Chapter 3

Experiments

This chapter describes the experiments performed on the data described in Chapter 2. The objectives (in other words, what we are trying to achieve) are described and then the techniques used to meet these objectives are explained. Finally we discuss the results obtained, show graphs of the processed data and discuss our interpretation of the experimental results.

3.1 Objectives

Our primary objective with this study is to create acoustic models of Afrikaans vowels and diphthongs as spoken by mother-tongue speakers. We then want to create acoustic models of Afrikaans vowels and diphthongs for mother-tongue English speakers and compare these models with the Afrikaans models. We would then like to determine whether there are significant differences between the two accent groups.

A further objective which follows from the first is to add South African English vowels and diphthongs to the models and also compare these with the Afrikaans models of the same sounds. This will help us to determine how much of an influence Afrikaans has had on South African English in this respect and vice versa.

Our third objective is to determine if intonation (the change in pitch [one of the prosodic effects] over time) has a large influence on how vowels and diphthongs are perceived between the two accents as was found to be the case for French, German and English by Grover, Jamieson and Dobrovolsky[17]. They found that adult French, English and German speakers differ in the slopes of their continuative intonation, and that, dependent on the age at which the language was acquired, a speaker would use either native (if learned at a young age, say 10) or foreign intonation (if learned at an older age, say 16). We do not perform perceptual test here, but simply analyse the intonation curves of the accent groups.

Analysts such as Rousseau[39] and Flege[18] have noted that second-language speakers often substitute phonetically "close" sounds from their first-language when they do not possess the sounds in their personal phoneme space. This phenomenon is known as equivalence classification, as explained in Chapter 2.

Thus, for example, Flege proposes that because English does not possess the <y> sound phone which occurs in French, L1 English speakers will classify an L1 French speaker's <y> as his <u> and pronounce a <u> when trying to articulate a <y>, even though there are significant differences in the F2 frequencies of the two vowels when spoken by French speakers.

Rousseau goes further to suggest that if such a substitution is heard often enough and seen as acceptable then the substitution may become permanent, ironically enough, in the second language. For example, Rousseau theorises that the use of <i> instead of <ə> in Afrikaans words such as : "*ignoreer*" [ixnuriər], "*imbesiel*" [imbəsiəl], "***Indië***" [indiə] and "*industrie*" [indəstri] are all as a result of the influence of English.

Our first objective then can be seen as a desire to determine where we observe elements of equivalence classification in South African English and Afrikaans speech. This process is complicated by the bilinguality of the speakers.

One of our objectives in this study is also to determine whether <e:> and <o:> are long vowels or diphthongs and ideally suggest a means of measurement which will make it possible to qualify a voiced sound as either a vowel or diphthong.

3.2 Data

It was found by Peterson and Barney[10] that female speech formant frequencies differ from mens' by significant, yet consistent amounts. Knowing that this would unnecessarily complicate our study, and as only the cross accent trends are of interest to us, we concentrated this study exclusively on men. We also know that age has a distorting effect. Children generally have voices which are even higher pitched than those of women, and older people often display a change in timbre. Peterson and Barney observed these factors with their study. We therefore once again constrained the database to men aged between 20 and 40 years rather than to have to perform complex speaker normalisation and/or run the risk of biasing the data.

De Villiers[40] states:

> "As is explained with the discussion of the speech organs, there are differences between speakers, and especially between the three big groups: children, men and women. It is understandable that the three groups speak the same speech sounds with different fundamental frequencies and formants, but the ratios between the formants of the different groups are not so entirely different and this probably explains why people understand each other in spite of the differences which are observed in the individuals or groups."

We argue therefore that it is also important to analyse the formant ratios, not only due to the seemingly inherent speaker normalisation effect, but to further investigate the validity of the formant ratio theory.

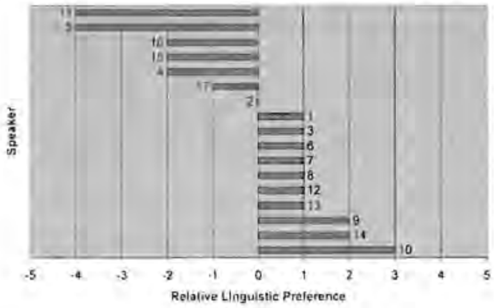


Figure 3.1: A bar graph displaying the linguistic preference of the speakers in the database.

