# Tongue positions corresponding to formant values in Australian English vowels

*Arwen Blackwood Ximenes*[1], *Jason A. Shaw*[1], *Christopher Carignan*[1]

[1]The MARCS Institute, Western Sydney University

a.blackwoodximenes@westernsydney.edu.au

## Abstract

A common assumption about vowel formants is that F1 inversely correlates with tongue height and F2 inversely correlates with tongue backness. This study compared vowel formants and corresponding lingual articulation in Australian English (AusE) for nearly all of the AusE monophthongs. Simultaneous acoustic and electromagnetic articulography (EMA) recordings are reported for four speakers producing multiple repetitions of ten monophthongs. Results show that, while in general formants correspond to the articulatory data, there are also cases in which the typically assumed correspondence breaks down. Consistency in Tongue Dorsum position was observed despite variation in F2.

**Index Terms**: speech production, acoustics, Electromagnetic articulography, Australian English

## 1. Introduction

Assumptions are commonly made regarding the articulatory nature of vowels based on their formant values: F1 is assumed to be inversely correlated with tongue height and F2 is assumed to be inversely correlated with tongue backness. This paper assesses this correspondence, reporting the tongue position of Australian English (AusE) vowels and corresponding formant values.

There is an abundance of acoustic studies of AusE vowels but comparative articulatory data are lacking. Some recent studies on vowel articulation focus on a small subset of vowels. Tabain [1] investigates the articulatory and acoustic properties of one vowel in different prosodic contexts. Watson, Harrington and Palethorpe [2] compared the acoustic and articulatory vowel spaces of AusE and New Zealand English (NZE). Their analysis covered four vowels, those in the words *hid*, *head*, *had*, and *herd*. Lin, Palethorpe and Cox [3] looked at a larger number of AusE vowels in the /CVl/ context. They focused on how vowel height influences lateral production (/CVl/) rather than on the phonetic properties of the vowels themselves. The degree to which the following /l/ influences the preceding vowel is not clear. The most comprehensive articulatory study of AusE vowels was undertaken over four decades ago [4]. Bernard reports on the results of an X-ray study investigating all the AusE vowels but does not report any quantitative measurements of the data. Bernard's qualitative description of X-ray data still constitutes the most comprehensive analysis of Australian vowel articulation to date.

This paper aims to address the lack of quantitative data on AusE vowels by describing the AusE vowel space articulatorily and comparing the results to formant values.

## 2. Method

### 2.1. Subjects

Articulatory and acoustic data were analysed from four Australian English speakers (two males and two females) ranging in age at time of recording from 20 to 42. All participants were recruited from the Western Sydney University community.

### 2.2. Materials

Stimuli comprised a list of lexical items and nonce words containing 15 vowels, including 10 monophthongs, in the sVd context. This paper focuses on analysis of the monophthongs. We list the stimulus items below. Each item is followed by, in parentheses, the reference word for the vowel devised by Wells [5]. The reference word disambiguates the spelling, which is particularly useful for nonce words: *said* (DRESS), *seed* (FLEECE), *sood* (FOOT), *sued* (GOOSE), *sid* (KIT), *sod* (LOT), *sawed* (THOUGHT), *surd* (NURSE), *sud* (STRUT), *sad* (TRAP). This set of monophthongs covers the whole AusE acoustic vowel space. The only AusE monophthong missing is START, which according to Cox [6] does not differ in its formants from STRUT.

### 2.3. Procedure

The movements of the articulators were tracked using an Northern Digital Inc. Wave EMA system at a sampling rate of 100Hz. This system uses an electromagnetic field to track the movement of small receiver coils or sensors (3 mm in size) glued or taped to the articulators. The electromagnetic field induces an alternating current in the sensors, and the strength of this current is used to determine the position of the sensors in relation to the transmitter. Articulatory movements are captured in the vertical, horizontal and lateral dimensions with high spatial-temporal resolution (< 0.5 mm rms error). In this study, we focused on movements in the horizontal and vertical dimension, since these are the dimensions typically assumed to correspond to formant values. The sensor trajectories were synchronized to the audio signal during recording by the NDI system. EMA sensors were glued to the following articulators along the midsagittal plane: jaw (below the lower left incisor), lips (at the vermillion edge of the upper and lower lip), tongue tip (TT), tongue blade (TB) and tongue dorsum (TD). The TD sensor was placed as far back as comfortable for the participant. The TT sensor was placed near the tip and the TB sensor was placed midway between the TT and TD sensors.

