An acoustic study of the realisation of KIT in the conversational speech of young English speakers in Australia

Gerard Docherty[1], Simón Gonzalez[2], Paul Foulkes[3].

Griffith University[1], Australian National University[2], University of York[3]
gerry.docherty@griffith.edu.au; simon.gonzalez@anu.edu.au; paul.foulkes@york.ac.uk

ABSTRACT

This study sheds new light on the under-investigated Weak Vowel Merger (WVM) in Australian English whereby an historical contrast between unstressed KIT and SCHWA has been lost in favour of a central schwa-like variant. We address the following questions: (a) to what extent are unstressed vowels derived historically from KIT realised with a centralised quality as per conventional accounts of WVM; (b) does the status of the lexical item as grammatical/non-grammatical impact significantly on the realisation of the unstressed syllables; (c) to what extent is any centralisation correlated with the duration of the unstressed vowels? We present acoustic analyses of the quality and duration of 2745 tokens of historical KIT by 40 speakers (20M/F) aged 18-22 from Perth recorded for c. 30 minutes engaged in an unscripted conversational interaction. While our findings are largely consistent with the conventional WVM analysis, they point to additional complexity that warrants further investigation.

Keywords: Australian English, Weak Vowel Merger, conversational speech

1. INTRODUCTION

Descriptions of English spoken in Australia [1, 2, 3] note that it is characterised by a process of Weak Vowel Merger (WVM) whereby the historical contrast between KIT and SCHWA vowel categories is lost in unstressed syllables, resulting in a centralised realisation and homophonous pairs such as *Rosa's*/*roses* or *Lennon*/*Lenin*. This centralisation does not, however, apply uniformly to unstressed tokens of the historical KIT vowel. Exceptions include the so-called 'happY-tensing' context where a closer variant approximating [i] is the norm (e.g. *movie, cities*), and in unstressed pre-velar and pre-postalveolar contexts where centralisation does not occur (e.g. in *frantic, rubbish, singing*).

For all that WVM is cited uncontroversially as a feature of English in Australia, Cox & Palethorpe's [3] is the only study to date providing relevant quantitative data in relation to this feature. The authors of that study focused on the acoustic characteristics (F1/F2) of word-final SCHWA realisations, comparing across a range of contexts including tokens of historical KIT. The specific comparisons made were between word-final lexical tokens of SCHWA (e.g. *Rosa*), and tokens that were followed by a possessive suffix (e.g. *Rosa's*), and others that were the nucleus of a plural suffix (e.g. *Roses*), the latter being the historical KIT context. The study documented variation across the various experimental conditions, but importantly noted that in line with the predictions of the WVM there was no significant difference between the latter two contexts. However, the findings also suggest that the realisation of the nucleus in these latter two contexts yielded a more [ɨ]-like realisation that differed significantly from that of the word-final lexical SCHWA tokens. This in turn suggests that the WVM may not be straightforwardly captured as an absorption of the historical KIT tokens within the distribution of SCHWA more generally, and is reminiscent of findings from other varieties of English displaying WVM that also point to the merger of historical KIT with SCHWA being more complex than might be assumed [4, 5].

In this context, the aim of the present study is to shed new light on the acoustic properties of WVM in Australian English, departing from the sole previous study [3] in a number of important ways. First, we approach the putative merger from the point of view of the historical KIT vowel. Second, we focus on data from conversational speech. Finally, we include tokens from the high-frequency unaccented grammatical item *it* in order to test the extent to which grammatical items participate in the WVM-related variability. It was decided to focus on a single grammatical item at this point of the investigation simply to test how such items pattern in relation to the WVM, and thus to explore whether it would be important to sample from a wider range of grammatical items containing historical KIT at a future stage of this project.

The three questions that we address are (a) to what extent are unstressed vowels derived historically from KIT realised with a centralised quality as per conventional accounts of WVM; (b) does the status of the lexical item as grammatical/non-grammatical impact significantly on the realisation of the unaccented syllables; (c) to what extent is any centralisation correlated with the duration of the unstressed vowels? The findings are discussed in

relation to the specifics of WVM in English in Australia and in relation to what we know of vowel mergers more generally.