To ensure that erroneous data (data which was incorrectly recorded or incorrectly spoken) could be excluded while minimising the effect on the statistical significance of our measurements, we decided to create a sizeable database with a significant number of utterances for each vowel per person.

3.2.1 Data structure

The data used in this study was collected from 17 speakers, 10 of whom are L1 (first language) Afrikaans speakers and 7 are L1 South African English speakers. Most of the speakers were relatively bilingual, but during the data recording process they were asked to indicate their language preference. The results of this query can be seen in Figure 3.1.

We make note here that we are aware that the research only holds true for a particular group of speakers in South Africa. There are a couple of L1 accents for both Afrikaans and South African English. The Cape Coloured community in the Cape Province generally speak L1 Afrikaans with an markedly different accent to that of the white Afrikaner population of the Gauteng Province.

The data was recorded in an anechoic chamber at the University of Pretoria using a ROSS RMA-102 Boom Microphone Headset and a Creative Labs Sound Blaster 16 at a sampling frequency of 16384 Hz. The data was recorded and written in 16 bit WAV format.

Using various recognised phonetic sources[23][41][40][1][20][42] we compiled a list of candidate words which contain most of the long vowels and diphthongs found in Afrikaans and South African English.

Please note that we have made an error in requesting that the speakers base their pronunciation of the vowel <ɛː> on the Afrikaans word "*êrens*". In the Cape Province amongst a section of the population this would have resulted in the correct vowel being uttered. Phonetically, in the Cape Province, we could write the word as [ɛːrəns] whereas in Gauteng Province where the data was recorded, a more correct phonetic transcription would be [æːrəns]. As a result we have no examples of <ɛː> and can therefore not determine its location in formant space.

A further error has been made in the data recording process, where we have requested that speakers utter the long vowel <ɔː>. We have in fact indicated example words which are based on the short vowel <ɔ>. We should have in fact used example words like "*sôe*" (sows) and "*rôe*" (rays [fish]). This mistake was, however, spotted too late to re-record the required long vowel. Further studies should attempt to study this vowel as we believe it is also prone to diphthongization.

Using Tables 3.1 and 3.2 we produced a sub-list (given in Table 3.3) which was used for the data recording. Tables 3.1 and 3.2 are lists of vowels and diphthongs respectively, drawn up from examples cited by some commonly referenced phoneticians. The lists are drawn up using the phonetic symbols used by the respective authors and we make no attempt at this point to distinguish between phonetic or phonological labelling used by the authors. As a result we find in Table 3.1 that Coetzee and de Villiers annotate <ɛː> as being a common vowel in all their example words on that line. This is in fact not true for all accent groupings of Afrikaans and this resulted in an error when the final reduced list was created. We choose the words for the reduced word list based on their familiarity and unambiguity (both of meaning and intended vowel/diphthong

SYMBOL	Afrikaans (Coetzee)[23]	Afrikaans (Wissing)[41]	Afrikaans (De Villiers)[40]	English (De Villiers)	English (Rabiner)[1]	English (Ward)[20]
a:		*baat*				
ɑ:	*aan, klaar, are, snare*					
æ:	*ver, sê*	*ver*				
e:	*eensaam, leen, bene, see*	*beet*	*bees*			
ɛ:	*êrens, bêre, lë*		*sê, ver*			
ə	*is, rit, middel, tevrede*	*bid*	*wit*	bird	about	about
i:	*Ier, mier*	*fier*	*dier*	heat		
o:	*oor, boom, bore, glo*	*boot*	*kool*			
ɔ:	*op, klop*	*bot*	*dom*	hot	bought, all	bought
ϕ:	*Europa, kleur*	*neus*	*reus*			Fr.peu
œ:	*brûe*	*lus*	*hut*			Fr.sœur
u:	*oer, vloer*	*voer*	*boer*	too		soon
y:	*uur, muur*	*vuur*	*uur*			

Table 3.1: A list of long vowels and words which contain these sounds.

