Ethnic variation in perceptual sensitivity to stress-based f0 in Singapore English

Jasper H. Sim[1], Julia Schwarz[2], Katrina Kechun Li[3], Adam J. Chong[4]

[1]Nanyang Technological University, Singapore
[2]University of Cambridge, UK
[3]Leiden University, The Netherlands
[4]Queen Mary University of London, UK

jasper.sim@nie.edu.sg

Abstract

Lexical word stress, where one syllable in a word is specified as prosodically prominent or distinctive, is typically cued by multiple prosodic correlates. Whereas in mainstream English varieties stressed syllables may be marked by a pitch accent, in Singapore English (SgE), f0 typically peaks on the final syllable of content words, regardless of underlying stress, with f0 alignment distinguishing stressed syllables instead. Previous work has further shown that SgE speakers do not consistently produce this contrast. This raises the question of whether SgE listeners are perceptually sensitive to stress-based f0 differences. In this exploratory study, we tested 16 early bilingual native Singaporeans on typicality judgements using resynthesised stimuli with swapped f0 contours (e.g., carTOON ↔ CARpet). When stimuli were presented individually, both f0-matched and f0-mismatched stimuli were judged equally 'Singaporean', suggesting limited sensitivity. In a paired comparison task that directly contrasted f0-matched and f0-mismatched stimuli, only Mandarin-speaking listeners showed higher preference for f0-matching stimuli, but only for stress-initial words. These findings suggest that f0 is a weak perceptual correlate of stress for many speakers of SgE, with sensitivity shaped by prosodic position and language background—potentially enhanced by tone language experience but attenuated in speakers of languages without lexical stress.

Index Terms: lexical stress, stress perception, language contact, bilingualism

1. Introduction

Major class words in mainstream varieties of English are lexically specified for stress, such that every lexical word has at least one syllable marked for metrical prominence [1], and words that differ in lexical stress (e.g., carTOON vs. CARpet) contrast in prominence through a range of prosodic cues. One such cue is f0, with lexically stressed syllables typically showing higher f0 if pitch accented, relative to unstressed syllables [2]. Both segmental and prosodic cues support lexical access and spoken word recognition, though cross-linguistic differences exist in whether suprasegmental or segmental cues predominate [3].

Owing to long-term language contact, the relationship between f0 and word-level prominence in Singapore English (SgE) markedly differs from that observed in mainstream varieties. Descriptive accounts of SgE intonation report that major class words in SgE exhibit an f0 peak on the final syllable, regardless

of underlying word stress [4]. Chong and German [5][6], working within the autosegmental-metrical framework, describe LH rises recurring across an utterance within an accentual phrase (AP), which generally consists of a content word and any preceding function words. In each AP, the L aligns with the left edge and the H with the right; that the H typically falls on the right edge, i.e., on the final syllable of the content word, contributes to the perception of word-final prominence [7]. Consequently, f0 is thought to be a weak correlate of lexical stress in SgE. Despite this, some have argued that f0 relates to stress position in some ways [8]. In particular, Chong and German [9] found that, while f0 contours generally conform to AP-level phasing (i.e., f0 rises through the target word), stress-initial words show a higher mean f0 at onset and an earlier f0 elbow than non-stress-initial words (see Figure 1). These could suggest that SgE-speaking listeners may rely on f0 cues differently in word recognition.
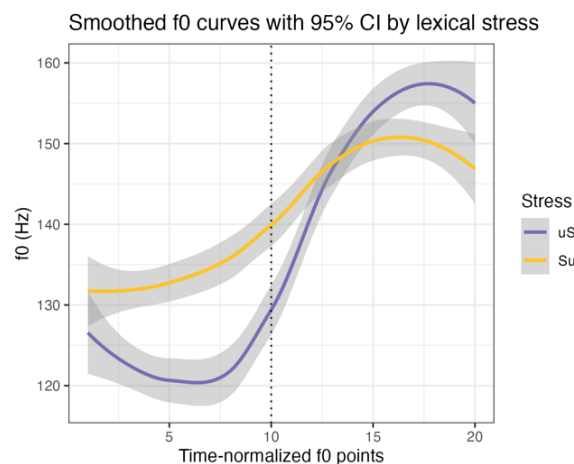

Figure 1: *Time normalised plots of f0 averaged over each stress condition (Su: stress-initial; uS: stress-final) for the stimuli used in the current study. Dotted line represents syllable boundary.*

Furthermore, consistent with past production studies showing that SgE, though a stabilised contact variety, remains internally variable [10][11], [9] reported variability in the degree to which stress-based contrasts were produced by SgE speakers. Likewise, [12][13] found differences in the intonation patterns among SgE speakers from different language backgrounds, despite all being native L1 speakers, raising the question of

whether their other L1 (e.g., Mandarin, Malay, or Indian languages) might also shape their perception.

