Reconciling Stress and Tone in Singaporean English

Abstract

This paper discusses the proposal that the Colloquial Singaporean English (CSE) of Chinese speakers has word-level tone, predictable from stress placement within the phonological word. Some aspects of CSE prosody are consistent with the presence of tone, for instance weak marking of stress, the prevalence of level steps in pitch, and the use of a wider pitch range than in British English (BrEng). Pitch tracks show that CSE distinguishes three tone heights and that tone is independent of context to a large extent. However, CSE cannot be characterised as possessing lexical tone because tone contrasts are linked to stress contrasts. The core of the paper evaluates three independent proposals describing this relationship between CSE tone and stress, reconciling them by showing that word-medial unstressed syllables may be unspecified for tone, and that CSE has highly significant acoustic correlates of stress in addition to tone.

Introduction

It is well-known that Singaporean English (SgEng) prosody is quite distinctive among world Englishes. Several studies have found that stress is less strongly marked than in British English (BrEng): specifically, that SgEng is more syllable-timed than stress-timed, and that unstressed syllables undergo less vowel reduction than in BrEng (Tongue, 1974, p. 32; Platt & Weber, 1980, p. 55; Tay, 1982; Brown, 1988; Low & Grabe, 1999; Low, Grabe & Nolan, 2000; Deterding, 1994, 2001, 2007). High pitch has been identified as an especially unreliable perceptual cue for SgEng prominence (Tay, 1982; Tan, 2006).

SgEng sentence intonation is similarly resistant to description within a strictly British English framework. Though SgEng may give the impression of varying little in pitch (Tongue, 1974, p. 32; Date, 2005), empirical studies show that SgEng speakers actually use a larger pitch range than BrEng speakers (Deterding, 1994; Low, 2000). This may perhaps be explained by
Platt and Weber’s (1980, p. 58) clarification that the set of possible pitch patterns in SgEng is highly circumscribed, as well as Goh’s (1998, 2005) observation that sentence pitch contours seem to be better characterised as a series of level tones than nuclear accents and contours. Several studies, including phonetic and corpus-based work, have noted that high pitch may occur on the final syllable of a declarative utterance like “Finish!” (Platt & Weber, 1980, p. 58; Low & Grabe, 1999; Zhu, 2001, p. 94; Date, 2005; Lim, 2008). Can it be a coincidence that a similar rising movement turns out to be the most common within the SgEng word, especially non-finally (Deterding, 1994)?

The features described above would be more easily explained if tone played a role in SgEng prosody. This has in fact generally been acknowledged in the case of SgEng pragmatic particles (Kwan-Terry, 1978; Bell & Peng, 1983; Loke & Low, 1988; Platt & Ho, 1989; Wong, 2004; Lim, 2004, p. 45; Lee, 2007; Lim, 2008).

In this paper I will discuss the extent to which SgEng may be described as possessing tone, focusing on characterising pitch contours at the word level and their relationship with stress, then briefly outlining some questions which remain to be answered, such as the most likely origin of these tone patterns. The specific variety I am concerned with is the informal speech often termed Colloquial Singaporean English (CSE), also Singlish in popular parlance. So far my analysis is limited to the CSE of Chinese speakers.

What makes a tonal language?

The traditional categories of word-level prosody are (a) stress-accent, (b) lexical tone, and (c) pitch-accent. Tone has at least surface manifestations for both of the latter categories: lexical tone is contrastive and independent of context, but in a pitch-accent language it is usually said that tone is predictable based on the location of an accent within the word. This three-way classification has, however, been shown to be problematic given that pitch-accent languages do not form a coherent class (Hyman, 2006). Furthermore, tone and stress coexist and interact in very diverse ways in Mandarin and Shanghai Chinese or the Neo-Stovakian dialect of Serbian/Croatian, for instance (Duanmu, 1999; Zec, 1999).

SgEng has previously been described simply as a stress-accent language (e.g. Bao, 2006). Challenging that view, this paper was originally titled Singlish as a Tonal Language, Thanks to Bazaar Malay when presented at ESEA 13. However, this title was misleading because I do not claim that SgEng possesses lexical tone, only that there is consistent use of categorical
tone in CSE. Since tone assignment is predictable, based on the location of the (stress) accent within the phonological word, some readers may prefer to think of CSE as a pitch-accent language, though I have avoided this term for the reasons outlined above.

The stress-tone relationship

**Surface tone patterns.** Wee (2008), Ng (2008a) and Siraj (2008) have independently proposed that CSE exhibits tone which remains relatively stable in different contexts. This is particularly noticeable with the word-final high, which can be found in different sentence positions, controlling for segmental effects (Figure 1).

