An acoustic analysis of short front vowel realisations in the conversational style of young English speakers from Western Australia

Gerard J. Docherty[+], Simon Gonzalez[*], Nathaniel Mitchell[+], Paul Foulkes[#]

Griffith University[+], Australian National University[*], University of York[#]
gerry.docherty@griffith.edu.au

ABSTRACT

This study presents the findings of an analysis of short front vowel (SFV) realisations in a corpus of unscripted conversational speech generated by 40 young speakers from Perth. As well as providing the first comparative account of SFV realisations in that location, we consider the extent to which the realisational variability observed is associated with properties of the continuous unscripted speech style that are known to influence spectral and temporal properties of vowels.

Keywords: Sociophonetic variation, vowel realisations, English in Australia, conversational speech.

1. INTRODUCTION

This paper presents findings from a study of the sociophonetic properties of vowel realisations in unscripted conversations between young speakers from Perth, Western Australia. Unscripted conversational interaction is by far the most common speech style that speakers engage in, and is therefore the natural environment in which to test for socially-structured variability. It is, however, the style that we know least about for Australian varieties of English, since, with few exceptions [e.g. 16], the majority of acoustic phonetic studies have focused on isolated word tokens (mostly controlling for consonantal context by using an /hVd/ frame).

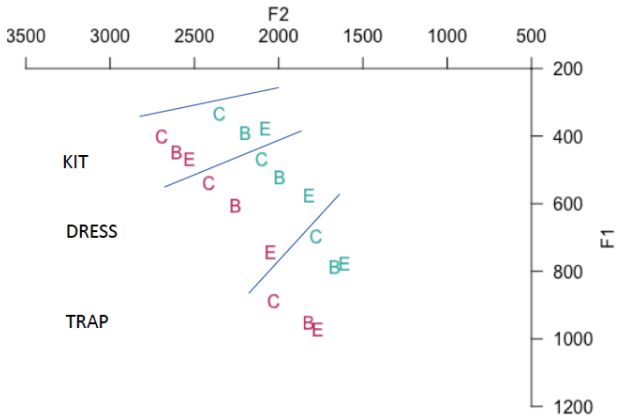
Beyond addressing this particular lacuna in our knowledge of Australian English (AusE), our focus on conversational speech style is also motivated by the fact that models of phonological representation and theories of sound change are increasingly hinged on listeners' processing and representation of the ambient speech that they experience as participants in a speech community. This is not a new development (cf. Ohala's model of the listener's contribution to change [24,25]). However, it has moved centre-stage as a consequence of the development and testing of models of processing and representation with an episodic dimension (e.g. [15,17,18]), and is reflected in recent work on topics such as perceptual sensitivity to sociophonetic variability ([28]), speech accommodation ([1,26]), and on modelling of the prior information that listeners bring to speech processing tasks [19]. It is therefore important that we understand the nature of listeners' routine experience as participants in a speech community, something which is not adequately captured by accounts of isolated word realisations.

Previous work [6,22,29] suggests we should expect conversational style realisations of vowels to be quite different from those found in controlled isolated words tasks, for example in the form of spectral and temporal reduction, and greater variability arising from variations in tempo, context, prosody, speaker, etc. In this study, we focus in particular on varying degrees of vowel reduction. Research stretching back over 40 years suggests that at least the following factors may be associated with greater levels of reduction [5,6]: shorter vowel duration; high frequency words; words with lower neighbourhood density; more predictable words; repeated mentions of words; words in a more casual speaking style. Of particular interest is evidence pointing to sociophonetic features gravitating towards those contexts favourable to greater reduction [5,21,23]. Therefore, these factors are more than simply potential confounds in our characterisation of vowel realisations in conversational speech – they might in fact need to be closely woven in to an account of any sociophonetic variability that surfaces. Hence, our current analysis sets outs to consider the extent to which factors such as these are relevant in accounting for the realisations that we have captured in our Perth sample. Specifically, our question is what factors (social and contextual) impact on the degree of vowel reduction in the short front vowel lexical sets in the unscripted performance of our Perth speakers?

