evidence sheds light on how timing is determined in SgE. A few studies have sought to understand the prosodic organization of SgE in terms of rhythm-based metrics that estimate the overall amount of variation in timing across syllables (Low, Grabe & Nolan, 2000; Deterding, 2001; Grabe & Low, 2002). As Arvaniti (2009) has argued, however, such metrics are "unreliable predictors of rhythm" (p. 46) and they "cannot reflect the origins of the variation they measure and thus cannot convey an overall rhythmic impression" (p. 55). In short, she suggests that rhythm-based metrics only describe the superficial variation in timing in a language without the possibility of linking that variation to any particular feature of the underlying linguistic system. Arvaniti specifically cites phrasal position as one potentially important structural predictor of such variation.

The goal of our study is thus two-fold: in addition to corroborating the relevance of the AP as an abstract unit of prosodic structure, it also seeks to test whether the AP provides an explanatory source of timing variation across syllables in an utterance. Through a consideration of how the number and distribution of AP boundaries in an utterance predicts timing variability, the results can shed light on earlier findings which struggled to place SgE within a cross-linguistic spectrum of prosodic systems, especially in relation to rhythmic class (Deterding, 2001; Grabe & Low, 2002; Low et al., 2000).

2.1 Intonation in SgE

The intonational pattern of a typical declarative sentence in SgE involves a series of rises, each encompassing a single content word and any preceding function words. Such sentences typically end in a rise-fall pattern (Chong, 2013; Deterding, 1994; Lim, 2004). An example of this common intonational contour is shown in Figure 1. One notices f0 peaks that tend to coincide with the ends of content words, with an especially high peak on the first content word.

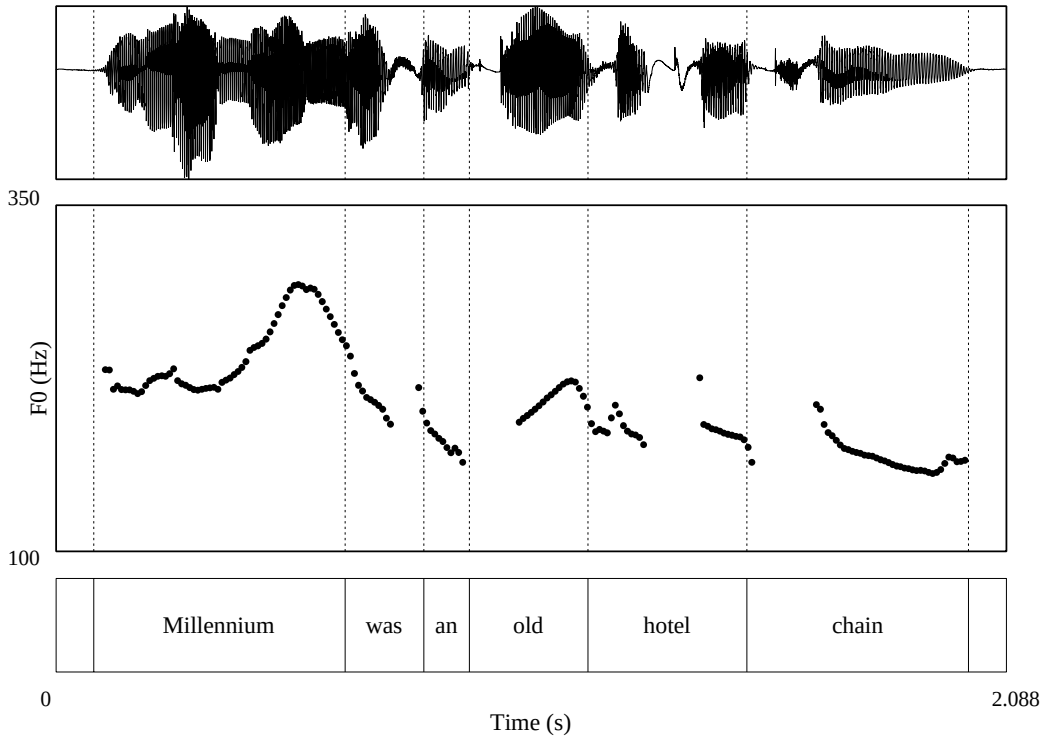

Figure 1. F0 contour on a standard declarative sentence: *Millennium was an old hotel chain*.

A number of previous studies have described the intonational system of SgE, primarily within phonetic frameworks of intonational transcription (Deterding, 1994, Low, 1994; Lim, 2004). Deterding (1994), in particular, presents the first quantitative investigation

of intonation in SgE, arguing that the domain of tone assignment, usually involving a rising tone, is a single stressed word rather than a specific syllable. In a more recent study, Ng (2011) sought to characterize how word prominence could be characterized by the realization of sequences of level tones on each syllable in a word. That study focused on the characterization of word-level prosody, particularly in terms of the relation between tone and stress. Her account, however, does not take sentence-level phrasing and context into consideration.

Chong (2013) analyzed SgE sentence-level intonation within the autosegmental-metrical framework (AM: Pierrehumbert, 1980; Beckman & Pierrehumbert, 1986). He argued for a *phonological* model in which tone alignment is largely edge-based (Jun, 2005) such that tones align to the edge of a phrasal domain larger than a prosodic word. This domain usually contains a single content word and any preceding function words. Chong also argued, building on Deterding's (1994) observations, that each phrasal unit is marked by a rising contour, with a high tone aligned to the right edge of a content word.

Across these different studies, several points of consensus emerge, two of which are relevant here. The first is the general characterization of tonal melody (i.e. a series of rises) of SgE declaratives, including the boosted pitch range of initial phrases (Chong, 2013; Deterding, 1994; Low 2000; Low & Brown, 2005). The second concerns the difficulty in identifying a prominent syllable, or nucleus, within phrases (Chong, 2013; Deterding, 1994; Lim, 2004; Low, 2000), a subject that we return to in the section 2.2. With the exception of Chong (2013), however, previous investigations have largely ignored the possibility that an utterance can be characterized in terms of hierarchically organized structure. While that study presents a preliminary phonological analysis, it did not provide quantitative evidence for the proposed model. Given this background, one major goal of the current paper then is

to provide quantitative data in support of one key aspect of Chong's phonological model regarding the existence of a level of prosodic structure above the lexical word, namely the Accentual Phrase.

