STATIC vs DYNAMIC PERSPECTIVES ON THE REALIZATION OF VOWEL NUCLEII IN WEST AUSTRALIAN ENGLISH

Gerard Docherty, Simón Gonzalez & Nathaniel Mitchell

Griffith University, Australia
gerry.docherty@griffith.edu.au, s.gonzalez@griffith.edu.au, n.mitchell@griffith.edu.au

ABSTRACT

This paper reports on an exploratory study of the application of different types of analysis method to the characterization of the acoustic properties of vowel realization in the performance of speakers of West Australian English (West AusE). Tense monophthongs and diphthongs produced in a word list by 18 speakers of West AusE were analysed using three different methods, two static and one dynamic. Results differ across the three methods with the dynamic analysis yielding substantially more detail and differentiation between and within vowel categories. Our findings enhance knowledge of a variety which has received scant attention in existing phonetic studies of AusE, and more generally contribute to the on-going discussion in the literature about which approach to acoustic analysis provides the best means of capturing the properties of vowel realization and variability.

Keywords: Vowels, sociophonetics, Australian English, acoustic analysis.

1. INTRODUCTION

The findings reported below arise from the first study to be undertaken of sociophonetic variability within the West Australian variety of English (West AusE) as produced by speakers from the Perth metropolitan area. A fundamental question to be addressed in any study of this sort is how best to capture and represent vowel realizations and the variability in these realizations within the speech sample concerned. Of course the analysis and representation of vowel realization and variability as fixed points in F1/F2 space is long-established [12]. This is relatively uncontroversial for monophthongs with researchers adopting a range of metrics for determining the point in time at which F1/F2 best represents the vowel target – for example, vowel midpoint, or the point of maximum F2. But the practice of classifying a vowel in respect of a single static measurement point is somewhat more controversial for diphthongs which by definition are characterised as a vector in vowel space rather than as a single point. While this has led some investigators to sample F1/F2 properties of diphthongs at more than one time-point in order to capture their dynamic aspects (e.g. [2], [6]), it is noteworthy that many investigators have adopted a single measurement point approach for both diphthongs and monophthongs, thereby characterising the former as a static point in F1/F2 space. This practice is particularly evident in work on social-indexical variability in vowel realizations, being present in the earliest work of this sort ([8]) and enduring to the present (e.g. [9]) and now embedded in the commonly-used FAVE (Forced Alignment and Vowel Extraction) programme [14]. Thus, the FAVE analysis deployed in [9] leads to F1/F2 for the FACE lexical set being sampled at the location of maximum F1.

In parallel to the above, there has been a growing strand of work which takes as a point of departure the premise that all vowels have inherent dynamic characteristics and that it is not sufficient to characterise them in terms of a point in F1/F2 space. Examples of studies which take this approach are [10, 15, 17, 18]. Note that by adopting this approach, the phonetic difference between a monophthong and diphthong becomes one of degree rather than of category. Note too that there is greater variability in the method of analysis which is adopted for capturing the dynamic acoustic properties of vowels than is evident within the 'static' approach.

The aim of the present study is to investigate the merits of different methods of capturing the acoustic properties of vowels, with a view to informing our approach to the analysis of a large-scale corpus of West AusE recorded in a natural conversational style. Since this is an exploratory study geared to refining our analysis methods, we focus on a subset of vowels (including monophthongs and diphthongs) drawn from a word-list produced by the participants. We compare two approaches to deriving 'static' measures of vowel quality with one method for determining the vowels' dynamic properties. Our test-bed for this study is the extent to which the various methods investigated reveal an effect on the realization of the vowel arising from whether it is embedded within a checked or unchecked syllable. This focus was chosen as a result of pilot work for the project [4] which revealed for the NEAR lexical set substantially greater diphthongisation in unchecked than in checked syllables and informal auditory analysis which suggested that similar differences may exist in other lexical sets.