![Figure 1: Normal (presented as a brand name) in different sentence positions](image)

In my recordings this word-final high was very often lost amidst sentence-final lowering, but as aforementioned, it is robustly attested in a significant minority of cases by many different studies of SgEng prosody (Platt & Weber, 1980, p. 58; Low & Grabe, 1999; Zhu, 2001; Date, 2005; Lim, 2008).

**Stress-dependent.** Does CSE also have contrastive tone? It is possible to construct an apparent minimal triplet for tone with [aj.si]\(^1\) as shown in Figure 2 below, controlling for sentence position as well as segmental effects. Three tone heights may seem a great many, but perhaps this should be less surprising given the rich tonal inventories of the main Chinese dialects natively spoken in Singapore, which use three level tones or more.

---

\(^1\) For simplicity, IC has been omitted; it is essentially indistinguishable from icy for most speakers. Also note that I see was not as consistently produced as the other two items; I suggest that in such cases speakers may have placed focus on I, resulting in I sée (HH).
However, these three items also contrast for stress in BrEng: *I sée*, *icy*, and *éye sea*. This situation repeats itself in other apparent near-minimal pairs for tone such as *móney*, *my knée*. Unfortunately CSE has few test cases controlled for segmental effects, since BrEng contrastive stress is so often levelled in part-of-speech distinctions such as *increase* (n.), *incréease* (v.) (Tay, 1982; Bao, 1998). Where contrastive stress is preserved, the vowels are often quite different, e.g. *présent* as a noun [prezən], *presént* as a verb [prizən]. We do, however, see alternation for some function words. For instance, the preposition *in* normally has low tone, but it may sometimes receive high tone when in focus, and always receives high tone when it is a stressed particle, e.g. *côme ín*. In cases where BrEng and SgEng stress patterns differ, I have recorded multiple tone patterns for the same speaker, e.g. *photógraphy* (LMMH), *phótográphy* (MMMH), and Siraj (2008) has found a similar state of affairs for *renáissance*, *rénnaissance*. Similarly, object pronouns (e.g. *them*, *me*, *it*) are said to be stressed in Singaporean English (Platt & Weber, 1980) and do in fact usually receive high tone, unlike subject pronouns (e.g. *they*, *I*, *it*). The data clearly point to a relationship between stress and tone.

Ng (2008a) proposes that the stress-tone relationship in CSE can be described as follows:

1. High tone (H) is assigned to the final syllable of the word.
2. Mid tone (M) starts on the first stressed syllable and spreads rightward to the penultimate syllable.
3. Low tone (L) is assigned to initial unstressed syllables.

We see these principles operating in the following morphologically simple words:


Two other accounts of CSE tone have been independently proposed and are discussed below.
Wee’s (2008; personal communication, 7 Aug 2008) tone transcriptions are identical to Ng (2008a), but he has argued for different principles of tone assignment:

(5) High tone is assigned to the final syllable of the word.
(6) Low tone is assigned to the initial syllable if it is unstressed in BrEng.
(7) Mid tone is default.

Wee’s generalisations cannot, however, account for the pronunciation of *électronics* (LLMH), since it allows only the first syllable to be specified for tone. Low tone on two initial syllables is however completely possible in Ng’s (2008a) account. The differences between the predictions are indicated with underlining in (8) below.

(8) Predicted by Wee (2008): *électronics* (LMNH)
    Predicted by Ng (2008a): *électronics* (LLMH)
    Observed: *électronics* (LLMH)

As Wee has pointed out, one word by itself is not a very strong counter-example. However, two initial unstressed syllables are rare in English to begin with (Selkirk, 1995), and even rarer in the everyday vocabulary of CSE. They usually result from stress shift in morphologically complex words, a process which is not characteristic of CSE (Bao, 1998). Some weak secondary stress is normally present, e.g. *chandeliér, pèrsevére, Quasimodó*. *Electronics* therefore provides a crucial test case militating against Wee’s analysis.

**Acoustic correlates of stress.** Another strength of Wee’s position is that it depends only on CSE speakers’ perception of BrEng stress as tone, instead of requiring CSE to possess stress in addition to tone. Certainly, as aforementioned, stress is not as strongly marked in SgEng (especially CSE) as in BrEng. However, my own mini-analysis tends to confirm that CSE stress has significant acoustic correlates.