SYMBOL	Afrikaans (Coetzee)[23]	Afrikaans (Wissing)[41]	Afrikaans (deVilliers)[40]	Afrikaans (Combrink)[42]	English (Rabiner)[1]	English (Ward)[20]
œu / əu	*bou, blou, oud, troue*	*bout*		*lou*		poor
əi	*by, ry, bly, eier, rys*	*byt*		*ly*		
œy	*bui, trui, uit, buite*	*buit*	*uit, ruik*	*lui*		
ɔi	*hondjie*	*bodjie, botjie*	*boikot*			boy, noise
ɑi	*matjie*	*badjie*	*aits, aitsa*			my, time
o:i	*ooi, nooit, mooi*	*looi*	*sooi, nooit*	*looi*		
e:u	*eeu, speeus, leeu*	*leeu*	[eu]*leeu*	[Eu]*leeu*		
ui	*moeite, koei*	*loei*	*moeite, boei*	*loei*		
ou			*oud, gou*			go, home
ei			*peil, pyl, ryk*			play, lady, make
ɑ:i	*aai, saaier, blaai*		*raai, laai*	[Ai]*laai*		[ai]my
iu			*leeu, spreeu*			
ɑu			cum laude			now, round
æu			*Crouse*			
iə			*weer*			here, beard, idea
uə			*koor*			pure, your
yə			*neus*			
ɛə			*werk*			there, fair, scarce
iɛ			*elke*			
əi	*litjie*					
ϕ:i	*neutjie*					
ɔə						more, board
ɑ^y					buy	
ɔ^y					boy	
ɑ^w					down	
e ^{y}					bait	

Table 3.2: A list of diphthongs and words which contain these sounds. Included in square brackets are alternative (yet similar) notations used by some phoneticians.

Phonetic Symbol	AFR.	ENG.
Long vowels		
a:	*klaar*(finished)	father
æ:	*werk*(work)	hat
e:	*bees*(cattle)	
ɛ:	*êrens*(somewhere)	
ə:	*wîe*(wedges)	about
i:	*dier*(animal)	heat
o:	*kool*(coal)	
ɔ:	*dom*(dumb)	bought
u:	*boer*(farmer)	soon
y:	*uur*(hour)	
oe:	*brûe*(bridges)	
Diphthongs		
œu	*blou*(blue)	
əi	*bly*(happy)	
œy	*trui*(jersey)	
ɔi	*hondjie*(small dog)	boy
o:i	*mooi*(pretty)	
a:i	*haai*(shark)	time
ou	*gou*(quickly)	home
ei	*ryk*(rich)	play

Table 3.3: The reduced long-vowel and diphthong word list used in the database recording.

intended to be pronounced).

A summary of possible recording structures is given in Table 3.4. The database consists of two main sections namely long vowels and diphthongs. Each of these main sections is divided into three sub-sections called isolated, context and pseudo-context. "Isolated" means the vowel or diphthongs were recorded in isolation as nothing more than a vowel or diphthong. For example, just the <a:> in father. By "context" we mean the vowel or diphthong was recited as part of a word, for example "father". Lastly, by "pseudo-context" we are referring to the h-vowel-t structure, similar to the one used by Peterson and Barney. "h" and "t" were chosen for their limited influence on the articulation of vowels, thus for example the person had to say [hi:t] as in "heat". Quite often though, no h-vowel-t word with that vowel, or especially diphthong exists. Nevertheless, the

Number of Utterances	Sound	Placement	Playback
2	Long vowel	Isolated	No
2	Long vowel	Isolated	Yes
2	Long vowel	Context	No
2	Long vowel	Context	Yes
2	Long vowel	Pseudo-context	No
2	Diphthong	Isolated	No
2	Diphthong	Isolated	Yes
2	Diphthong	Context	No
2	Diphthong	Context	Yes
2	Diphthong	Pseudo-context	No

Table 3.4: The various ways is which the data was recorded.

speakers were instructed to attempt, for example, to articulate sounds such as [hət] (which is a non-existant word) using the <ə> vowel found in the word about. By "playback" we mean that in that particular section the sound was either played back to the speaker for him to evaluate and re-record if desired, or not played back at all with no chance of altering the sound once recorded. The reasoning behind this is that some speakers may alter their pronunciation when hearing themselves and we wanted examples of both possibilities.

This gives us a potential database of ten utterances per long vowel of which we have seventeen cases (counting SA English and Afrikaans individually) and ten utterances per diphthong of which we have twelve cases (again counting SA English and Afrikaans individually). A potential two hundred and ninety words were thus recorded per speaker. This gives a total of 4930 words.

The raw WAV data comprises about 350 megabytes.