2. METHODS

The material analysed in this study was drawn from a corpus of recordings of pairs of 40 young (18-22) speakers from Perth (Western Australia) engaged in same-sex unscripted conversations each lasting around 30 minutes. All of the participants had been fully-schooled in Perth (from age 5) and there were equal numbers of males and females. While social class is not relevant to the present study, it can be noted that twenty of the speakers were residents of neighbourhoods ranked by the Australian Bureau of Statistics to be in the top socio-economic decile, and the remaining twenty were from neighbourhoods with a lower socioeconomic ranking. The majority of pairs of speakers knew each other in advance but to varying degrees. A fieldworker was present in the same room as the participants in order to initiate and conclude the conversation recording process but only intervened (rarely) if the participants' conversation subsided and was in need of a prompt.

The recordings (44Khz, 16 bit) were made using Sennheiser EW112-P-G3 lapel microphones and an Edirol R44 digital recorder. Conversations were transcribed with *Elan* [6] (starting 5 minutes in to each recording) and force-aligned within *LaBB-CAT* [7] using *HTK* [8], with a follow-up manual correction of misaligned segment boundaries. Using default settings in *Praat* [9], F1/F2 values were estimated for each vowel at the midpoint of each token (see [10] for caveats on this static approach to vowel description). Pre-/l, w, j/, pre-nasal, and post-/w, j, r/ environments were excluded, as well as post-nasal/lateral tokens where segmentation was problematic. The F1/F2 measurements were not normalised, in order to allow for comparison with previous studies (notably [3]), and consequently the findings for male and female results are reported separately below.

In order to address the research questions we created a subset of the corpus with tokens of historical KIT occurring in a number of different contexts (each identified henceforth by the relevant acronym from the list below). Three of these were contexts where descriptions of the WVM suggest that centralisation of the vowel should not be anticipated:

(a) MONO - nucleus of monosyllabic content word (e.g. *big, trip*)

(b) POLYAC - accented nucleus of a polysyllabic content word (e.g. *different, issue*)

(c) PREVEL - unaccented nucleus in polysyllabic lexical items with a following velar context (e.g. *allergic, public*)

One context was where centralised realisations were anticipated as per the WVM:

(d) UNAC - unaccented nuclei contained within a polysyllabic lexical item (e.g. *massive, habit, places*)

Two subsets of *it* were generated, the latter differentiated as a result of initial auditory analysis of the data suggesting that phrase-final position might be associated with great levels of reduction.

(e) PSINIT - tokens of grammatical item *it* occurring phrase-internally

(f) PSFINIT – tokens of grammatical item *it* occurring phrase-finally

In addition to the above, two further subgroupings of the data were identified:

(g) HAPPY_T – tokens of unstressed word-final KIT providing a context where we might anticipate that the vowel nucleus would be raised and fronted as per the pattern of happY-tensing reported for this and other varieties of English (e.g. *city, easy, movie*)

(h) FLEECE – as a point of reference we identified tokens of the FLEECE lexical set realised in monosyllabic content words. Note that in this case, in order to avoid the effect of the on-glide that can characterise the quality of this vowel for many speakers of Australian English, the formant measurements were taken at the 80% point of the vowel interval.

Table 1 provides a summary of the dataset, itemising the number of tokens in each subgroup.

	Females	Males
MONO	173	244
POLYAC	212	207
PREVEL	122	99
UNAC	356	288
PSINIT	86	79
PSFINIT	268	227
HAPPY_T	207	177
FLEECE	146	179
Total	1570	1500

Table 1: Number of tokens of each of the subgroupings of historical KIT included in the acoustic analysis (see main text for further explanation).

3. FINDINGS

In order to provide a single quantitative metric of the degree of centralisation for each token, we adopted a

method previously deployed by Labov et al. [11] and Grama et al. [12], calculating F2-(2*F1) as a derivative indicator of relative location along the front diagonal of the vowel space. We refer to this measure henceforth as F2DERIV (Hz). A higher value of F2DERIV equates to a relatively closer and fronter vowel quality.