While there have been relatively few apparent- or real-time studies of sound change in AusE, there is compelling evidence of the short front vowel (SFV) series (KIT, DRESS TRAP) having participated in a complex series of changes over the last 100 years or so. In a study drawing on a range of sources (the oldest of which was a set of speakers born between 1885 and 1895), Cox & Palethorpe [8] note that change in the realisation of SFVs has altered direction over the past 40-50 years. In the earlier material, speakers of AusE were characterised by a general raising of the SFVs, which Cox & Palethorpe suggest was triggered by a raising of TRAP in a "push-chain" type of shift. In more recent recordings, however, it is clear that following a reversal of the TRAP trajectory,

such that it is now produced with a much lower realisation, all of the SFVs have lowered or are continuing to lower compared to their earlier realisations. Cox & Palethorpe note that this is akin to what would be expected with a drag-chain model of vowel shift. Other recent studies of vowel realisations in AusE [3,7,12] confirm the overall finding of SFV lowering, but they also reveal more accentuated lowering of DRESS and TRAP in particular (Figure 1).

Figure 1: Mean F1/F2 frequencies for KIT, DRESS TRAP lexical sets reported by Cox [7] (C), Billlington [3] (B), and Elvin, Williams & Escudero [12] (E). Female/male speakers in darker/lighter shade respectively.



The differences evident in Fig. 1 could be because the studies shown are of different varieties (B's speakers are from Melbourne, E's are from Western Sydney, which the authors argue cannot be conflated with C's material from Sydney [13]). However, what they reveal is considerable variability in the realisation of a series of vowels that have been undergoing change over many generations.

Our analysis enhances our understanding of this aspect of AusE by providing data from a location that has previously not been subject to investigation. By analyzing SFV realisations in conversational speech we are able to establish the extent to which conventional assumptions about the configuration of SFVs also apply in that style, and we are able to test for the effects of some of the factors inherent to that style on patterns of realisation.

2. METHODS

The participants in the study were all young people living in Perth (aged 18-22), having been entirely schooled (from age 5) in the city (n= 40, 20 females and 20 males). Speakers were classified by whether or not they resided in neighbourhoods ranked by the Australian Bureau of Statistics to be in the top socio-economic decile. This classification acts as a proxy for social class, allowing us to test the hypothesis, arising from anecdotal comments, that speakers from higher ranking neighbourhoods are identifiable by their phonetic characteristics. Neighbourhood selection was balanced equally through the sample; the higher SES neighbourhoods are referred to as NhoodA, and the others as NhoodB.

Speakers participated in same-sex conversational dyads in which they were invited to converse without a script for c. 30 minutes. Most speakers were known to each other in advance but with varying degrees of familiarity. A fieldworker was present in the room to manage the recording process but only intervened on the rare occasion that the participants struggled to maintain the conversation.

The recordings for each speaker (44Khz/ 16 bit) were segmented in Elan (starting 5 minutes in to each recording) and force-aligned within LaBB-CAT [14] using HTK [30], with manual correction of misalignments. Findings are reported below for realisations of the KIT, DRESS and TRAP lexical sets (a total of 2,168 tokens). The vowels analysed were produced in a range of consonantal environments, but pre-/l, w, j/, pre-nasal, and post-/w, j, r/ environments were excluded from the present analysis, as were tokens that were located in grammatical words.

Using default settings in Praat [4], F1/F2 tracks were estimated for each vowel (all lexical words whether or not in accented position). For the present analysis, the frequencies of F1 and F2 were calculated at the midpoint of each token of the KIT, DRESS and TRAP lexical sets (see [11] for caveats on this static approach to vowel description). For the present purposes, it was decided not to normalise the F1/F2 measurements, in order to provide a basis for comparison with previous studies.