In this exploratory study, we ask whether and how SgE speakers are sensitive to the stress-based f0 patterns found in production studies. Specifically, we ask:

(i) whether SgE-speaking listeners are perceptually sensitive to the f0 alignment differences that distinguish SgE words with different underlying stress; and

(ii) whether their ability to discriminate is modulated by individual bilingualism.

2. Methodology

Participants were 16 native Singaporeans, all early bilinguals of SgE and an ethnic mother tongue [EMT; Mandarin ($n = 6$), Malay ($n = 8$), or Tamil ($n = 2$)]. They were aged 21–36 years ($M_{age} = 28$, $SD_{age} = 4.8$; males = 10) and reported no hearing or language impairments. Ethics approval was obtained from NTU-IRB [2023-852]. Participants completed the tasks on Gorilla.sc following a headphone screening test [14].

The stimuli consisted of 15 SgE word pairs (Table 1), matched as closely as possible in the position of their voiced/sonorant segments, but differing in stress pattern (i.e., stress-initial vs. stress-final). They were produced by a 32-year-old male English-Mandarin bilingual Singaporean using the carrier phrase '[WORD] is what he said'. Smoothed f0 contours of the pre-synthesised target words were overall consistent with the previously reported production patterns (Figure 1; [9]).

Each target word, its carrier phrase, and the voiced/sonorant segments of each syllable in the target word were manually segmented in Praat [15]. Syllable-level, time-normalised pitch contours (30 points/syllable) were extracted from each target word and swapped across word pairs (e.g., the two syllables of carTOON were replaced with the f0 contour of the two syllables of CARpet, and vice versa). Both the original and pitch-swapped target words were resynthesised using *overlap-add* method in Praat. To ensure continuity in the f0 trajectory, the resynthesised target words were then concatenated with either (a) their own original carrier phrase, or (b) the unmanipulated carrier phrase of the paired word. This produced two stimulus conditions: (i) **Matching** – target word with its original pitch contour and original carrier phrase; and (ii) **Mismatching** – target word with a pitch-swapped contour combined with the corresponding unmanipulated carrier phrase of the other word. The f0 of the carrier phrases themselves was not altered; they were only reassigned across conditions to align with the swapped contours of the target words. Duration and intensity were left unchanged throughout. Overall intensity across stimuli was normalised by scaling to ensure that no stimulus was noticeably louder than the others.

The same stimuli, along with their orthographic forms, were presented in a typicality judgement task (Task 1) followed by a paired comparison/forced choice task (Task 2). In Task 1, stimuli were presented individually in random order, and listeners rated how typically Singaporean it sounded on a 7-point Likert scale (1 = very untypical; 7 = very typical). In Task 2, the Matching and Mismatch stimuli of each target word were presented together, and listeners selected the one they judged to be more typically Singaporean sounding to them.

Table 1: *Stimuli (stress-initial, stress-final)*

Pair	Words	Pair	Words
1	*archive, arcade*	9	*demon, demean*
2	*balling, balloon*	10	*distance, distaste*
3	*camping, campaign*	11	*parrot, parade*
4	*carton, cartoon*	12	*retail, retain*
5	*column, cologne*	13	*salad, salon*
6	*concept, conceit*	14	*surplus, surprise*
7	*concert, concern*	15	*table, taboo*
8	*concord, concoct*		