Figure 1 shows separately for males and females the distribution of F2DERIV for each of the eight subgroupings of the dataset described above.

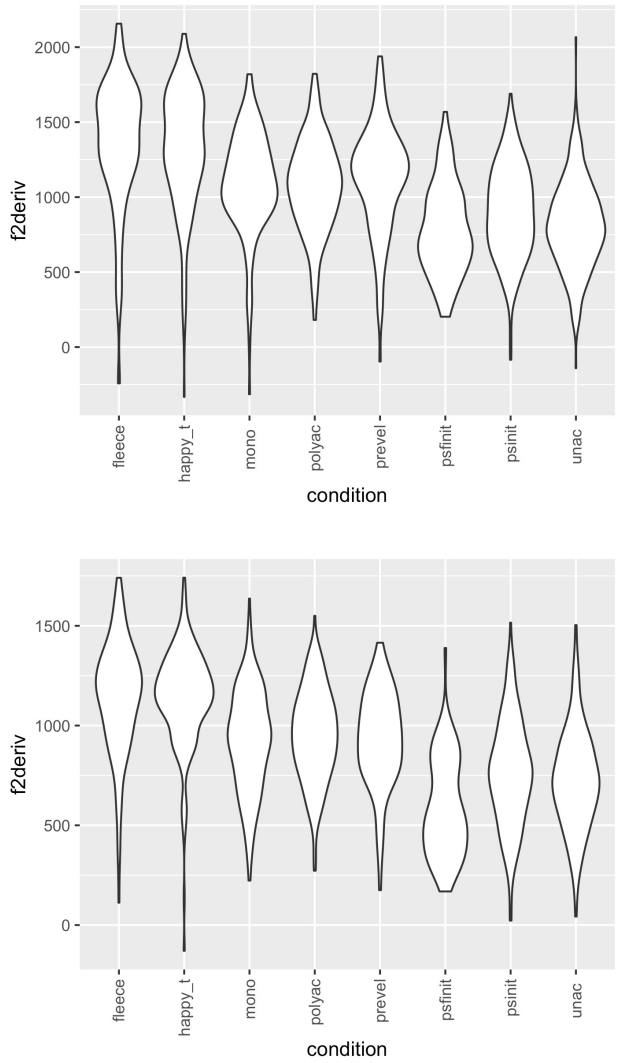

Figure 1: distribution of F2DERIV (Hz) for each of the eight subgroupings of the dataset described: females top panel; males lower panel.

The HAPPY_T and FLEECE tokens have the highest distribution of F2DERIV values, with a second cluster being formed by MONO, POLYAC and PREVEL tokens and the least close/front realisations being found for UNAC, PSINIT and PSFINIT tokens.

In order to gauge the extent to which the various distributions across the subgroupings were quantitatively different, mixed-effect models were calculated using the lmer function as part of the lme4 package [13] in R [14]. P-values were calculated using the lmerTest package [15]. For all models, F2DERIV was the dependent variable, speaker and word were included as random effects, and the subgrouping was included as a fixed effect. The data for males and females were modelled separately.

The quantitative analysis largely reflected the clustering evident from the plots. Using FLEECE as the reference grouping, the highest model estimates were found for HAPPY_T, which was not significantly different from FLEECE but different from all the other groupings. For both males and females, the UNAC grouping proved to be significantly different from MONO, POLYAC and PREVEL, but not from either of the *it* groupings. For female speakers, neither of the *it* groupings were differentiated statistically from the MONO/POLYAC/PREVEL set, whereas for male speakers, the phrase-final PSFINIT grouping was, thereby patterning with the UNAC grouping.

A similar picture is evident if the data are plotted in a conventional F1/F2 chart, although the differentiation is a little less clear to the eye (see Figure 2). This reinforces the validity of using the single F2DERIV measure as a metric for the variation that is targeted in this study.

We also examined the full set of historical KIT tokens for any relationship between F2DERIV values and the duration of vowel tokens. There was no compelling evidence of any correlation between the two. The vast majority of tokens had durations of less than 100 ms, and while there was a tendency for there to be only a few relatively long tokens in the low F2DERIV range, a subsequent statistical analysis yielded non-significant correlations for both male and female speakers.

