**TITLE:** Prominence and intonation in Singapore English

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

*Corresponding author: a.chong@qmul.ac.uk, Tel: +44 20 7882 6638

1. Queen Mary University of London, London, UK

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

**Abstract**

Previous work on Singapore English prosody has focused largely on establishing the acoustic correlates of lexical stress and examining where the language falls within a rhythm-class typology. Little attention, however, has been paid to how lexical prominence, if present, interacts with phrasal prominence. In this study, we examine the extent to which f0 realizations vary across lexical items with differing stress patterns, while taking into account that prosodic phrasing requirements necessitate an f0 rise to the phrase-final syllable. We show that across target types of varying stress placement, syllable length, and constituency, f0 realizations are highly consistent, involving a rise from the start of the target word or phrase which culminates with a peak on the phrase-final syllable. The location of lexical prominence is the primary influence on the scaling of f0 across the entire target, with stress-initial targets having a higher mean f0. Exploratory analysis of duration and intensity measures further corroborates the prominence-lending nature of the phrase-final syllable, with some evidence for marking of prominence on non-final lexically stressed syllables. The findings support the primarily post-lexical role that f0 plays in marking phrase edges, instead of lexical heads, in Singapore English, in line with a previously proposed AM model of Singapore English intonation. The implications of these

findings for the study of prosodic typology and sociolinguistic variation in Singapore English are also discussed.

**Keywords:** Singapore English, intonation, prosodic typology, lexical stress, prominence, edge-marking, Autosegmental Metrical Theory

**Prominence and intonation in Singapore English**

## 1. Introduction

Research on English intonation has primarily focused on the examination of 'Mainstream' English varieties spoken in North America or the British Isles, and closely related varieties in South Africa and Australia/New Zealand (Grice, German & Warren 2020). The intonation systems of "Contact" or "New" varieties of English, however, have received comparably less attention. In this article, we follow Grice et al. (2020) in drawing a distinction between Mainstream and Contact varieties of English. As noted by Grice et al., the former not only pattern closely together typologically in terms of the general phonological characteristics of their intonation systems (e.g., presence of postlexical stress accents and a two-tiered phrasing hierarchy), but they also have received greater attention both in terms of description and in terms of their role in shaping various theoretical frameworks. The latter have in common a history of close contact with varieties that are typologically distinct from Mainstream varieties, and consequently they present as a group a much higher degree of typological diversity in their intonation systems. While this distinction closely resembles that of Inner versus Outer Circle varieties proposed by Kachru (1985), we agree with Grice et al. that the terminology we adopt more closely reflects the emphasis of our study on synchronic structural aspects of those varieties as opposed to their historical context.

"New" varieties of English present rather distinct intonational systems from Mainstream varieties. Due to their development through language contact in often highly multilingual environments, these varieties often show significant influence from local indigenous languages, often developing hybrid systems that incorporate prosodic features

of both British English and the indigenous contact languages (e.g. Maltese English: Vella 1994; Indian English: Fuchs 2016, Maxwell 2014, Maxwell & Payne 2021). Some varieties like Nigerian and Ghanaian English have also been argued to be tone languages with the edges of lexical words being marked by tone (Gussenhoven 2017, Gussenhoven & Udofo 2010, Gut 2005). Further, these systems themselves often exist synchronically within communities that have high degrees of multilingualism. Maxwell and Payne (2021), for example, found that Indian English speakers from different L1 backgrounds showed simultaneous convergence on some prosodic features (e.g. durational marking of lexical stress), but divergence in other aspects, particularly, at the level of phonetic implementation (e.g. rise timing). Examining these varieties, therefore, presents an opportunity to not only address a major gap in typological coverage, but also to test and enrich existing theoretical frameworks for modeling intonation.

Singapore English (SgE) is one example of these Contact varieties (Foley 1988). As is common with many Contact varieties of English, SgE developed and exists within a complex linguistic ecology, with high levels of historical contact with typologically divergent languages, in particular, Southern Chinese languages (e.g., Hokkien, Teochew), Straits Malay, and Tamil. Moreover, there exists significant variation within the umbrella of SgE that is conditioned by various social factors including ethnicity, language background, and socioeconomic status (e.g., Brown 2000, Gupta 1998, Tay 1982). Differences in syntax, lexical choice, code switching, and the use of sentence-final discourse particles, in particular, have been discussed in the literature (see Lim 2004, Deterding 2007, Leimgruber 2013 for an overview). There has been considerable work on describing the segmental patterns of SgE, e.g., the substitution of /θ/ with [f] or [t]

(Moorthy & Deterding 2000), the lack of a tense/lax distinction in vowels (e.g., Bao 1998, Deterding 2005), final consonant cluster reduction (Gut 2005), or final stop glottalization (Deterding 2007). There is relatively little work, however, that systematically addresses prosody and intonation in SgE, especially with reference to within-speaker differences. One major reason for this is that, unlike segmental features, far less is known about which prosodic and intonational features represent core aspects of the standardized variety versus lectal variation, since a comprehensive *phonological* model of SgE intonation is still lacking. We review the previous work on SgE intonation in Section 1.3.

The issue of lexical word stress, in particular, presents a challenge for understanding the intonational system of SgE. As a starting point, the term *lexical word stress* will refer here to any mechanism by which one syllable in a word is lexically specified to be prosodically distinctive in *some way*. This could mean, for example, that due to its lexical status, one syllable within a word is realized with greater acoustic prominence, whether through greater intensity, greater duration, a more forceful or more canonical segmental articulation, etc. In many systems, however, having a lexically distinctive prosodic status may entail that a syllable's location in the word has consequences for the organization of local prosodic structure. In Mainstream English varieties, for example, a lexically stressed syllable phonologically attracts a pitch accent when one happens to be assigned to the word. In that case, the effects of the distinctive lexical status are over and above merely specifying greater acoustic prominence, which also takes place in the absence of pitch accents. In other systems, such as Hong Kong English (Luke 2000; Wee 2008) or Ghanaian English (Griper-Friedman 1990; Gussenhoven 2017), there are lexically privileged syllables which by rule are assigned a H

tone, but which may also block the rightward spreading of non-H tones or trigger the spreading rightward of additional H tones. In such systems, in other words, a lexically distinctive syllable need not necessarily bear greater acoustic prominence as compared to its neighbours within the word. In that sense, the present study leaves open the possibility that to the extent that in SgE there are syllables which are attributed a special status in the lexicon (which we refer to throughout as "lexical word stress"), the consequences of that status may involve greater prominence along one or more acoustic dimensions and/or may have consequences for prosodic structure, such as being assigned specific tones or attracting tones that are present for structural or functional reasons (e.g., phrasal tones).

Studies on lexical word stress in SgE present conflicting findings. In this study, we report on the results of a production study that examines whether post-lexical f0 is sensitive to lexical word stress, and we do so as part of an effort to develop a comprehensive model of the intonational phonology of SgE. In particular, we test the predictions of an existing phonological model of tone-to-stress assignment (Ng 2009, 2011). In the sections that follow, we first discuss the linguistic context in Singapore, and previous work on SgE prosody and intonation, before outlining the approach adopted in the current study. We examine whether SgE intonation can be modeled in terms of sparsely distributed tonal targets, in contrast to Ng (2011), as has been demonstrated for numerous other intonation languages. We also address the question of how these tonal targets are distributed relative to the segmental string (i.e., by phonological association) and whether this distribution is sensitive to lexical word stress. In doing so, we seek to illustrate how a due consideration of a *phonological* model of intonation, and more specifically, one that provides explicitly

for the phonetic interface, can yield significant insights into the intonation system of a Contact English variety.

### 1.1 Singapore English in context and models of variation

SgE is spoken in a highly multilingual context, alongside languages such as Mandarin Chinese, Tamil and Malay. The general consensus is that the variety has its origins in English-medium schools during the British colonial period (Gupta 1998, Ho & Platt 1993). Currently, SgE is the working language of government and the primary medium of instruction in the education system. Increasingly, English is also becoming more commonly used as the predominant household language. In fact, between the 2010 and 2020 Census, the number of residents who reported English (whether alone or with another language) as their most frequently spoken language at home rose from 32.29% to 48.25% (Singapore Department of Statistics 2010, 2020: Table 1). The rate of English use at home also differs by race/ethnicity as well as education levels. In the 2020 Census, for the ethnic Chinese population as a whole, 47.64% use English most frequently at home compared to just 39.03% of ethnic Malays as a whole and 59.25% of ethnic Indians. Notably, as seen in Table 1 below, the use of English as the most spoken home language increased across all three major ethnic groups between the last two census periods. The degree of English use is also dependent on educational background. Within the Chinese population, which is the focus of the current study, more than 60% of university-educated ethnic Chinese use English most frequently at home, compared to 39% with just a secondary school qualification (Singapore Department of Statistics 2020: Table 48). A similar rate of English use is also observed amongst university-educated Malays and Indians.

| | Total | | Chinese | | Malay | | Indian | |
|---|---|---|---|---|---|---|---|---|
| **Language most spoken at home** | **2010** | **2020** | **2010** | **2020** | **2010** | **2020** | **2010** | **2020** |
| **English** | **32.29** | **48.25** | **32.62** | **47.64** | **17.01** | **39.03** | **41.61** | **59.25** |
| Mandarin Chinese | 35.64 | 29.9 | 47.74 | 40.23 | 0.09 | 0.06 | 0.11 | 0.06 |
| Other Chinese languages | 14.33 | 8.71 | 19.22 | 11.76 | 0.04 | 0.02 | 0.02 | 0.02 |
| Malay | 12.19 | 9.24 | 0.25 | 0.2 | 82.66 | 60.69 | 7.93 | 6.01 |
| South Asian Languages | 4.44 | 3.16 | 0.00 | 0 | 0.06 | 0.02 | 49.91 | 34.45 |
| Others | 1.11 | 0.74 | 0.17 | 0.16 | 0.15 | 0.16 | 0.41 | 0.21 |

**Table 1.** Singapore Census data from 2010 (Table 48) and 2020 (Table 42). Reported rates of language most spoken at home by ethnic group.

Not surprisingly there exists substantial within-speaker variation depending on the context of use, and the pattern of such variation can differ between populations of speakers in Singapore. Traditional models of variation in SgE share in common some notion of an acrolectal Standard "high" variety, sometimes suggested to be indistinguishable from other standard varieties of English around the world, and a basilectal colloquial "low" variety, often referred to as "Singlish" (Gupta 1994, Brown 1999). Differences between these sub-varieties have been described in terms of differences in lexical choice (e.g., Gupta 1992a, Wee 1998), the use of sentence-final discourse particles (e.g., *lah*) (Gupta 1992b, 1994), syntactic and morphological differences (e.g., copula deletion, Ho & Platt 1993), as well as a number of phonological differences with emphasis on segmental differences (e.g., the absence or presence of tense-lax vowel contrasts; see Lim 2004 for an overview).

Researchers, however, diverge on the way in which this variation should be modelled, with some adopting a diglossic approach (Gupta 1994), and others arguing for the two sub-varieties as end points on a lectal continuum (Pakir 1991, Platt 1975, Platt & Weber 1980). It is generally agreed, however, that speakers can exploit different sub-varieties for stylistic goals, though those higher on the socio-economic status (SES) scale tend to command a wider range of styles (Gupta 1998).

More recent approaches to SgE variation have adopted more of a cultural orientation or indexicality approach. Alsagoff (2007, 2010) for example argues that the variation in SgE is the result of cultural orientation, either towards a global target (referring to Standard SgE as "International SgE"), or local target (Singlish as "Local SgE"). While the distinction between varieties is still maintained in Alsagoff's model, it is less important within an indexical approach (Leimgruber 2013) in which specific features (or sociolinguistic variables) are used to index certain social stances. What is common across these models is the reliance on an existing understanding of what the relevant features or variables are that distinguish between different varieties or lects. As we discuss below, however, while there is a wealth of existing work on lexical, syntactic, and segmental features of SgE and their associated intra- and inter-speaker variation, comparatively little is known about the prosodic, and in particular the intonational, system of SgE.

For the purposes of the present study, we are agnostic as to how inter- and intra-speaker variation in SgE should be modelled. Our study focuses on examining the speech of one group of SgE speakers, namely ethnic Chinese speakers. Our speakers were all university-educated and were all within a similar age range. While this population is admittedly relatively homogenous, we have chosen this approach as a means of

establishing a conclusive initial model which can then serve as a basis for future work aimed at examining the complex dynamics of population differences more directly. Our materials (Section 2.2) include sentences consisting of Standard SgE lexical items, syntax, and morphology, without any discourse particles that are more commonly used in Colloquial SgE. Given that the study took the form of a reading task and occurred in a laboratory setting, we are confident that the variety we are capturing is representative of Standard SgE.

### 1.2. Prosody and lexical word stress in SgE

Most existing work on SgE prosody has primarily focused on the existence and nature of lexical word stress, as well as on situating SgE within a rhythm class category (Ramus et al. 1999, Grabe & Low 2002). The notion of lexical word stress, in particular, is controversial for SgE. Early impressionistic descriptions (e.g., Platt & Weber 1980) suggested that stress in SgE had "shifted" to the final syllable, that is to say, that, diachronically, it had departed from the location of lexical word stress in British English (BrE). Later work, however, argued that the perception of final stress, by primarily BrE researchers, was likely due to the use of greater phrase-final lengthening in SgE compared to BrE, which can contribute to the perception of final prominence in SgE (Low & Grabe 1999). While many previous studies assume some notion of lexical word stress (e.g., Bao 2006), typically involving a mapping from the location of stress in BrE, some have also argued for a model of SgE prosody without stress at all (e.g., Chow 2016). Functionally, it has been reported that unlike for BrE, verbs and nouns are not distinguished on the basis of stress-based prominence (e.g., noun vs. verb forms of *record*; Bao 1998). It should be

noted, however, that BrE listeners are reported to use primarily vowel reduction cues for this contrast rather than suprasegmental cues (e.g. Cutler 1986). It has also been shown, however, that SgE does not evince vowel reduction in putatively stressless syllables (Deterding & Poedjosoedarmo 1998, Tay 1982, Low et al 2000). Relatedly, studies examining the degree of variability in the duration of successive syllables, as measured by the Pairwise Variability Index (PVI), have found that SgE has lower PVIs (Low 1994, 1998, Low et al. 2000, Deterding 2001) than British English, indicating that it is more "syllable-timed" than British or American English with respect to the rhythm class hypothesis. This is consistent with the fact that stress placement is more difficult to perceive in SgE (Deterding 1994, Tan 2006).

A number of studies have sought to establish the acoustic correlates of lexical word stress, or word-level prominence more generally, in SgE. The emerging picture from this line of work is that fundamental frequency (f0) is not a strong correlate of stress; instead, intensity and duration seem to be more important cues (Lim & Tan 2001, Tan 2003, 2006). Part of the reduced importance of f0 might be explained by the tendency for all content words to have an f0 peak on the final syllable regardless of stress placement (Deterding 1994, Chong & German 2017). In fact, Tan (2016) recently showed that this word-final f0 peak reliably engenders a percept of prominence, though whether this is at the lexical or post-lexical level is unclear.

Despite the finding that f0 is not a strong cue for stress, it has nonetheless been argued that f0 does in fact relate to stress position in some way. A number of authors have argued for SgE as a "tone" language, in the sense that tone is assigned to every syllable at the lexical level (e.g., Wee 2008; Lim 2004, 2009; Ng 2011), with some models positing a

link between tone assignment and stress. Lim (2004), for example, argued that tone is assigned at the syllable level, resulting in a surface pattern consisting of a series of level tones. Lim's conclusions, however, are primarily drawn from impressionistic observations of pitch tracks, and furthermore they overlook the possibility that surface f0 patterns may be determined by factors other than syllable-by-syllable tone assignment. In two related approaches, Ng (2011) and Wee (2008) propose phonological models of tone assignment that rely on stress placement. Ng's model, for example, is based on the rules in (1).

(1) Ng's (2011) tone-stress model

(i) Assign a H tone to the final syllable of a word

(ii) Assign a M tone to the leftmost stressed syllable

(iii) Assign a L tone to any preceding unstressed syllable

(iv) Remaining unstressed syllables receive an M tone as a result of rightward spreading from the stressed syllable, or may remain unspecified for tone.

Wee's model differs slightly in that M is the default tone for all syllables while L is "optionally" assigned to initial syllables and must be lexically-specified for each case. The wording of the rule concerning the low tone ("Assign [L] to initial syllable in specific cases": p. 21, ex. (23)) along with the sample derivations provided (i.e., the rule applies to *origin* but not to *managing*) suggest a lexical source for this difference. Wee does not elaborate on the issue, however, except for a footnote (p. 38, footnote 19) which tentatively suggests that the assignment of L may coincide with unstressed initial syllables in Mainstream English varieties.

Ng provides quantitative acoustic results of a small production study involving three speakers. In order to determine which syllable in a word was stressed, however, Ng assumed that the "first syllable in a word with M tone must be stressed" (Ng 2011: p. 34), following the classification assumptions in (1) above. In other words, the diagnostic used for the a priori determination of stress location was drawn from assumptions internal to the model being tested. Interestingly, Ng found no significant difference in f0 between adjacent stressed and unstressed syllables, although precisely which comparisons were made is unclear (Ng 2011: 34). Since Ng's model specifically predicts a difference in f0 between stressed, M-assigned syllables and unstressed, L-assigned syllables that precede them, these results do not support the claim that stressed syllables are marked through tone in SgE.

Common to Ng's and Wee's analyses outlined above are the assumptions that SgE has three levels of phonological tone and that each syllable must be specified for tone. In a wide variety of languages, however, tonal specifications are sparse, in that they are either (a) features of phonological structures larger than a syllable, or (b) post-lexical, and therefore do not necessarily occur on every word or syllable. In such languages, a set of implementation rules explains how f0 should vary over stretches of syllables that are unspecified for tone. As a consequence, intermediate values of f0 do not necessarily reflect the realization of an underlying mid-level tone but can be merely the result of an interpolation between a low and a high tone. It is certainly true that the surface realization of f0 in languages with dense tonal specification can vary in connected speech due to phonetic implementation rules (see, e.g., Downing and Rialland (2017) for examples of the interaction of intonation and lexical tone in African tone languages). Ng's model, while

technically not a model of intonational phonology, in principle, does allow for some interaction with higher levels of prosodic organization. This is primarily discussed, however, in terms of pitch range, presumably to allow for the maintenance of tonal contrasts at the word level (see Ng 2011: 37). Ng's discussion of tone assignment does nevertheless acknowledge the possibility of phonetic realization rules, namely, interpolation is proposed to account for cases where an observable rising f0 through a syllable spanned by M on the left and H on the right precludes the need to assume tone assignment for that syllable. We return to a discussion of these details below.

### 1.3. Intonation in SgE

SgE declarative utterance typically involve a series of f0 rises, with the domain of the rising contour usually involving a single content word and any associated function words that precede it. This is illustrated in Figure 1: the sentence *Millennium was an old hotel chain* can be parsed into the following units in parentheses based on the domain of each rise: (*Millennium*)(*was an old*)(*hotel*)*(chain).* Typically, the rise on the first word or group of words shows the largest f0 range, as can be seen on the word *Millennium* in Figure 1 (Deterding 1994, Low 2000, Low & Brown 2005, Chong 2013). All subsequent rises show a much more compressed pitch range, as can be seen on (*was an old*)(*hotel*) in Figure 1. These utterances typically end with a falling tone (e.g. on *chain* in Figure 1) or a rising-falling pattern on the utterance-final word (Chong 2012, 2013; Chow 2016; Deterding 1994; Lim 2004; see German & Chong (2018) for further examples).