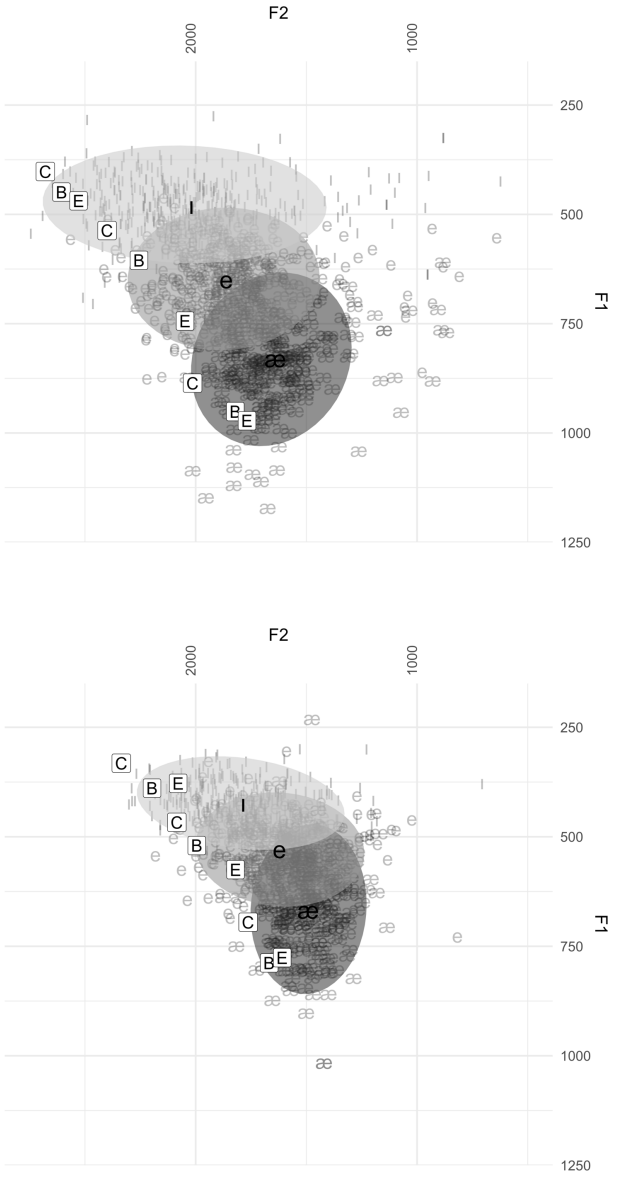
3. FINDINGS

3.1. SFV distributions

Figure 2 shows the distribution of realisations of the three SFVs separately for females and males. For comparison, Fig. 2 also shows the mean values for each vowel as reported by Cox [7], Billington [3], and Elvin et al [12] reproduced from Fig. 1.

It can be seen that the mean F1/F2 values for each of the SFVs are more reduced than those reported for the isolated word tokens in previous studies; i.e. more centralised and somewhat more compressed within F1/F2 space. The three mean values are fairly evenly spread, but it is clear that the mean values themselves are not an adequate representation of the significant variability associated with the realisation of each vowel, nor of the substantial overlap between the three categories. In particular, for both males and females, DRESS realisations appear to overlap substantially with the other two categories. This suggests that the configuration of the SFV series in this variety is not particularly driven by separation of contrasting vowels in routine conversational speech.

Figure 2: Mean F1/F2 frequencies (Hz) for KIT, DRESS & TRAP within the Perth corpus. Ellipses centred around the means at 95% CI. Top panel: females; bottom panel: males. Isolated word SFV means reported by [7] (C), [3] (B), and [12] E reproduced from Fig. 1 shown in small squares.



3.1. Degrees of vowel reduction

In order to shed light on factors influencing the degree of reduction in SFV realisations, we adopted an approach previously used by DiCanio et al [10]. For each speaker we identified a vowel realisation centroid based on the grand mean of monophthong realisations (across the full range of monophthong lexical sets with the exception of schwa and FLEECE – the later often being realised as a diphthong by speakers of AusE). The reduction of each monophthong token was equated to the Euclidean Distance (ED) in Hz between the midpoint F1/F2 for that token and the speaker-specific centroid; i.e. the larger the ED the less the reduction.