3. Results

3.1. Task 1 – Typicality rating

We predicted that, for Task 1, listeners would rate the stimuli with the expected f0 contour (Matching) as more typically Singaporean than those with the Mismatching f0 contour. Findings revealed that both Matching and Mismatching stimuli were rated as highly typically Singaporean (Figure 2), with very similar ratings between Matching, $M(SD) = 5.47(1.42)$, $Mdn = 6$, and Mismatching stimuli, $M(SD) = 5.53(1.30)$, $Mdn = 6$. Mixed-effects ordinal regression was conducted using CLMM [16] in R [17] with ratings as response. The maximal model included Condition*Stress + Condition*EMT as fixed effects, while the reduced model included Condition + Stress + EMT, with random intercepts for Listener and Word. The analyses revealed no significant differences in ratings between conditions ($b = 0.21$, $SE = 0.13$, $p = .09$) or any effect of listeners' EMT (Malay: $b = 1.10$, $SE = 1.48$, $p = .46$; Mandarin; $b = 1.33$, $SE = 1.54$, $p = .39$).
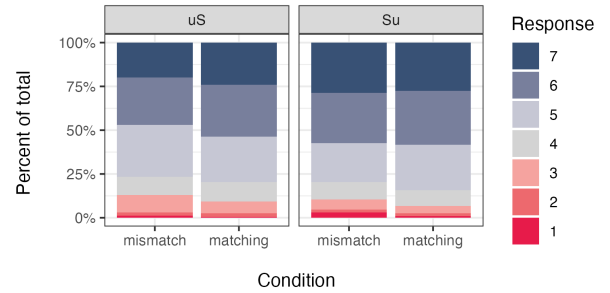

Figure 2: *Typicality judgement ratings by Stress*

3.2. Task 2 – Paired comparison / forced choice

Similarly, for Task 2, we predicted that listeners would show a preference for the Matching stimuli over the Mismatching ones as being more typical. Findings revealed that, overall, Mandarin-speaking listeners chose the Matching stimuli more frequently (65%) compared to speakers of Malay (54%) and Indian languages (50%). Logistic mixed-effects regression was conducted using GLMER [18] in R [17] with preference as response (1 = Matching). The model included Stress*EMT as fixed effects, with random intercepts for Listener and Word. The analyses, illustrated by the marginal means in Figure 3, revealed that Mandarin-speaker listeners were more likely to prefer Matching stimuli than speakers of other EMTs, but only for stress-initial (Su) words (Mandarin v. Malay: $OR = 2.77$, $SE = 0.91$, $p = .005$; Mandarin v. Indian.Lang: $OR = 3.64$, $SE = 1.72$, $p = .02$).
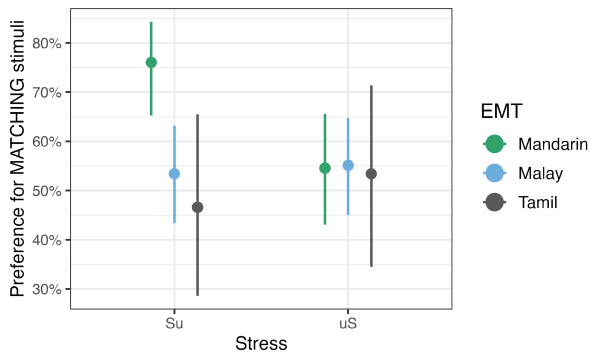
Figure 3: *Predicted probabilities of listener preferences for Matching stimuli by Stress and EMT*

4.   Discussion

When SgE-speaking listeners—regardless of their language background—were presented with SgE words with f0 contours that did not conform to the expected stress patterns observed in past production studies (e.g., carTOON produced with the contour of CARpet), they showed no preference for the prosodically Matching stimuli than for Mismatching ones. However, when the Matching and Mismatching stimuli were presented together in a paired comparison / forced choice task, only Mandarin-speaking bilingual listeners appeared sensitive to these cues, showing a stronger preference for the Matching stimuli, but only for stress-initial words.

The preliminary findings overall suggest that stress-based f0 differences may not be particularly salient to many SgE-speaking listeners, or that they were highly tolerant of small f0 deviations. Perceptual sensitivity seems to depend on prosodic position and the language background of the listeners, which we discuss in turn below.

4.1.   Effect of language background

It is unexpected that Mandarin-speaking listeners were more sensitive to stress-based f0 cues than speakers of other ethnic mother tongues, since all participants are native speakers of SgE. Based on the data, it is unclear whether the other Malay- and Tamil-speaking listeners were simply less able to perceive the f0 differences, or whether they were more tolerant than the Mandarin-speaking group of atypical patterns. It also remains uncertain whether those who were not sensitive to f0 modulations were instead relying on other preserved cues, such as intensity or duration.