2. METHOD

2.1 Participants and materials

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. Recordings of 18 participants (12 females and 4 males) were used in the present study. In a sound-treated recording studio, participants read a 165-item isolated word list designed as a probe to elicit carefully produced tokens of certain key social-indexical variables (the West AusE project is primarily focused on a corpus of natural conversations between pairs of speakers). As such, unlike most previous work on AusE vowels, which has typically investigated /hVd/ environments, the segmental environments for vowels within the word-list were diverse. For the present study, we analysed a subset of tense vowels and diphthongs, focusing on those which had tokens in the word-list in both checked and unchecked syllables: FACE, PRICE, CHOICE, GOAT, NEAR, SQUARE, FLEECE, BATH and GOOSE. This provided 66 vowel tokens per speaker. The small number of tokens per speaker is counterbalanced by the eventual 200 speaker sample size for the main project.

2.2 Acoustic Measurements

The recordings for each speaker were first segmented in *Elan* and subsequently force-aligned within *LaBB-CAT* [5] using *HTK* [19] with subsequent manual correction of misalignments. Using *Praat* [1], F1/F2 tracks were estimated for each vowel, extracting 50 points for each vowel token spread equidistantly through the segmented vowel interval. Three methods were then deployed to capture the acoustic properties of the vowels, as follows:

2.2.1 Static – Fixed Point

The first approach based the analysis on a single fixed point in the vowel interval irrespective of the configuration of the formants at that point in time. For monophthongs, formant values were measured at the temporal midpoint of the vowel. For diphthongs, formant values were extracted at 25% of the vowel interval (as previously carried out, for example, by [16]).

A script was written in *R* ([13]) for an automatic calculation of the significant differences between checked and unchecked contexts using a Wilcoxon test. The statistical test examined differences for each of F1 and F2 pooled across speakers but with male and female speakers analysed and reported separately.

2.2.2 Static – Target Estimate

The second approach also involved measuring F1 and F2 at a single point within each vowel, but in this case the definition of the point varied in order to best capture what might be deemed the vowel 'target'. A similar methodology has been applied in previous studies of AusE [2, 17], and is widely used in analyses of vowel variation and change more generally (e.g. [9]). For high front vowels, the formants were measured at the point where F2 peaks. For high back vowels the measurement point was located at the trough of F2. For low vowels, it was located at the peak of F1. For diphthongs, the target was determined as for monophthongs depending on the quality of the initial phase of the diphthong – thus for NEAR our measurement point was located at the peak of F2, for PRICE, at the peak of F1, etc. We also carried out a Wilcoxon test to determine whether there were differences between the checked/unchecked conditions. Male and female token were analysed and reported separately.

2.2.3 Dynamic - SSANOVA

The third approach analysed the trajectories of F1/F2 across the duration of the vowel. SSANOVA ([3]) was used to calculate statistical differences between the best fit of multiple formant trajectories across the two conditions. In this way the technique can be used to determine statistically whether "the shapes of multiple curves are significantly different from one another" ([3]: 411). The statistical analysis is based on a comparison of 95% Bayesian confidence intervals associated with each smoothed spline. This technique has been used in previous vowel formant analyses studies ([6], [11]). These studies have looked at variability in the realization of formant trajectories ([11]), and at the impact on this of factors such as speaker sex and age ([6]). In this study, we compare formant trajectories of the same vowel in the checked and unchecked conditions.

For each vowel token we imported into *R* the 50 points generated by Praat for each formant. The next step was to extract the middle 80% of each trajectory in order to reduce the effects of immediate segmental context. With the data organized by vowel and context, we carried out a SSANOVA analysis based on 40 equidistant points for each vowel token. Previous analyses have used fewer points (7 in [11] and 9 in [6]). We chose 40 points to have an accurate tracing of vowel formants since the greater the number of points in the trajectory, the most reliable is the best fit obtained. Following [3], we used smoothing parameters for SSANOVA sourced from the *R* package *gss*.