2.2 Prominence and timing in SgE

The difficulty of distinguishing between stressed and unstressed syllables in SgE (Deterding, 1994; Tan, 2006) is a longstanding problem. For one, fundamental frequency does not seem to be a reliable cue to lexical stress (Tan, 2006; although see Ng, 2011) as is the case for British English. A related thread of investigations has sought to classify SgE's rhythm within the stress- and syllable-timed typology (Grabe & Low, 2002; Low, Grabe & Nolan, 2000; Ramus, Nespor & Mehler, 1999). The primary interest of these studies centers on the claim that SgE is a syllable-timed language (Bao, 2006; Deterding, 2001; Low & Brown 2005; Low et al, 2000; Platt & Weber, 1980; Tay, 1982; Tongue, 1979) compared to British English (BrE) which is considered a stress-time language. Quantitative studies investigating this typology usually utilize a measure, the "Pairwise Variability Index" (PVI), which captures the degree of variability in duration of successive syllables. A higher PVI is taken as characteristic of a stress-timed language, and a lower PVI a syllable-timed language. Across a number of studies, SgE has been found to have lower PVIs than other languages, in particular BrE, in both read (Low 1994, 1998; Low et al, 2000) and conversational speech (Deterding, 2001). Such findings support the notion that SgE is more syllable-timed, thus making lexical stress placement more difficult to pinpoint.

What all of these investigations lack, however, is a consideration of the phonological organization of the intonational system. Attempting to classify SgE as a stressed- or syllable-

timed language based on low-level phonetic detail, such as the PVI, overlooks the potentially important role of phonologically-driven variation. It is possible that durational variation is a consequence of any of a number of aspects of prosodic organization, and in particular the language's phrasal phonology. Moreover, it is also possible that the difficulty in identifying prominence may be due to a confluence of a number of factors. For one, it seems that while lexical "stress" from BrE may have been preserved at an abstract level, it has been remapped onto phonetic correlates in other ways (Tan, 2003, 2006; also see Chong & German, 2015).

Low and Grabe (1999) hint at the possible role of higher-level phrasal structure in their investigation of lexical stress placement in SE, examining whether or not lexical stress placement in SgE is truly different from that in BrE. Previous authors (e.g. Tay, 1982; Tongue, 1979) suggested that stress falls on the final syllable in SgE. Low and Grabe, however, point out that those observations were based on differences in position relative to the end of the intonation phrase (IP). In that positional context, acoustic cues to phrase boundaries and lexical stress are confounded. To address this confound, Low and Grabe (1999) conducted a production study in which both SgE and BrE speakers produced trisyllabic words (with the *–ly* suffix, e.g. *hopelessly*) in sentence-final and sentence-medial position. They then compared the durations of the final and penultimate syllables in target words in both phrasal positions. They found that the degree of phrase-final lengthening in IP-final position, as measured by differences in vowel duration between the penultimate and final syllable, was larger in SgE than in BrE. They found no durational differences, however, between the two varieties in IP-medial position. Low and Grabe also found that compared to BrE, SgE was characterized by smaller f0 differences between an initially stressed syllable and following unstressed syllable in IP-final position, but not in IP-medial position. The

authors argue that together, the smaller f0 differences and more substantial final lengthening in SgE contribute to the perception of final stress in SgE by BrE listeners.

While Low and Grabe's study highlights the potentially important role of phrasal structure in determining durational variation, it does not consider the effects of phrasal structure at levels of phrasing below the IP, nor does it address the implications that this class of effects has for the findings of rhythm-based approaches. If present, however, effects of lower-level phrasing are crucial for a general understanding of durational variability, since they influence a much higher proportion of syllables in each utterance than the IP-level does. Given recent evidence supporting the existence of the AP, in this paper we therefore consider durational differences across a wider range of phrasal contexts. In doing so, we adopt an approach that views existing rhythm-based findings for SgE as incidental to its phonological structure. In particular, we argue that most of the variability in duration can be explained by the density of phrasing units across speech samples and by the degree of lengthening that these induce on specific positions. By investigating the *phonological* organization SgE and its associated phonetic implementation rules, we believe that this provides a superior explanatory basis for comparing the prosodic system of SgE against those of other English varieties.

2.3 Singapore English intonation in a social context

"Singapore English" does not refer to a single language variety, since substantial and systematic variation exists both between speakers across different populations and within speakers across contexts of use. Most characterizations of variation in SgE focus on within-speaker variation; in other words, the set of linguistic features used by a given speaker varies

194 depending on the social context and the speaker's social objectives on any given occasion.
195 Traditionally, this situation has been characterized with reference to two subvarieties: a
196 "high", *acrolectal*, or standard-conforming variety that tends to be used in more formal
197 contexts, and a "low", *basilectal*, or colloquial variety that tends to be used in informal
198 contexts or to mark solidarity between speakers (Gupta, 1994; Platt, 1975, 1977; inter alia).
199 Depending on the researcher, these subvarieties may represent nearly discrete linguistic
200 systems in a diglossic situation (Gupta, 1994), or merely endpoints along a continuum (e.g.,
201 Platt 1975). There is general consensus, however, that individual speakers of Singapore
202 English typically command more than one variety. Differences between the varieties have
203 been described in terms of differences of lexis, morphology, syntax, the use of sentence-final
204 particles, or phonology, with discussion of phonology emphasizing segmental variation such
205 as the tendency to distinguish between long and short variants of vowels (esp., /i/ - /ɪ/) or
206 to produce [f] for the phoneme /θ/. No study that we are aware of has discussed prosodic
207 form with reference to the classic within-speaker subvarieties.
208 Research on between-speaker variation in SgE has emphasized differences among
209 three major ethnic groups (Chinese, Malay, Indian). At least a few studies have shown that
210 listeners can reliably identify the ethnicity of an individual from their speech (e.g., Deterding
211 & Poedjosoedarmo, 2000), though the findings of Deterding (2007) suggest that this is
212 unlikely to be due to segmental differences. Instead, a number of studies point to prosody
213 and intonation as the source of inter-ethnic differences. Tan (2010), for example, describes
214 contour shapes which appear to be linked to intonational features of speakers' mother
215 tongue languages. Additionally, Tan (2002, 2006) shows inter-group differences in how
216 sentence-level stress (nuclear and emphatic) is produced and perceived (see also Lim & Tan,

2001). While those studies found differences in how various phonetic correlates were prioritized, qualitatively the groups were very similar.

In our study, we nevertheless controlled for inter-ethnic differences by analyzing speakers from just one group, namely, ethnically Chinese speakers whose mother tongue is Mandarin. The homogeneity of our participant population was further ensured by the fact that participants were recruited from the student population of a university campus and fell in a rather narrow age range of just 9 years. Finally, the sentences in our materials involved only standard (i.e., SSE) lexical items, syntax and morphology. Given that the study used a reading task and was conducted in the rather formal context of university laboratory, it is likely that our speakers were producing a phonologically and phonetically standard variety. The systematicity and robustness of our results speaks to the fact that our study characterizes the prosodic system of a single variety, and we can be reasonably confident that this variety corresponds to what is most commonly referred to as Singapore Standard English (SSE).