The findings corroborate evidence from studies on stress 'deafness' (e.g., [19][20]), which show that speakers of languages with lexically marked stress or tone outperform speakers of languages without contrastive lexical prominence on tasks requiring perceptual sensitivity to lexical prominence. In a 'stress deafness' sequence recall task [19], [21] found that on the whole SgE listeners seemed to behave like listeners from languages argued to lack contrastive lexical stress. However, Mandarin-speaking SgE listeners were on the whole more sensitive to the prosodic contrast than Malay- and Tamil-speaking listeners. We postulate that speakers' other language(s)—that is, the effects of individual bilingualism— could contribute to (the lack of) perceptual sensitivity in SgE. Singapore Malay [22] and Tamil [23] (spoken by some ethnic Indians) do not have contrastive lexical stress, and if there is cross-linguistic influence, speakers may likewise be less

sensitive to lexically contrastive f0 cues in their SgE. By contrast, the greater sensitivity of Mandarin speakers to f0 differences could plausibly stem from their lexical tone language background. This interpretation aligns with evidence from [24] and [25], which demonstrates a pitch-processing advantage in tone language speakers, suggesting that their enhanced attunement to f0 modulations may carry over into non-tonal contexts.

As aforementioned, language-background-related differences have also been observed in previous production studies. While Singaporeans generally exhibit similar global intonation patterns, [12][13], in their examination of child-directed and mother-directed speech, found differences in f0 scaling and alignment among Malay-speaking Singaporeans. Because the present study focused only on perception, our findings cannot speak directly to whether the listeners themselves produce such stress-based f0 patterns. A follow-up production study is therefore planned to examine whether the lack of sensitivity to f0 cues observed here may be related to listeners' own production patterns.

4.2.   Asymmetrical effect

The findings also revealed that for Mandarin-speaking listeners who were sensitive to these cues, a stress-initial contour applied to stress-final words was tolerated, whereas a stress-final contour applied to stress-initial words significantly heightened perceived unnaturalness. We note that, in SgE, it is primarily the first syllable of a disyllabic word where the two stress conditions diverge most clearly, with the initial unstressed syllable showing lower f0 and later f0 alignment (see Figure 1). Given the data, it remains unclear whether the heightened unnaturalness in stress-initial words was driven by: (i) the lowered f0; (ii) the later f0 elbow alignment; (iii) interactions between f0 modulations with intensity or durational cues that have been preserved; or a combination of these factors. One possibility is that based on production data, in a SgE AP, Ls can occur *before* the stressed syllable, e.g., on an unstressed function word, resembling an L*, but Ls rarely fall on or after the stressed syllable. In the current study, in the Mismatching condition, L falls after the stressed syllable of the pitch-swapped stress-initial stimuli, which could have heightened the perception of unnaturalness for the listeners.

5.   Conclusion

In this study, we investigated the extent to which SgE-speaking listeners are sensitive to the variety-specific stress-based f0 differences reported in past production work, and whether their ability to discriminate is modulated by language background. Preliminary findings indicate that, overall, listeners show limited sensitivity to these cues in perception. However, perceptual sensitivity is influenced by language background, with Mandarin-speaking bilinguals exhibiting enhanced sensitivity. Together with recent work on stress 'deafness' in Singaporeans, these results point to a notable disjunction between production and perception of lexical stress in SgE. They also underscore the value of examining contact varieties, where long-term multilingual exposure can give rise to complex interactions between production patterns, perceptual sensitivity, and language background, even in 'stabilised' L1 varieties.

6. Acknowledgements

7. References

[1] L. M. Hyman, "*Word-prosodic typology*," *Phonology*, vol. 23, no. 2, pp. 225–257, Aug. 2006, doi: 10.1017/S0952675706000893.

[2] D. R. Ladd, *Intonational Phonology*, 2nd ed. Cambridge University Press, 2008. doi: 10.1017/CBO9780511808814.

[3] J. M. McQueen and L. Dilley, "Prosody and Spoken-Word Recognition," in *The Oxford Handbook of Language Prosody*, C. Gussenhoven and A. Chen, Eds., Oxford University Press, 2020, pp. 508–521. doi: 10.1093/oxfordhb/9780198832232.013.33.

[4] D. Deterding, "The intonation of Singapore English," *Journal of the International Phonetic Association*, vol. 24, no. 2, pp. 61–72, Dec. 1994, doi: 10.1017/S0025100300005077.

[5] A. J. Chong and J. S. German, "The Accentual Phrase in Singapore English," *Phonetica*, vol. 74, no. 2, pp. 63–80, 2017, doi: 10.1159/000447429.

[6] A. J. Chong, "Towards a model of Singaporean English intonational phonology," *The Journal of the Acoustical Society of America*, vol. 133, no. 5, pp. 3570–3570, May 2013, doi: 10.1121/1.4806536.

[7] Y. Tan, "'Native' and 'non-native' perception of stress in Singapore English," *World Englishes*, vol. 34, no. 3, pp. 355–369, Sep. 2015, doi: 10.1111/weng.12151.