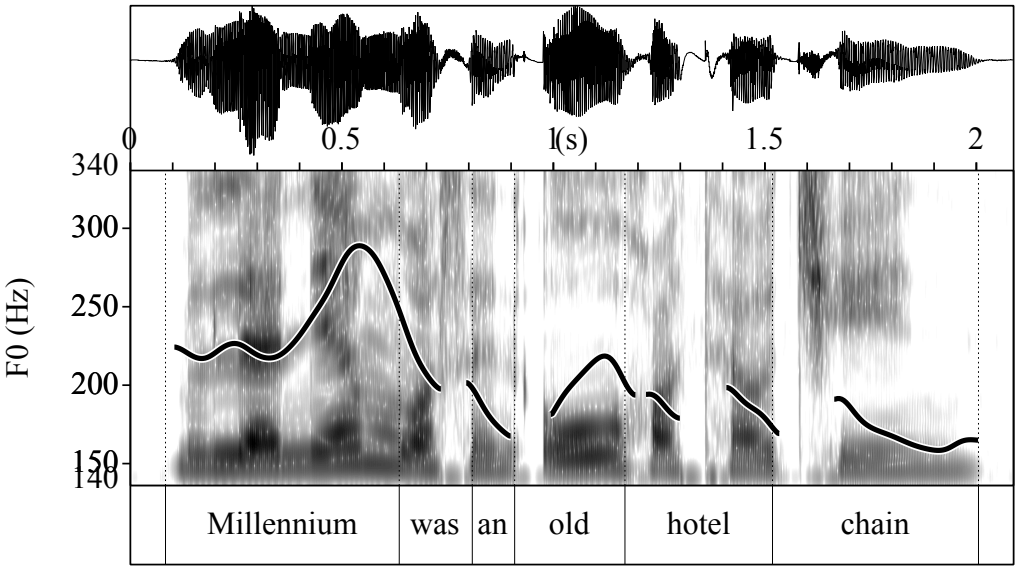



**Figure 1.** A typical declarative f0 contour in SgE.

Previous characterizations of SgE intonation have relied heavily on phonetic frameworks (e.g., Deterding 1994, Lim 2004, Low 1994, Tan 2010). Applying the British model (O'Connor & Arnold 1973, Cruttenden 1986), Deterding (1994), for example, argues that the domain of tone assignment in SgE is a single word, rather than a specific syllable, and that it is not always possible to identify a single most prominent syllable in a word. Deterding also observes that, regardless of the expected stress pattern, a rising contour over each entire word is the most common non-final pattern, with rise-falls most common utterance-finally. Such observations have led some researchers to conclude that the British model, which assumes a central role for pitch movements on nuclear (i.e., utterance-prominent) syllables, is not suitable for characterizing SgE (see also Lim 2004). Tan (2010) similarly eschews the application of a Tones and Breaks Indices (ToBI)

annotation system (Beckman & Ayers 1997, Beckman & Hirschberg 1994, Pitrelli et al. 1994, Silverman et al. 1992) in favour of a purely phonetic transcription, based on the related assumption that word-level stress plays too important a role in the ToBI system. These phonetic approaches to intonation rely primarily on impressionistic transcription of tonal movements and tonal groups/units, which can span an entire utterance. The crucial difference between these frameworks and the phonological approach we adopt here is the principled distinction between an abstract phonological level of representation and the phonetic implementation rules that determine their surface realization (see Arvaniti 2011, Jun 2005, Ladd 1998 for an overview of differences between these approaches).

The arguments against such models, however, overlook the fact that the Autosegmental-Metrical Framework (AM; Beckman & Pierrehumbert 1986, Ladd 1983 Pierrehumbert 1980) on which the American English ToBI annotation system is based allows for metrical structure that arises from phrasing and not only from lexically determined word-level stress (see Arvaniti 2022, Jun 2014 for recent overviews; Section 1.2). An AM model of SgE intonation has been proposed by Chong (2012, 2013). According to that model, SgE has at least one level of phrasing above the prosodic word and below the intonational phrase (IP), called the Accentual Phrase (AP), which usually consists of a single content word and any preceding function words (see Figure 2). Tones are associated primarily to the edges of the AP (aL at the left edge and Ha at the right), while an L* pitch accent can optionally occur on stressed syllables. The optionality of the L* in Chong's original model reflects the fact that it was often difficult to perceive prominence on a specific syllable in the target words/sentence elicited by Chong (2012, 2013), and that the f0 valley, especially in medial phrases, did not always coincide with a

stressed syllable (see also German & Chong 2018; c.f. Chow 2016 for an SgE AM model without lexical stress and pitch accent).

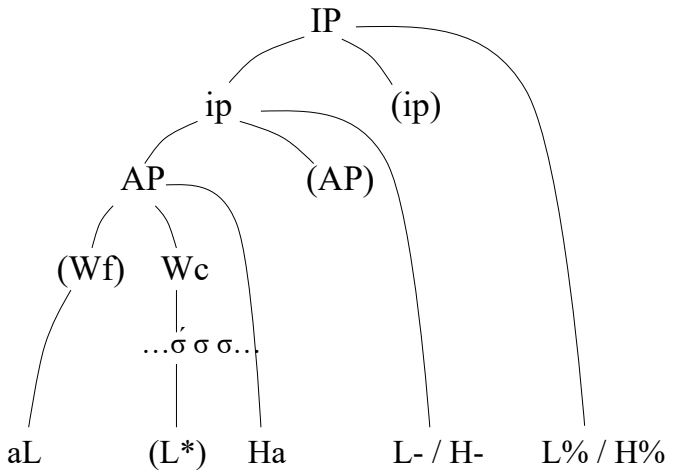



**Figure 2.** A abstract schema of the prosodic structure of SgE as proposed by Chong (2012).

In a production experiment, Chong & German (2017) compared the acoustic correlates associated with word-final versus AP-final positions and found that AP-final syllables were longer and had a higher f0. The authors demonstrate that together, duration and peak f0 provide robust cues for discrimination between these two levels of phrasing. This finding clearly points to the relevance of phrase-level structure for determining pitch contours in SgE.

By comparison, accounts like Ng (2009, 2011), Lim (2004), and Wee (2008) do not explicitly distinguish between abstract phonological representations and the rules that map them onto phonetic realisations (i.e., phonetic implementation rules; Arvaniti & Ladd 2009, Pierrehumbert 1980, Pierrehumbert & Beckman 1988). As a consequence, they are unable to generate predictions concerning which types of variation do or do not contribute to

meaningful differences between ostensibly contrasting forms. This issue is particularly relevant given that virtually all descriptive characterizations of SgE intonation take for granted that the overall non-final word or phrase level pattern is "rising". In AM-based models, phonetic implementation rules determine how f0 varies over syllables or longer temporal spans that are unspecified for tone. Such models therefore readily account for rising or falling patterns on the basis of sparsely distributed tonal targets. The introduction of mid-level or other tones assigned to syllables internal to the rise is thus less parsimonious in the absence of further motivating evidence. In the next section, we outline an approach which proposes that intonational patterns in SgE can be explained by phrasal structure along with an inventory of abstract and sparsely distributed phonological units whose phonetic realization through f0 is governed by systematic mapping rules.

It should be noted here that the previous studies cited in Section 1.2 were aimed primarily at modelling the prosodic patterns of Colloquial SgE (CSE; Lim 2004, Ng 2011, Wee 2008; cf. Deterding 1994, Low & Grabe 1999, Low et al. 2002, which emphasize laboratory speech as we do here). Given the target population (university students) and setting (read speech in a laboratory) of the present study, the data likely reflect a more "standard" variety of SgE. As mentioned previously, there is currently no existing research that addresses how the intonational features of SgE vary as a function of social context (e.g., formal vs. informal) socioeconomic background, or language dominance (though see Tan (2003, 2010) for an overview of differences associated with ethnic and language background). Nonetheless, the *phonetic* patterns observable in the pitch tracks presented in Ng (2009, 2011) and Lim (2004) suggest a high degree of similarity with those found in our own data, most notably in the prevalence of rises spanning a similar-sized lexical

domain. We are confident, therefore, that many aspects of our model extend to the patterns found in CSE as well. While future research will need to explore the modelling of intra- and interspeaker variation in intonation, the results of the present study provide a solid empirical and theoretical basis for doing so.

### 1.4 Current approach and caveats

The previous section shows that the nature of lexical word stress, and its relation to f0, in particular, is far from a resolved issue in SgE. In the present study, we explore whether lexical word stress plays a role in SgE by examining whether different lexical items exhibit distinct patterns of f0 realization when other factors are held constant.

To avoid circularity, we start with the hypothesis that, if lexical word stress exists in SgE, then the word-by-word specification of stressed syllables is correlated with that of BrE. As described above, it is often suggested that SgE "inherited" its stress patterns from BrE. While the status of SgE lexical stress has been widely debated, if it is the case that SgE has no lexical word stress at all, then there is no reason to expect that our f0 measures will differ systematically according to the classification system (i.e., based on BrE) that we apply to our target items. In effect, we expect a null result in that case. Alternatively, SgE may have its own system of lexical stress, either because it has diverged from the BrE system over time or because it has developed one independently. In that case, we would expect either a null or weak result, since the pattern of variation in the SgE system would be either partly or completely uncorrelated with the one used to classify our materials. If, however, our assumption is correct, then two outcomes are possible. We might observe no differences in f0 patterns because, although SgE has BrE-like lexical stress, this is not

manifested through f0 patterns. Alternatively, we might observe a systematic difference in f0 patterns according to the assumed classification system. In that case, the results of our study will provide a direct test of Ng's (2009, 2011) tonal account of lexical stress described in the previous section, an issue which we return to at the end of this section.

The current approach is also informed by the fact that phonetic prominence (which is often taken as a cue for lexical stress) can arise for reasons other than lexical word stress, in particular, as a function of the position of a syllable within a prosodic phrasal unit, such as at a right edge (Gordon 2014, Gussenhoven 2004, Jun 2014). For example, previous work has shown that in stress-accent languages, including English, syllables preceding prosodic boundaries show consistent lengthening (e.g., Wightman et al 1992, Turk & Shattuck-Hufnagel 2007), over and above the syllable's status as stressed, accented, or nuclear accented. Similarly, in languages like French (Jun & Fougeron 1995, 2000) and Korean (Oh 1998, Cho & Keating 2001), syllables at the end of smaller phrases (Accentual Phrases) show similar lengthening in addition to prominent f0 marking.

For the present study, we seek to build on Chong's (2012, 2013) AM-based phonological model of SgE intonation described in the previous section (Figure 2). Given previous findings showing the importance of phrase-level structure for determining pitch contours and syllable length in SgE (Chong & German 2017), our study explores the possibility that the SgE phonological system may have developed different cues for lexical word stress that do not necessarily involve phonetic or perceptual prominence (see discussion in Section 1). More specifically, we address the possibility that lexical word stress is manifested in the scaling and shape characteristics of f0 contours over and above

those required by AP-level phrasing, which at a minimum involve a phrase-initial low target and a phrase-final high target.

In the first part of our analysis, we specifically seek to assess Ng's (2011) claim that lexical stress in SgE is realized by the assignment of a mid-tone to a specific syllable. To do this, we examine f0 means on disyllabic and trisyllabic target words that differ in the location of stress (i.e., according to the assumed BrE patterning). Figure 3 provides schemas of the predicted tonal patterns under Ng's model for two- and three-syllable words each under two patterns of stress. While all words in both length contexts are predicted to end with a high f0, differences in stress should determine the distribution of low and mid tones pre-finally. In particular, Ng's model predicts different patterns for the initial syllable (in bold; stressed syllables are underlined) of stress-initial versus stress-medial and stress-final single word targets, respectively, **M**H vs. **L**H for two syllable targets and **M**MH vs. **L**MH for three syllable targets. Under a strict level tone interpretation of Ng's model, then regardless of word length, the f0 of the initial syllable should be clearly higher when stress is on the initial syllable than otherwise. On the assumption that interpolation is provided for, then these differences should minimally appear as differences in the scaling of the starting f0 and, for three-syllable targets, in the alignment of the turning point up to the final H target, such that it occurs earlier when stress is on the initial syllable than otherwise. In very narrow terms, Ng's model clearly predicts stress-based differences in f0 for the initial syllable, whereas such differences in other syllables are not expected. The design of the present study allows us to test these predictions directly.

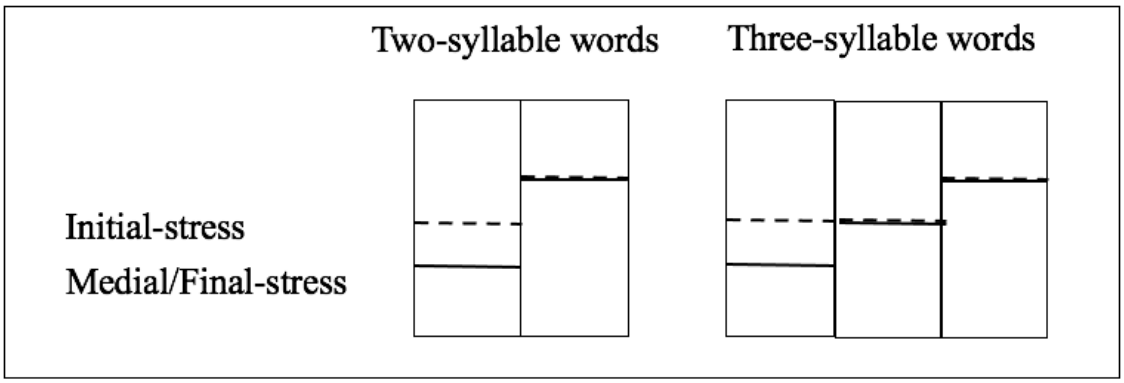



**Figure 3.** Predicted (from Ng 2011) tonal patterns based on stress placement. Dashed = initial stress. Solid = medial stress (three syllable words) or final stress (two-syllable words).

By comparison, Chong's model only posits two tonal levels: L tones that can be either attributed to the aL at the left edge of the AP or to an L* pitch accent, and the AP-final H tone (see Figure 2). Under the general assumption that f0 interpolates between tonal targets, the predictions for f0 *means* on each syllable within an AP under Chong's model are less straightforward. This is because a single syllable can, in principle, have multiple tones associated to it. In disyllabic targets, for example, final syllables when stress is final (σσ́) can host both an L* pitch accent and an AP-final H. With stress on the initial syllable (σ́σ), L* would be associated with the first syllable, and the second syllable would host only H. In the latter case then, the rise from L* would begin earlier leading to a less steep slope and potentially an earlier alignment of the AP-final H. Similarly, for trisyllabic targets, since L* is expected to be phonologically associated to the stressed syllable, then the rise to AP-final H should begin earlier in initially stressed words (σ́σσ) as compared to words with stress on the medial syllable (σσ́σ). In the latter case, we also expect an

interpolation from AP-initial L to L* on the medial syllable followed by a later and possibly steeper rise to the final H. Regarding measures of mean f0, it is expected to be generally *lower* on all non-final syllables when stress is non-initial. This is because the later starting point of the rise means that f0 stays lower for longer, even if a similar final H target is eventually reached on the final syllable.

It is also possible that any stress-based differences in f0 realization may not be straightforwardly observable based on f0 *means* alone, and that more dynamic measures of f0 alignment may better reveal them. In order to more directly address existing proposals, the current study focuses on examining f0 means on each syllable within an AP, leaving the issue of f0 alignment and stress for future investigation. We do, however, provide a visualization of time-normalised f0 curves from the two- and three-syllable single word data set as a means for observing broad trends which may serve as the basis for such investigation.

In addition to the two- and three-syllable single word targets used for the comparisons described above, our study includes monosyllabic targets (e.g. *Lin*) as well as three and four syllable targets containing a content word preceded by one or more function words (e.g. *He murmured, he might mail*). Together, these materials allow us to explore how the shape of the f0 contour (e.g., the scaling and timing of the rise) is influenced by the composition of the AP (i.e., the number of syllables and number of words). Ng's model (see also Lim 2004) predicts different f0 contour shapes depending on the composition of the utterance. Specifically, monosyllabic targets are predicted to be realized as a high plateau as compared to the typically rising contour for multisyllabic targets. As mentioned, while Ng's model in principle allows for some interaction with higher level prosody, this

is primarily discussed in terms of pitch range. Ng's discussion of tone assignment also allows for interpolation between tonal targets though this is stipulated only when observed pitch tracks do not support the attribution of level tones, namely, when there is a "straight-line rise from mid to high" in, for example, the medial syllable of *minimum* (Ng 2008: 26). This is in contrast to a separate example of the same word in which the medial syllable is level and a tone is therefore attributed to it. As a consequence, monosyllabic targets are predicted to be realized as a high plateau as compared to the typically rising contour for multisyllabic targets.

Other specific cases for which Ng allows for interaction with higher levels of prosodic structure include the higher pitch range associated with utterance-initial rises and the fact that some utterances end with either a downstep-like sequence of plateaus or a highly compressed pitch range. Given that more generally, Ng proposal describes "level steps in pitch" (Ng 2011: 29, 38), we can only speculate about what the model would predict for monosyllabic words and other polysyllabic words. Overall, Ng does not provide sufficiently explicit details regarding phonetic implementation, leaving the issue of contour shape over and above tone specification largely to speculation, especially in cases where level realizations of f0 are not observed.[1]

[1] Lim's (2004) analysis of intonation is couched within an INSTINT framework (Hirst & Di Cristo 1998) which is a not a phonological framework (see Ladd 1998, Jun 2005, Arvaniti 2011, for a general discussion of phonetic vs. phonological approaches). It is therefore difficult to make direct comparisons. Lim (2004) suggests that a "striking feature of CSE intonation is the tendency for the pitch to move in terms of *sustained level steps* [emphasis our own], rather than gliding more gradually from one pitch level to another" (p. 42). Elsewhere, tunes are described as being level over each word (e.g. in an utterance of "You told me.").

By contrast, AM-based models make a principled distinction between abstract phonological tonal specification and phonetic implementation, with "systematic variation" (Arvaniti & Ladd 2009) in surface tunes reflecting similar underlying tonal specifications that are modulated by, among other factors, utterance length. In line with this, Chong's model predicts that a general LH pattern should be apparent for all targets regardless of their internal composition (e.g. stress pattern, distribution of word boundaries) and length (i.e., number of syllables). As mentioned above, one aim of the present study is to provide a visualization of f0 contours across different patterns of stress, length, and syntactic composition in order to identify broad differences which may serve as the basis for future investigation.

Finally, our study includes an exploratory analysis of two other typologically common acoustic correlates of stress, namely duration and intensity, and we discuss how stress might interact with phrasal edge prominence in SgE. While vowel quality is known to be a salient cue to lexical word stress in Mainstream varieties of English (e.g. Fry 1958, Beckman 1986, Beckman & Edwards 1994), we leave an investigation of that issue for future research.