2.4 The Current Approach

Through the present study, we seek first of all to accrue evidence for a particular abstract (i.e., phonological) unit of prosodic phrasing. We do this by testing explicitly whether pre-boundary lengthening occurs at the right edge of this unit, a location which we identify independently based on f0 correlates that have been previously linked to that unit. Following Chong (2013), we refer to this unit as the Accentual Phrase (AP). This nomenclature reflects the fact that the AP superficially resembles abstract units in other languages which have that name (esp. French (Jun & Fougeron, 2000, 2002) and Korean (Jun, 1996, 1998)), in that it

241 represents a level of phrasing larger than the word but smaller than the largest unit (i.e., the
242 intonational phrase). Our choice of terminology therefore does not represent any typological
243 claim that would have implications for the phonetic or phonological characteristics of the
244 AP beyond those which we present here.

245 Our predictions and analyses involve several assumptions which share much in
246 common with those of the Autosegmental-Metrical framework (Pierrehumbert, 1980;
247 Pierrehumbert & Beckman, 1988). These include, first of all, a distinction between tones,
248 which are abstract phonological units, and observable f0 characteristics of syllables, words
249 and utterances. These two levels of representation are related by a set of *realization rules*,
250 which are characteristics of a language-specific model of intonation. This is closely related to
251 the second assumption, namely, that syllables may be underspecified for tone. The fact that
252 realization rules can describe the behavior of f0 across toneless syllables makes this possible.
253 Finally, we assume that phrasing and tone assignment are *a priori* independent theoretical
254 choices. Any necessary relationship between them represents part of a language-specific
255 theory.

256 Finally, given the diversity of findings concerning the role of prominence in SgE, our
257 study promises to clarify certain issues concerning the explanatory source of stress as well as
258 rhythmic alternation in that variety. If significant lengthening occurs at the AP-level, then
259 this could explain why earlier studies reported "stress" occurring word-finally. Such a finding
260 would provide a stark contrast with Low and Grabe's (1999) claim that lengthening does not
261 occur internally to an IP. This could also explain why in some studies, SgE appears to be
262 difficult to classify rhythmically or at best argued to have "mixed" rhythm (Grabe & Low,
263 2002; see also Arvaniti 2009). In short, we hope through this study to improve the model of

SgE prosody by identifying the regularities in the system which best explain differences in duration from syllable to syllable and from word to word.

3. Methods

3.1 Participants

26 native speakers of Singapore English (15 male, 11 female) were recruited from the campus of Nanyang Technological University to participate in the study. All participants were undergraduate or graduate students at the university at the time of the study, with an age range of 20 to 28 years (mean = 24.2, SD = 2.4). All participants were residents of Singapore since birth, and were selected based on their self-identification as being ethnically Chinese. All reported having Mandarin as an official "mother tongue" language, which means they would have received substantial exposure to that language from a young age through education and possibly also in the home. Most reported having significant experience with at least one other language, typically including at least one other Chinese variety. 7 other participants were recruited but were excluded from the analysis either because they did not identify as being ethnically Chinese (n = 6) or because they produced overall disfluent tokens (n = 1).

3.2 Materials

In order to isolate the effects of phrasal position on lengthening, sets of target sentences were created such that individual target words appeared either (i) internally to an AP, (ii) at

the right edge of an AP, or (iii) at the right edge of both an AP and an IP boundary. Since nouns, verbs and other content words nearly always trigger the presence of an AP boundary, it is extremely rare for such words to occur internally to an AP. At the same time, function words like articles are virtually always phrased with a following content word. For this reason, it is not trivial to construct matched sets of target sentences that include target words in both positions. One exception appears to be certain types of functional heads, whose arguments may be omitted through processes of VP ellipsis (auxiliary verbs) or relative clause formation (prepositions). Examples of these two kinds of constructions are given in (1) and (2), respectively.

(1) a. He said he <u>will</u> go tomorrow.

b. He said he <u>will</u> tomorrow.

(2) a. He said that the prize was <u>for</u> Lin during dinner.

b. He said who the prize was <u>for</u> during dinner.

Since a phrase boundary usually occurs at the beginning of the following adverbial phrase (*tomorrow, during dinner*) regardless of whether the argument (*go, Lin*) is present, the relevant function word will occur at a boundary when the argument is omitted. Typical phrasing patterns for the sentences in (1) are illustrated in (3a) and (3b). Furthermore, when the argument and the adverbial phrase are omitted, as in (3c), the function word will fall at the end of the utterance and therefore also at an IP boundary.

(3) a. *AP-medial (word boundary)*

He said he will go tomorrow

( )$_{AP}$ ( )$_{AP}$ ( )$_{AP}$

( )$_{IP}$

b. *AP boundary*

He said he will tomorrow

( )$_{AP}$ ( )$_{AP}$ ( )$_{AP}$

( )$_{IP}$

c. *IP boundary*

He said he will

( )$_{AP}$ ( )$_{AP}$

( )$_{IP}$

Sets of three carrier sentences were created for each of 18 target words following the model in (3), such that each target word occurred in all three positional contexts. There were a total of 30 items with auxiliaries (10 target words) and 24 with prepositions (8 target words), for a total of 54 experimental items. This study was conducted in parallel with a study exploring stress in utterance-initial words. 60 additional sentences from that study, which involved no VP ellipsis or preposition stranding, were therefore included in the study and served as distractors.

335 3.3 Procedure

336

337 During the experiment, participants were seated in a sound-attenuated booth. Sentences
338 were presented on a computer screen one at a time, and subjects were instructed to say each
339 sentence aloud in a conversational style "as though talking to a friend". Target sentences
340 were preceded by a short context consisting of a single question (see Appenix A)[2], and
341 participants were asked to read both the context and the target silently before reading the
342 targets aloud. This step allowed us to control the pattern of focus and specifically to avoid
343 contrastive focus on words in the target region[3]. Item presentation was self-paced, and
344 participants were permitted to produce each sentence more than once in case of speech
345 errors or hesitations. Stimulus presentation was controlled using E-Prime (Psychology
346 Software Tools, Pittsburgh, PA). All experimental and distractor items were divided into five
347 approximately equal-sized and counterbalanced blocks. Between- and within-block
348 randomization was carried out separately for each participant using E-prime's built-in
349 randomization function. This procedure was done twice such that participants had two
350 repetitions of the task.

351

352

353 4. Results

354

[2] One item (Item 18, Appendix A) was not presented with a context due to an oversight.
[3] The contexts were designed to either (a) reinforce a broad focus reading in the case of the preposition targets, or (b) render the target region (i.e., for (3), *he will*, *he will go*) as *given* in the case of the auxiliary targets. However, Lim (2004), Low (2006) and others have noted a lack of a prosodic marking difference between given and new information in SgE. For this reason, and because context was constant across conditions, it is not expected to play a role in our study.