The data were drawn from the experiment illustrated in Figure 2 above. Speakers were instructed to pretend the target word was a common brand name like Colgate. Focus effects

---

2 As pointed out by an anonymous reviewer, all pronunciations listed in Wells (2000, p. 254) and Jones (2006, p. 165) include a secondary stress: not *electronic*, but *électronic* and *élécétronic* (note that *electronics* itself was not listed). Some SgEng speakers do indeed have a pronunciation consistent with the latter: *électronics* (LMH, LM∅H), presumably by analogy with *élécétric*. However, the destressed pronunciation with two initial low tones also occurs in SgEng and would be consistent with the loss of secondary stress in frequent lexical items (for an overview, please see Ng, 2010, §5).
were minimised by asking subjects to emphasise the first word in reply to the experimenter’s question “Can order ah?”. Similarly, phrase-boundary effects were minimised by putting the target word in phrase-medial position within a short sentence frame. All variations of the frame are illustrated in (9) below.

<table>
<thead>
<tr>
<th>Frame:</th>
<th>ALWAYS</th>
<th>BETTER</th>
<th>CANNOT</th>
<th>order ____ from</th>
<th>Geylang</th>
<th>Jurong</th>
<th>Katong</th>
<th>Punggol</th>
<th>lah.</th>
</tr>
</thead>
</table>

For the mini-analysis, I compared pairs of adjacent syllables in order to control for speech volume. I also controlled for vowel, syllable type (CV), and the same tone (mid). This limited the data to 41 tokens recorded in 2008 by three speakers (21-year-old female, 21-year-old male, 23-year-old male). The words used are listed in (10) below, with the relevant syllable pairs underlined.

| (10) | economy, economics, economical, minimum, minimise, minimisation, psychology, psychological, psychologically, technology, technological, technologically, photography, wikipedia |

Syllables were classified as stressed or unstressed based on the stress-tone relationship given in (1)-(3) above, which predicts that the first syllable in the word with mid tone must be stressed. If neither syllable fitted this criterion, the third syllable with mid tone was considered stressed, based on the uncontroversial assumption that stress tends to alternate.

A paired t-test showed that stress was very significantly correlated with greater maximum and average vowel intensity, as well as greater vowel and consonant duration (p<0.001, p<0.001, p<0.001, p<0.005 respectively). Pitch as measured by fundamental frequency (F0)\(^3\) did not differ significantly; in fact the unstressed member of each pair had slightly higher F0 on average, possibly because more of them were adjacent to the final high tone. This is in keeping with Tan’s (2006) finding that “the most dominant perceptual cue for prominence in SgEng is intensity, not pitch.” As such it seems unnecessary to resort to analyses of CSE tone assignment which do not rely on stress, such as Wee’s.

---

\(^3\) Note that most current phonetic measurements of tone (e.g. Jun, 2005) rely on pitch targets, which are most easily spotted as inflexion points in fundamental frequency (F0). In this case I measured minimum F0 within the vowel even though the relevant syllables were not necessarily perceived with lower pitch than their neighbours. This helped to eliminate preceding segmental effects, i.e. high initial F0 caused by the release of a voiceless stop. There was relatively little interference from the following segmental context because the data were limited to CV syllables.
Siraj (2008) agrees with Ng (2008a) in giving stress a crucial role in CSE tone assignment, but there are slight differences in our data (indicated with underlining below), as demonstrated with the matched five- and six-syllable words in (11).

(11) a. Ng: minimisation (MMMMH) vs. Siraj: modification (MHHMH)
   b. Ng: psychologically (LMMMMH) vs. Siraj: repairability (LMHMHH)

Siraj gives the following algorithm for tone assignment (slightly paraphrased here), which diverges from Ng (2008a) in proposing a foot-final high in addition to a word-final high:

(12) Mid tone for each stressed syllable.
(13) High tone for the right edge of each foot.
(14) High tone for the right edge of each word.
(15) Low tone for any remaining unstressed syllables.

One source of differences is evidently inter-speaker variation, because Siraj also recorded one speaker who produced long spans of mid tone (personal communication, 17 Nov 2008), while conversely I also recorded a speaker who broke up long words with multiple high tones. It could be that some speakers have a foot-final high as Siraj proposes; alternatively it is possible that some speakers allow only one foot per phonological word and may therefore split some long words into multiple phonological words as suggested by Ng (2009). A point in favour of Siraj’s foot boundary tone proposal is that the phonological word is generally agreed to align only with morpho-syntactic boundaries (Raffelsiefen, 1999, p. 133). Further work with Siraj’s consultants would be necessary to clarify this point.

