HORIZONTAL DIPHTHONG SHIFT IN NEW ZEALAND ENGLISH

Márton Sóskuthy,[1] Jennifer Hay,[2] James Brand[2]

[1]University of British Columbia, [2]NZILBB – University of Canterbury
marton.soskuthy@ubc.ca

ABSTRACT

The diphthongs PRICE and MOUTH in New Zealand English have been characterised as showing 'diphthong shift' and 'glide weakening'. Like most analyses of diphthongs, previous studies have focussed on changes in the qualities of two temporally fixed targets: the nucleus and the offglide. We model changes in the full formant trajectories of PRICE and MOUTH over a period of 130 years. We first present results from Generalised Additive Models that reveal substantial 'horizontal' restructurings in the timing of the trajectories alongside expected 'vertical' shifts in nucleus and off-glide targets. We then use computational methods to capture these horizontal changes by identifying points of inflection in F1 and F2 trajectories. The timing of these inflection points moves significantly later for both vowels. We argue that changes in trajectory timing are a significantly overlooked aspect of changing diphthong production.

Keywords: Diphthong contours, New Zealand English, Sound change, Generalised Additive Models

1. INTRODUCTION

In principle, diphthongs may show variation in the targets for their component units, the timing of these targets, the shape and length of the transition between them, and possibly other dynamic features as well. In this paper, we are concerned with two specific aspects of this variation in the realm of acoustics: 'vertical' variation involving shifts in formant values at different target positions; and 'horizontal' variation relating to the timing of these targets (the analogy of vertical vs. horizontal variation comes from spectrographic displays of diphthongs with time on the horizontal axis and frequency on the vertical axis). We focus specifically on diachronic changes in the diphthongs PRICE and MOUTH in New Zealand English.

Traditional approaches to the study of diphthongal sound change do not capture information regarding temporal structure. The standard method is to study changing formant values at two target positions: the nucleus and the offglide (e.g. [5, 2, 12]). This approach is able to capture vertical changes in the production of diphthongs, and also changes in the degree of diphthongisation by considering the acoustic distance between the two targets [10, 13]. However, such work has not traditionally been concerned with potential horizontal changes.

There is increasing evidence that diphthongs are more than simply a sum of their parts. Peeters & Barry [14] found that Dutch, English and German listeners prefer different timing relations among the nucleus, transition and offglide in /aɪ/ and /aʊ/-type diphthongs. Watson & Harrington [16] presented evidence that dynamic information can significantly improve the discriminability of diphthongs, and reported substantial timing differences in the location of the target between tense and lax vowels. Fox & Jacewicz [6] found variation in formant dynamics across three American dialects, including timing differences in the realisation of PRICE. Finally, Cardoso [4] identified patterns of synchronic phonological variation and diachronic change in the location of the 'inflection point' in PRICE and MOUTH in Liverpool English.

While our understanding of the temporal dynamics of diphthongs has improved considerably over the last three decades, there is still little work analysing large data sets in terms of changes in diphthongal trajectory and – with the exception of [4] – none that looks at horizontal change over time. In this paper, we attempt to fill this gap by modelling the full trajectories of PRICE and MOUTH, and investigating changes in these trajectories over the history of New Zealand English. Both vowels have been affected by a process of 'diphthong shift' (cf. [17]). The nucleus of MOUTH has fronted, the nucleus of PRICE has backed, and the offglides for both have lowered [8, 15]. While the process of diphthong shift has traditionally been described as above, with reference to vertical changes in vowel quality, we are interested in the degree to which horizontal changes could also be observed. In this paper, we focus on PRICE and MOUTH in contexts where they are not followed by a voiceless consonant.

All data and code for this paper are avail-

able as an OSF repository at the following URL: `osf.io/74mza`. We encourage readers to consult the animated summary of our findings under `icphs > MOUTH_PRICE_in_NZE.mp4`.