We carried out three analyses of the trajectory data for each formant separately, again pooling across speakers but analysing males and females separately. First we tested whether the trajectories were significantly different across the checked/unchecked conditions. If our analysis yielded significant differences at any portion in the trajectory (as determined by non-overlapping Bayesian 95% confidence intervals), then the formant trajectories were classified as significantly different from one another. Second, we examined the directionality of the trajectory classifying them as either rising or falling. Third, we considered the extent and location of any overlap between trajectories across the two conditions. These three dimensions were chosen to enable us to capture some of the principal parameters of the dynamic representations which our analysis generated.

3. RESULTS

Figure 1 provides an overview in F1/F2 space of the acoustic properties of the vowels focused on in this study (due to space constraints we show the male speakers only here). The results of the statistical analysis undertaken for each vowel and across all three measurement methods are shown in Table 1. An asterisk indicates that for a particular vowel and analysis method a significant difference was found across checked and unchecked conditions.

A first observation is that in the comparisons based on static measurement points, only a minority of cells yield a significant difference. Comparing the two 'static' approaches, it can be seen that statistical differences are more prevalent for F2 than for F1, suggesting that differences which the static measures are identifying are more related to relative anteriority of the tongue body rather than to overall apertures of the vocal tract. It is also evident from Table 1 that finding a difference using one particular static analysis method does not mean that the same difference will be found with the other static measure tested in this study. Thus, both GOOSE and (in line with our pilot work [4]) NEAR both show differences in either F1 and/or F2 across the two conditions, but the findings across the two static measures are not consistent.

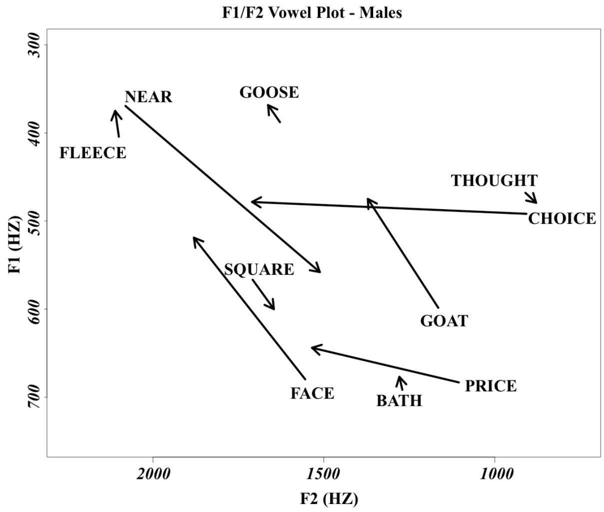

Figure 1. Mean F1/F2 Vowel Plot for the male speakers pooling across checked and unchecked conditions. The values are based on measurements taken at the 10% and 90% points of the vowel intervals.

The second main finding is that when the checked/unchecked conditions are compared using the dynamic approach to characterising F1/F2, a very different picture emerges with significant differences yielded in 31/40 cells. Further analysis of the outcomes of the dynamic analysis sheds light on differences in vowel realization that the static approaches fail to capture. In particular, results from

	F1						**F2**					
	fixed		***Target***		***dyn.***		***fixed***		***target***		***dyn.***	
Vowel	*M*	*F*	*M*	*F*	*M*	*F*	*M*	*F*	*M*	*F*	*M*	*F*
Bath	-	-	-	-	*	*	-	-	-	-	*	*
Choice	-	-	-	-	*	-	-	-	*	*	*	*
Face	-	-	-	*	*	*	-	-	-	-	*	*
Fleece	-	-	-	-	*	*	-	-	-	-	*	*
Goat	-	*	-	-	-	*	-	-	-	-	*	*
Goose	-	-	-	-	*	*	*	*	-	*	*	*
Near	*	*	-	*	*	*	*	*	-	-	*	-
Price	-	-	-	-	-	*	-	-	-	-	-	*
Square	-	-	-	-	*	*	-	*	-	-	*	-
thought	-	-	*	-	-	-	-	-	*	*	-	*

Table 1. Results of the statistical comparison of F1/F2 trajectories across checked/unchecked conditions (* in the fixed/target columns = significantly different ($p < 0.05$), * in the dynamic column = non-overlapping 95% Bayesian CIs)

the application of the dynamic method to comparing across checked/unchecked conditions reveal a range of traits which could be associated with a statistical difference between two trajectories. We now exemplify these in relation to the four examples shown in Figure 2 (F2 for BATH-males, FACE-males, and PRICE-males and F1 for FLEECE-males).