The target stimulus words were displayed on a computer monitor placed outside of the magnetic field. One word was presented per trial. There were 15 trials (one per vowel) per block and eight blocks in the experiment. This resulted in 15 (vowels) x 8 (repetitions) = 120 vowel tokens per participant.

Of the recorded data, the monophthongs consist of 10 (vowels) x 8 (repetitions) = 80 tokens per participant, 320 monophthong tokens in total. Speech acoustics were recorded using a shotgun microphone at a sampling rate of 22 kHz. Technical problems due to data acquisition, analysis, and mispronunciation, resulted in three tokens (less than 1% of the total data) being excluded from the analysis.

Head movements were corrected computationally, with reference to sensors glued to the nasion and mastoids. The articulatory data were rotated relative to the occlusal plane so that the origin of the coordinate system corresponds to the front teeth. The occlusal plane was established by having the participant bite down on a protractor with 3 sensors affixed in a triangular formation.

### 2.4. Articulatory measurements

Measurements were extracted from sensor trajectories using the labelling procedure, *findgest*, an algorithm developed for the Matlab-based software package, "Multi-channel visualization application for displaying dynamic sensor movement" (MVIEW), by Marke Tiede at Haskins Laboratories. This program was used to detect the nearest tangential velocity minimum of the TD sensor (taken from movement in the horizontal and vertical dimensions together) during the interval corresponding to the vowel. We then extracted positional coordinates from all the lingual sensors at this vowel target landmark.

In some cases the parse MVIEW provided for the TD velocity minimum was impacted on by the surrounding consonants, i.e. TD reached its positional target for the vowel during the preceding /s/ (velocity minimum), making it difficult to differentiate the velocity peaks of the vowel and adjacent consonant. TB was used in cases where it showed more controlled movement towards vowel constriction.

### 2.5. Acoustic measurements

Formant data (F1 and F2) for each vowel was extracted from the sound files at the time point of the articulatory measurements (i.e., the vowel target, as described above). Using the time points extracted from the articulatory measures for the acoustic analysis enables a direct comparison between articulation and acoustics. Our method of parsing vowel targets using the point of minimum velocity in the articulatory data follows similar general principles used to identify formants in Cox [6] and Harrington, Cox and Evans [7]. In these papers vowel targets were identified based on formant displacement patterns, e.g., max/min F1/F2, depending on vowel. Max/min formant values correlate closely to the minimum velocity of articulator movement in our data. Other acoustic studies have used the acoustic midpoint of the vowel, which did not consistently correspond to the velocity minimum of the TD or TB sensors in this data.

### 2.6. Analysis

One of the challenges of analyzing speech production across speakers is that anatomical differences influence both the formant values and EMA positional coordinates. In the case of formants, differences in vocal tract length influence the average formant values. In articulatory data, differences in tongue shape, volume, and sensor placement lead to different average values. In both cases, because of differences in anatomy, between-speaker differences for the same vowel can be larger than within-speaker differences across vowels. In order to facilitate comparison across our four speakers, we normalized both the formant values and the positional coordinates by calculating z-scores of the horizontal and vertical dimensions for the TD sensor and of F1 and F2. Z-scoring preserves the within-speaker structure of the data but allows for a direct comparison across speakers by controlling for interspeaker vocal tract differences.

## 3. Results

We report the acoustic results first followed by the articulatory results. After describing the general pattern in both sets of data and the correspondence between them, we take up some exceptions to the main pattern in the discussion section.

### 3.1. Acoustic data

The distribution of normalized formant values (F1 and F2) across the acoustic vowel space for all speakers is presented in Figure 1. The ellipses contain 95% of the data for each vowel, and are centered on the mean of each vowel category.