We undertook mixed-effect regression modelling of our data in order to test for an association between the extent of vowel reduction and a range of factors hypothesised to be relevant in accounting for variability in degree of reduction. After a preliminary analysis to identify the predictors that improved the model, the following predictors were used as fixed factors in the final model: vowel duration, speaker sex, neighbourhood, syllable count (# syllables in the item), first mention (vs. 2nd or subsequent mention) and word frequency. The random intercepts were speaker and words; no interactions were tested. Factor significance was determined using lme4 [2] in R [27] and calculated using the Satterthwaite likelihood t-tests. Corpus neighbourhood density was calculated using an R script which identifies distances between words in the corpus based on their similarities using an approximate string matching function. Word frequencies were determined using CELEX; we used the CobLog frequency, which is the logarithmic frequency of the COBUILD corpus, comprising over 17m words.

No significant effects were found for syllable count, first mention and word frequency. The overall model showed that the significant predictors were duration ($\chi2(1) = 11.566$, $p < 0.001$), sex ($\chi2(1) = 29.367$, $p < 0.001$), Nhood ($\chi2(1) = 4.5553$, $p = 0.0328$), and vowel ($\chi2(2) = 181.78$, $p < 0.001$). The analysis shows that longer vowels were less reduced ($\beta = 296.36$, $df = 2116.92$, $p = 0.0007$), males showed more reduction than females ($\beta = -90.99$, $df = 35.41$, $p < 0.0001$), speakers of Nhood B showed more reduction than those in Nhood A ($\beta = -30.25$, $df = 34.77$, $p = 0.0407$) and KIT showed overall less reduction than DRESS ($\beta = -197.12$, $df = 370.05$, $p < 0.0001$) and TRAP ($\beta = -169.45$, $df = 465.44$, $p < 0.0001$). The latter effect is a by-product of the method adopted for measuring degree of reduction; all else being equal, KIT realisations are collectively further from speakers' centroid values than DRESS or TRAP, and show less dispersion in the F1 plane. Fig. 3 shows a summary of the modelling analysis.

We subsequently ran individual models separately for each SFV. This allowed us to observe what factors were significant for each vowel. DRESS yielded significant differences for Sex ($\beta = -74.72$, $df = 33.69$, $p < 0.0001$) and Nhood ($\beta = -31.01$, $df = 33.37$, $p = 0.048$), with Males and Nhood B showing more reduction than Females and Nhood A speakers (see panel A in Fig 4). KIT showed significant differences for Sex ($\beta = -129.41$, $df = 35.89$, $p<0.0001$) and Duration ($\beta = 615.23$, $df = 641.84$, $p = 0.0025$), with Males and shorter vowels showing more reduction (see panel B in Fig. 4). TRAP also showed significant differences for Sex ($\beta = -0.07681$, $df = 4392.77$, $p < 0.0001$) and Duration ($\beta = -0.07681$, $df = 4392.77$, p

< 0.001), with Males and shorter vowels showing more reduction (panel C in Fig. 4).

Figure 3: Summary of the significant findings of the overall mixed effects model (see text for details), plotted using the R package sjPlot [20].

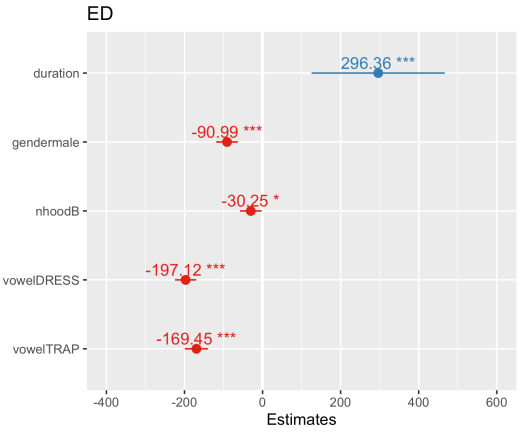
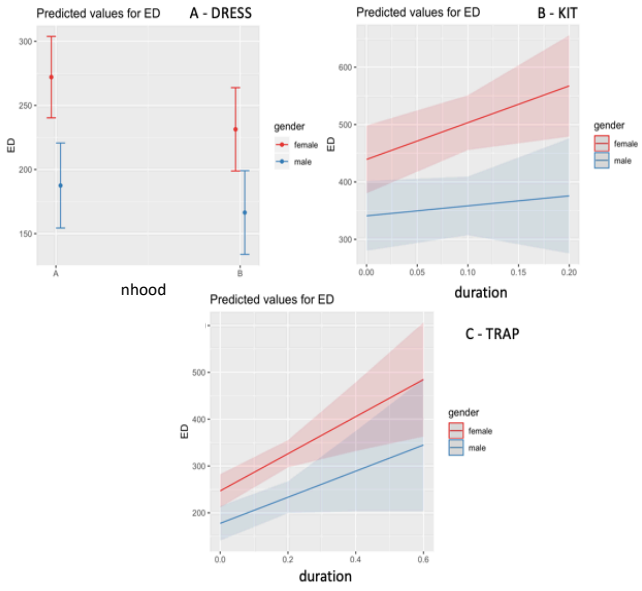