2. GENERAL METHODS

2.1. Materials

We extracted dynamic F1 and F2 measurements for all PRICE and MOUTH tokens in the forced-aligned [18] Origins of New Zealand English (ONZE) corpus [9]. The core ONZE corpus contains recordings from over 500 speakers of New Zealand English born between 1857–1988, thus spanning 130 years of sound change in apparent time. The formant measurements were extracted automatically using a combination of the software package LaBB-CAT [7] and Praat [3]. Each vowel token was sampled at 11 evenly spaced time points, yielding time-normalised formant measurements at 0%, 10%, 20%, etc. of the vowel duration.

2.2. Data processing

Since the formant measurements were generated automatically using unsupervised methods, they inevitably include errors. To remove problematic tokens, we heavily filtered the raw data. First, we manually excluded 41 speakers due to systematic errors in formant tracking. We established reasonable absolute bounds for F1 and F2 for males and females using the 1st and 99th percentiles of the formant data in [11] and excluded all formant measurements outside these bounds. Within each speaker, we also excluded F1/F2 measurements that were further from the lower/upper quartiles than 1.5 times the interquartile range. After the exclusion of individual measurement points, we discarded all vowel tokens that had more than 4 measurement points missing for either F1 or F2. This leaves an overall 31,025 vowel tokens for PRICE and 20,123 vowel tokens for MOUTH from 539 speakers. Although these filtering methods capture many errors in the data, it should be noted that even the cleaned data set contains a large amount of noise. We deal with this issue by employing smoothing techniques in our statistical analyses, which are designed to infer smooth underlying curves from potentially noisy observations.

2.3. Data analysis

We conducted two analyses for each vowel. First, we used generalised additive models (GAMs) to track changes in the shapes of F1 and F2 contours as a function of year of birth. This analysis allows us to observe key trends in changing formant dynamics. The second analysis focuses specifically on the temporal dynamics of PRICE and MOUTH. We use computational methods to track changes in the position of *inflection points* along the trajectories. More detail about the specifics of our methods is provided alongside the results in the next section.

Our analyses are exploratory (as opposed to confirmatory; see e.g. [1]). Conventional measures of statistical significance such as p-values are known to be unreliable in such modelling settings. Therefore, they are best seen as 'indicators of surprise and should not be taken at face value as exact probabilities' ([1], p. 227). All the major patterns that we describe are extremely robust and emerged as significant in all the models that we fit. Nonetheless, to discourage a confirmatory interpretation of our results, p-values are only used in a supportive role in this paper.