In line with previous acoustic studies of AusE (e.g., [6]), the vowels are fairly evenly distributed across the vowel space and can be classified as "front", "central", and "back" on the basis of the formants. There are four vowels with high F2 (i.e. "front" vowels) that differ in F1: FLEECE, KIT, DRESS and TRAP. There are also differences in F2 amongst the front vowels, but part of these differences are due to general properties of formant spaces, e.g., as F1 increases, F2 of front vowels decreases. We assume that the differences in F2 are at least in part attributable to this relationship and may not be under speaker control. There are three "central" vowels that have intermediate F2 values, GOOSE, NURSE and STRUT, and also differ in increasing F1. The remaining "back" vowels have low F2: FOOT, LOT and THOUGHT. We now turn to the articulatory data to observe how the differences in formant values correspond to tongue position.

### 3.2. Articulatory data

In order to assess whether the observations of formants correspond to "front", "central", and "back" articulatory positions, we focus on the TD sensor. The mapping from articulation to acoustics is of course impacted by differences in vocal tract diameter across the entire length of the vocal tract. Nevertheless, we have found that even data from the TD sensor alone reveals a general correspondence to formant values in line with expectations. Figure 2 shows the normalized values (z-scores) of the TD sensor for all four subjects. The TD data represents the range of motion with which that fleshpoint on the tongue varies across vowels. The *y*-axis shows the vertical position, and the *x*-axis shows horizontal position from front (positive z-scores on the left side of the figure) to back (negative z-scores on the right side of the figure). As with the formant data, ellipses contain 95% confidence intervals for each vowel distribution and are centered on the mean.

The distribution of vowels in the articulatory data generally follows the distribution of vowels in formant space. More specifically, F1 tends to be inversely correlated with tongue height, and F2 tends to be inversely correlated with tongue backness. Of the "front", "central", and "back" vowels determined on the basis of the formants, the back vowels show the least overlap at the TD sensor. The "back" vowels, FOOT, THOUGHT, and LOT, have a TD position more posterior than the other vowels, as indicated by the negative z-score.

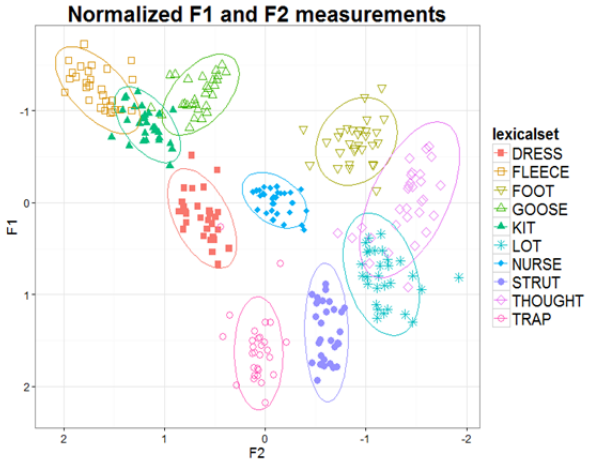


Figure 1. *Normalised F1 and F2 for Australian English vowels*

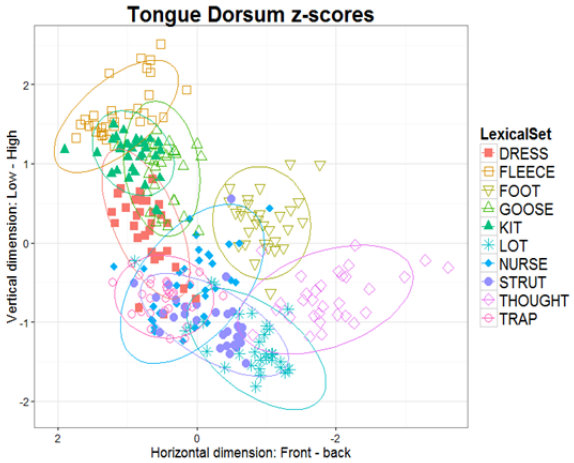


Figure 2. *Z-scores of the Tongue Dorsum sensor position for Australian English vowels*

The center of the ellipses for STRUT and NURSE are closest to zero on the *x*-axis, indicating that they are at the average level of backness in the data. These vowels, in addition to GOOSE, are the "central" vowels: they all had intermediate F2 values. Of these three central vowels, GOOSE is the most front TD position. Presumably, rounding of GOOSE lowers F2, compensating for TD frontness. The front vowels FLEECE, KIT, DRESS, and TRAP have horizontal positions that are higher than all the other vowels, indicating that they have the lowest degree of backness.