Recordings of target sentences were segmented and extracted automatically using Praat. Phonetic segmentation and labeling was carried out automatically using the SPPAS force-alignment tool (Bigi, 2015), and subsequently, the target regions were checked for alignment errors and corrected manually using visual inspection of the spectral and intensity characteristics. The target regions, including the target word and target nucleus, were labeled manually in accordance with the phonetic labeling. All utterances were then verified for naturalness by one of the authors, who is a native speaker of SgE. Finally, the two measures of interest were extracted from the vowel nucleus of each target automatically using Praat: duration and mean F0. In the case of the one disyllabic target word, *about*, only the vowel nucleus in the second and final syllable was measured.

In the analyses that follow, we assess the differences between mean duration values as well as f0 values with linear mixed-effects models using the *lmer()* function in the *lme4* package (Bates, Maechler, Bolker & Walker, 2015) in R (R Core Development Team). Statistical significance of fixed factors was determined by likelihood ratio tests using the *anova()* function in R, in which a model with the particular factor is compared to one without (all else being constant). We only analyzed participants' productions from the second repetition of the task, where productions were overall much more fluent. There were a total of 1404 possible tokens (26 speakers X 18 words X 3 phrasal contexts). 39 were excluded due to recording difficulties, leaving a total of 1365 tokens for analysis.

4.1 Phrasal position and nucleus duration

Nucleus duration of target words by different phrasal positions (word vs. AP vs. IP) are shown in Fig. 2. The linear mixed effects model included phrasal position as a fixed factor

(reference group = AP) as well as random intercepts for subject and item, with random slopes by phrasal position for each of the random intercepts. The main effect of phrasal position was significant ($\chi^2(2) = 29.12$, $p < 0.001$). Post-hoc pair-wise comparisons were conducted using the *glht()* function from the *multcomp* package (Horton, Bretz & Westfall, 2008). The full model results are shown in Table 1. These indicate that target nuclei were significantly longer in AP position compared to word position ($p < 0.001$), but significantly shorter in AP position than in IP position ($p < 0.001$). Not surprisingly, target nuclei in IP position are also significantly longer than those in word position ($p < 0.001$). Thus, target nuclei in AP position have longer durations than those in word position, but shorter duration than in IP position, revealing the presence of durational correlates for at least two levels of phrasal structure above the prosodic word.

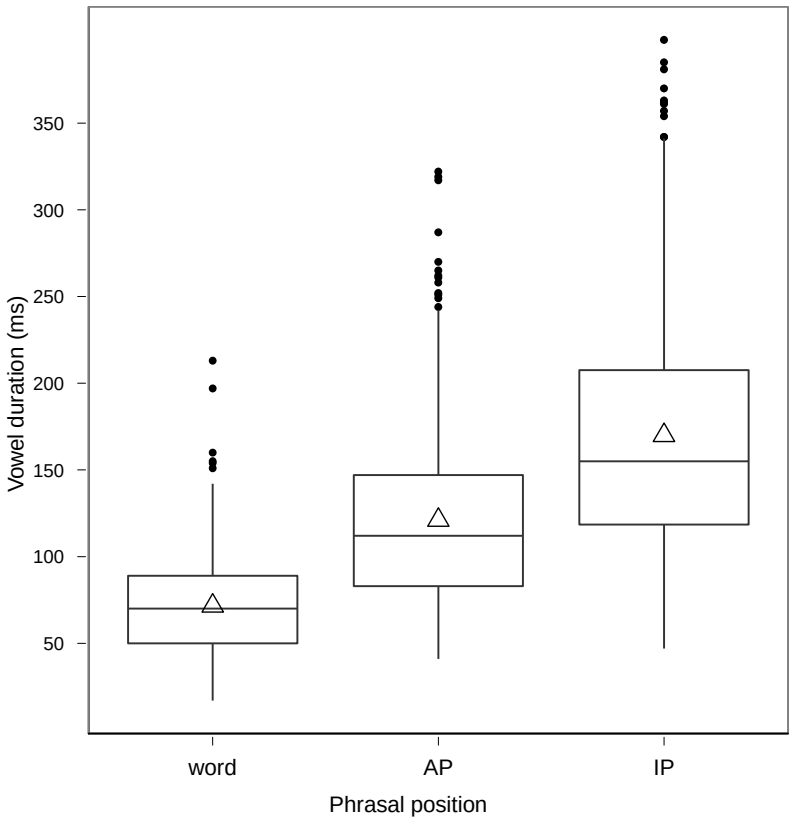

Figure 2. Boxplot of nucleus duration by phrasal position. Triangles indicate mean values.

Table 1. Model results from multiple comparisons of nucleus duration by phrasal position

	Estimate	Std. Error	z value	Pr(>|z|)	
AP vs. Word	-0.05	0.01	-7.24	< 0.001	***
AP vs. IP	0.06	0.01	6.07	< 0.001	***
Word vs. IP	0.12	0.02	7.62	< 0.001	***

4.2 Phrasal position and f0

The phonological model in Chong (2013) posited that APs are marked by a H tone, thus this predicts independently of durational differences that AP-final target words should have a higher f0 than non-AP final target words. In order to confirm our assumptions regarding the expected locations of AP boundaries in our materials, we therefore tested whether f0 differed between targets in word-final vs. AP-final position. To test this, a linear mixed effects analysis was applied to a subset of the overall data, including f0 measurements in word and AP positions. 12 more tokens were not analyzed due to poor f0 tracking, leaving a total of 894 tokens for analysis. These models include mean f0 as the dependent variable and phrasal position (reference = AP) as a fixed factor, as well as random intercepts for subjects and items as well as random slopes for phrasal position. The effect of phrasal position was significant ($\chi^2(1) = 29.56$, $p < 0.001$) such that mean f0 of target nuclei was higher in AP position (mean = 162.27, SD = 22.68) than word position (mean = 150.50, SD = 21.53). The full model results are presented in Table 2. While the overall amount of variance can be

explained by large inter-speaker differences in f0 range, the model estimate suggests that AP-final targets were approximately 12 Hz higher than AP-medial targets. This small effect size is consistent with the observation that strong f0 range compression occurs after the first AP in an utterance (Chong, 2013; Deterding, 1994; Low, 2000). Thus, the result of the f0 comparison supports our assumption that AP boundaries occur at the expected locations.