Another consistent difference is that Siraj often transcribes a non-foot-final high tone where Ng (2008a) and Wee (2008) would transcribe a mid tone. These mismatches are underlined in (16) below.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a. éléphant (MMH)</td>
<td>mánaging (MMH)</td>
<td>Pánama (MHH)</td>
</tr>
<tr>
<td>b. América (LMMH)</td>
<td>original (LMMH)</td>
<td>rhinóceros (LMHH)</td>
</tr>
</tbody>
</table>

Note that these mismatches always appear after a stressed syllable with mid tone, and before

---

4 My thanks to Renate Raffelsiefen for pointing this out to me.
a word-final high tone. Ng (2008b) concedes that in these cases we often see a straight-line rise from mid to high as illustrated with \textit{minimum} in Figure 3 below:

![Figure 3: Interpolation between M and H tones](image)

Ng (2009) proposes that the medial syllable in Figure 3 should not be transcribed with a high tone, but rather should be considered unspecified for tone (Ø), permitting interpolation between the preceding mid tone (M) and the word-final high tone (H). However, there are still cases where it seems more appropriate to transcribe an actual mid tone, for instance in Figure 4 below:

![Figure 4: MMH specified tones](image)

Confirmation of these transcriptions awaits perception experiments, but in the meantime I would suggest that it is unnecessary to posit a foot-final high. All extant data can be satisfactorily reconciled based on Ng’s (2009) proposal that medial unstressed syllables are only optionally specified for mid tone, e.g. \textit{éléphant} (MMH or MØH), \textit{América} (LMMH or LMØH).

\textbf{Summary.} The relationship between stress and tone seems best described by the following generalizations:

\begin{itemize}
  \item (17) High tone (H) is assigned to the final syllable of the phonological word.
  \item (18) Mid tone (M) spans all non-final stressed syllables.
  \item (19) Low tone (L) is assigned to initial unstressed syllables.
  \item (20) Remaining unstressed syllables may remain unspecified for tone or receive mid tone by rightward spreading.
\end{itemize}
In my recordings, morphologically complex words often appear with multiple high tones, e.g. NTUC (HMMH). Ng (2010) analyses these cases as multiple phonological words, each with its own final high tone.

**Implications and future work**

**Perception tests.** As aforementioned, perception tests are needed to confirm the transcriptions and generalizations discussed above. One assertion I have drawn back from is that CSE uses three *level* tones, partly because certain stressed monosyllables (e.g. *don’t*, *damn*, sometimes also “Can!”) generally give the impression of a distinct rise, even though a high level pronunciation may also be acceptable. It would also be interesting to find out how much CSE relies on the word-final high compared to boundary glottalisation (Tongue, 1974, p. 32; Brown, 1988; Ng, 2008b) for judgments on word segmentation.

**Interaction with higher-level prosody.** Categorical tone within the CSE word must interact with sentence-level intonation in some way. For instance, if both tone and intonation are to be preserved, it is hardly surprising that CSE makes use of a greater pitch range than BrEng. Categorical tone also helps to explain “boosted pitch” (Deterding, 1994), the unusually high pitch which occurs chiefly on the first stress-bearing word of a sentence in SgEng: this would be expected if a word-final high tone was interacting with either an utterance-initial high boundary tone or wide pitch range. Another question to be answered is why CSE utterances sometimes seem to terminate with level pitch across the word or a declining series of level tones, suggesting an utterance-final low boundary tone or a compressed pitch range. The word-final high tone is also clearly relevant to the controversial issue of whether sentence-final syllables as in “Finish!” actually attract stress or merely lengthened (Tongue, 1974, p. 20; Platt & Weber, p. 58; Low & Grabe, 1999).

**Origins.** Whereas Chinese CSE speakers reserve high tone for word-final position, most speakers of tonal languages map English stress to high tone, as in Hong Kong English (Wee, 2008; Cheung, 2008), Mandarin-influenced English (Zhang, Nissen & Francis, 2008), Nigerian English (Gut, 2005), and numerous West African creoles (Devonish, 2002). An affinity between stress and higher tone is a common state of affairs cross-linguistically (De Lacy, 1999, 2002), and this relationship holds to some extent in CSE, since stressed syllables receive mid tone rather than low tone. The crucial point that needs explaining, then, is the
Recent work argues that its ultimate source is the phrase-final high of Malay (e.g. Lorentz, 1997) via Bazaar Malay (Ng, 2008a) or Baba Malay (Lim, to appear), probably reinforced by Indian English (Ng, to appear). Note that the word-final high is the first distinctive feature of CSE (apart from loanwords) which cannot be explained by Chinese contact alone; as such it confirms the important role played in CSE’s development by non-Chinese Singaporeans.