Although Figure 2 shows data from just a single fleshpoint on the tongue, articulatory differences that correspond to those in the formants can be observed. In particular, the relative height and backness of vowels at the TD sensor is preserved in the F1 and F2 values. The vowel space expressed in terms of TD position is more compact than the vowel space expressed in formants. Consequently, there is more overlap in the articulatory data compared to the acoustic data. From this we can ascertain that other aspects of vowel articulation function to enhance the differences observable from TD position, so TD does not capture all of the articulatory change.

## 4. Discussion

The AusE vowels in this study can be clearly differentiated on the basis of F1 and F2, and a similar partitioning of the vowel space can be observed in the position of the TD sensor in vertical and horizontal dimensions. The AusE vowel space can be viewed as taking a 4:3:3 configuration, whereby there are four "front" vowels differing in height, three "central" vowels differing in height and "three" back vowels also differing in height. For the most part, the acoustic and articulatory data are in correspondence, as is expected from the assumption that F1 is inversely correlated with tongue height and F2 is inversely correlated with backness. This is remarkable given that the articulatory data come from a single fleshpoint, the TD. It is important to note that this relationship suggested requires more precise quantification. Incorporating other aspects of articulation, in particular jaw height, and tongue curvature may provide a more dispersed view of the articulatory vowel space. The overlap seen at the TD for some vowels may be unimportant if those vowels are differentiated in another part of the vocal tract, such as is suggested by Wood [8], where four different constriction locations are proposed.

Alongside the general correspondence observed across Figure 1 and Figure 2 in relative position of vowels, we have also identified some mismatches. When we zoom in on individual speakers, we observe some cases in which a change in F2 does not correspond with differences in TD position or in the position of other lingual sensors, as expected

For one male speaker, we found a mismatch in backness for vowels LOT and THOUGHT. The general trend in the data is for the following correspondence: F2, lower for THOUGHT than for LOT, corresponds to TD position, which is also further back for THOUGHT than for LOT. One speaker shows the group pattern in TD position (TD further back for THOUGHT than for LOT) but does not show the group pattern in F2. Rather, for this speaker, the F2 for THOUGHT was not lower than for LOT (leading to some overlap between the THOUGHT and LOT ellipses in Figure 1). It is possible that this is due to reduced rounding in THOUGHT for this speaker. Although we have not as yet been able to quantify the effect, rounding is expected to be greater for THOUGHT than for LOT and should contribute to the separation in F2. Cases such as this underscore the indeterminacy of interpreting formant values in terms of articulation, or at least on a single fleshpoint. Because they are shaped by multiple articulatory constrictions in the vocal tract, it is not always possible to map changes in formants to changes in TD position. In this case, articulation shows consistency across speakers while formant values show variation, which is likely attributable to degree differences in rounding.

Another mismatch in the data is less easy to explain. In the other of our male speakers, we observed an inconsistency in the acoustic-articulatory relation in the "central" part of the vowel space. Although this speaker shows the same level of correspondence as other speakers in the front and back

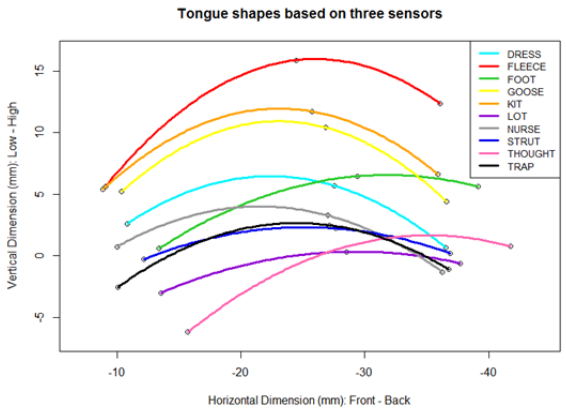


Figure 3. *Averages of a male speaker's three lingual sensors, with polynomials fit to the averaged sensor points for each vowel. The tongue tip is on the far left and the tongue dorsum is on the far right.*

sections of the vowel space, the central vowels NURSE, GOOSE and STRUT all have a similar level of backness (i.e., horizontal position of the TD), but there are large differences in F2. Despite similar horizontal positions of the TD (and TB and TT), F2 is highest for GOOSE, followed by NURSE, then STRUT. The averages of this speaker's lingual sensors are represented in Figure 3, with a polynomial fit to the three sensor points for each vowel.