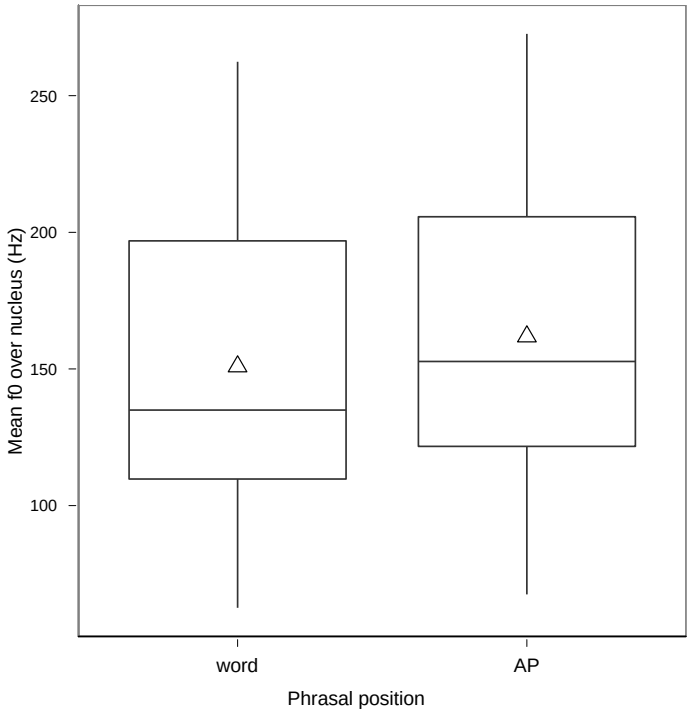

Figure 3. Boxplot of mean f0 of nucleus by phrasal position. Triangles indicate mean values.

Table 2. Model results for mean F0 by phrasal position.

	Estimate	Std. Error	*t*-value
Intercept	162.827	9.201	17.697
Phrasal position (= Word)	-11.847	1.651	-7.176

4.4 Duration and f0 together

In many prosodic systems, the distinction between different boundary categories are not found in any single phonetic parameter, but are manifest through a combination of parameters (e.g. Streeter, 1978; Price, Ostendorf, Shattuck-Hufnagel & Fong, 1991). If the distinction between word-level and AP-level boundaries represents a perceivable categorical distinction in SgE, then we can expect the two categories to be well-separated in a space defined by some combination of their various phonetic correlates. To test this, we first plotted the tokens from our corpus in a two dimensional plane defined by nucleus duration and f0. Since large interspeaker differences give rise to substantial overlap in the distributions, we used the speaker-standardized z-scores of both duration and f0 for each token. Figure 4 shows that the word-level and AP-level tokens are distributed in two large clusters with a moderate amount of overlap, suggesting that they represent two distinct categories.

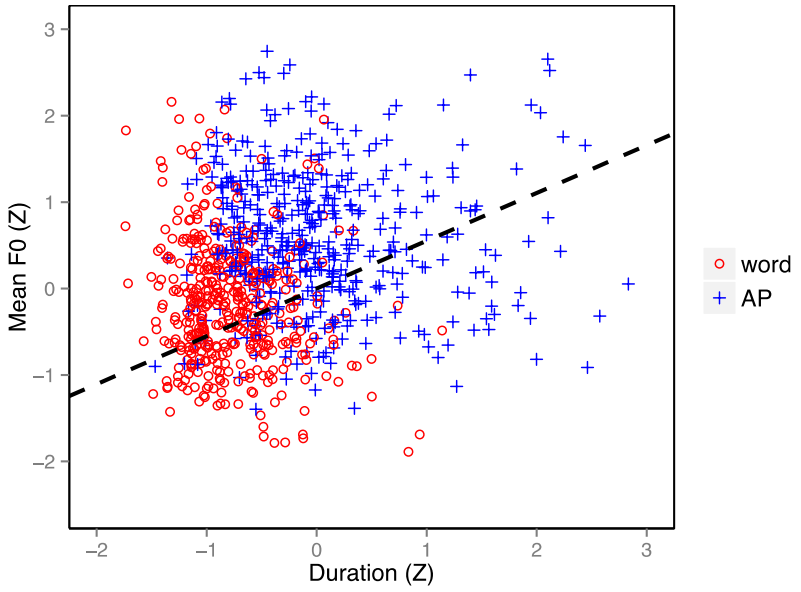

Figure 4. Scatterplot of nucleus duration and nucleus mean f0 by boundary type. All values are standardized by speaker (z-score). The dashed line represents the eigenvector of the linear discriminant.

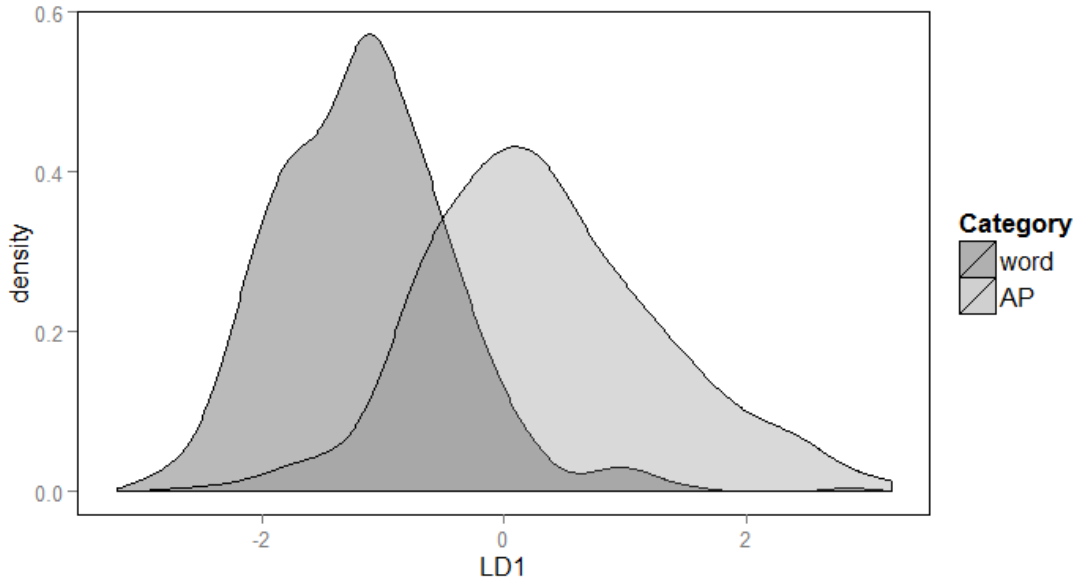

Figure 5. A density plot of the orthogonal projection of the datapoints in Fig. 4 onto the linear discriminant (or equivalently, its eigenvector).

We explored this issue quantitatively using linear discriminant analysis (LDA), which is an unsupervised machine learning method that provides an objective measure of differentiation between two or more groups. For different groups occuring in some feature space of one or more dimensions, it identifies a set of linear combinations of features which maximizes the between-group means while minimizing the within-group variances. In the two-dimensional case, this method seeks a single vector combination of the two axis parameters (i.e., duration and f0) which accomplishes this. The resulting linear discriminant can be used to assess the goodness of separation of the groups by comparing the predicted group membership of each observation based on the linear discriminant against its actual group membership, yielding an overall accuracy score, where higher accuracy scores are associated with better separation of the groups. In short, this method allowed us to assess whether word- and AP-boundaries are well-separated in terms of nucleus duration, nucleus mean f0, or both. A leave-one-out cross-validation was used to assess the accuracy of the resulting discriminants. 12 tokens for which f0 was unavailable were excluded from the analysis.