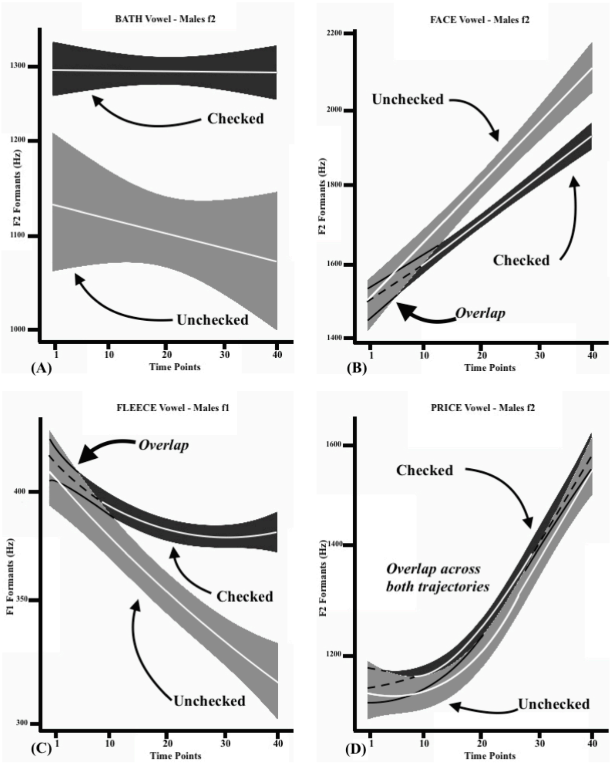

Figure 2. Dynamic analysis results for F2/BATH/males (A), F2/FACE/males (B), F1/FLEECE/males (C), and for F2/PRICE/males (D): solid lines show SSANOVA trajectories across checked/unchecked conditions together with 95% Bayesian CIs (shaded)

The dynamic analysis reveals that while two trajectories might be significantly different this does not mean that they are necessarily statistically different throughout their duration. Our results showed differences in the extent of overlap which was found in formant trajectories across different conditions. In some cases, such as A in Figure 2, trajectories were entirely non-overlapping, but in the majority of the comparisons undertaken using SSANOVA some extent of trajectory overlap was found. Examples B, C and D in Figure 2 show cases of overlap but to different degrees. In B and C, the overlap is present for a relatively short portion of the overall trajectories, a pattern which was found to be the most frequently occurring within our dataset. By contrast, in D, there is observable overlap along the trajectories of both formants reflecting the fact that, in this instance, SSANOVA yields a non-significant difference in comparing across the two conditions. In the cases where trajectory overlap was found across the two conditions, the SSANOVA analysis also allowed an analysis of where the overlap was located within the trajectories being compared. B and C in Figure 2 show examples of overlap in the early part of the F1 trajectory (the pattern which was found most frequently within the dataset), whereas D, as previously indicated, shows an F2 trajectory overlap for PRICE extending through the full vowel interval.

Inspection of the SSANOVA results also revealed differences in the directionality of formant trajectories across conditions. For example, panel A in Figure 2 shows a flat trajectory for F2 in the BATH/male vowels in the checked context, but a falling trajectory in the unchecked context. Figure 2-B shows an example where the formants in both conditions have a rising trajectory, and C shows a case where the F1 tracks adopt increasingly divergent trajectories as the vowel realization progresses. Results showed that the majority of the comparisons (75%) involved trajectories which did not differ in respect of directionality (as in B, for example).

The dynamic analysis also provides a sense of the relative consistency of overall formant trajectories across tokens and conditions; e.g. the relatively narrow 95% CI bands in panel B suggest much greater consistency overall for F2 trajectories in FACE than is found (in A) for F2 trajectories in BATH.