## 2. Methods

### 2.1 Participants

24 native speakers (11 F, 13 M; Mean age = 22.5, range = 19-28) of Singapore English were recorded for this study. All participants self-reported as being ethnically Chinese and most were students at Nanyang Technological University (NTU) at the time of the study. All recordings took place at NTU. An additional three ethnically Chinese speakers

participated in the study but whose data are not included in the analysis since more than half their tokens were excluded based on criteria outlined below (n = 2) and general disfluency in productions (n=1).

### 2.2 Materials

Previous studies suggest that the rising contour associated with the utterance-initial word (or Accentual Phrase in Chong's model) shows the greatest overall change in f0 (Deterding 1994, Low & Brown 2005), while subsequent f0 movements tend to be compressed by comparison. As a consequence, any effects of stress differences on the realization of f0 are likely to be more readily detected in utterance-initial position. For this study, we therefore placed all target items, including single and multi-word targets, in the initial position of their associated carrier sentences. Critical target items included 16 two-syllable and 16 three-syllable monomorphemic words. These words were selected such that the expected location of stress based on standard BrE pronunciation was on the first syllable for half of the items in each group, and on the second syllable for the other half.

It is well-established that the prosodic constituent associated with the basic rising contour in SgE can comprise multiple lexical items including a content word along with associated function words to its left. In at least some languages, the morpho-syntactic composition of a prosodic phrase can have important consequences for its phonological structure, such as in French, where an L tone associated with the Accentual Phrase aligns to the left edge of the first content word in the phrase (Welby 2002, 2006). In order to explore if, and how, AP-internal word boundaries influence the *shape* of the f0 contour (e.g., through timing and alignment differences as opposed to merely raising or lowering

the mean f0 of entire syllables), we also manipulated the internal constituency of APs by including 20 multi-word targets that consisted of mono-, bi-syllabic and trisyllabic verbs preceded by one or more function words (either a personal pronoun or a personal pronoun followed by a modal verb). Finally, we also included 8 monosyllabic targets for a combined total of 60 target sentences. Although targets were not limited to those containing only sonorants, they were paired across stress conditions so as to maximize the similarity of segmental patterns. This step ensured a degree of control for segmental and microprosodic effects (see, e.g., Vander Klok et al. 2018). Examples of each target types are shown below in (2) and (3) with stressed syllable underlined. In the main quantitative analysis that follows in Section 3.1, we focus on a statistical analysis of single-word di- and tri-syllabic targets that differ in stress position (2b, c). A full list of target items is provided in Appendix A. Together, these materials make it possible to assess whether and how the f0 contour of an AP depends on both its internal morphosyntactic structure as well as the lexical stress pattern of words contained within it. We provide a data visualization of the f0 patterns of all target types in the corpus in Section 3.2.

(2) Single word targets:

a) Monosyllabic (S): *e.g. Lin*

**b) Disyllabic (Su, uS): *e.g. Alan, Elaine***

**c) Trisyllabic (Suu, uSu): *e.g. Oliver, Manila***

(3) Multiword targets (# = word boundary):

a) Disyllabic (u#S): *e.g. He won, She ran*

b) Trisyllabic (u#Su, u#uS, u#u#S): *e.g. He murmured*, *They denied, He might mail*

c) Quadrisyllabic (u#Suu, u#uSu): *e.g. He delivered, He verified*

## 2.3 Procedure

Participants were recorded in a sound-attenuated booth using a Shure SM81 microphone linked to a FocusRite Saffire PRO40 audio interface with a sampling rate of 48kHz. Recordings were made in .wav format with 16-bit quantization. Target sentences were displayed one at a time on a computer monitor, with presentation order controlled by E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA). Participants were asked to read each sentence as naturally as possible "as though they were talking to a friend". The 60 target items were combined with an additional 54 items from a separate experiment described in Chong and German (2017). Participants produced the entire list of 114 items twice. The items were grouped into five blocks which were balanced for condition. The order of blocks and the order of items within blocks was randomized separately for each participant and for each repetition of the experiment within participants. Participants were able to take a short break between repetitions, and no time-limit for the completion of trials was imposed.

## 2.4 Pre-analysis and data processing

Only recordings from the second repetition of the list were analysed due to overall higher degrees of fluency relative to the first repetition. These recordings were first phonetically segmented and labelled automatically using the SPPAS force-alignment tool (Bigi 2015). The segmentation was then checked by the authors for alignment errors and corrected

manually using visual inspection of the spectrogram and waveform in Praat. The syllables and vowels of the target words were annotated based on the corrected segmentation.

Productions were excluded from analysis if they met one of the following three conditions: (i) participants produced the target disfluently or with the unintended number of syllables (e.g. the word *Emily* produced with two instead of three syllables: [ɛmli] instead of [ɛmili]) (n = 41), (ii) there was an observable pitch reset following the initial AP (n = 5), or (iii) the f0 peak was realized earlier than on the final syllable of the initial AP or there (n = 54) was a flat plateau over the entire AP (n = 47). The last criterion is motivated by the fact that existing descriptions of SgE intonation generally agree that the f0 peak on APs (or prosodic words) is realized on the final syllable (or right edge) of the content word, regardless of stress location. While exceptions to this pattern represent interesting cases of variation in their own right, they represent a small minority of tokens in our study, and therefore do not provide the means for a rigorous quantitative assessment of the influence of stress on f0. Additionally, many of these tokens sound intuitively to native listeners like uses of a Mainstream English variety (e.g., American English), and we indeed suggest that they may represent instances of social indexing or some other socio-performative behaviour which incorporates phonological elements of one or more of those varieties as markers. If that is the case, then it is possible that such tokens involve different ways of realizing stress or other postlexical prosody, which would interfere with the principal goals of our study. While a full account of all intonational patterns used by SgE speakers would necessitate a consideration of these tokens, here, we restrict our attention to the global pattern that is both most commonly discussed in descriptive accounts of SgE, and also the most frequent pattern in our study. We nevertheless discuss some of these non-

normative and previously undocumented patterns in Section 3.3 below. In total, 1,293 tokens were used for quantitative analysis.

In order to facilitate the qualitative visual analysis of f0 contours (Section 3.3), f0 was sampled at 3-ms time steps over each syllable in all targets (both single and multi-word) using a custom Praat script based on the built-in autocorrelation function. In order to assure good time resolution (i.e., shorter analysis windows) without missing datapoints for low f0 regions, the pitch floor was adjusted separately for each participant (approximately 10 Hz below the lowest f0 reached in the target range). After initial f0 extraction, pitch objects were manually inspected for obvious pitch tracking errors due to, for example, pitch doubling/halving, voice quality issues, or obstruent boundary effects, and the associated pitch points were either adjusted (i.e., in cases of halving/doubling) or removed. After manual correction, the pitch objects were smoothed using the built-in Praat smoothing function with a bandwidth of 15 Hz. The f0 contours of the target regions were then further processed by dividing each syllable into 10 equal intervals and calculating the average f0 of each interval. This yielded an f0 time-series for each token which was time-normalized to other tokens while preserving syllable boundaries.

Our dataset is divided into two sets for the purposes of analyses (the full set of data can be found here: https://osf.io/ahycv/) For the first set, our emphasis is on applying quantitative statistical analyses to disyllabic and trisyllabic single word targets (n = 700). Mean f0 values over the entire vowel were extracted from these smoothed pitch objects using a custom Praat script. The mean intensity and duration of vowels in each syllable of the target words were also extracted using a custom Praat script. Mean intensity was obtained using the "Energy" method in Praat. Upon visual inspection, the distributions of

raw duration measures were highly skewed and applying a log transformation substantially reduced skewness. These measures were therefore log10-transformed prior to analysis, though raw duration is used in the figures below for ease of interpretation. Separate linear mixed-effects models were fit using the *lme4* package (Bates, Maechler, Bolker, and Walker 2015) in R (R Core Team 2015) with mean f0, log duration and mean intensity over the nucleus of each syllable as the dependent variables, and stress pattern (Su vs. uS for disyllabic targets or Suu vs. uSu for trisyllabic targets) and syllable number as fixed factors, as well as the interaction between the two. Speaker gender was further included as a control factor. The models also included random intercepts by speaker and target. This was the maximal random effects structure that converged. The best fit model for each measure was first assessed using log-likelihood ratio tests by comparing a model with the interaction term to a subset model without it using the *anova()* function. We report *F*-statistics for significant effects from the best fit model below (derived from the *anova()* function), with *p*-values obtained using Sattherthwaite method from the *lmerTest* package (Kuznetsova, Brockhoff & Christensen 2014, 2017). Where post-hoc pair-wise comparisons are necessary to explore significant interactions, these were conducted using the *lsmeans()* function in the *emmeans* package (Lenth, Singmann, Love, Buerkner & Herve 2018) in R, with Tukey HSD correction for multiple comparisons.

In the second set of analyses in Section 3.2, we examine all target types in the corpus including the single word di- and trisyllabic targets above, as well as monosyllabic targets (e.g. *Lin;* n = 176) and multiword targets for which the internal syntactic composition of the AP was varied (e.g. *He murmured, I might mail;* n = 417). For these

comparisons, we provide a visualization of time-normalized, global f0 patterns across targets.

## 3. Results: f0

### 3.1 Two and three-syllable single word targets

We first examined the effect of lexical stress on mean f0 for each syllable in two- and three-syllable single word targets. Mean raw f0 by syllable (i.e., first, second, or third) and stress pattern are shown in Figure 4 for both disyllabic (upper) and trisyllabic (lower) targets. Overall, f0 rises throughout the entire target word as expected, and the qualitative shape of the f0 trajectory does not differ by stress position. For both types of targets, the f0 trajectories are parallel with an overall slightly higher f0 for targets with stress on the initial syllable.

For two-syllable targets, the inclusion of the two-way interaction between syllable and stress position did not improve model fit ($\chi^2(1) = 3.21$, $p = 0.07$). The final optimal model therefore included only main effects. There was a significant main effect of syllable number ($F(1, 696) = 747.4$, $p < 0.001$), reflecting the fact that f0 rises throughout the word, as expected. There was also a significant main effect of stress position ($F(1, 14) = 9.83$, $p = 0.007$), with f0 being higher overall when stress was on the initial syllable in Su vs. uS targets. There was also, unsurprisingly, a main effect of gender with male speakers having a lower f0 than female speakers ($F(1, 22) = 117.87$, $p < 0.001$).

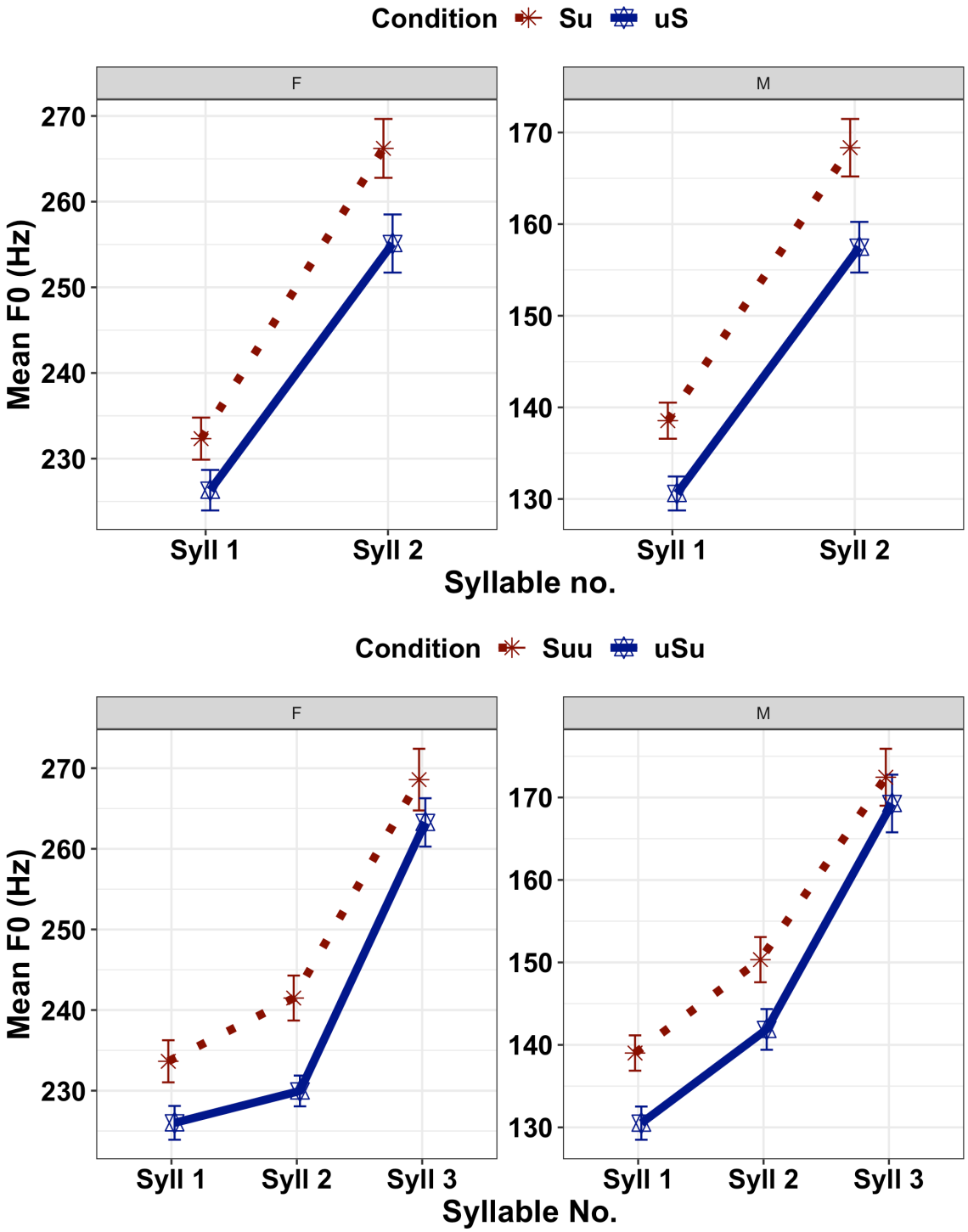



**Figure 4.** Mean raw f0 in Hz with 95% CI by syllable number, word type, and speaker gender for two-syllable (upper) and three-syllable (lower) targets.

For three syllable targets, the best fit model included the two-way interaction of syllable by stress position ($\chi^2(2) = 8.03$, $p = 0.02$). There was a significant effect of syllable ($F(2, 949) = 630.9$, $p < 0.001$), as well as a significant main effect of stress position ($F(1,14) = 28.0$, $p < 0.001$), with initial-stressed targets having an overall higher f0 than medial-stressed targets. Male speakers also had a significantly lower f0 than female speakers ($F(1,22) = 117.5$, $p < 0.001$). Post-hoc comparisons revealed that while all syllables in Suu words had a significantly higher f0 than their counterparts in uSu words, the difference for the final syllable was smaller (Table 2). Post-hoc comparisons further confirmed that f0 on each successive syllable was significantly higher than the preceding one ($p < 0.0001$) (i.e., rising through the target. See supplementary materials for full details).

| Suu vs. uSu | est. | SE | df | t.ratio | *p* |
|---|---|---|---|---|---|
| Syll 1 | 8.10 | 1.87 | 42.1 | 4.31 | 0.0001 |
| Syll 2 | 10.20 | 1.88 | 41.6 | 5.44 | <0.0001 |
| Syll 3 | 4.39 | 1.88 | 41.8 | 2.29 | 0.03 |

**Table 2.** Post-hoc comparisons for f0 in three syllable target words.

To summarize, our results for both two- and three-syllable targets corroborate previous reports of an overall rising f0 contour as the predominant pattern in utterance-initial words in SgE. We also found f0 patterns that have not previously been reported, which we return to in Section 3.3. We did not, however, find any evidence that the location of lexical stress influences the mean f0 of specific syllables in different ways, which would be expected if stress is associated with a specific tonal level. Instead, words with initial stress showed higher mean f0 values across all syllables, suggesting that lexical stress influences the scaling of f0 across the entire word domain. This result is not predicted by a

model like Ng's (2009, 2011) in which tone is specified on each syllable. We further note that the difference in f0 on the first syllable as a function of stress is rather small in magnitude (roughly 10 Hz). We return to these points in the General Discussion in Section 6.

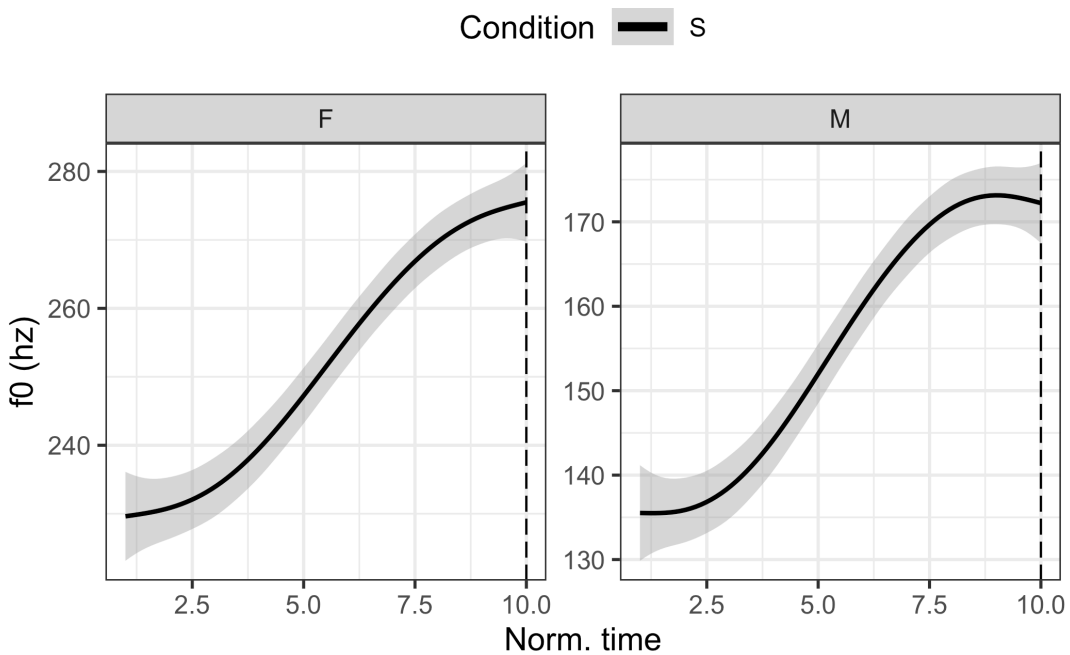
### 3.2 f0 contours across all AP types

The previous section assesses statistically how the mean f0 of individual syllables varies as a function of hypothesized stress patterns, and it focuses on disyllabic and trisyllabic AP targets consisting of a single monomorphemic word. While syllable by syllable averages are useful for exploring whether and how tones are distributed across syllables, they can obscure more fine-grained differences in f0 realization, such as those related to temporal alignment and f0 scaling, which are also potential correlates of lexical word stress. In this section, we therefore present graphical summaries of averaged, time-normalized f0 contours for each type of target included in our study. While our materials were not specifically designed to allow for detailed statistical analysis of contour shape differences across targets with different composition, these visualizations nevertheless make it possible to identify broad differences in contour shape across target types, and on that basis to generate concrete hypotheses which may serve as the basis for future studies involving materials that are controlled differently from the present study.