When both duration and f0 were included in the model, the predictive accuracy of the resulting linear discriminant was 81.5%. A binomial test confirmed that this rate is significantly different from chance, which is 50.6% ($p<0.0001$). The dashed line in Figure 4 represents the eigenvector of the linear discriminant, which means that the two groups had the best separation when projected onto a line having this particular slope ($m = 0.553$). Figure 5 shows the distribution of the two groups after each datapoint is projected orthogonally onto this line.

A one-dimensional generalization of LDA can reveal how well-separated the groups are along specific phonetic parameters. When either duration or f0 alone are used, the

predictive accuracies of the associated discriminant models are 74.5% and 70.2%, respectively. While these values are significantly different from chance, the fact that they are lower than 81.5% shows that the two phonetic parameters contribute jointly to the separation of the groups. In other words, duration and f0 work complementarily to distinguish word-level from AP-level boundaries. This is also reflected in the fact that the linear discriminant of the two-dimensional model has an intermediate slope value (i.e., $0 < m < 1$). Together, these results strongly support (i) the presence of two different categories for the word and AP data, and (ii) the fact that this distinction is manifest through a combination of both f0 and duration.

5. Discussion

Our results revealed robust changes in duration at locations corresponding to the right boundary of the AP. Thus, our findings provide strong support for the contribution of the AP to variability in duration across syllables in SgE. The three-way difference between AP-internal, AP boundary, and IP-boundary positions provides strong corroborating evidence for the fact that the AP represents an abstract level of prosodic organization intermediate to the word-level and IP-level. This key aspect of the phonological model was further confirmed by our finding that duration and f0 contribute jointly to a categorical distinction between word-level and AP-level boundaries.

Our study follows recent investigations of the intonational systems of new Englishes within an Autosegmental-Metrical framework (see Gussenhoven 2015 for an overview), without imposing a BrE intonational structure onto SgE (see Tan, 2006, Lim 2009). The focus in these studies has predominantly been on characterizing the tonal inventory and

tonal association rules in new English varieties, particularly what some have argued to be tonal varieties (Gussenhoven 2014, also see Lim 2009). Our current investigation differs from this thread in focusing primarily on prosodic constituency (vs. intonational melody) in investigating durational correlates to the Accentual Phrase. This level of structure has yet to be suggested for other English varieties.

The finding that the AP is a major determinant of variability in duration has important implications for the analysis of timing and rhythm in SgE. As an abstract unit that is realized through multiple, independent phonetic events, the AP provides an explanatory basis for timing differences from syllable to syllable and across utterances. Given a reasonably detailed account of how AP boundaries are distributed based on syntax and other structural influences, it is possible to estimate differences in syllable timing based on the textual content of an utterance.

Given similarly detailed models for different language varieties, it is also possible to predict rhythmic differences among those varieties directly from their respective phonological structure. Specifically, the tendency for syllable duration to alternate can be estimated from how densely distributed the phonological positions are that give rise to lengthening, as well as from the degree of lengthening that is typically contributed by those positions. For BrE, the relevant positions are (i) syllables that are lexically marked for stress (including primary and secondary stress), (ii) the subset of lexically stressed syllables that also bear pitch accents, and (iii) syllables near an intermediate phrase boundary. For SgE, the relevant positions are primarily syllables occurring at an AP boundary, with the possibly weaker role of lexically-determined stress still to be determined. Although AP boundaries in SgE are more densely distributed than ip boundaries in BrE, the combined distribution in BrE of ip boundaries, lexical stress, and pitch accents taken together is expected to be much

519 denser. This difference would then give rise to a much higher degree of inter-syllable timing
520 variability, and can therefore explain the finding that BrE has scores higher than SgE on the
521 various variability metrics.

522 The density of APs within a larger IP also gives rise to a stronger macro-rhythm (Jun
523 2005, 2015), a parameter of phrasal prosody which refers to the perceived rhythm due to
524 changes in f0. Jun's (2015) parameter of macro-rhythm is an attempt to capture global
525 phase-medial tonal patterns that are independent of the type of prominence marking (head
526 vs. edge-prominence), and crucially is defined purely in terms of tonal alternations without
527 reference to smaller prosodic units like syllables or feet. Because each content word or AP is
528 marked tonally, and each AP is generally marked with the same tonal contour (rising), SgE
529 achieves a strong macro-rhythm in comparison with other inner-circle varieties of English
530 (e.g. BrE or American English; see Jun 2015), which often have a larger set of possible
531 phrase-medial pitch accents as well as less evenly spaced phrasal units.

532 Our results contrast somewhat with those of Low and Grabe (1999), which did not
533 reveal phrase-final lengthening in IP-medial positions. Based on the lexical and syntactic
534 composition of their materials, most of these target positions were very likely AP-final.
535 However, the design of that study only allowed for comparison of neighboring syllables that
536 were in the same word (i.e., the penultimate vs. the final syllable). This means, first of all,
537 that the role of AP-finality and word-finality were confounded. Second, the syllables being
538 compared were not identical in terms of segmental composition, a fact which may have
539 contributed a substantial source of noise in the study. By contrast, our study avoided these
540 issues by isolating the role of AP-finality and controlling for identical segmental structure
541 across the syllables being compared.

Finally, we wish to comment on the extent to which our results are expected to generalize to other ethnic varieties of SgE as well as to more naturalistic conversational speech. We have chosen here to focus on the speech produced by ethnically Chinese speakers of SgE, in particular university-educated Chinese SgE speakers. While one possible concern is that our findings do not comprehensively represent that variety as a whole, a number of factors make this unlikely. First of all, ethnic Chinese make up 74.1% of the country's population (Singapore Census, 2010), and in that` sense represent the "dominant" variety of SgE. Second, although Tan (2010) has suggested that SgE speakers of different ethnicities show differences in the tonal shapes of their utterances, the substantial contact that occurs between these groups, as well as the necessity for a high degree of inter-intelligibility between them makes it relatively unlikely that there are broad differences in the overall phonological organization. Moreover, our full set of collected recordings included three ethnically Indian speakers, one Malay speaker, and two Eurasian (mixed-ancestry) speakers. Preliminary examination of this data reveals a very similar pattern of results to that of the Chinese-only subset presented here.