[8] E.-C. Ng, "Reconciling stress and tone in Singapore English," in *Asian Englishes: changing perspectives in a globalised world*, L. J. Zhang, R. Rubdy, and L. Alsagoff, Eds., Singapore; New York: Pearson Prentice Hall, 2011, pp. 48–59.

[9] A. J. Chong and J. S. German, "Prominence and intonation in Singapore English," *Journal of Phonetics*, vol. 98, p. 101240, May 2023, doi: 10.1016/j.wocn.2023.101240.

[10] J. H. Sim, "Influence of bilingualism or caregiver input? Variation in VOT in simultaneous bilingual preschoolers in Singapore," in *Proceedings of the 20th International Congress of Phonetic Sciences*, R. Skarnitzl and J. Volín, Eds., Guarant International, 2023, pp. 2344–2348.

[11] J. H. Sim and B. Post, "Influence of caregiver input and language experience on the production of coda laterals by English–Malay bilingual preschoolers in multi-accent Singapore," *J. Child Lang.*, pp. 1–26, Jul. 2023, doi: 10.1017/S0305000923000375.

[12] J. H. Sim and B. Post, "Variation in pitch scaling in English of young simultaneous bilinguals in Singapore," presented at the Phonetics and Phonology in Europe 2021, Barcelona, Jun. 2021.

[13] A. J. Chong, J. H. Sim, and B. Post, "Ethnicity and intonational variation in Singapore English child-directed speech," in *Proceedings of the Second International Conference on Tone and Intonation*, Singapore: Chinese and Oriental Languages Infomation Processing Society, Nov. 2023, pp. 41–42. [Online]. Available: https://www.colips.org/conferences/tai2023/proceedings/pdf/2023.tai-abstract.21.pdf

[14] K. J. P. Woods, M. H. Siegel, J. Traer, and J. H. McDermott, "Headphone screening to facilitate web-based auditory experiments," *Atten Percept Psychophys*, vol. 79, no. 7, pp. 2064–2072, Oct. 2017, doi: 10.3758/s13414-017-1361-2.

[15] P. Boersma and D. Weenink, *Praat: doing phonetics by computer*. (2023). [Online]. Available: http://www.praat.org/

[16] R. H. B. Christensen, *ordinal—Regression Models for Ordinal Data*. (2019). [R package]. Available: https://CRAN.R-project.org/package=ordinal.

[17] R Core Team, *R: A language and environment for statistical computing*. (2022). R Foundation for Statistical Computing, Vienna, Austria. [Online]. Available: https://www.R-project.org/

[18] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using **lme4**," *Journal of Statistical Software*, vol. 67, no. 1, 2015, doi: 10.18637/jss.v067.i01.

[19] H. Rahmani, T. Rietveld, and C. Gussenhoven, "Stress 'Deafness' Reveals Absence of Lexical Marking of Stress or Tone in the Adult Grammar," *PLoS ONE*, vol. 10, no. 12, p. e0143968, Dec. 2015, doi: 10.1371/journal.pone.0143968.

[20] E. Dupoux, N. Sebastián-Gallés, E. Navarrete, and S. Peperkamp, "Persistent stress 'deafness': The case of French learners of Spanish," *Cognition*, vol. 106, no. 2, pp. 682–706, Feb. 2008, doi: 10.1016/j.cognition.2007.04.001.

[21] A. J. Chong, "'Stress deafness' variation in Singapore English listeners," presented at the Phonetics and Phonology in Europe 2025, Palma, Spain, Jun. 2025.

[22] D. Hamzah and J. S. German, "A phonetically-motivated model of intonational phonology in Singapore Malay," 2019.

[23] E. Keane, "Prominence in Tamil," *Journal of the International Phonetic Association*, vol. 36, no. 01, p. 1, Jun. 2006, doi: 10.1017/S0025100306002337.

[24] E. D. Bradley, "Phonetic dimensions of tone language effects on musical melody perception.," *Psychomusicology: Music, Mind, and Brain*, vol. 26, no. 4, pp. 337–345, Dec. 2016, doi: 10.1037/pmu0000162.

[25] C. Y. Lin, M. Wang, W. J. Idsardi, and Y. Xu, "Stress processing in Mandarin and Korean second language learners of English," *Bilingualism*, vol. 17, no. 2, pp. 316–346, Apr. 2014, doi: 10.1017/S1366728913000333.