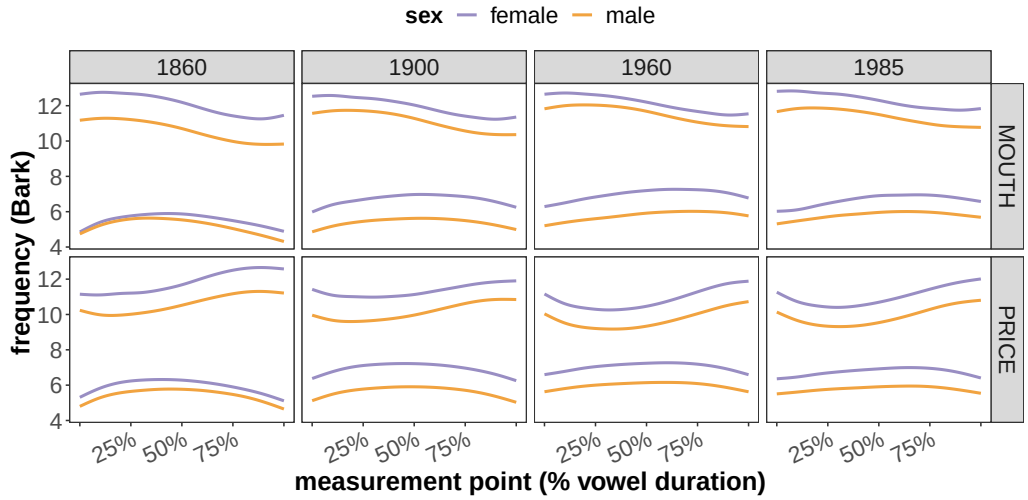
3. MODELLING RESULTS

3.1. GAM analysis

The model presented in this section was fit to an aggregated version of the raw data consisting of pointwise by-speaker averages for F1 and F2. Thus, each speaker is represented by 11 measurement points for F1 and F2 for each vowel (which obviates the need for random effects in our model). We fit a single large fixed effects GAM to the F1 and F2 values for both vowels, including separate tensor product smooths over year of birth and measurement point (i.e. the location of the measurement along the vowel trajectory) for each combination of vowel, F1/F2 and sex. The model also controlled for potentially non-linear effects of average log vowel duration within speakers. Our tensor product smooths are analogous to interactions between year of birth and measurement point in a conventional regression model, but they allow for non-linearities in the main effect of measurement point, the main effect of year of birth, and their interaction. In practice, this means that our model is sufficiently flexible to capture complex – and potentially decelerating or accelerating – changes in the shapes of formant contours. The decision to fit a single model to all our data was motivated by practical concerns; fitting separate models to each combination of vowel, F1/F2 and sex would yield essentially identical results.

While p-values from exploratory models must be taken with a pinch of salt, all terms relating to changes in the height and shape of formant contours are significant at a level well beyond $p < 0.0001$

Figure 1: Predicted F1 and F2 contours for MOUTH (top) and PRICE (bottom) for females (purple) and males (orange).



(with the exception of changes in the shapes of F2 contours for MOUTH, which are significant at $p = 0.034$ for females and $p = 0.0009$ for males). In practical terms, this suggests that we can be confident that the observed changes are robust.

The main results of the model are summarised in figure 1, which shows snapshots of the model's predictions for F1/F2 separately for the two vowels and two sexes at different time points. The snapshots are not linearly spaced in terms of year of birth. They were chosen strategically to show key milestones in the changes. For a fuller presentation of our results, consult the accompanying OSF repository, which includes animations that cover the entire time period (`icphs > MOUTH_PRICE_in_NZE.mp4`). These animations also include audio generated by filtering a synthetic source using the predicted formant trajectories. The IPA transcriptions in this section are based on these synthetic stimuli.

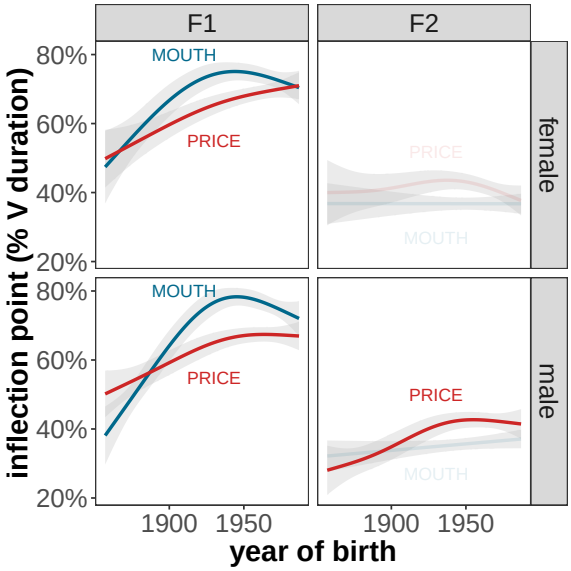
Let us first discuss MOUTH. The F1 contour shows a fairly substantial overall increase for both the females and the males, along with a general flattening of the trajectory shape. Most of this increase occurs between 1860 and 1900, with some further incrementation between 1900 and 1960. There is a slight reversal of this change after 1960 that is somewhat more pronounced for the female speakers. Another change in F1 that follows roughly the same timeline relates to its maximum: while the F1 maximum is timed near the beginning of the contour for speakers born around 1860, it shifts to the middle of the contour by 1900 and to the end of the contour by 1960 for both males and females. This change also shows a slight reversal near the end of the observed time period. The changes in F2 are much less pronounced: we see a slight overall rise in F2 accompanied by a flattening of the contour with no obvious changes in timing. Our auditory impressions based on the resynthesised stimuli can best be summarised as follows: [ə̹ʉ̹] (1860) > [ɜ̹ə̹] (1900) > [ɛ̹ˑə̹] (1960) for the females; and [ɐ̹ʊ̹] (1860); [ɛ̹ə̹] (1900); [ɛˑə̹] (1960) for the males.