Figure 4: Predicted Euclidean Distance (ED)-based measures of vowel reduction arising from the independent models for KIT, DRESS, and TRAP (>ED = < reduction, see text for details), plotted using the R package sjPlot [20].



Our quantitative analysis therefore points to realisational patterns of the SFVs being influenced by the overall duration of the vowel as expected, and also by sex and neighbourhood. However, these factors are not at play equally throughout the SFV set; for example, while duration appears to be the predictor with the greatest weight, it is not significant in the realisation of DRESS. Sex is the one predictor that is consistently significant across all three SFVs, with female speakers showing less reduction than male speakers, but caution and further investigation are called for before it can be concluded that this reflects a male vs. female stylistic difference. While the use of an ED-based measure of reduction anchored on speaker-specific centroids goes a long way to capturing the intrinsic acoustic differences between male and female vowels, it may not have eliminated differences arising from the overall area covered by the male vs. female vowel spaces (evident in Fig. 2).

3. DISCUSSION

The findings of our analysis shed new light on the realisation of SFVs by English speakers in Australia. The mean F1/F2 measures for the conversational tokens of our Perth speakers are more reduced and compressed than those of the isolated word studies previously reported. Our analysis has also revealed a wide range of variation in each of the SFVs, together with substantial areas of overlap between KIT and DRESS and DRESS and TRAP (and for male speakers between all three). The factors associated with reduction and compression in the connected speech style appear to outweigh any drive to maximise the distinctiveness of the three vowels. The experience of listeners in the Perth speech community is therefore not one in which the SFVs can be neatly partitioned into relatively distinct distributions in vowel space, which in turn has potential implications for how we theorise about the nature of any changes under way in this variety (and potentially other varieties of AusE).

Our findings also highlight the importance of ensuring that when testing for the impact of social factors on the realisation of vowels in connected speech materials, it is vital to build in an assessment of the full range of contextual factors that could be influential in relation to those vowel distributions; in this case, vowel duration proved to be significant for the SFVs overall and for two of our vowels independently. The fact that duration proved to be significant in vowels that are by definition "short" is quite striking, as short vowels are inherently less capable of varying in duration. Our expectation is that this would be even more marked in other lexical sets. In this regard though, we note though that it is important to assess the extent to which duration operates independently of lexical accent as a predictor of degree of reduction. Likewise, it would be useful to test for the influence of some of the other contextual factors (such as lexical frequency, neighbourhood density etc.) on lexical sets that are less temporally constrained. These are all matters for further investigation.

(Research funded by the Australian Research Council DP130104275).

4. REFERENCES

[1] Babel, M. 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *J. Phonetics* 40, 177-189.

[2] Bates, D., Machler, M., Bolker, B. M., Walker, S. C. 2015. Fitting linear mixed-effects models using lme4. *J. Statistical Software* 67, 1–48.

[3] Billington, R. 2011. Location, location, location: Regional characteristics and national patterns of change in the vowels of Melbourne adolescents. *Australian Journal of Linguistics* 31, 275-303.

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

[5] Clopper, C. G., Mitsch, J. F., Tamati, T. N. 2017. Effects of phonetic reduction and regional dialect on vowel production. *J. Phonetics* 60, 38-59.