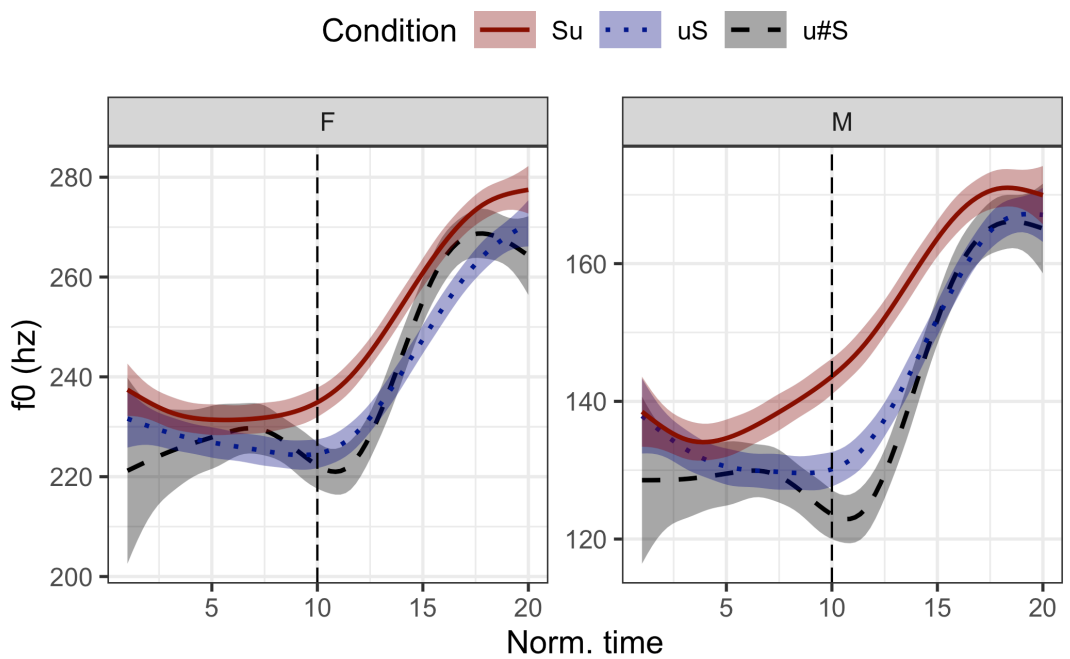
Figure 5 shows averaged, time-normalized f0 curves for all target types elicited in the study. These include (a) monosyllabic targets (e.g., *Lin*), (b) single word disyllabic targets with two hypothesized stress patterns (e.g., Su: *Alain,* uS: *Elaine*) along with two word disyllabic targets with stress on the second syllable (u#S: *He won*), (c) single word

trisyllabic targets with two hypothesized stress patterns (e.g., Suu: *Oliver,* uSu: *Manila*), (d) two word trisyllabic targets with two hypothesized stress patterns (e.g., u#Su: *He murmured*, u#uS: *They denied*) along with three word trisyllabic targets with stress on the final syllable (e.g., u#u#S: *He might mail*), and finally (e) two word quadrisyllabic targets with two hypothesized stress patterns (e.g., u#Suu: *He verified*, u#uSu: *He delivered*). The single word targets in (b) and (c) were designed specifically to test the principal hypotheses addressed in this study and were analyzed statistically in the previous section, while the other target types were included for primarily exploratory purposes as discussed in this section.

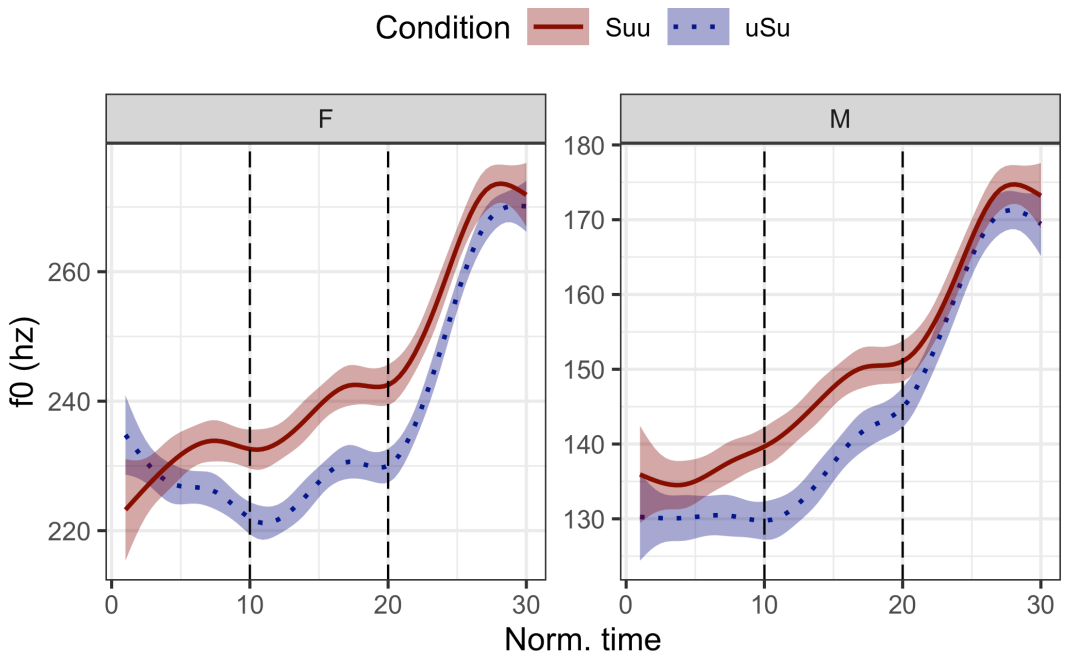
To generate the curves, each syllable of each token was divided into 10 intervals of equal length, and the mean f0 of each interval was calculated from the pre-processed pitch objects described in Section 2.4. The number of intervals for a token therefore depends on the number of syllables in the target type, such that one-syllable targets have 10 intervals per token, two-syllable targets have 20 intervals per token, etc. The data in Figure 5 therefore represents summarized data (notably the mean f0 per interval number) across all tokens for each target type, broken down by gender. Curves of estimated means were fitted to these data using the *geom_smooth()* function from the *ggplot2* package (Wickham 2016) with the GAM (generalized additive model) method with 95% confidence intervals.


Condition
S
F
M
f0 (hz)
280
260
240
170
160
150
140
130
2.5
5.0
7.5
10.0
Norm. time

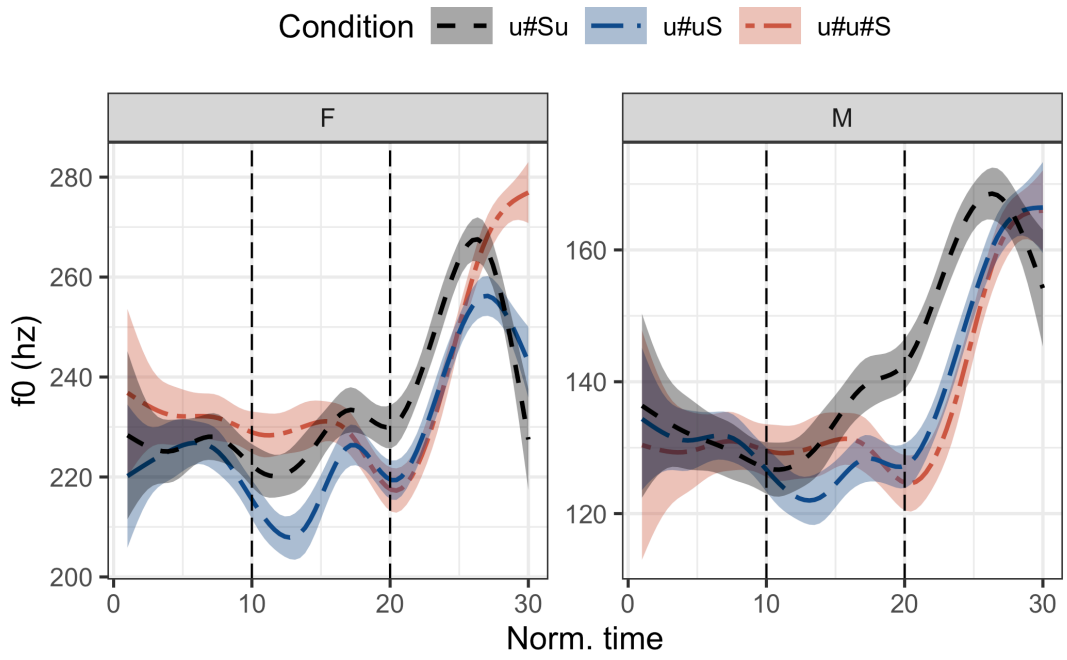
a)


Condition
Su
uS
u#S
F
M
f0 (hz)
280
260
240
220
200
160
140
120
5
10
15
20
Norm. time

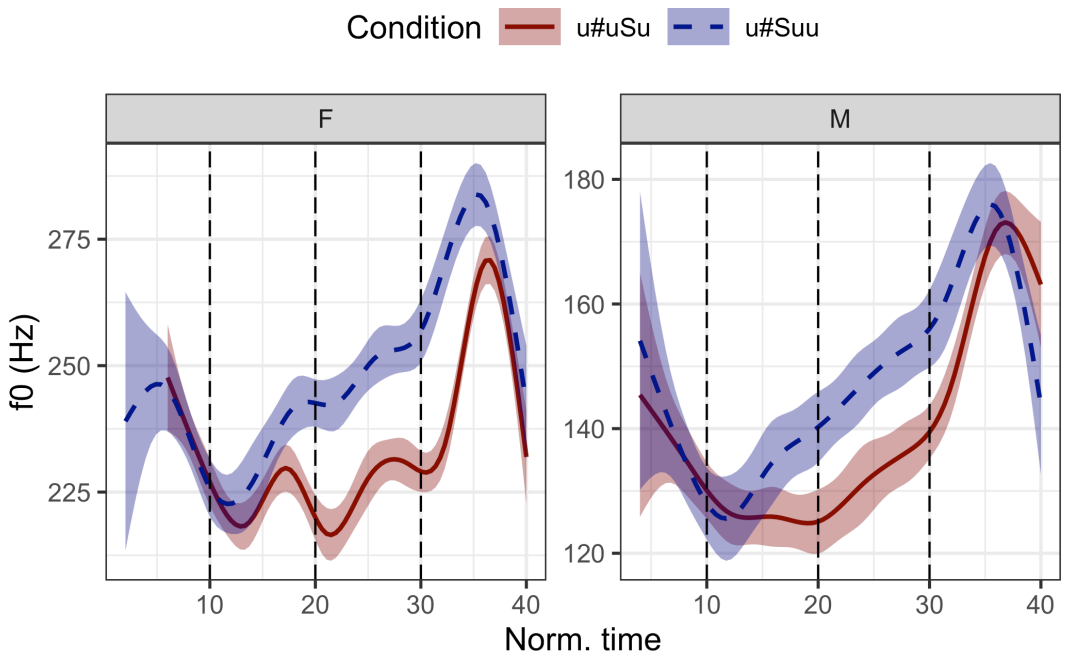
b)


Condition
Suu
uSu
F
M
f0 (hz)
Norm. time

c)


Condition
u#Su
u#uS
u#u#S
F
M
f0 (hz)
Norm. time

d)




e)

**Figure 5.** Time normalized plots of f0 averaged over each condition for (a) monosyllabic targets (n = 176), (b) disyllabic single-word targets (n = 368) and multi-word targets (n = 91), (c) trisyllabic single-word targets (n = 332), (d) trisyllabic multi-word targets (n = 246), and (e) four-syllable targets (multi-word only, n = 80). Dashed vertical lines indicate syllable boundaries and "#" indicates word boundaries.

These plots reveal, first of all, that regardless of the length or the constituency of the APs, the overall shape of the f0 contour can be characterized as starting from a low f0 close to the beginning of the target word/phrase and moving towards a high f0 close to the end of the target. We stop short of characterizing the general pattern as "rising", since for some target types, f0 remains level or dips downward before rising to the f0 peak (notably, uS, u#S, uSu, as well as all three- and four-syllable multiword targets). Additionally, in all but a few cases, f0 turns downward just after reaching a peak in the final syllable. While

this phenomenon has been observed in a previous study by the authors (Chong & German 2015), further research is needed to determine whether it is due to specific timing requirements of the AP-final high target or whether it is better explained by the fact that Accentual Phrases ending with a vowel or sonorant tend to be produced with a final glottal closure. Certain target types also include additional small dips in f0 near syllable boundaries (esp., u#Su, u#uS, and u#uSu for male speakers). Since our materials were not specifically controlled for the sonorance of onsets, we can only speculate that these dips are artefacts of microprosodic perturbations. In spite of these variations, our results suggest a highly general tendency for initial targets to be produced with a low f0 target toward the left and a high f0 target toward the right.

Crucially, this general pattern extends to monosyllabic targets, which is not predicted if monosyllabic targets are assigned only an H tone as in Ng's (2011) model. In fact, any characterization of SgE which assumes that each syllable is assigned one tone or that syllables bearing a single tone are realized with a level f0 (e.g., Ng 2011) fails to account for the consistent low-to-high pattern we observe for different sizes of APs, which would at a minimum require an L and an H in the case of monosyllables. Additionally, a tone-to-syllable account cannot explain rises that begin and end on the same syllable, or rises that continue across multiple syllables as is observed for Suu targets in the present study. Moreover, a qualitative visual inspection of Figures 4a through 4e suggests that the f0 excursions for monosyllabic targets are highly similar in magnitude to those for longer targets, regardless of stress pattern. Even if it is assumed that phonetic implementation rules allow for non-level realizations of individual tones, this fact is difficult to explain in a tone-to-syllable account.

Regarding the role of stress, the results in Figure 5 largely reinforce the results based on by-syllable f0 averages: differences in hypothesized stress pattern are associated with a change in the overall height of f0 across the entire target and not with a difference in the relative f0 of adjacent syllables as predicted by Ng's model. Additionally, the greater resolution of contour shape afforded by these results suggests that differences in stress may be cued partly by f0 alignment, and more specifically, by the timing of the f0 elbow. This effect is most apparent for two- and three-syllable single word targets (i.e., uS, Su, uSu, Suu) where the elbow occurs (and the rise begins) in the first syllable when stress is initial and in the second syllable when stress is on the second syllable. Crucially, this effect reflects a difference in the alignment of f0 targets, as opposed to the height of f0 associated with specific syllables (cf. Ng 2011). It is left to future work to assess the extent to which these alignment differences are systematic.

### 3.3 Other tonal patterns within the AP

In addition to the dominant pattern involving a rising f0 through the AP, our speakers produced a variety of less frequent, though recurring, patterns. As explained in Section 2.4, these were excluded in the quantitative analyses since they would tend to interfere with a rigorous assessment of predictions that specifically concern the most dominant global pattern. Ultimately, any phonological model of SgE intonation needs to account for such variation, and as suggested above it is possible that some of the factors that influence the choice of intonational tune may be sociolinguistic in nature. Here we describe three commonly recurring non-dominant patterns that appeared in our dataset.

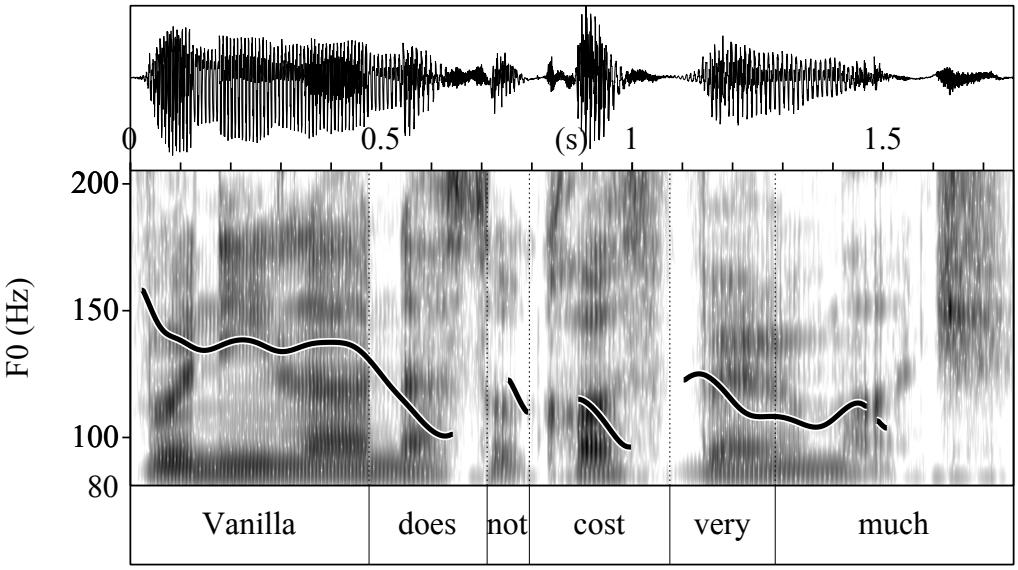



**Figure 6.** Waveform, spectrogram, and f0 track illustrating an initial AP produced with a high plateau (medial stress).

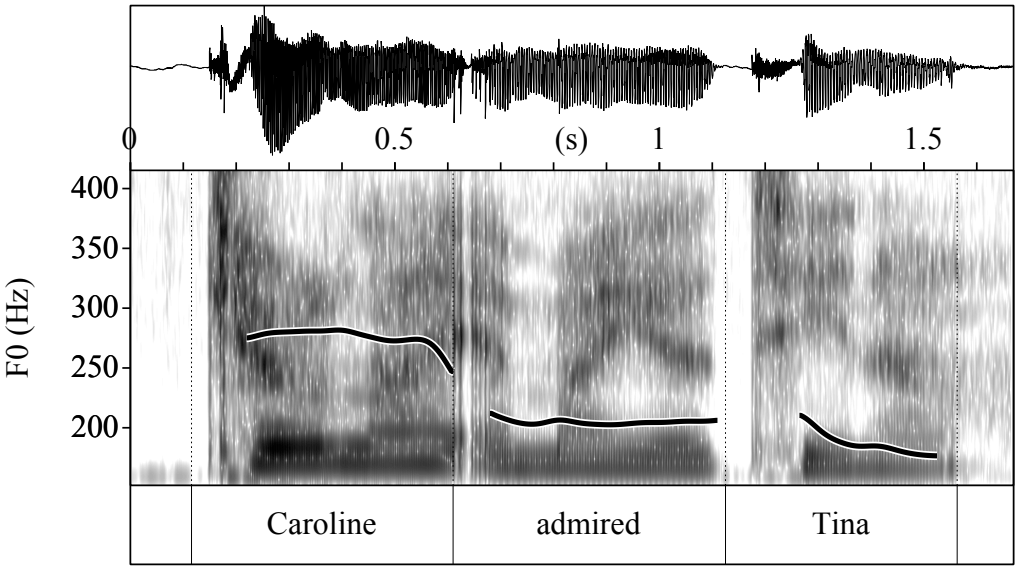



**Figure 7.** Waveform, spectrogram, and f0 track illustrating an initial AP produced with a high plateau (initial stress).