4. DISCUSSION

In general, the findings are in line with what was anticipated, given the conditions under which WVM is said to apply and the previous acoustic study by Cox & Palethorpe [3]. The historical KIT vowels in HAPPY_T contexts are closer and fronter than those in MONO words, which in turn are fronter and closer than those found in unaccented syllables in the UNAC content words. Likewise there is a good deal of consistency in the realisation of historical KIT across MONO, POLYAC and PREVEL contexts, and more reduced variants overall are found in the UNAC and the two "*it*" groupings PSINIT and PSFINIT.

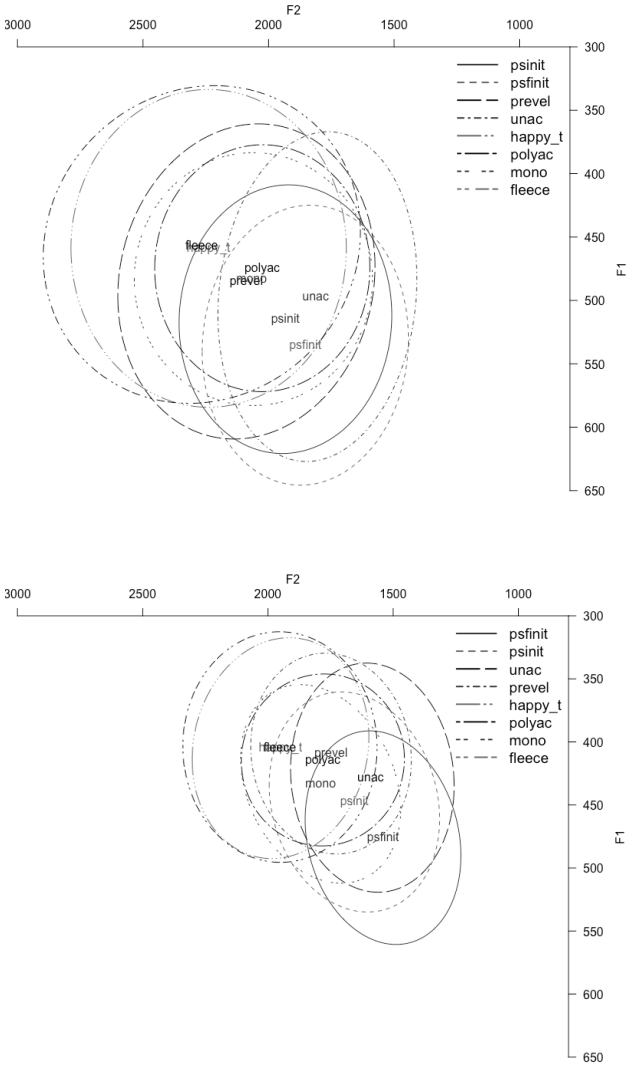

Figure 2: Distribution of realisations in F1/F2 space (Hz) for each of the eight subgroupings of the dataset described: females top panel; males lower panel; labels centred on the mean F1/F2 value with ellipses at +/- 1s.d.

The F1/F2 values for the more reduced variants, together with the extensive distributional overlap of all of the different subgroupings with each other (and in fact with FLEECE), are consistent with Cox & Palethorpe's [3]'s suggestion that WVM often results in a vowel which is more of an [ɨ] quality than a schwa. Of course, it is a point of debate whether schwa itself has a definable target that could act as a common point of reference for this sort of comparison [16], but the results reported here, together with those reported in [3], suggest that the conceptualisation of WVM as a simple merger of KIT with SCHWA in unstressed syllables may not be the best way to capture the patterns of realisation observed across this dataset. This interpretation could be tested by adding to the data sample tokens in which SCHWA figures as part of the phonological specification in order to provide the full FLEECE-to-SCHWA range in which to situation the historical KIT realisations. This forms part of the next stage of this line of investigation.