Speakers outside the group studied so far. Because the current work by Ng (2010), Wee (2008) and Siraj (2008) has been limited to Chinese speakers, it is not yet clear to what extent this description of CSE applies to speakers with other home languages, though I have encountered Malay and Indian Singaporeans whose CSE gives the impression of including the same consistent stress-tone relationship. Selvarani Suppiah’s doctoral dissertation on primary school children (in progress) may throw light on this area. On the one hand it seems unlikely that a prosodic system complicated by stress and three tones should spread by contact, but on the other hand the numerical domination of the Chinese would be expected to encourage levelling in the direction of Chinese CSE.

International intelligibility. Another question of immediate interest to educators is to what extent word-level tone interferes with non-Singaporeans’ comprehension of CSE. Speaking purely anecdotally, it seems that when tone persists in Standard Singaporean English speech, it causes far greater difficulties for native speakers of Chinese and Japanese (e.g. Date, 2005, p. 182) than for native English speakers. I would suggest that these speakers must rely more heavily on pitch in their perception of English stress (Wang, 2008), because it is such an important prosodic cue for Chinese lexical tone and Japanese pitch accent.

Conclusion

Some distinctive aspects of SgEng prosody, such as weak marking of stress, the prevalence of level steps in pitch, and the use of a wider pitch range than in BrEng, can be more easily explained when CSE is analysed as possessing tone. In my recordings three tone heights can be distinguished, and tone may be preserved despite sentence-final lowering. At the same time, CSE cannot be described as possessing lexical tone because tone contrasts are

---

5 Note that CSE is not the only language with word boundary tones: Standard Serbo-Croatian (Godjevac, 2005) has also been analysed in this way. A number of languages have also been analysed with an Accentual Phrase (a prosodic unit slightly larger than a word) marked by rising intonation, e.g. Bengali, Korean and Georgian. (Sun-Ah Jun, personal communication, 23 Feb 2009; Sun, 2005)
predictable from stress. In fact controlling for categorical tone makes it possible to prove that CSE possesses not only underlying stress contrasts, but also highly significant acoustic correlates of stress, i.e. duration and intensity.

The three original papers which independently proposed that CSE has tone (Wee, 2008; Ng, 2008a; Siraj, 2008) largely agree in transcription, and all relate tone to stress. I have argued that Wee’s account leaves a crucial data point unexplained, while Siraj’s justly corrects Ng (2008a) but introduces some unnecessary high tones which may be better analysed as interpolation over syllables unspecified for tone. The generalizations which seem to fit the data best are listed in (17)-(20) above: essentially, high tone marks the final syllable in the phonological word, mid tone spans all non-final stresses, and low tone appears on initial unstressed syllables. The most unusual aspect of CSE tone is the word-final high, which probably originates in Chinese speakers’ interpretation of a Malay phrase-final high (Ng, to appear).

Author
Ng E-Ching is currently a Ph.D. student in Yale’s Linguistics department. After her undergraduate degree in English, also at Yale, she taught General Paper and English Literature at Raffles Junior College in Singapore for five years. Her research interests include laboratory phonology and language contact.

Acknowledgements
My presentation at ESEA 13 developed ideas first presented at the Language Transfer Workshop, UWE (Bristol), 9-11 July 2008, and the Workshop on Prosodic Alignment at the Word Level at Institut für Deutsche Sprache, 20-21 Nov 2008. It was nurtured by my advisor, Darya Kavitskaya, as well as Gaja Jarosz, Jelena Krivokapic, Claire Bowern, Stanley Insler, Maria Piñango and Stephen Anderson; also my fellow students, especially Erich Round, Leandro Bolanos and Scott McClure. Beyond Yale, many useful comments came from Wee Lian Hee, Pasha Siraj, Mary Beckman, Sun-Ah Jun, Low Ee Ling, Diana Apoussidou, Julie McGory, Tom Conners, Uri Tadmor, Devyani Sharma, Sakina Suffian Sahuri, Bao Zhiming, K. P. Mohanan and an anonymous reviewer. My thanks also to Raffles Junior College for the use of their music facilities, Lisa Lim for access to the Grammar of Spoken Singapore English Corpus (GSSEC), and the 21st Century Center of Excellence (Japan), Tokyo University of Foreign Studies for access to the Multilingual Spoken Corpus (Malay). Finally, heartfelt thanks must go to Lim Tse Yang, Huang Zhipeng, Tan Wen Sze, my extended family and numerous consultants for their good humour and stamina.
References


Ng, E.-C. (2008b, Nov 21). Affixation, compounding and the prosodic word in Singaporean English. Paper presented at the workshop on Prosodic Alignment at the Word Level,
Institut für Deutsche Sprache, Mannheim.

Ng, E.-C. (2010). Reduction, frequency and morphology in Singaporean English prosody. Unpublished manuscript, Yale University. ROA 1198.


Phonology, 16, 225-264.