One commonly recurring pattern can best be characterized as a high f0 plateau stretching across the entire target (n = 47). Impressionistically, perceivable lengthening on the final syllable suggests that, as with the dominant pattern, this domain corresponds to an AP. An example of this is shown in Figure 6. This type of pattern has previously been observed in utterance non-initial position (see Chong 2012, 2013, German & Chong 2018, Chong & German 2019). Chong & German (2019) found that such plateaus in non-initial position are associated with APs that are shorter in duration, suggesting that they are possibly the result of an undershoot of the AP-initial L due to temporal compression. German & Chong (2018) also found for non-initial APs that plateaus occurred more often when stress was on the initial syllable. A consideration of the count of excluded tokens reveals that this latter point is corroborated in the current data, with most "plateau" exclusions involving single-word targets (n = 34, out of the total 47 such exclusions) arising from targets with a stressed initial syllable (28 out of 34, or 82%; Figure 7).

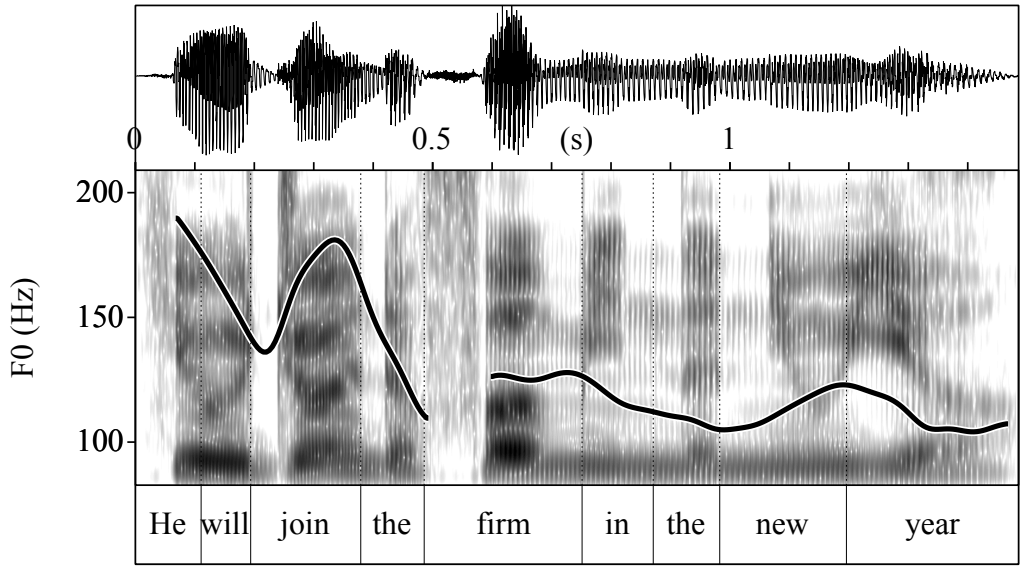


**Figure 8.** Waveform, spectrogram, and f0 track illustrating an initial AP produced with a falling-rising pattern.

We also observe two other tonal patterns that have not been previously reported in the literature, and to the best of our knowledge are not generally considered typical of this variety given previous descriptions. Figure 8 shows an example of a falling-rising contour realized within the target (n = 10). While it is plausible that this pattern corresponds to two separate APs (e.g., *[he][will join]*), there are two facts which suggest that it corresponds to a single AP. First, as with the pattern in Figure 6, there is perceivable lengthening only on the final syllable of the target (*join*) in Figure 8, thus overall, the rhythmic pattern is highly comparable to that of the "default" rising pattern. Second, in all other cases we have observed, the peak of the second AP (i.e., following the target) is subject to substantial downstep relative to the peak of the first AP (i.e., at the end of the target). In this case, however, the second peak in the target occurs at a very similar height to the first peak. We therefore propose instead that the initial high f0 either corresponds to an IP-initial boundary tone (similar to %H in Mainstream American English), or alternatively that HLH is a distinct tonal realization of the AP.

A third recurring pattern involved an f0 peak early in the target (n = 44). As shown in Figure 9, this occurred predominantly in single word targets, which again suggests that only a single AP is involved. Crucially, the peaks in all such examples occur towards the end of the second syllable. When lexical stress is expected on the second syllable, as with *Meridien* in Figure 9, this creates the rather marked impression that the speaker is producing a head-marking pitch accent as the result of switching to a Mainstream English-like system. However, in a small number of cases where lexical stress is expected on the

initial syllable (e.g., *Melanie*; Figure 10), the peak nevertheless falls at the end of the second syllable. Since the current data does not allow us to rule out the possibility that different occurrences of this pattern correspond to different behaviors, its status as either an instance of head-marking (possibly imported from Mainstream varieties) or as a distinct tonal realization of AP is unclear.

Previous models and descriptions of SgE intonation have dealt almost exclusively with the basic rising pattern (Chong 2012, 2013, Chow 2016, Deterding 1994). As a result, the occurrence of these less typical intonational patterns in our controlled sample poses an analytical challenge for any intonational phonological model of SgE intonation. An important question for future research, therefore, is the extent to which these patterns arise due to methodological factors (e.g., a university environment, laboratory speech) or to sociolinguistic factors, such as the degree to which participants orient towards Western varieties of English. We return to these points in the General Discussion in Section 6.

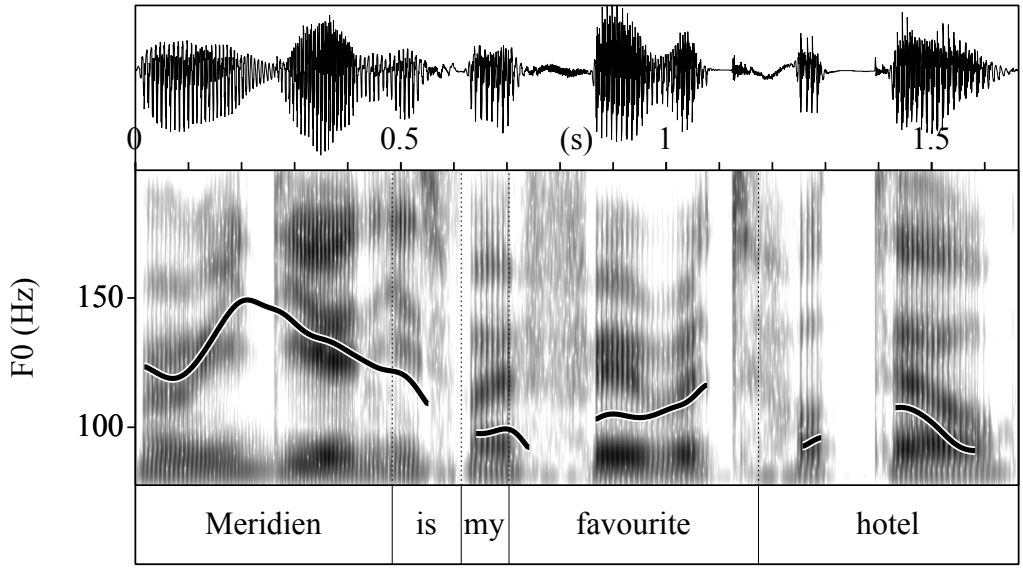


**Figure 9.** Waveform, spectrogram, and f0 track illustrating an initial AP produced with an f0 peak on the second, stressed syllable.

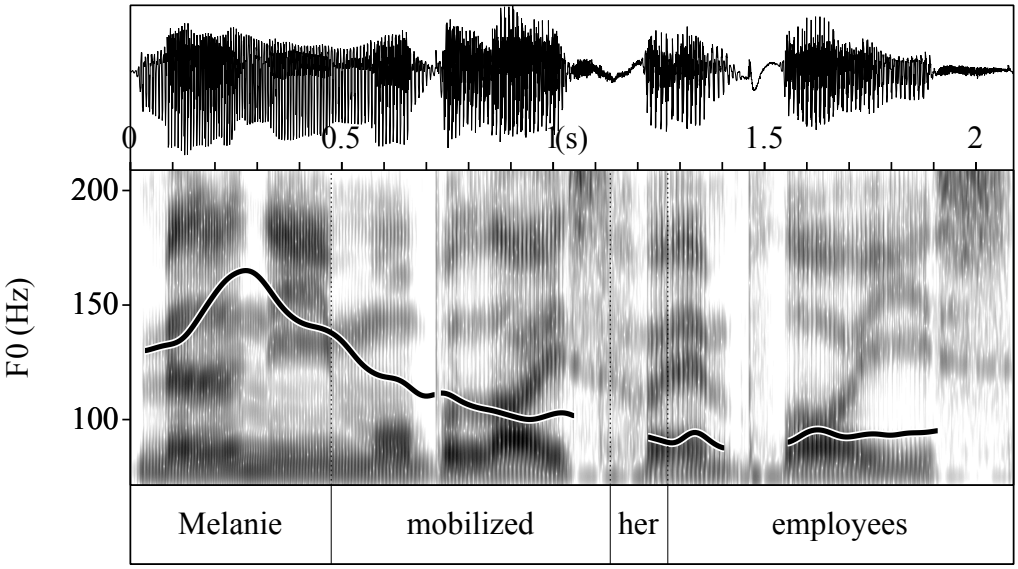



**Figure 10.** Waveform, spectrogram, and f0 track illustrating an initial AP produced with an f0 peak at the end of the second, unstressed syllable (stress-initial target).

### 3.4 Summary of f0 patterns

To summarize, our examination of f0 patterns as a function of putative stress placement revealed that overall, f0 is higher across an entire target when stress is on the initial syllable. Contrary to the predictions of Ng's (2009, 2011) model, we found no evidence for a specific tonal target associated with a stressed syllable. In other words, there is no evidence for a contrast between L and M on the initial syllable based on stress. Furthermore, a comparison of f0 contours across targets involving different numbers of syllables, differences in stress placement, and different syntactic composition revealed a highly general tendency for f0 to rise through the targets. The consistency of this pattern across

target types supports the central role of the AP as the domain of the f0 rise, as opposed to the prosodic word (cf. Chow 2016; Ng 2009, 2011), even if the internal structure of the contour might be sensitive to stress in a more fine-grained way. We return to this possibility in the General Discussion. These results also further corroborate Chong's (2012, 2013) proposal that the right edges of the AP are associated with a H tone, with the left edge marked by an L tone.

## 4. Other correlates of lexical stress

In the previous section, we established that lexical word stress plays a somewhat limited role in determining f0 patterns, in the sense that it does not determine the locations of major pitch events as in Mainstream varieties. To assess whether stress may be marked by other acoustic correlates, we conducted an exploratory analysis of duration and intensity as a function of putative stress placement. Variation in duration and intensity may be driven by a number of prosodic factors other than lexical stress, including, in particular, position within a prosodic phrase. In the case of SgE, Chong & German (2017) showed that vowels in AP-final syllables are associated with longer durations relative to those in word-final AP-medial syllables. Intensity was not examined in that study. For the present study, we therefore expect that AP-final syllables undergo significant lengthening independently of where stress occurs. The important question then is whether the lexically stressed syllable leads to increases in duration and intensity over and above that which is conditioned by phrasal position. For purposes of better comparison with the analysis of f0 above, and to more directly assess the relationship between stress-based and positional effects, our

analysis focuses on the two- and three-syllable single word targets (n = 700) examined in Section 3.1 above.

### 4.1 Results: Duration

For two-syllable targets, as can be seen in the two upper panels of Figure 11, for both males and females, stressed syllables are longer than unstressed syllables, though the magnitude of the within-word difference was smaller for Su words than uS words. The best fit model for duration in two-syllable targets was first obtained using model comparison, which indicated that the inclusion of the two-way interaction significantly improved model fit ($\chi^2(1) = 138.7$, $p < 0.001$). There was no significant main effect of stress pattern ($F(1, 14) = 2.92$, $p = 0.11$), though there was a significant effect of syllable number ($F(1, 695) = 153.1$, $p < 0.001$). There was no significant effect of speaker gender ($F(1,22) = 1.44$, $p = 0.24$). To explore the interaction, post-hoc pairwise comparisons of syllable duration within words were carried out for each stress pattern (Table 3). The results confirm that successive syllables differed significantly in duration, though the direction of this effect was different for the two stress patterns. For Su words, the first, stressed syllable was longer than the second unstressed syllable, while the reverse was true for uS words. While our study does not control specifically for the effect of phrasal position, AP-final lengthening appears to be limited in Su words, in that the second syllable in those words was comparable in duration to the first syllable of uS words. Further, we compared the difference in duration within syllable, but across target type (and therefore stress position). The first, stressed vowel in Su words was significantly longer than the first, unstressed vowel in uS words,

and the reverse was true for the second, stressed vowel in uS words compared to the second, unstressed vowel Su words (Table 3).

As shown in the two lower panels of Figure 11, for three-syllable words the phrase-final (unstressed) syllable had the longest duration regardless of stress pattern. The most parsimonious model included the interaction of stress pattern by syllable number ($\chi^2(2) = 116.6$, $p < 0.001$). There was no main effect of stress pattern ($F(1, 13.8) = 3.36$, $p = 0.088$) but there was a main effect of syllable number ($F(2, 949) = 227.0$, $p < 0.0001$). As with two-syllable words, there was no significant effect of gender ($F(1,22) = 2.77$, $p = 0.11$). Due to the significant interaction of stress pattern and syllable number, we performed post-hoc pairwise comparisons (Table 4). The results indicate firstly that in both stress patterns, the final vowels were longer than all preceding vowels, indicative of phrase-final lengthening. For both stress patterns, stressed vowels were longer than unstressed vowels, though the magnitude of the difference between the first and second syllables was smaller in Suu words than in uSu words. Further, when we compare the difference in duration of each syllable, across target type (and therefore stress position), we find that the duration of vowels in the final syllable in Suu and uSu words was not significantly different. However, vowels in the first stressed syllable in Suu words were significantly longer than those in the first unstressed syllable in uSu words. The reverse was true for the second syllable in Suu vs. uSu words, as here vowels in the second syllable of the latter were longer.

To summarize, our exploratory analysis of duration as a correlate of prominence corroborates the prediction based on Chong and German (2017) that phrase-final syllables are in general longer than non-final syllables. We also found that stressed syllables are generally longer than unstressed syllables, though at least for three-syllable words, this

tendency is overridden by phrase-final lengthening. Interestingly, the final syllables in Su targets were not longer than the initial syllables, suggesting that final lengthening may be limited or absent in that context. Overall, our analysis provides further evidence for prominence-driven durational effects, which are conditioned primarily by phrasal position, and to a lesser extent by lexical word stress.

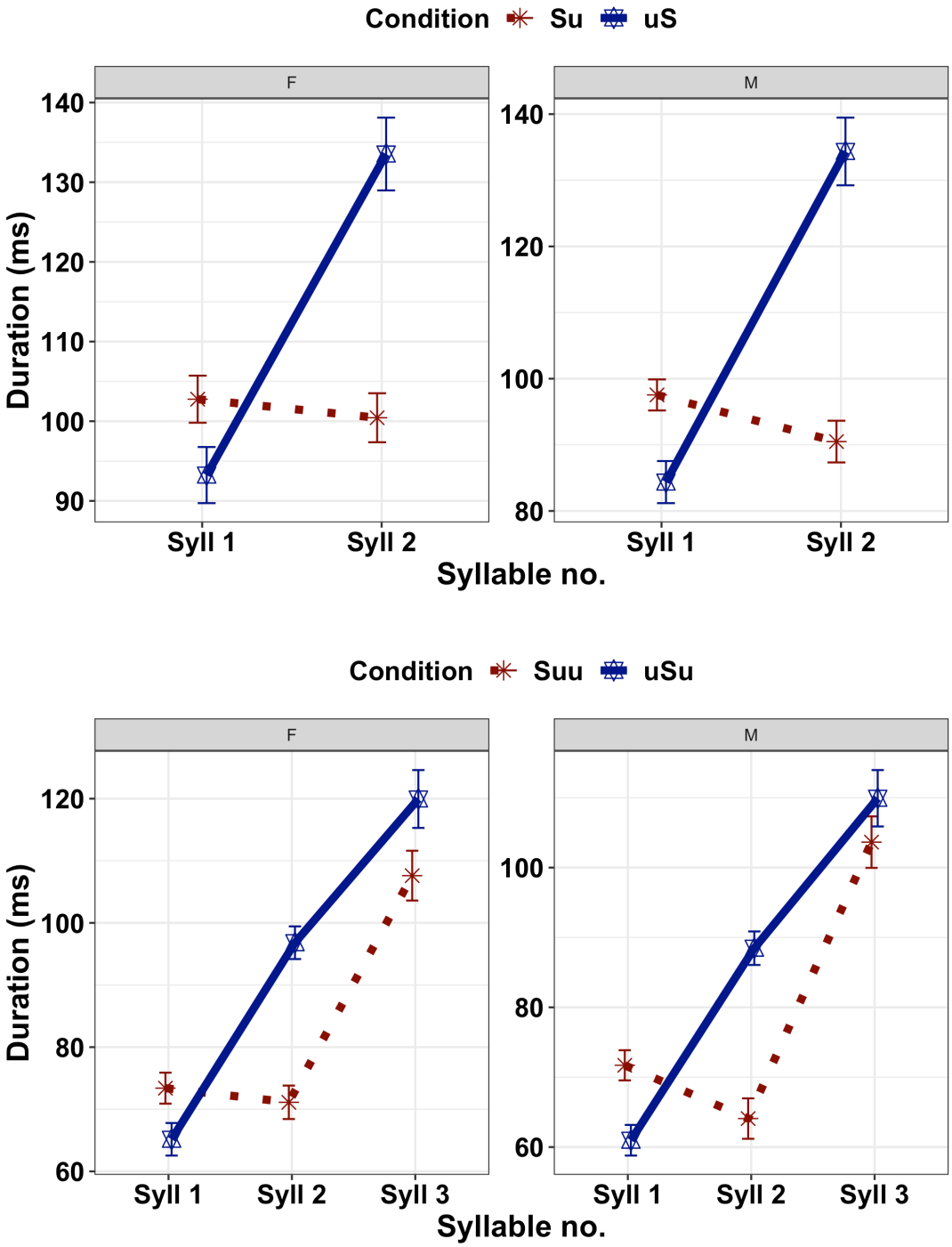



**Figure 11.** Mean raw untransformed duration (ms) with 95% CI by syllable number, stress pattern, and speaker gender for two- (upper) and three- (lower) syllable targets.

| | Est. | SE | df | t.ratio | *p* |
|---|---|---|---|---|---|
| Su: Syll 1 vs. Syll 2 | -0.03 | 0.012 | 695 | -2.42 | 0.02 |
| uS: Syll 1 vs. Syll 2 | 0.18 | 0.012 | 695 | 15.08 | <0.0001 |
| | | | | | |
| | est. | SE | df | t.ratio | *p* |
| Syll 1: Su vs. uS | -1.15 | 0.024 | 18.4 | -6.02 | < 0.0001 |
| Syll 2: Su vs. uS | 0.07 | 0.024 | 18.4 | 2.83 | 0.01 |

**Table 3.** Post-hoc comparisons of log duration for two-syllable targets.

| | est. | SE | df | t.ratio | *p* |
|---|---|---|---|---|---|
| Suu: Syll 1 vs. Syll 2 | 0.05 | 0.015 | 949 | -3.21 | 0.0039 |
| Suu: Syll 2 vs. Syll 3 | -0.21 | 0.015 | 949 | -14.19 | <0.0001 |
| Suu: Syll 1 vs. Syll 3 | -0.16 | 0.015 | 949 | -10.93 | <0.0001 |
| uSu: Syll 1 vs. Syll 2 | 0.18 | 0.014 | 949 | -2.42 | <0.0001 |
| uSu: Syll 2 vs. Syll 3 | -0.09 | 0.014 | 949 | 15.08 | <0.0001 |
| uSu: Syll 1 vs. Syll 3 | -0.27 | 0.014 | 949 | -18.60 | <0.0001 |
| | | | | | |
| | est. | SE | df | t.ratio | *p* |
| Syll 1: Suu vs uSu | 0.70 | 0.025 | 23 | 2.77 | 0.01 |
| Syll 2: Suu vs uSu | -0.16 | 0.025 | 22.9 | -6.24 | <0.0001 |
| Syll 3: Suu vs uSu | -0.04 | 0.025 | 22.9 | -1.39 | 0.18 |

**Table 4.** Post-hoc comparisons of log duration for three-syllable targets.

### 4.3 Results: Intensity

Intensity profiles for two- and three-syllable target words are presented in Figure 12. For two-syllable targets (Figure 12 upper), intensity increases across the target regardless of stress pattern, though intensity in the first syllable is higher for Su words. The final model included the two-way interaction of stress pattern and syllable number ($\chi^2(1) = 14.08$, $p = 0.0002$). There was a significant main effect of stress pattern ($F(1, 14) = 6.43$, $p = 0.02$) and syllable number ($F(1, 695) = 23.31$, $p < 0.001$). There was, however, no effect of gender ($F(1,22) = 1.06$, $p = 0.30$). Post-hoc comparisons revealed that the interaction is driven by the fact that the unstressed first syllable in uS had a significantly lower intensity than the final stressed syllable, while there was no difference between syllables in Su words.