Given that our data was read laboratory speech, we are confident that our findings reflect general facts about this more formal register, and we acknowledge that the extent to which they generalize to more naturalistic conversational speech and to Colloquial Singapore English will need to be determined by further studies. In spite of this, our data show that the prosodic system of SgE differs typologically in terms of overall phonological organization from other more "standard" varieties that might be available to the speakers. Moreover, the qualitative generalizations regarding the rising contours across words fit closely with previous descriptions of tonal contours in Colloquial Singapore English (Deterding, 1994; Lim, 2004; Ng, 2011), which suggests that the two varieties share a common phonological

566 organization.

567

568 ## 6. Conclusion

569

570 The driving questions of this study concerned the number and nature of abstract phrasing
571 levels in SgE and how those contribute to timing variability. Evidence for at least two levels
572 above the word highlights the need for a shift in the approach to cross-varietal comparisons
573 of timing variability. Our study did not test for additional levels of phrasing, such as the
574 intermediate phrase, though we do not rule out the possibility that more exist. If present,
575 further studies will establish whether these other levels are also realized through f0 and/or
576 timing differences. Given the stress-like nature of the AP-final position, an important
577 question that remains concerns the extent to which lexical stress is phonetically realized in
578 SgE, and how this interacts with phrasal position.

579

580

581 References

582

583 Arvaniti A (2009): Rhythm, timing and the timing of rhythm. Phonetica 66: 46-63.

584 Bates D, Maechler M, Bolker B, Walker S (2015): Fitting linear mixed-effects models using
585 lme4. Journal of Statistical Software 61: 1-48.

586 Beckman ME, Edwards J (1990): Lengthenings and shortenings and the nature of prosodic
587 constituency; in Kingston J, Beckman ME (eds.): Papers in Laboratory Phonology:
588 Between the grammar and the physics of speech. Cambridge, Cambridge University Press,
589 pp. 1-16.

590 Beckman M, Pierrehumbert J (1986): Intonational structure in Japanese and English.
591 Phonology Yearbook 3: 255-309.

592 Bao Z (2006): Clash avoidance and metrical opacity in Singapore English. STUF-
593 Sprachtypologie und Universalienforschung 59: 133-147.

594 Bigi, B (2015): SPPAS - Multi-lingual Approaches to the Automatic Annotation of Speech.
595 The Phonetician: International Society of Phonetic Sciences 111-112 / 2015-I-II: 54-69.

596 Chong AJ (2013): Towards a model of Singaporean English intonational phonology.
597 Proceedings of Meetings on Acoustics 19.

598 Chong AJ, German, JS (2015): Prosodic phrasing and F0 in Singapore English. Proc 18th
599 ICPhS, Glasgow.

600 Deterding D (1994): The intonation of Singapore English. J Int Phon Assoc 24: 61-72.

601 Deterding D (2001): The measurement of rhythm: A comparison of Singapore English and
602 British English. J Phonet 29: 217-230.

603 Deterding D (2007): The vowels of the different ethnic groups in Singapore; in Prescott D
604 (ed.): English in Southeast Asia: Varieties, Literacies and Literatures. Newcastle:
605 Cambridge Scholars Publishing, pp. 2-29.
606 Deterding D, Poedjosoedarmo GR (2000): To what extent can the ethnic group of young
607 Singaporeans be identified from their speech? in Brown A, Deterding D, Low E-L (eds.):
608 The English Language in Singapore: Research on Pronunciation. Singapore, Singapore
609 Association for Applied Linguistics, pp. 1-9.
610 Singapore Department of Statistics (2010): Census of population 2010. Statistical release 1:
611 Demographic characteristics, education, language and religion. Retrieved from:
612 http://www.singstat.gov.sg/publications/publications-and-
613 papers/cop2010/census10_stat_release1.
614 Grabe E, Low E-L (2002): Acoustic correlates of rhythm class; in Gussenhoven C, Warner
615 N (eds.): Laboratory Phonology, vol. 4(1). Berlin, Walter de Gruyter, pp. 515-546
616 Gupta AR (1989): Singapore Colloquial English and Standard English. Singapore Journal of
617 Education 10: 33-39.
618 Gupta AR (1994): The Step-tongue: Children's English in Singapore. Clevedon, Multilingual
619 Matters.
620 Gussenhoven C (2014): On the intonation of tonal varieties of English, in Filppula M,
621 Klemola J, Sharma D (eds.): The Oxford Handbook of World Englishes online.
622 Publication Date: Dec 2014. DOI: 10.1093/oxfordhb/9780199777716.013.29.
623 Horton T, Bretz F, Westfall P (2008): Simultaneous Inference in General Parametric Models.
624 Biometrical Journal 50: 346-363.
625 Jun S-A (1996): The Phonetics and Phonology of Korean Prosody: Intonational phonology
626 and prosodic structure. New York, Garland Press.

627 Jun S-A (1998): The accentual phrase in Korean prosodic hierarchy. Phonology 15: 189-226.
628 Jun S-A (2005): Prosodic Typology; in Jun S-A (ed.): Prosodic Typology: The phonology of
629 intonation and phrasing. Oxford: Oxford University Press, pp. 430-458.
630 Jun S-A, Fougeron C (2000): A phonological model of French intonation; in Botinis A (ed.):
631 Intonation: Analysis, Modelling and Technology. Dordrecht, Kluwer Academic Publishers,
632 pp. 209–242.
633 Jun S-A, Fougeron C (2002): Realizations of accentual phrase in French intonation. Probus
634 14: 147–172.
635 Lim L (2004): Singapore English: A grammatical description. Amsterdam, John Benjamins.
636 Lim L (2004): Sounding Singaporean, in Lim L (ed.): Singapore English: A grammatical
637 description. Amsterdam, John Benjamins, pp. 19-56.
638 Lim L (2009): Revisiting English prosody: (Some) new Englishes as tone languages? English
639 World-Wide 30: 218-239.
640 Lim L, Tan YY (2001): How are we stressed?! Phonetic correlates and stress placement in
641 Singaporean English, in Maidment, JA, Estebas i Vilaplana E. (eds.): Proceedings of
642 Phonetics and Teaching Learning Conference 2001. London, University College London,
643 pp. 27-30.
644 Low E-L (1994): Intonation patterns in Singapore English; MPhil thesis, University of
645 Cambridge.
646 Low E-L (1998): Prosodic prominence in Singapore English; PhD thesis, University of
647 Cambridge.
648 Low E-L (2000): Is lexical stress placement different in Singapore English and British
649 English? in Brown A, Deterding D, Low E-L (eds.): The English language in Singapore:

650 Research on pronunciation. Singapore: Singapore Association for Applied Linguistics, pp.
651 22-34.
652 Low E-L (2006): A cross-varietal comparison of deaccenting and given information:
653 Implications for international intelligibility and pronunciation teaching. TESOL Quarterly
654 4: 739-761.
655 Low E-L, Brown A (2005): English in Singapore: An introduction. Singapore, McGraw Hill.
656 Low E-L, Grabe E (1999): A contrastive study of prosody and lexical stress placement in
657 Singapore English and British English. Lang Speech 42: 39-56.
658 Low E-L, Grabe E, Nolan F (2000): Quantitative characteristics of speech rhythm: Syllable-
659 timing in Singapore English. Lang Speech 43: 377-401.
660 Ng E-C (2011): Reconciling stress and tone in Singaporean English. in Zhang L, Rubdy R,
661 Alsagoff L (eds.): Asian Englishes: Changing perspectives in a globalised world. Singapore:
662 Pearson Longman, pp. 48-59.
663 Pierrehumbert JB (1980): The phonology and phonetics of English intonation; PhD
664 dissertation, Massachusetts Institute of Technology.
665 Platt J (1975): The Singapore English Speech continuum and its basilect 'Singlish' as a
666 'creoloid'. Anthropological Linguistics 17: 363-374.
667 Platt J (1977): A model for polyglossia and multilingualism (with special reference to
668 Singapore and Malaysia). Language in Society 6: 361-378.
669 Platt J, Weber H (1980): English in Singapore and Malaysia. Status, features functions. Kuala
670 Lumpur, Oxford University Press.
671 Price PJ, Ostendorf M, Shattuck-Hufnagel S, Fong C (1991): The use of prosody in syntactic
672 disambiguation. J Acoust Soc Am 90: 2956-2970.
673 Psychology Software Tools Inc. (2012): E-Prime 2.0. Retrieved from http://www.psnet.com

674 R Core Team (2015): R: A language and environment for statistical computing. R
675 Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-
676 project.org/
677 Ramus F, Nespor M, Mehler J (1999): Correlates of linguistic rhythm in the speech signal.
678 Cognition 72: 1 –28.
679 Streeter L (1978): Acoustic determinants of phrase boundary perception. J Acoust Soc Am
680 64: 1582-1592.
681 Tan YY (2003): Acoustic and perceptual properties of stress in the ethnic subvarieties of
682 Singapore English, Ph. D. dissertation, National University of Singapore.
683 Tan YY (2006): Is the stressed syllable stressed? The perception of prominence in Singapore
684 English; in Hashim A, Hassan N (eds.): Varieties of English in Southeast Asia and Beyond.
685 Kuala Lumpur, University of Malaya Press, pp. 113-152.
686 Tan YY (2010): Singing the same tune? Prosodic norming in bilingual Singaporeans; in
687 Ferreira MC (ed.): Multilingual Norms. Frankfurt, Peter Lang, 173-194.
688 Tay MWJ (1982): The phonology of educated Singapore English. English World Wide 3:
689 135-145.
690 Tongue RK (1979): The English of Singapore and Malaysia. Singapore, Eastern Universities
691 Press.
692 Wightman CW, Shattuck-Hufnagel S, Ostendorf F, Price PJ (1992): Segmental durations in
693 the vicinity of prosodic phrase boundaries. J Acoust Soc Am 91: 1707-1717.

694
695

Appendix A: Experimental sentences (target words are underlined)

1. Prompt: *Will Sam go shopping?*
    a. Word: He said he <u>will</u> go tomorrow.
    b. AP: He said he <u>will</u> tomorrow.
    c. IP: He said he <u>will</u>.

2. Prompt: *Can you stay?*
    a. Word: I think I <u>can</u> stay for a while
    b. AP: I think I <u>can</u> for a while
    c. IP: I think I <u>can</u>.

3. Prompt: *Was Elaine ill?*
    a. Word: She said she <u>was</u> ill this morning.
    b. AP: She said she <u>was</u> this morning
    c. IP: She said she <u>was</u>.

4. Prompt: *Were Arsenal losing?*
    a. Word: I think they <u>were</u> losing before Ramsey scored.
    b. AP: I think they <u>were</u> before Ramsey scored.
    c. IP: I think they <u>were</u>.

5. Prompt: *Has Esther asked for help?*
    a. Word: I think she <u>has</u> asked already.

720 b. AP: I think she has already.

721 c. IP: I think she has.

722

723 6. Prompt: *Did Peter say if I should visit?*

724 a. Word: He said that you should visit next week.

725 b. AP: He said that you should next week.

726 c. IP: He said that you should.

727

728 7. Prompt: *Have you two met?*

729 a. Word: I think we have met before.

730 b. AP: I think we have before.

731 c. IP: I think we have.

732

733 8. Prompt: *Will Daryl come to the party?*

734 a. Word: He said that he might come later.

735 b. AP: He said that he might later.

736 c. IP: He said that he might.

737

738 9. Prompt: *Is the porridge stall closing?*

739 a. Word: I heard that it is closing next month.

740 b. AP: I heard that it is next month.

741 c. IP: I heard that it is.

742

743 10. Prompt: *Are the children finished drawing?*

744 a. Word: They said they <u>are</u> finished for now.

745 b. AP: They said they <u>are</u> for now.

746 c. IP: They said they <u>are</u>.

747

748 11. Prompt: *Why is Dan upset?*

749 a. Word: He knows that Ernest is <u>with</u> Mel at the party.

750 b. AP: He knows who Ernest is <u>with</u> at the party.

751 *c.* IP: He knows who Ernest is <u>with</u>.

752

753 12. Prompt: *What happened last night?*

754 a. Word: They said that the prize was <u>for</u> Lin during dinner.

755 b. AP: They said who the prize was <u>for</u> during dinner.

756 c. IP: They said who the prize was <u>for.</u>

757

758 13. Prompt: *Why is Charmaine so happy?*

759 a. Word: She found out that the visitors are <u>from</u> France just now.

760 b. AP: She found out where the visitors are <u>from</u> just now.

761 *c.* IP: She found out where the visitors are <u>from</u>.

762

763 14. Prompt: *What happened in Lit (literature) class today?*

764 a. Word: We learned that the poem is <u>about</u> love in tutorial.

765 b. AP: We learned what the poem was <u>about</u> in tutorial.

766 c. IP: We learned what the poem was <u>about</u>.

767

15. Prompt: *Where are the presentation files?*

    a. Word: Sam asked if he should send them to Jill by email.

    b. AP: Sam asked who he should send them to by email.

    c. IP: Sam asked who he should send them to.

16. Prompt: *I fell asleep. What happened in the movie?*

    a. Word: Joker revealed that the trap was for Batman at the end.

    b. AP: Joker revealed who the trap was for at the end.

    c. IP: Joker revealed who the trap was for.

17. Prompt: none

    a. Word: Beth has wondered if the poem is by Kipling for awhile.

    b. AP: Beth has wondered who the poem is by for awhile.

    c. IP: Beth has wondered who the poem is by.

18. Prompt: *What did Sue want?*

    a. Word: She asked if the photos were in boxes last time.

    b. AP: She asked what the photos were in last time.

    c. IP: She asked what the photos were in.