4. DISCUSSION

This study is exploratory in nature, and caveats need to be applied before generalizing from the findings described above. In particular, while the SSANOVA analysis focused only on the central 80% of the vowel interval, further analysis is needed to ascertain the extent to which differences across checked/unchecked conditions might be confounded by the fact that consonantal context was not necessarily consistent across the two. Further analysis is also required to investigate the extent to which inter-speaker variability may be a relevant factor in respect of the differences across condition reported above. However, the findings do indicate that there is a good deal of significant variability across conditions which the static analyses regularly used in studies of this sort have been unable to identify. SSANOVA seems to be well-adapted to the task of identifying and characterising the nature of these differences, although the fact that a large number of the dynamic comparisons proved to be significant calls for caution in interpreting those differences and suggests there would be merit in comparing the findings arising from the particular approach to dynamic formant analysis used in study with that adopted by other investigators (e.g. [7, 10, 18]).

Research funded by the Australian Research Council (DP 130104275).

5. REFERENCES

[1] Boersma, P., Weenink, D. 2014. Praat: doing phonetics by computer [Computer program]. Version 5.4.04, retrieved 28 December 2014 from http://www.praat.org/

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

[3] Davidson, L. 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *J. Acoust. Soc. Am.* 120, 407-415.

[4] Docherty, G.J. 2013. Sources of variation in the realization of vowels in West AusE. Presentation to the 1st LVC-A Conference, Melbourne, July 2013.

[5] Fromont, R., Hay, J. 2012. LaBB-CAT: An annotation store. University of Otago, Dunedin, New Zealand: *Australasian Language Technology Workshop (ALTA),* 4-6 Dec 2012. In Proceedings 10: 113-117.

[6] Haddican, W., Foulkes, P., Hughes, V., Richards, H. 2013. Interaction of social and linguistic constraints on two vowel changes in northern England *Language Variation and Change*, 25, 371–403.

[7] Jacewicz, E., Fox, R., Salmons, J. 2011. Cross-generational vowel change in American English *Language Variation and Change*, 23, 45–86.

[8] Labov, W., Yaeger, M., Steiner, R. 1972. A quantitative study of sound change in progress: Volume 1. Report on National Science Foundation Contract NSF-GS-3287. Philadelphia, PA: University of Pennsylvania.

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

[10] McDougall, K. 2006. Dynamic Features of Speech and the Characterisation of Speakers: Towards a New Approach Using Formant Frequencies. *International Journal of Speech, Language and the Law,* 13, 89-126.

[11] Nycz, J., De Decker, P. 2006. A New Way of Analyzing Vowels: Comparing Formant Contours Using Smoothing Spline ANOVA. Presentation at the 35th NWAV Conference, Ohio State University, Columbus, Ohio.

[12] Peterson, G. E., Barney, H. L. 1952. Control methods used in a study of the vowels. *J. Acoust. Soc. Am*. 24, 175–184.

[13] R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

[14] Rosenfelder, I., Fruehwald, J., Evanini, K., Jiahong, Y. 2011. *FAVE (Forced Alignment and Vowel Extraction) Program Suite*. http://fave.ling.upenn.edu.

[15] Van der Harst, S., Van de Velde, H., Van Hout, R. 2014. Variation in Standard Dutch vowels: The impact of formant measurement methods on identifying the speaker's regional origin. *Language Variation and Change,* 26, 247–272.

[16] Van Heuven, V., J., Edelman, L., van Bezooijen, R. 2002. The pronunciation of /ɜI/ by male and female speakers of avant-garde Dutch. In Broekhuis, H., Fikkert, P. (eds.), *Linguistics in the Netherlands 2002*. viii, 222 pp. (pp. 61–72)

[17] Watson, C., Harrington J. 1999. Acoustic evidence for dynamic formant trajectories in Australian English vowels. *J. Acoust. Soc. Am*. 106, 458-468.

[18] Williams, D. Escudero. P. 2014. A cross-dialectal acoustic comparison of vowels in Northern and Southern British English. *J. Acoust. Soc. Am*. 136, 2751-2761.

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