Unlike the case of THOUGHT and LOT discussed above, it is unlikely that the difference in F2 across GOOSE, NURSE and STRUT is due to a degree difference in rounding. Lip rounding is expected to be the greatest degree for GOOSE, followed by NURSE. Bernard [4] reported smaller lip aperture for GOOSE than NURSE). However for this speaker, GOOSE and NURSE show unexpectedly high F2 values. Rounding would be expected to lower F2, the opposite pattern of what we observed. One hypothesis for this mismatch between acoustic and articulatory data is that the relation between F2 and backness is nonlinear in this portion of the vowel space, i.e. sometimes small differences in backness may have a large influence on F2 [9]. Given the particular anatomy of this speaker, central vowels may have unstable relations between F2 and TD backness. An alternative hypothesis is that something else is influencing F2 other than TD backness. One possibility may be the differences in tongue curvature which can be seen in Figure 3, and which has been shown to differentiate the vowels of English [10]. Another suggestion is that height influences F2 to a greater degree than backness in the central vowel space for this speaker. A third possibility is that aspects of lingual articulation outside of the mid-sagittal plane, e.g., tongue grooving, may be playing a role. A fourth possibility is lip rounding. Further investigation would be needed to discover why F2 varies despite similar degrees of TD backness for these central vowels, and why this is the case in this part of the vowel space and for this speaker in particular.

## 5. Conclusions

Generally speaking, the relationship between acoustics and articulation as previously described, such that F1 is inversely related to vowel height and F2 is inversely related to backness, was confirmed in our report of Australian English monophthongs. Moreover, the relationship was apparent from a single fleshpoint on the tongue, attached to the Tongue Dorsum, although a more precise quantification of the relation will require incorporating other dimensions of articulation. There were also a few corners of the data in which the assumed correspondence between acoustics and articulation broke down. For one speaker, the backness of the central vowels did not correspond to F2. For another, the backness of back vowels did not correspond to F2. In both cases, we observed consistency in TD position across vowels despite variation in F2. In the latter case but not the former, it is likely that rounding perturbs the relation between TD backness and F2. We conclude that formant values offer a heuristic for diagnosing TD position on the basis of acoustic data which is largely valid, particularly for front vowels and where vowel rounding is not at issue. For some speakers, F2 may not provide a valid indication of TD backness for central vowels, although additional research is needed to understand the precise conditions under which the normally assumed correspondence between F2 and TD backness breaks down.

## 6. Acknowledgements

This research was supported in part by Australian Research Council grant DE120101289 to the second author. We would like to thank all the speakers for their participation and colleagues at the MARCS Institute for their technical assistance with EMA experiments.

## 7. References

[1] Tabain, M. (2008). Production of Australian English language-specific variability. *Australian Journal of Linguistics, 28*(2), 195-224.

[2] Watson, C. I., Harrington, J., & Palethorpe, S. (1998). A kinematic analysis of New Zealand and Australian English vowel spaces. In ICSLP.

[3] Lin, S., Palethorpe, S., & Cox, F. (2012). An ultrasound exploration of Australian English /CVl/ words. *Proceedings of the 14th Australasian International Conference on Speech Science and Technology*, p.105-108.

[4] Bernard, J. B. (1970). A cine-X-ray study of some sounds of Australian English. *Phonetica, 21*(3), 138-150.

[5] Wells, J. C. (1982). *Accents of English* (Vol. 1). Cambridge University Press.

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

[7] Harrington, J., Cox, F., & Evans, Z. (1997). An acoustic phonetic study of broad, general, and cultivated Australian English vowels. *Australian Journal of Linguistics*, *17*(2), 155-184.

[8] Wood, S. (1979). A radiographic analysis of constriction location for vowels. *Journal of Phonetics, 7,* 25-43.

[9] Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, *17*, 3-45.

[10] Dawson, K. M., Tiede, M. K., & Whalen, D. H. (2016). Methods for quantifying tongue shape and complexity using ultrasound imaging. *Clinical Linguistics & Phonetics*, *30*(3-5), 328-344.