PRICE shows very similar changes in F1, though the shift in the F1 maximum seems to be implemented slightly more slowly, and does not reach completion until the last snapshot. F2 changes more dramatically than it does for MOUTH: the portion between 15%–40% of vowel duration shows gradual lowering, moving from a relatively flat shape to a dipped curve. The main inflection point (which becomes the F2 minimum after 1860) shifts gradually towards the centre of the vowel. This movement is more obvious for the males than it is for the females. The changes can be transcribed roughly as follows: [ɜ̹ɪ̹] (1860) > [ɑ̹ə̹] (1900) > [ɑˑə̹] (1960) for the females; and [ʌ̹ɪ̹] (1860) > [ɑ̹ə̹] (1900) > [ɑˑə̹] (1960) for the males.

3.2. Analysis of inflection points

We referred to points of interest in the formant contours as the 'F1 maximum' and the 'F2 inflection point' / 'minimum'. In some cases (such as the F2 minimum for PRICE), these points of interest correspond well with the conventional division of diphthongs into nucleus and offset; in other cases (such as the gradually shifting F1 maximum for PRICE and MOUTH), it is harder to find equivalents in conven-

Figure 2: Predicted location of inflection point for F1 (left) and F2 (right) contours for MOUTH (blue) and PRICE (red) for females (top) and males (bottom). Non-significant smooths are shown using faint lines.



tional units of analysis. In this section, we follow [4] in using the term 'inflection point' to refer to them.

We define inflection point as a prominent change of direction (or 'bend') in a formant trajectory. For F1 contours and MOUTH F2 contours, we were primarily interested in the location of ∩-shaped bends; for PRICE F2 contours, we were interested in the location of ∪-shaped bends. We operationalised this definition of inflection points using the first and second derivatives of smooth curves fitted to each speaker's data. ∩-shaped bends were identified by looking for (i) a point along the trajectory where the first derivative reached 0 and the second derivative had a negative value, or, if no such point was found, (ii) the minimum of the second derivative. ∪-shaped bends were identified by looking for (i) a 0 along the first derivative with a positive corresponding value along the second derivative or (ii) a maximum in the second derivative. This operationalisation closely matched our own intuitions about inflection points in the majority of cases.

We extracted a single inflection point per formant per vowel for each speaker using the method described above, and fitted a GAM with the location of the inflection point as the outcome variable, and separate smooths over year of birth for each combination of vowel, sex and formant. Figure 2 summarises the results of the model in the form of a prediction plot. Smooths that did not show significant changes are shown using faint lines.

The results from this model largely confirm our impressions from the previous section. The inflection point shows a substantial shift over time toward the end of the vowel for F1 for both vowels and both sexes. The change is more rapid for MOUTH than it is for PRICE, reaching completion just before 1950, and reversing slightly after 1950. PRICE shows a continuous shift throughout the entire period for both sexes. The final location of the inflection point is essentially the same for PRICE and MOUTH around 70%. There is also a somewhat more subtle shift in the F2 inflection point for PRICE for males (starting at around 30% and ending at 40%), but no similar shift is observed for females.

4. DISCUSSION AND CONCLUSIONS

Analysis of changing trajectories for the PRICE and MOUTH vowels in New Zealand English reveals that the inflection point of the vowel has moved significantly later in both vowels. We strongly suspect that these changes have been crucial in creating diphthongs that are auditorily distinct from corresponding vowels in other varieties. This suggests that analyses which are restricted to formant measurements at static target points may overlook the timing dimension as an important aspect of within-vowel diphthongal variation. Our findings are also in line with previous accounts of changes in NZE diphthongs: the initial portion of PRICE shifts towards the back (and lowers), while the initial portion of MOUTH shifts slightly towards the front (and lowers); the final portion of both diphthongs shows heavy centralisation.