[6] Clopper, C. G., Turnbull, R. 2018. Exploring variation in phonetic reduction: Linguistic, social, and cognitive factors. In Cangemi, F., Clayards, M., Niebuhr, O., Schuppler, B., Zellers, M. (eds), *Rethinking Reduction: Interdisciplinary perspectives on conditions, mechanisms, and domains for phonetic variation*. Berlin: Mouton, 25—72.

[7] Cox, F. 2006. The acoustic characteristics of /hVd/ vowels in the speech of some Australian teenagers. *Australian Journal of Linguistics* 26, 147-179.

[8] Cox, F., Palethorpe, S. 2008. Reversal of Short Front Vowel Raising in Australian English. *Proc. 2008 SST Conference.*

[9] Delvaux, A., Soquet, A. 2007. The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica* 64, 145-173.

[10] DiCanio, C., Nam, H., Amith, J., Castillo García, R., Whalen, D.H. 2015. Vowel variability in elicited versus spontaneous speech: Evidence from Mixtec, *J. Phonetics* 48, 45–59

[11] Docherty, G. J., Gonzalez, S. & Mitchell, N. 2015. Static vs dynamic perspectives on the realisation of vowel nuclei in West Australian English. *Proc. ICPhS,* Glasgow UK.

[12] Elvin, J., Williams, D., Escudero, P. 2016. Dynamic acoustic properties of monophthongs and diphthongs in Western Sydney Australian English. J. *Acoust. Soc. Am.* 140, 576-581.

[13] Elvin, J., Escudero, P. 2014. Comparing acoustic analyses of Australian English vowels from Sydney: Cox (2006) versus AusTalk. *Proc. 2014 SST Conference.*

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

[15] Goldinger, S. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105, 251-79.

[16] Grama, J., Gonzalez, S. & Travis, C. 2018. Phonological Conditioning in the rotation of short front vowels in Australian English. *Proc. 2014 SST Conference.*

[17] Hay, J., Foulkes, P. 2016. The evolution of medial (-t-) over real and remembered time. *Language* 92, 298-330.

[18] Hay, J., Pierrehumbert, J., Walker, A., LaShell, P. 2015. Tracking word frequency effects through 130 years of sound change, *Cognition* 139, 83-91.

[19] Kleinschmidt, D. F., Jaeger, T. F. 2015. Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review,* 122, 148-203.

[20] Lüdecke, D. 2018. *sjPlot: Data Visualization for Statistics in Social Science*. R package version 2.4.1.9000, https://CRAN.R-project.org/package=sjPlot.

[21] Mendoza-Denton, N., Hay, J., Jannedy, S. (2003). Probabilistic sociolinguistics: Beyond variable rules. In Bod, R., Hay, J., Jannedy, S. (eds.), *Probabilistic Linguistics,* Cambridge MA:MIT Press, 97–139.

[22] Moon, S-J., Lindblom, B. 1994. Interaction between duration, context, and speaking style in English stressed vowels. *J. Acoust. Soc. Am*. 96, 40–55.

[23] Munson, B., Solomon, N. 2004. The effect of phonological neighbourhood density on vowel articulation. *J. Speech, Lang, Hearing Res*. 47, 1048-1058.

[24] Ohala, John J. 1981.The listener as a source of sound change. *Chicago Linguistic Society (Parasession on language and behavior)* 17, 178–203.

[25] Ohala, John J. 1993.The phonetics of sound change. In Jones, C. (ed.), *Historical linguistics: Problems and Perspectives,* London: Longman, 237-78.

[26] Pardo, J.; Gibbons, R., Suppes, A., Krauss, R. 2012. Phonetic convergence in college roommates. *J. Phonetics* 40,190–97.

[27] R Core Team. R: A language and environment for statistical computing, 2018. Version 3.4.2.

[28] Sumner, M., Kim, S. K., King, E., and McGowan, K. 2014. The socially-weighted encoding of spoken words: A dual-route approach to speech perception. *Frontiers in Psychology* 4, 1-13.

[29] Tucker, B., Ernestus, M. 2016. Why we need to investigate casual speech to truly understand language production, processing and the mental lexicon. *The Mental Lexicon,* 11, 375-400

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