The first unstressed syllable in uS words also had lower intensity than the first stressed syllable in Su words (see Table 5).

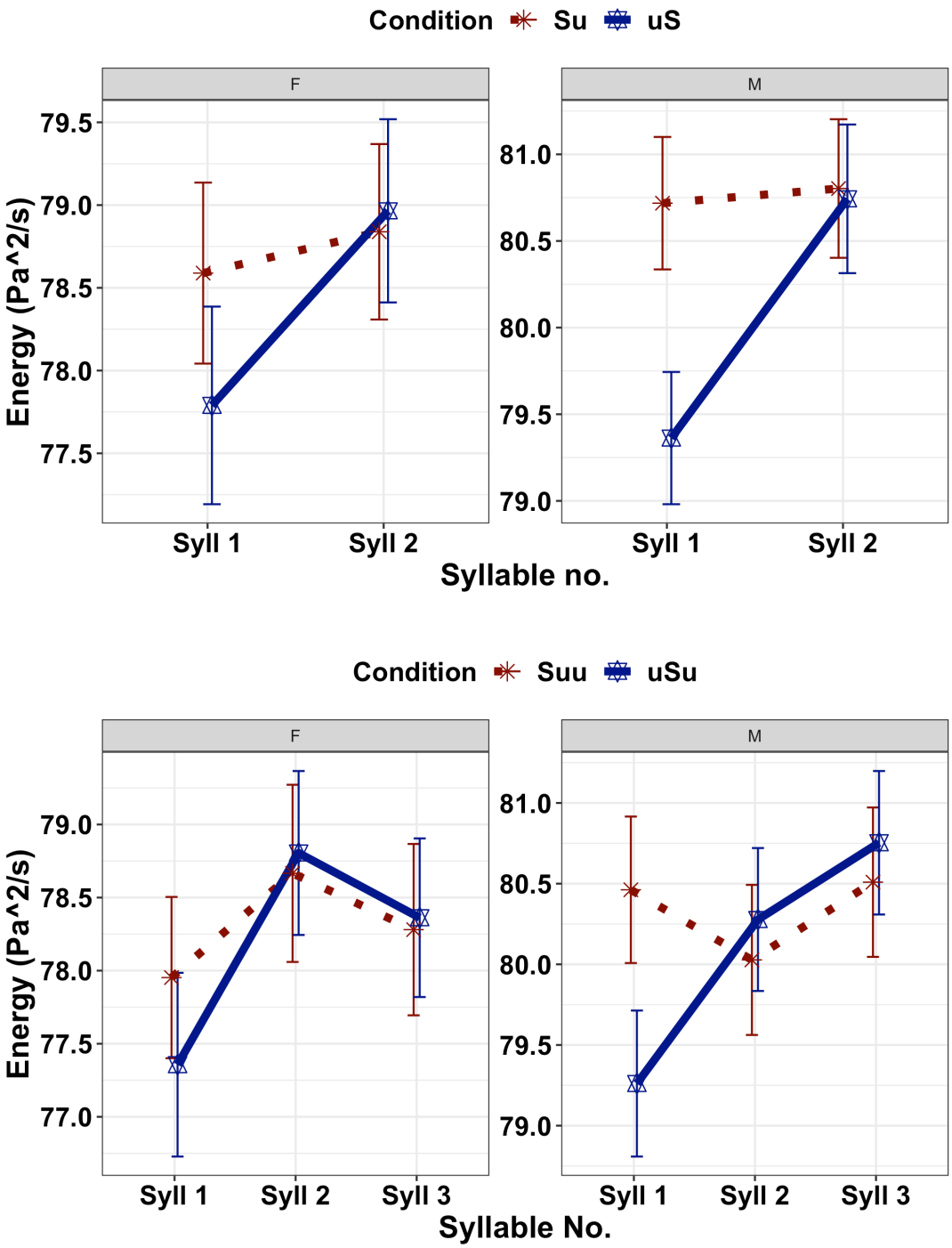



**Figure 12.** Mean intensity with 95% CI by syllable no. by word type by speaker gender for two- (upper) and three- (lower) syllable targets.

As with two-syllable targets, the initial syllable in three-syllable targets (Figure 12, lower) shows a higher intensity when stressed (Suu) than when unstressed (uSu). In both cases, intensity rises to the second syllable, though much more sharply for uSu targets, resulting in similar intensity values for the two patterns in the second and third syllables. The most parsimonious model included the two-way interaction of stress pattern and syllable number ($\chi^2(2) = 17.10$, $p = 0.0002$). There was no significant main effect of stress pattern ($F(1, 14) = 1.64$, $p = 0.22$), but there was a significant effect of syllable number ($F(2, 949) = 11.83$, $p < 0.001$). There was no significant effect of gender ($F(1, 22) = 0.98$, $p = 0.33$). Post-hoc comparisons revealed that the interaction was driven by the fact that the unstressed first syllable in uSu had a significantly lower intensity than both the medial stressed syllable and final syllable, while there was no difference between syllables in Suu words. As with two-syllable targets, intensity was significantly lower on the first unstressed syllable in uSu words than the first stressed syllable in Suu words (see Table 6).

| | est. | SE | df | t.ratio | *p* |
|---|---|---|---|---|---|
| Su: Syll 1 vs. Syll 2 | 0.16 | 0.21 | 695 | 0.75 | 0.45 |
| uS: Syll 1 vs. Syll 2 | 1.29 | 0.21 | 695 | 6.08 | <0.0001 |
| | | | | | |
| | est. | SE | df | t.ratio | *p* |
| Syll 1: Su vs. uS | 1.15 | 0.28 | 28.1 | 4.18 | 0.0003 |
| Syll 2: Su vs. uS | 0.02 | 0.28 | 28.1 | 0.08 | 0.94 |

**Table 5.** Post-hoc comparisons for intensity for two-syllable targets.

| | est. | SE | df | t.ratio | *p* |
|---|---|---|---|---|---|
| Suu: Syll 1 vs. Syll 2 | 0.06 | 0.22 | 949 | 0.28 | 0.96 |
| Suu: Syll 2 vs. Syll 3 | -0.07 | 0.22 | 949 | -0.30 | 0.95 |
| Suu: Syll 1 vs. Syll 3 | -0.13 | 0.22 | 949 | -0.58 | 0.83 |
| uSu: Syll 1 vs. Syll 2 | 1.18 | 0.22 | 949 | 5.50 | <0.0001 |
| uSu: Syll 2 vs. Syll 3 | -0.04 | 0.22 | 949 | -0.18 | 0.98 |
| uSu: Syll 1 vs. Syll 3 | -1.22 | 0.22 | 949 | -5.68 | <0.0001 |
| | | | | | |
| | est. | SE | df | t.ratio | *p* |

| | | | | | |
|---|---|---|---|---|---|
| Syll 1: Suu vs uSu | 1.03 | 0.29 | 36 | 3.55 | 0.001 |
| Syll 2: Suu vs uSu | -0.09 | 0.29 | 36 | -0.31 | 0.76 |
| Syll 3: Suu vs uSu | -0.06 | 0.29 | 36 | -0.21 | 0.83 |

**Table 6.** Post-hoc comparisons for intensity for three-syllable targets.

Taken together, these results suggest that intensity plays a role in cueing prominence related to stress, and given that intensity is consistently high in the final syllable of the AP, also phrasal position.

## 5. Individual differences

Before turning to the General Discussion, we present a brief review of individual differences that emerged in our study, focusing on f0 and duration. The general pattern in our findings involved an overall higher f0 across the entire AP when stress occurs on the initial syllable. Furthermore, stressed syllables were generally longer than unstressed syllables, though the strongest influence on duration was related to phrasal position, namely, the AP-final syllable. On an individual level, however, there was some variation not only in the extent to which speakers' productions showed sensitivity to the pattern of stress, but also in terms of how they made use of different acoustic cues to realize differences in stress pattern. As shown in Figure 13, for example, Speaker 4 exhibits a very small effect of stress pattern on f0 scaling for both two- and three-syllable targets. With respect to duration, this speaker exhibits no clear effect of stress for three-syllable targets, and while two-syllable targets follow the same overall pattern of interaction as the general finding, duration for Su words actually increases across the target rather than decreasing.

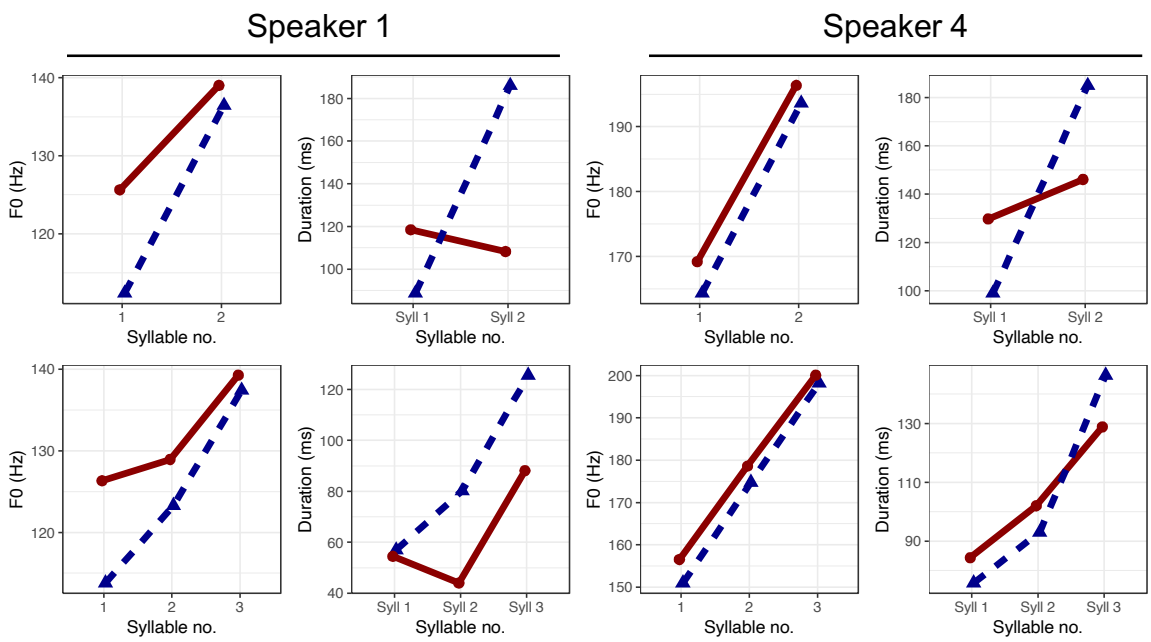



**Figure 13.** F0 and duration by syllable and stress pattern for two-syllable (upper) and three-syllable (lower) targets for two speakers. Red solid = Su, Suu, Blue dashed = uS, uSu.

By comparison, Speaker 1 shows an effect of stress on the f0 of the first syllable that is much stronger than the general pattern, while on subsequent syllables the differences are much smaller than in the general pattern. Such examples suggest that at least some speakers use f0 to mark the stress status of individual syllables, at least for initial syllables. For duration, however, Speaker 1 closely follows the general pattern for both two- and three-syllable targets.

An examination of all individual speakers in our data set, however, reveals that the general rising pattern in f0 across the AP is highly consistent, with a final peak on the AP-final syllable, though subtle differences in the alignment of the start of the rise, as well as the slope, occur across individuals and as a function of stress pattern. While we do not find any evidence for f0 peaks or valleys that align with putatively stressed syllables, we leave

open the possibility that the exact timing of the f0 rise is sensitive to stress pattern (see Section 3.2). Regarding duration, it is clear that AP-final syllables are consistently lengthened, with less consistent lengthening of pre-final lexically prominent syllables. In some cases, speakers show no clear evidence of lengthening associated with the location of stress. We discuss the implications of these results for a model of lexical and phrasal prominence in SgE below in the General Discussion.

## 6. General Discussion

In this study, we sought to assess the extent to which details of f0 contours can be explained by differences in lexically determined word stress in SgE. We further explored the extent to which lexical word stress influences other acoustic parameters, namely duration and intensity. In the sections below, we discuss the implications of our findings for the acoustic realization of both stress and phrasal prominence in SgE (Section 6.1), as well as for an intonational phonological model of SgE more generally (Section 6.2). We then discuss the implications of these findings for prosodic typology (Section 6.3) and for the study of intonation in Contact varieties of English more broadly (Section 6.4).

### 6.1 Sources of prominence in SgE

The present study was primarily concerned with assessing whether the details of f0 contours are influenced by differences in the position of lexically distinctive, or "stressed", syllables within words. To do this, we first measured the mean f0 of syllable nuclei in two- and three-syllable target words occurring in utterance-initial position which differed in the assumed position of lexical word stress – on either the first or second syllable. Our main

finding is that regardless of the position of stress, f0 rises through the target word, with some exceptions showing a non-rising pattern. Crucially, the position of stress had no discernible effect on the relative height of f0 in adjacent stressed and unstressed syllables. Instead, mean f0 was higher across the entire target when stress was on the initial syllable as compared to on the second syllable. In other words, for stress-initial targets, mean f0 begins at a higher value and continues to rise over the target at a similar rate with the result that the f0 contour shape of both first- and second-syllable stress targets is highly comparable. Stress, therefore, appears to influence primarily the scaling of f0 within the target region. Our study also examined whether the relationship between stress position and f0 is conditioned by target length (in terms of number of syllables) and syntactic composition (i.e. a content word with or without preceding function words). Our results suggest that the rising pattern consistently occurs regardless of these differences in the composition of the target region.

Previous proposals, particularly by Ng (2009, 2011; see also Lim 2004), suggest that f0 patterns in SgE are determined by tonal specifications on each syllable in a word-level prosodic domain, and that these specifications are realized as level tones. Ng (2009 2011) further proposes that the distinction between initial and non-initial stress in a word is marked by a phonological contrast in the value of the tone that is assigned to the initial syllable (i.e., **L**H versus **M**H for two-syllable words, and **L**MH versus **M**MH for three-syllable words). Our findings are incompatible with these proposals for several reasons. First, while mean f0 was overall higher on the initial syllable for words with initial stress, this difference continued throughout the target word, which would be unexpected if LH contrasts with MH and LMH with MMH. Additionally, the average size of the difference

in f0 for the first syllable as a function of stress location was approximately 10 Hz. This is well below the Just Noticeable Difference (JND) required for the maintenance of a phonological contrast in tone, which is 20-30 Hz in the general case ('t Hart 1981), and even higher in sentential contexts (Harris & Umeda 1987).

Our findings are also problematic for earlier proposals in that f0 was not realized as level tones on individual syllables. Instead, the dominant pattern involved a contour which rises from a low f0 early in the target word or phrase to a high f0 late in the target. This was true even for monosyllabic targets, a fact which would be surprising if those targets included only a single tonal specification. Our findings are therefore much more compatible with a model in which f0 patterns are determined by tonal specifications at the level of a phrasal domain, as proposed by Chong (2012, 2013) and Chong and German (2017). More specifically, the generality of the rising pattern supports a model in which L and H tonal targets are assigned to the left and right edges of an Accentual Phrase, and language-specific phonetic implementation rules determine how f0 interpolates between them, resulting in systematic variation (Arvaniti & Ladd 2009) dependent on the length and syntactic composition of targets.

Differences in the location of stress appear then not to manifest through phonological contrast *per se*, but rather through effects on phonetic implementation that affect the scaling of f0 at the level of a prosodic domain (here, the AP). One possible source of this scaling effect is the generally higher intensity associated with the first syllable when it is stressed. The positive correlation between intensity and f0 is well-documented (Gramming, Sundberg, Ternström, Leanderson, & Perkins 1988). It is possible, therefore, that higher intensity on the first syllable leads to a higher starting point for f0 which then

remains higher as a consequence of fully general implementation rules. A second possibility is that this apparent scaling effect is caused by differences in the alignment of L*. The visualizations of f0 contours across conditions in Figure 5 suggest that the rise to the target final peak begins earlier when stress is on an earlier syllable, a tendency which is consistent with Chong's (2012, 2013) proposal that L* associates to a stressed syllable. Assuming that there is interpolation between the L* and the AP-final H, then it is expected that f0 would tend to be higher over all syllables following L* when stress is earlier in the AP. Additional research is needed, however, to more conclusively establish the source of this scaling effect.

Our study also explores whether and how lexical and phrasal prominence is cued by duration and intensity. Consistent with previous findings by Chong and German (2017), our results showed that AP-final syllables are generally associated with longer duration. Our results also revealed some evidence that stress patterns are cued by differences in duration over and above those determined by phrasal position. Specifically, stressed syllables were longer in duration than adjacent unstressed syllables. The only exception to this was in uSu words, where final lengthening resulted in a final unstressed syllable that was longer than the preceding stressed syllable. This finding can be explained if the effect of final lengthening is greater in magnitude than stress-based lengthening, such that the contrast associated with the latter is maintained even if its effects are "overridden" by the former. In other words, a pre-final syllable is longer when stressed than when it is unstressed, even if in absolute terms it is shorter than following unstressed final syllable.