The relative independence of horizontal changes in F1 and F2 poses challenges for accounts assuming a simple binary division between nucleus and offglide. For instance, the F2 contours for MOUTH in the 1900 snapshot of figure 1 would suggest the existence of two targets around 30% and 75%, while the F1 contour suggests a single target around 50%. One possible interpretation is that diphthongal targets are inherently dynamic, and that speakers retain fine-grained information about their temporal coordination. An alternative interpretation is that the horizontal movement of the F1 peak is an artefact of asynchronous vertical shifts in the nucleus and the offglide. Our findings do not allow us to clearly distinguish between these two possibilities. The notion of horizontal diphthong shift provides a descriptively appealing characterisation of the observed changes, but the cognitive bases of such horizontal shifts clearly require further investigation.

5. REFERENCES

[1] Baayen, R. H., Vasishth, S., Bates, D., Kliegl, R. 2017. The cave of shadows. addressing the human factor with generalized additive mixed models. *Journal of Memory and Language* 94, 206–234.

[2] Blake, R., Josey, M. 2003. The /ay/ diphthong in a Martha's Vineyard community: What can we say 40 years after Labov? *Language in Society* 32(4), 451–485.

[3] Boersma, P., Weenink, D. 2009. Praat: doing phonetics by computer (version 5.0.38) [computer program]. Version 5.1.17; Retrieved on 29/09/2009 from http://www.praat.org/.

[4] Cardoso, A. 2015. Dialectology, phonology, diachrony: Liverpool English realisations of PRICE and MOUTH. Doctoral dissertation, University of Edinburgh.

[5] Cox, F., Palethorpe, S. 2001. The changing face of Australian English vowels. *Varieties of English around the world: English in Australia* 17–44.

[6] Fox, R. A., Jacewicz, E. 2009. Cross-dialectal variation in formant dynamics of American English vowels. *The Journal of the Acoustical Society of America* 126(5), 2603–2618.

[7] Fromont, R., Hay, J. 2008. ONZE Miner: the development of a browser-based research tool. *Corpora* 3(2), 173–193.

[8] Gordon, E., Campbell, L., Hay, J., Maclagan, M., Sudbury, A., Trudgill, P. 2004. *New Zealand English: its origins and evolution*. Cambridge: Cambridge University Press.

[9] Gordon, E., Maclagan, M., Hay, J. 2007. The ONZE Corpus. In: Beal, J. C., Corrigan, K. P., Moisl, H. L., (eds), *Models and methods in the handling of unconventional digital corpora: Volume 2, Diachronic Corpora* volume 2. Basingstoke, Hampshire: Palgrave Macmillan 82–104.

[10] Haddican, B., 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(3), 371–403.

[11] Hillenbrand, J., Getty, L. A., Clark, M. J., Wheeler, K. 1995. Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America* 97, 3099–3111.

[12] 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.

[13] Maclagan, M., Hay, J. 2007. Getting fed up with our feet: Contrast maintenance and the New Zealand English short front vowel shift. *Language Variation and Change* 19(01), 1–25.

[14] Peeters, W. J. M., Barry, W. J. 1989. Diphthong dynamics: production and perception in Southern British English. *First European Conference on Speech Communication and Technology*.

[15] Sóskuthy, M., Hay, J., Maclagan, M., Drager, K., Foulkes, P. 2017. *The closing diphthongs in early New Zealand English* 529–561. Cambridge University Press.

[16] Watson, C. I., Harrington, J. 1999. Acoustic evidence for dynamic formant trajectories in australian english vowels. *The Journal of the acoustical society of America* 106(1), 458–468.

[17] Wells, J. 1982. *Accents of English. 3 Volumes*. Cambridge: Cambridge University Press.

[18] Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X., Moore, G., Odell, J., Ollason, D., Povey, D., et al., 2002. *The HTK book*. Cambridge University Engineering Department.