One obvious limitation of this study is that it has only tracked a single grammatical item, *it*. Further work will be needed to ascertain if *it* is typical of other unstressed grammatical words. While it appears to be patterning alongside unstressed KIT vowels in content words (excluding those with a following velar consonant), some further scrutiny is needed of what appears to be a bimodal distribution for males' realisation of *it* in phrase-final position possibly mapping to the auditory impression of occasional phrase-final tokens having particularly centralized realisations.

If the WVM is not straightforwardly a merger with SCHWA there is a question about the nature of the phonological representation of those lexical items which tend to have the [ɨ] realisation. When we refer to the KIT lexical set in English spoken by Australian speakers, do we simply need to specify that we are referring only to accented contexts? But it is a moot point whether for speakers and listeners, the vowels in *it, is, this, massive, radical* are identified in an exemplar-based representation as "the same" as the vowels in *delicious, Christmas, gifts*, in turn begging the question of whether the full range of contexts need to be included in accounts of ongoing variation and change in the vowel system. This is a significant issue to address given the high frequency of many of the relevant grammatical items. In order to resolve this, it would be instructive to undertake some auditory/ perceptual experimentation with speakers of English in Australia to gauge how they categorise those same vowels in comparison to how they categorise KIT variants in accented syllables.

Finally, we should note some non-canonical realisations that were characteristic of the conversational speech style represented by our sample; in particular, widespread reduction of *it's* to [ts] or [s], and regular elision of the vowel in [s_z] plural affix contexts (e.g. *places, faces*) with the unstressed syllable carried by continuous friction and a switch of voicing (noting the high frequency of the [s_z] plural affix). This underscores the value of investigating phonological variation within natural connected speech styles and the need to develop models of production and processing that are consistent with what we know of the phonetic properties of those styles.

Acknowledgment
Research funded by the Australian Research Council (DP130104275).

6. REFERENCES

[1] Wells, J. 1982. *Accents of English*. Cambridge University Press.

[2] Cox, F., Palethorpe, S. 2007. Australian English (Illustrations of the IPA). *Journal of the International Phonetic Association,* 37(3), 341-350.

[3] Cox, F., Palethorpe, S. 2018. Rosa's roses – unstressed vowel merger in Australian English. *Proc. 17th Australasian Int. Conf. on Speech Science and Technology* (SST17). Australasian Speech Science and Technology Association (ASSTA), 89-92.

[4] Gordon, E., Campbell, L., Hay, J., Maclagan, M., Sudbury, A., Trudgill, P. 2004. *New Zealand English: Its Origins and Evolution.* Cambridge University Press.

[5] Flemming, E., Johnson, S. 2004. Rosa's roses: reduced vowels in American English. *Journal of the International Phonetic Association,* 31, 83-96.

[6] ELAN (Version 6.4) [Computer software]. 2022. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan

[7] Fromont, R., Hay, J. 2012. LaBB-CAT: An annotation store. *Proc. Australasian Language Technology Workshop (ALTA),* 113-117.

[8] Young, S., Evermann, G., Gales, M. et al. 2006. *The HTK Book (for HTK Version 3.4)*, Cambridge University Engineering Department.

[9] Boersma, P., Weenink, D. Praat, 2018. Version 6.0.40.

[10] Cox, F. Docherty, G.J. Forthcoming. Sociophonetics and vowels. In Strelluf, C. (ed.). *The Handbook of Sociophonetics.* Routledge.

[11] Labov, W., Rosenfelder, I., Fruehwald, J. 2013. One hundred years of sound change in Philadelphia: linear incrementation, reversal, and reanalysis. *Language* 89(1). 30–65.

[12] Grama, J., Travis, C.E., Gonzalez, S. 2019. Initiation, progression, and conditioning of the short-front vowel shift in Australia. *Proc. 19th Int. Congress Phon. Sci (ICPhS),* 1769-1773.

[13] Bates, D, Mächler, M., Bolker, B., Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software,* 67, 1-48.

[14] R: A language and environment for statistical computing. (2020). R Foundation for Statistical Computing, Vienna, Austria. [Online]. Available: http://www.R-project.org/

[15] Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13), 1–26.

[16] Cohen Priva, U., Strand, R. 2023. Schwa's duration and acoustic position in American English. *Journal of Phonetics,* 96, 101198.