One exception to the otherwise general pattern of phrase-final lengthening occurred in two-syllable targets with stress on the first syllable (i.e., Su targets), since in that case,

the final syllable was actually shorter than the first. This pattern closely matches that found in the first two syllables of three-syllable targets, suggesting that unstressed syllables immediately following a stressed syllable are shortened, perhaps as part of a mechanism to reinforce the contrast. If this analysis is correct, then our findings suggest that, in disyllabic words at least, the effects of stress pattern on duration override those of phrasal position for final unstressed syllables immediately following stressed syllables. Further studies are needed to assess this apparent interaction between stress and position on duration.

Our findings in relation to intensity revealed that AP-final syllables are always associated with higher intensity, at least in utterance-initial APs. The main locus of stress-related intensity differences is on the first syllable, in that initial stressed syllables showed a higher intensity than initial unstressed syllables, while intensity in other positions was highly similar regardless of stress status. Interestingly, in stress-initial tri-syllabic targets, intensity is relatively constant across the entire word, starting at a higher level than stress-medial targets.

Taken together, the results of the current study support a core aspect of the intonational model proposed by Chong and German (2017), in that f0 in SgE is used primarily to mark the edges of phrasal units. Across AP targets which differ in terms of syllable number, stress pattern, and syntactic composition, the clearly dominant pattern involves a rising f0, with some evidence that the f0 turning point is sensitive to the position of stress (see Figure 5). Furthermore, and contrary to Ng's proposal, our study found no evidence that f0 height on *individual* syllables is sensitive to stress, at least for the utterance-initial targets examined in the current study. Instead, it appears that the lexical status (as stressed or not) determines the starting height of f0 at the left edge of the phrase,

which then rises to the AP-final boundary tone according to language-specific phonetic implementation rules. Overall, this pattern of variation suggests that f0 in SgE is used primarily for post-lexical prominence marking, marking edges of prosodic domains rather than heads. Together with the findings of Chong & German (2017) the results of the present study also corroborate the important role of a phrasal unit, the Accentual Phrase, which lies hierarchically between the word and the intonation phrase (cf. Chow (2016) which does not posit any role for a phrasal structure between these levels).

All in all, our examination of f0, duration, and intensity revealed that the AP-final syllable is the principal locus of prominence marking, in that it is associated with a high f0 (a H tone), significantly longer duration, and higher intensity (or plateau intensity). Stressed syllables are most reliably marked by a marginally longer duration and higher intensity. Even so, with the exception of duration in two-syllable APs, the effects of phrasal position on these cues override those of stress pattern. The confluence of acoustic cues on the AP-final syllable may therefore explain Tan's (2015) finding that SgE listeners often perceive prominence on the final syllable of words.

### 6.3 Variation in SgE intonation

No discussion about SgE can ignore the extent of inter- and intra-individual variation that exists within that variety. While the current study does not address the issue of how best to model this variation (see the Introduction for an overview of existing proposals), common to all existing models is the need to identify linguistic features that vary either by code (in a diglossic approach, e.g., Gupta 1998), location on a lectal continuum (e.g., Pakir 1991), or as a function of indexical stance (under an indexicality approach, e.g., Leimgruber 2013).

Our current study sheds light on several features of the prosodic system that are potentially relevant.

First, while our study involved a relatively well-controlled set of participants with similar language and education backgrounds, patterns emerged which are unexpected or previously unobserved for the variety (cf. Ng 2009, 2011; Lim 2004). This includes, in particular, cases where the f0 peak appears to be realized on a stressed syllable rather than AP-finally (e.g., Figure 8). Such evidence for potential head-marking when the predominant pattern involves edge-marking poses an analytical challenge for modelling the intonational phonology of SgE. For one, it is unclear whether such patterns should be regarded as part of the SgE system proper or as being drawn from some other variety. Here, an indexical fields approach (Eckert 2008) towards SgE (following Leimgruber 2013) provides an alternative to an approach which seeks to artificially force all features into specific varieties. Given the head-marking nature of intonation in Mainstream varieties of English, it is plausible that the use of a peak aligned to a lexically stressed syllable is the consequence of a socially-motivated choice to index a particular cultural orientation, namely a Global one (Alsagoff 2007, Leimgruber 2013) as compared to a Local one. Leimgruber (2013) and others discuss features primarily in comparing Colloquial SgE vs. Standard SgE. Here, however, we suggest that this approach could also be used to explain the use of intonational features typical of Mainstream English varieties.

Second, by adopting an approach grounded in AM, we are able to consider whether observed variation in prosodic patterns occurs at the phonological (i.e., structural or representational) level or at the level of phonetic implementation. Our results show that a central *phonological* feature of SgE is that phrasal structure is the primary determinant of

f0, and that in the dominant pattern this involves a rise to a H edge tone associated with the end of the AP regardless of the stress pattern. Some types of variation, such as the difference between the general rising pattern, the high plateau (Figure 6), the falling-rising pattern (Figure 8), and the head-aligning pattern (Figure 9) will most likely turn out to be best explained by phonological differences in the tonal sequence associated with AP (c.f. Chong & German 2019). Other types of variation, such as differences in the timing or curvature of f0 or in the specific ways in which stress pattern affects the scaling and slope of the rise, are best explained by differences in phonetic implementation (e.g. Figure 5). As discussed in Section 1.3, while most previous work has sought to model the prosody of Colloquial SgE, as opposed to the more "standard" SgE productions used in the present study, we are nonetheless confident that many of the observations as well as many aspects of the phonological model of intonation argued for here may be extended to Colloquial SgE. The global phonetic patterns of intonation in Colloquial SgE (e.g. Ng 2009, 2011) and Standard SgE (the current work) are impressionistically quite similar: the overall tonal melody is a rise, with the domain being roughly a single content word (though Ng (2009, 2011) does not compare multiword targets). Nevertheless, the findings presented here provide a basis for future research aimed at comparing intra- and inter-speaker intonational variation in SgE. For example, ongoing work by Sim and Post (2021) similarly find support for distinguishing between variation at the level of phonology versus phonetic implementation. That study explored the acquisition of intonational features by SgE children from different L1 backgrounds (Mandarin Chinese and Malay) and found that while there is a high degree of consistency in global contour shape used by these children (rising), the primary source of variation involves differences in relative scaling and height.

Finally, we note that a large number of studies on the prosody of Contact English varieties emphasize either the characterization of rhythm using various rhythm metrics (e.g., Deterding (2001) for SgE) or the description of global phonetic parameters such as pitch span, dynamism, etc. (e.g., Meer & Fuchs 2022). Such metrics not only yield different findings depending on the specific corpora they are applied to (Arvaniti 2009), but more importantly they provide only descriptive adequacy in that they cannot reveal differences in the phonological patterning that underlies variation in these measures across varieties. In other words, while such measures may shed light on *some* aspect of the linguistic system, they do not necessarily identify the structural *sources* of these differences (see Arvaniti (2009) for a discussion). By directly addressing the underlying phonological structure of the intonation system of a particular variety, it is not only easier to identify the sources of observed variability, but it also makes it possible to more precisely situate SgE within a typological framework of prosodic systems, both cross-varietally and cross-linguistically.

### 6.4 Implications for intonational typology

Finally, our results have implications for our understanding of intonational typology. We found that lexical word stress is largely associated with increased duration (with a lesser role for intensity). This suggests that, at least for a subset of the speakers in our study, word-level representations include a lexical specification for prosodically distinctive syllables, even if the acoustic prominence of the associated surface cues are weaker compared to Mainstream varieties of English (though see Cutler (1986) for some evidence that suprasegmental cues do not necessarily play a role in Mainstream varieties as well). By comparison, variation in f0 was not consistently associated with specific lexically

determined syllables, but instead is determined by prosodic phrasing (see also Chong & German 2017). If correct, this analysis suggests that SgE can be added to the purportedly rare class of languages that have been characterized as possessing lexical word stress but without pitch accents on stressed syllables. For languages like Kuot (Lindström & Remijsen 2005), Wolof (Rialland & Robert 2001), Uyghur (Major & Mayer 2018), and according to some studies, certain varieties of Indonesian (van Zanten & ven Heuven 1998, van Zanten et al. 2003), f0 has been analyzed as only performing an edge-marking function, with no evidence for post-lexical pitch accents associated to stressed syllables (see Gordon (2014), Roettger & Gordon (2017) for a discussion on disentangling word-level stress from pitch movements attributable to phrasal units larger than a word).

Our current study also highlights how the examination of Contact varieties of English can contribute to intonational typology (see Grice, Warren & German 2020). To date, relatively less attention has been paid to such Contact varieties as compared to Mainstream varieties. Crucially, however, Contact varieties have complex histories of language contact resulting in the high level of diversity that we observe across their synchronic systems. Furthermore, many Contact varieties exist within complex and highly multilingual linguistic ecologies, a condition which provides fertile ground for examining the effects of language contact (e.g., Baltazani, Przedlacka & Coleman 2020 on Greek) and multilingualism (e.g., Maxwell & Payne 2021 on L1 effects in Indian English) on intonational grammars and on the details of intonational phonetic implementation.

### 6.5 Conclusion

In this study we examined the extent to which lexical prominence in SgE is marked by f0, duration and intensity. We found that stressed syllables are typically marked by a longer duration and higher intensity. We did not find conclusive evidence for f0 valleys or peaks associated with stressed syllables, although there was suggestive evidence for the role of stress in determining L* alignment. Instead, our principle finding was that f0 was scaled higher across the AP when stress was on the initial syllable of a target word. Furthermore, word-final (and AP-final) syllables were also associated with higher f0, longer duration, and higher intensity, with the magnitude of the f0 and duration rise being larger than the difference between non-final stressed and unstressed syllables. We interpret this finding as evidence for the central role of post-lexical prosodic prominence associated with the edges of phrasal units over more weakly cued lexical prominence relations. Further evidence that f0 is predominantly determined by phrasing was found by visualizing and qualitatively examining global f0 contours over APs of different sizes, stress placement and constituency. Given the differences in duration and f0 found in the current study, future work is needed to determine the degree to which these differences are perceptually meaningful to speakers of SgE. Further, given the evidence for weakly cued lexical word stress, the study of how focus is realized in SgE (i.e. whether specific syllables are targeted for focal enhancement) should aid in further clarifying the role of head vs. edge prominence in SgE. Finally, the results of our study provide a foundation for further exploration the ways in which speakers of different sociolinguistic backgrounds vary in terms of intonation.

**Disclosure statement**

Neither authors have any potential conflicts of interest.

**Acknowledgements**

This research was supported by grants ANR-16-CONV-0002 (ILCB) and the Excellence Initiative of Aix-Marseille University (A*MIDEX), by the A*MIDEX-funded SIRL project (n° ANR-11-IDEX-0001-02), by a grant from the Singapore Ministry of Education Academic Research Fund Tier 1 (2013–T1–002–169) to the second author, as well as funds from the UCLA Department of Linguistics to the first author. We would like to thank Marc Garellek, Sameer Khan, 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, UCL Speech Science Forum, Cambridge Phonetics & Phonology Seminar, National University of Singapore, Queen Mary University of London, Lancaster Phonetics Lab and ICU Prosody Seminar for helpful comments and discussion. We would also like to thank Andy Xu, 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. An earlier version of this work appeared as part of the Proceedings of the 18th ICPhS (Chong & German 2015). This research was conducted with the approval of the Institutional Review Board at Nanyang Technological University, Singapore (NTU IRB: 2013-08-001). We would finally like to thank Jeff Mielke, and three anonymous reviewers for their helpful comments and feedback. All remaining faults are our own.

**References**


Alsagoff, L. (2007). Singlish: Negotiating culture, capital and identity. In Vaish, Viniti, Gopinathan, Saravanan, & Liu, Yongbing (eds.), *Language, Capital, Culture* (pp.24-46). Rotterdam: Sense Publishers.

Alsagoff, L. (2010). English in Singapore: culture, capital and identity in linguistic variation. *World Englishes, 29,* 336-348.

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica, 66*, 46–63.

Arvaniti, A. (2011). The representation of intonation. In Marc van Oostendorp, Colin Ewen, Elizabeth Hume and Keren Rice (eds), *The Blackwell Companion to Phonology*. Wiley-Blackwell.

Arvaniti, A. (2022). The autosegmental metrical model of intonational phonology. In Shattuck-Hufnagel, S. & Barnes, J. (eds.). *Prosodic Theory and Practice*. Cambridge: MIT Press.

Arvaniti, A. & Ladd, D. R. (2009). Greek wh-questions and the phonology of intonation. *Phonology*, 26, 43-74.

Baltazani, M., Przedlacka, J., & Coleman, J. (2020). Intonation of Greek-Turkish contact: A real-time diachronic study. *Proceedings of the 10th International conference on Speech Prosody,* 730-734.

Bao, Z. (1998). The sounds of Singapore English. In J. A. Foley et al., *Englishes in New Cultural Contexts: Reflections from Singapore* (pp. 152-174). Singapore: Oxford University Press.

Bao, Z. (2006). Clash avoidance and metrical opacity in Singapore English. *Sprachtypologie und Universalienforschung, 59*(2), 133-147.

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

Beckman, M. E. (1986). *Stress and non-stress accent.* Dordrecht: Foris.

Beckman, M. E., & Ayers, G. M. (1994), *Guidelines for ToBI Labelling*. Online MS and accompanying files. Available at http://www.ling.ohio-state.edu/phonetics/E_ToBI.

Beckman, M. E., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In P. A. Keating (ed.), *Phonological structure and phonetic form: Papers in Laboratory Phonology III* (pp. 7-33). Cambridge: Cambridge University Press.

Beckman, M. E., & Hirschberg, J. (1994), *The ToBI Annotation Conventions.* Online MS. Available at http://www.ling.ohio-state.edu/~tobi/ame_tobi/annotation_conventions.html.

Beckman, M. E. & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook, 3*, 255-309.

Bigi, B. (2015). SPPAS – multi-lingual approaches to the automatic annotation of speech. *Phonetician* 111–112/I–II: 54–69.

Boersma, P,, &Weenink, D. (2015). *Praat: doing phonetics by computer* (Computer program). Version 5.4.19, retrieved 12 June 2015 from http://www.praat.org/.

Brown, A. (1999). *Singapore English in a Nutshell*. Federal, Singapore.

Brown. A., Deterding, D., & Low, E.L. (2000). *The English language in Singapore: Research on pronunciation*. Singapore: Singapore Association for Applied Linguistics.

Cho, T., & Keating, P. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics, 29*, 155-190.

Chong, A.J. (2012). A preliminary model of Singaporean English intonational phonology. *UCLA Working Papers in Phonetics, 111*, 41-62.

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

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

Chong, A.J., & German, J.S. (2017). The accentual phrase in Singapore English. *Phonetica, 74,* 63-80.

Chong, A.J., & German, J.S. (2019). Variation in tonal realization in Singapore English intonation. *Proceedings of the 19th International Congress of Phonetic Sciences.*

Chow, D. (2016). *The intonational structure of Singapore English*. Unpublished PhD Thesis. Ottawa, University of Ottawa.

Cruttenden, A. (1986). *Intonation*. Cambridge: Cambridge University Press.

Cutler, A. 1986. Forbear is a Homophone: Lexical Prosody Does Not Constrain Lexical Access. *Language and Speech, 29*(3), 201-220.

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

Deterding, D. (2001). The measurement of rhythm: a comparison of Singapore English and British English. *Journal of Phonetics*, *29*, 217–230.

Deterding, D. (2005). Emergent patterns in the vowels of Singapore English. *English Worldwide,* 26(2), 179-197.

Deterding, D. (2007). *Singapore English*. Edinburgh: Edinburgh University Press.

Deterding, D., & Poedjosoedarmo G.R. (1998). *The sounds of English: Phonetics and phonology for English teachers in Southeast Asia.* Singapore: Prentice Hall.

Deterding, D., & Poedjosoedarmo G.R. (2000). To what extent can the ethnic group of young Singaporeans be identified from their speech? In A. Brown, D. Deterding, E.-L. Low (eds), *The English Language in Singapore: Research on Pronunciation* (pp. 1-9). Singapore: Singapore Association for Applied Linguistics.

Foley, J. A. (1988). *New Englishes: The case of Singapore*. Singapore: Singapore University Press.

Foley, J. A., Kandiah, T., Bao, Z., Gupta, A.F., Alsagoff, L., Ho, C. L., Wee, L., Talib, I. S., & Bokhorst-Heng, W. (1998). *English in new cultural contexts*. Singapore: Oxford University Press.

Fry, D.B. (1958). Experiments in the perception of stress. *Language and Speech 1*, 126-152.

Fuchs, R. (2016). *Speech rhythm in varieties of English: Evidence from Educated Indian English and British English*. Berlin, Heidelberg: Springer.

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

Gordon, M. (2014) Disentangling stress and pitch accent: Toward a typology of prominence at different prosodic levels. In H. van der Hulst (ed.), *Word Stress: Theoretical and Typological Issues* (pp. 83-118). Oxford: Oxford University Press.

Grabe, E. & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 515-546). Berlin: Mouton de Gruyter

Gramming, P., Sundberg, J., Ternström, S., Leanderson, R., & Perkins, W. H. (1988). Relationship between changes in voice pitch and loudness. *Journal of Voice, 2*(2), 118–126.

Grice, M., German, J. S., & Warren, P. (2020). Intonation systems across varieties of English. In C. Gussenhoven & A. Chen (eds.), *The Oxford Handbook of Language prosody* (pp. 285-302). Oxford: Oxford University Press.

Griper-Friedman, L. (1990). The tone system of West African Coastal English. *World Englishes, 9*(1), 63-77.

Gupta, A. F. (1992a). Restrictive labels for a dictionary of Singapore and Malaysian English. In A. Pakir (ed), *Words in cultural context: Proceedings of the Lexicography Workshop* (pp. 139-142). Singapore: Unipress.

Gupta A. F. (1992b). The pragmatic particles of Singapore Colloquial English. *Journal of Pragmatics, 18*(1), 31-57

Gupta A. F. (1994). *The Step-tongue: Children's English in Singapore*. Clevedon, England: Multilingual Matters.

Gupta A. F. (1998). The situation of English in Singapore. In Foley et al. (eds.), *English in new cultural contexts* (pp. 106-126). Singapore: Oxford University press.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Gussenhoven, C. (2017). On the intonation of tonal varieties of English. In M. Filppula, J. Klemola & D. Sharma (eds.), *The Oxford Handbook of World Englishes* (pp.569-598). Oxford: Oxford University Press.

Gussenhoven, C. & Udofo, I. (2010). Word melodies vs. pitch accents: A perceptual evaluation of terracing contours in British and Nigerian English. *Proceedings of Speech Prosody 2010.*

Gut, U. (2005). The realisation of final plosives in Singapore English: phonological rules and ethnic differences. In D. Deterding, A. Brown, & E.-L. Low (eds), *English in Singapore: Phonetic research on a corpus* (pp. 14-25). Singapore: McGraw-Hill.

Harris, M. S. & Umeda, N. (1987). Difference limens for fundamental frequency contours in sentences. *Journal of the Acoustical Society of America, 81*(4), 1139-1145.

Hirst, D. & Di Cristo, A. (1998). A survey of intonation systems. In D. Hirst & A. Di Cristo (eds.) *Intonation systems: a survey of twenty languages* (pp. 1-43). Cambridge: Cambridge University Press.

Ho, M. L., & Platt, J. (1993). *Dynamics of a Contact Continuum: Singaporean English*. Clarendon, Oxford.

Jun, S.-A. (2014). Prosody typology: by prominence type, word prosody and macro-rhythm. In S.-A. Jun (ed), *Prosodic Typology II: The Phonology of Intonation and Phrasing* (pp. 520-539). Oxford: Oxford University Press.

Jun, S.-A., & Fougeron, C. (1995). The accentual phrase and the prosodic structure of French. In *the Proceedings of XIIIth International Congress of Phonetic Sciences. Vol. 2* (pp. 722-725). Stockholm, Sweden.

Jun, S.-A., & Fougeron, C. (2000). A Phonological Model of French Intonation. In A. Botinis (ed.), *Intonation: Analysis, Modeling and Technology* (pp.209-242). Dordrecht: Kluwer Academic Publishers.

Kachru, B. B. (1985) Standards, codification and sociolinguistic realism: the English language in the outer circle. In R. Quirk and H.G. Widdowson (Eds), *English in the world: Teaching and learning the language and literatures* (pp. 11-30). Cambridge: Cambridge University Press.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of statistical software*, *82*(1), 1-26.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2014). *lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package).* R package version 2.0-6.

Ladd, D. R. (1983). Phonological features of intonational peaks. *Language*, 59, 721-759.

Ladd, D. R. (1998). *Intonational phonology*. Cambridge: Cambridge University Press.

Leimgruber, J. R.E. (2013). *Singapore English: Structure, variation and usage*. Cambridge: Cambridge University Press.

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Emmeans: Estimated marginal means, aka least-squares means. *R package version*, *1*(1), 3.

Lim, L. (ed.)(2004). *Singapore English: A grammatical description*. Amsterdam: John Benjamins.

Lim, L. & Tan, Y.-Y. (2001) How are we stressed?! Phonetic correlates and stress placement in Singapore English. In J.A. Maidment & E. Estebas-Vilaplana (eds.), *Proceedings of PTLC2001: Phonetics Teaching & Learning Conference 2001* (pp.27-30). London: University College London.

Lindström, E. & Remijsen, B. (2005). Aspects of the prosody of Kuot, a language where intonation ignores stress, *Linguistics, 43*(4), 839-870.

Low, E.-L. (1994). *Intonation Patterns in Singapore English*. Unpublished MPhil thesis. Cambridge: University of Cambridge.

Low, E.-L. (1998). *Prosodic prominence in Singapore English*. Unpublished PhD thesis. Cambridge: University of Cambridge.

Low, E.-L. & Brown, A. (2005). *English in Singapore: An introduction*. Singapore: McGraw-Hill.

Low, E.-L. & Grabe, E. (1999). A contrastive study of prosody and lexical stress placement in Singapore English and British English. *Language and Speech, 42*(1), 39-56.

Low, E.-L., Grabe, E. & Nolan, F. (2000). Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language and Speech, 43*(4), 377-401.

Luke, Kang Kwong (2000). Phonological re-interpretation: the assignment of Cantonese tones to English words. *Proceedings of the 9th International Conference on Chinese Linguistics (ICCL)*. National University of Singapore.

Major, T. & Mayer, C. (2018). Towards a phonological model of Uyghur intonation. In K. Klessa, J. Bachan, A. Wagner, M. Karpiński, and D. Śledziński (Eds). *Proceedings of the 9th Speech Prosody International Conference.*

Maxwell, O. (2014). *The intonational phonology of Indian English: An Autosegmental-Metrical analysis based on Bengali and Kannada English*. Unpublished Ph.D. Dissertation. Melbourne: University of Melbourne.

Maxwell, O., & Payne, E. (2021/to appear). Investigating (rhythm) variation in Indian English: an integrated approach. In R. Fuchs (eds.), *Speech rhythm in L2 and learner varieties of English*. Berlin: Springer.

Meer, P., & Fuchs, R. (2022). The Trini Sing-Song: Sociophonetic variation in Trinidadian English prosody and differences to other varieties. *Language and Speech*, *65(4)*, 923-957.

Moorthy, S. M. & Deterding, D. (2000). Three or tree? Dental fricatives in the speech of educated Singaporeans. In Brown. A., Deterding, D., & Low, E.L. (eds.) *The English language in Singapore: Research on pronunciation* (pp. 76-83). Singapore: Singapore Association for Applied Linguistics.

Ng, E.-C. (2009). Non-plateaus, non-tonal heads: Tone assignment in Colloquial Singapore English. In *Proceedings of the Chicago Linguistics Society, 45,* 487-501.

Ng, E.-C. (2011). Reconciling stress and tone in Singapore English. In Lawrence J. Zhang, Rani Rubdy & Lubna Alsagoff (eds.), *Asian Englishes: Changing Perspectives in a Globalised World* (pp. 48-59). Singapore: Pearson Longman.

O'Connor, J.D. & Arnold, G.F. (1973). *Intonation of colloquial English*. London: Longman.

Oh, M. (1998). The prosodic analysis of intervocalic tense consonant lengthening in Korean. In D. Silva (ed), *Japanese/Korean Linguistics 8*, 317-330. Stanford: CSLI.

Pakir, A. (1991). The range of depth of English-knowing bilinguals in Singapore. *World Englishes, 10*, 167-179.

Pierrehumbert, J. B. (1980). *The Phonology and Phonetics of English Intonation*. Unpublished Ph.D. dissertation. Cambridge, MIT.

Pierrehumbert, J. B., & Beckman, M. E. (1988). *Japanese Tone Structure*. Cambridge, MA: MIT Press.

Pitrelli, J. F., Beckman, M. E. and Hirschberg, J. (1994), 'Evaluation of Prosodic Transcription Labelling Reliability in the ToBI Framework', *Proceedings of the 1994 International Conference on Spoken Language Processing* (PLACE), 123-126.

Platt, J. T. (1975). The Singapore English speech continuum and its basilect 'Singlish' as a 'creoloid'. *Anthroplogical Linguistics, 17,* 363-374.

Platt, J. T. & Weber, H. (1980). *English in Singapore and Malaysia*. Petaling Jaya: Oxford University Press.

Psychology Software Tools Inc (2012). *E-Prime 2.0.* http://www.psnet.com.

R Core Team (2015). *R: A language and environment for statistical computing*, R Foundation for Statistical Computing. Vienna: Austria. <https://www.R-project.org/>.

Ramus, F., Nespor, M. & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73*, 265-292.

Rialland, A,, & Robert, S. (2001). The intonational system of Wolof. *Linguistics, 39*(5), 893–939.

Roettger, Timo B. & Gordon, Matthew K. (2017). Methodological issues in the study of word stress correlates. *Linguistic Vanguard,* 3(1).

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J. and Hirschberg, J. (1992), 'TOBI: a Standard for Labeling English Prosody', *Proceedings of the 1992 International Conference on Spoken Language Processing* (PLACE), 867-870.

Sim, J. H. & Post, B. (2021). Variation in pitch scaling in English of young simultaneous bilinguals in Singapore. Poster Presentation. *Phonetics and Phonology in Europe 2021*, Barcelona.

Singapore Department of Statistics (2010). *Census of Population 2010.* https://www.singstat.gov.sg/publications/reference

Singapore Department of Statistics (2020). *Census of Population 2020.* https://www.singstat.gov.sg/publications/reference

‘t Hart, J. (1981). Differential sensitivity to pitch distance, particularly in speech. *The Journal of the Acoustical Society of America, 69*, 811-821.

Tan, Y.-Y. (2003). *Acoustic and perceptual properties of stress in the ethnic subvarieties of Singapore English*. Unpublished PhD dissertation. Singapore: National University of Singapore.

Tan, Y.-Y. (2006). Is the stressed syllable stressed?! The perception of prominence in Singapore English. In A. Hashim & N. Hassan (eds.), *Varieties of English in Southeast Asia and beyond* (pp.133-152). Kuala Lumpur: University of Malaya Press.

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

Tan, Y.-Y. (2015). ‘Native’ and ‘non-native’ perception of stress in Singapore English: ‘Native’ and ‘non-native’ perception of stress in Singapore English. *World Englishes, 34*(3), 355-369.

Tay, M. W. J. (1982). The phonology of educated Singapore English. *English World-Wide*, *3*(2), 135-145.

Tongue, R. K. (1974). *The English of Singapore and Malaysia*. Singapore: Eastern Universities Press.

Turk, A. E., & Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics*, *35*(4), 445-472.

Van Zanten, E. & van Heuven, V. J. (1998). Word stress in Indonesian; Its communicative relevance. In *Bijdragen tot de Taal-, Land- en Volkenkunde 154,* 129-149

Van Zanten, E., Goedemans, R., Pascilly, J. (2003). The status of word stress in Indoensian. In J. van de Weijer, V. J. van Heuven, & H. van der Hulst (eds.), *The phonological spectrum: Volume II: Suprasegmental structure* (pp. 151-175). Amsterdamn: John Benjamins.

Vander Klok, J., Goad, H., & Wagner, M. (2018). Prosodic focus in English vs. French: A scope account. *Glossa, 3*(1): 71, 1-47.

Vella, A. (1994). *Prosodic structure and intonation in Maltese and its influence on Maltese English*. Unpublished Ph.D. Thesis. Edinburgh: University of Edinburgh.

Wee, L. (1998). The lexicon of Singapore English. In Foley et al. (eds.), *English in new cultural contexts* (pp. 175-200). Singapore: Oxford University press.

Wee, L.-H. (2008). More or Less English: Two Phonological Patterns in the Englishes of Singapore and Hong Kong. *World Englishes, 27*(3/4), 480-501.

Welby, P. (2002). The realization of early and late rises in French: A production study. In B. Bel, & I. Marlien (Eds.), *Proceedings of the Speech Prosody 2002 conference*, Aix-en-Provence (pp. 695–698).

Welby, P. (2006). French intonational structure: Evidence from tonal alignment. *Journal of Phonetics*, *34*(3), 343–371.

Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wightman, C., Shattuck-Hufnagel, S., Ostendorf, M. and Price, P. J. (1992). Segmental Durations in the Vicinity of Prosodic Phrase Boundaries. *Journal of the Acoustical Society of America, 91,* 1707-1717.

**Appendix A: List of target sentences**

| Set | No. of Syll. | Stress Cond. | Target | IPA | Sentence | Single/Multi |
|---|---|---|---|---|---|---|
| 1 | 1 | S | *Lee* | [li] | Lee borrowed the book from the teacher. | Single |
| 2 | 1 | S | *May* | [me] | May needed to see the show. | Single |
| 3 | 1 | S | *Sue* | [su] | Sue mobilized her employees. | Single |
| 4 | 1 | S | *Faye* | [fe] | Faye does not sing very much. | Single |
| 5 | 1 | S | *Neil* | [nil] | Neil entertained Jennifer. | Single |
| 6 | 1 | S | *Lin* | [lin] | Lin is my favourite name. | Single |
| 7 | 1 | S | *Nan* | [nɛn] | Nan was annoyed by her boss. | Single |
| 8 | 1 | S | *Min* | [min] | Min is very famous. | Single |
| 1 | 2 | Ss | *Manny* | [mɛni] | Manny borrowed the key from the office. | Single |
| 1 | 2 | sS | *Renee* | [ɹene] | Renee borrowed the cup from the kitchen. | Single |
| 2 | 2 | Ss | *Lenny* | [lɛni] | Lenny needed to see the show. | Single |
| 2 | 2 | sS | *Lahore* | [lʌhɔ] | Lahore needed to change its laws. | Single |
| 3 | 2 | Ss | *Lina* | [linʌ] | Lina mobilized her employees. | Single |
| 3 | 2 | sS | *Amir* | [ʌmiə] | Amir mobilized his employees. | Single |
| 4 | 2 | Ss | *Milo* | [mailo] | Milo does not cost very much. | Single |
| 4 | 2 | sS | *Lenore* | [lɛnɔ] | Lenore does not work very much. | Single |
| 5 | 2 | Ss | *Daniel* | [dɛnjil] | Daniel admired Patrick. | Single |
| 5 | 2 | sS | *Elaine* | [ilen] | Elaine entertained Ronny. | Single |
| 6 | 2 | Ss | *Melon* | [mɛlʌn] | Melon is my favourite fruit. | Single |
| 6 | 2 | sS | *Milan* | [milʌn] | Milan is a European city. | Single |
| 7 | 2 | Ss | *Alan* | [ɛlʌn] | Alan was annoyed by his boss. | Single |
| 7 | 2 | sS | *Aileen* | [ailin] | Aileen was scolded by her boss. | Single |
| 8 | 2 | Ss | *Nolan* | [nolʌn] | Nolan is very busy. | Single |
| 8 | 2 | sS | *Irene* | [aiɹin] | Irene is very busy. | Single |

| | | | | | | |
|---|---|---|---|---|---|---|
| 1 | 3 | Sss | *Oliver* | [ɔlivə] | Oliver borrowed the mouse from the desk. | Single |
| 1 | 3 | sSs | *Melina* | [mʌlinʌ] | Melina borrowed the chair from the room. | Single |
| 2 | 3 | Sss | *Germany* | [dʒəmʌni] | Germany needed to change its laws. | Single |
| 2 | 3 | sSs | *Manila* | [mʌnilʌ] | Manila needed to change its laws. | Single |
| 3 | 3 | Sss | *Melanie* | [mɛlʌni] | Melanie mobilized her employees. | Single |
| 3 | 3 | sSs | *Helena* | [hɛlenʌ] | Helena mobilized her employees. | Single |
| 4 | 3 | Sss | *Emily* | [ɛmili] | Emily does not drive very much. | Single |
| 4 | 3 | sSs | *Vanilla* | [vʌnilʌ] | Vanilla does not cost very much. | Single |
| 5 | 3 | Sss | *Caroline* | [kɛɹolain] | Caroline admired Tina. | Single |
| 5 | 3 | sSs | *Sebastian* | [sɛbɛstiʌn] | Sebastian entertained his son. | Single |
| 6 | 3 | Sss | *Jonathan* | [dʒɔnʌtʌn] | Jonathan is my favourite name. | Single |
| 6 | 3 | sSs | *Meridien* | [mɛɹidiʌn] | Meridien is my favourite hotel. | Single |
| 7 | 3 | Sss | *Donovan* | [dɔnʌvʌn] | Donovan was scolded by his boss. | Single |
| 7 | 3 | sSs | *Millennium* | [milɛniʌm] | Millennium was an old hotel chain. | Single |
| 8 | 3 | Sss | *Cameron* | [kɛməɹʌn] | Cameron is very smart. | Single |
| 8 | 3 | sSs | *Desertion* | [dizəʃʌn] | Desertion is rare in the army. | Single |
| 9 | 2 | s#S | *He won* | [hi wɔn] | He won the race around the bay. | Multi |
| 10 | 2 | s#S | *She ran* | [ʃi ɹɛn] | She ran in the marathon today. | Multi |
| 11 | 2 | s#S | *I heard* | [ai həd] | I heard the news on the radio. | Multi |
| 12 | 2 | s#S | *They bought* | [de bɔt] | They bought the toys for the children. | Multi |
| 9 | 3 | s#sS | *He began* | [hi bigɛn] | He began to write his new book. | Multi |
| 9 | 3 | s#Ss | *He murmured* | [hi məməd] | He murmured the words of the song. | Multi |
| 9 | 3 | s#s#S | *He will join* | [hi wil dʒɔin] | He will join the firm in the new year. | Multi |
| 10 | 3 | s#sS | *She canoed* | [ʃi kʌnud] | She canoed along the Kallang river. | Multi |

| | | | | | | |
|---|---|---|---|---|---|---|
| 10 | 3 | s#Ss | *She laboured* | [ʃi lebəd] | She laboured in the kitchen all day. | Multi |
| 10 | 3 | s#s#S | *She can sing* | [ʃi kɛn siŋ] | She can sing the songs from the book. | Multi |
| 11 | 3 | s#sS | *I delayed* | [ai diled] | I delayed the printing of the newspaper. | Multi |
| 11 | 3 | s#Ss | *I hammered* | [ai hɛməd] | I hammered the nails into the wall. | Multi |
| 11 | 3 | s#s#S | *I might mail* | [ai mait mel] | I might mail the letter to the school. | Multi |
| 12 | 3 | s#sS | *They denied* | [de dinaid] | They denied the request of the student. | Multi |
| 12 | 3 | s#Ss | *They tallied* | [de tɛlid] | They tallied the scores of the players. | Multi |
| 12 | 3 | s#s#S | *They should learn* | [de ʃud lən] | They should learn the rules of the game. | Multi |
| 13 | 4 | s#Sss | *He verified* | [hi vɛɹifaid] | He verified the accuracy of the data. | Multi |
| 13 | 4 | s#Sss | *She terrified* | [ʃi tɛɹifaid] | She terrified the children of the school. | Multi |
| 13 | 4 | s#sSs | *He delivered* | [hi dilivəd] | He delivered the letter to the family. | Multi |
| 13 | 4 | s#sSs | *She recovered* | [ʃi ɹikʌvəd] | She recovered the data from her computer. | Multi |

**Appendix B: Speaker demographics**

| No. | Gender | Age | Highest Ed. | Occupation | Reported L1 | Reported L2 | Reported L3 |
|---|---|---|---|---|---|---|---|
| S01 | M | 25 | University-Ongoing | student | English | Mandarin | French |
| S03 | M | 24 | University-Ongoing | student | English | Mandarin | Hokkien |
| S04 | M | 28 | Junior College | IT executive | English | Mandarin | |
| S06 | F | 22 | University-Ongoing | student | English | Mandarin | Korean |
| S07 | M | 24 | University-Ongoing | student | Mandarin | English | |
| S08 | F | 23 | University-Ongoing | student | English | Mandarin | Hokkien |
| S10 | M | 24 | University-Ongoing | student | English | Mandarin | |
| S11 | F | 22 | University-Ongoing | student | English | Mandarin | Cantonese |
| S13 | M | 22 | University-Ongoing | student | English | Mandarin | |
| S14 | F | 21 | University-Ongoing | student | English | Mandarin | Hokkien |
| S15 | F | 20 | University-Ongoing | student | English | Mandarin | |
| S16 | F | 27 | University-Completed | case manager | English | Mandarin | |
| S17 | F | 20 | University-Ongoing | student | English | Mandarin | |
| S18 | M | 21 | University-Ongoing | student | English | Mandarin | |
| S19 | M | 21 | University-Ongoing | student | English | Mandarin | Teochew |
| S20 | F | 19 | University-Ongoing | student | English | Mandarin | Malay |
| S21 | M | 21 | University-Ongoing | student | English | Mandarin | |
| S22 | F | 21 | Junior College | student | English | Mandarin | |
| S23 | M | 21 | University-Ongoing | student | Mandarin | Hokkien | English |
| S25 | F | 21 | University-Ongoing | student | English | Mandarin | Cantonese /Hokkien/ Hainanese |
| S27 | M | 26 | University-Ongoing | student | Mandarin | English | Malay |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| S28 | M | 26 | University-Ongoing | student | English | Mandarin | |
| S30 | F | 20 | University-Ongoing | student | English | Mandarin | |
| S31 | M | 20 | Junior College | full-time National Serviceman | English | Mandarin | |

**S04 reported spending 6 months overseas in Brunei as part of Singapore's military service. No other speakers in the corpus reported time spent overseas.