3.3 Method

Now that we have described the data we used to meet the objectives of this study we will describe the methods employed to check and process the data.

3.3.1 Data recording and verification

The data for each person was recorded in a single thirty minute session. The data was recorded over a number of days with English and Afrikaans speakers randomly distributed. Recording instructions were given in the speaker's first language before the speaker entered the anechoic chamber.

The data recording session took place with the speaker alone in the anechoic chamber. They were prompted by text on a computer screen to recite the vowels and diphthongs one by one in ten stages (as layed out in Table 3.4). Each utterance was automatically detected and saved to a separate file before the next prompting took place. The order of the words was purely random with two utterances of each word/vowel/diphthong being recorded to ensure redundance. Where possible, both recorded instances were used. Afrikaans source words were highlighted in green and English source words in red. Before the commencement of each section a text paragraph was displayed and a voice recording explaining the next section was played with an example of what was expected. One speaker complained of difficulty reading some of the words due to colour blindness but stated that he did not think it had influenced the accuracy of his utterances. Playback of his data confirmed this. One other speaker was allowed to re-record his entire session due to multiple mistakes, the cause of which he could not explain.

The software was written and run under the Linux operating system in plain text mode and using alphanumeric colour codes to generate the desired red and green colours.

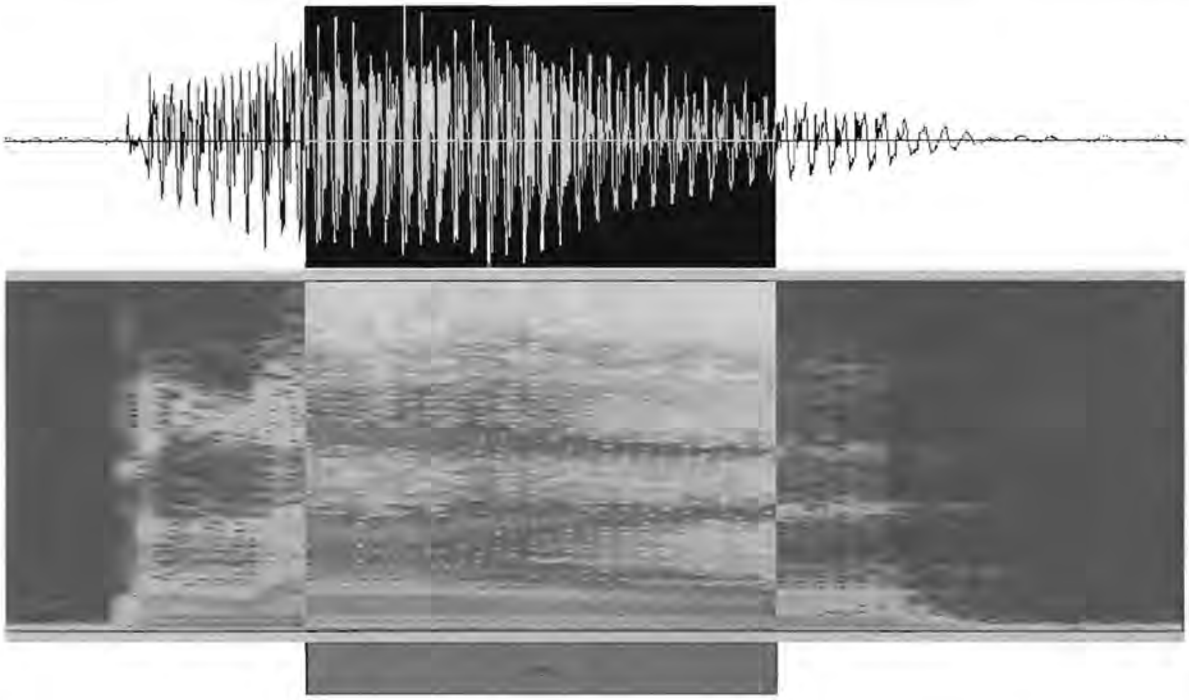


Figure 3.2: This figure demonstrates diphthong extraction. The top half shows the time-energy waveform, the bottom half the spectrogram and the very bottom shows the label (tag) given to the diphthong segment (in this case <œu> from "*blou*" (blue)).

The recorded "words" are then processed manually to extract the vowel or diphthong which we require. This process is called segmentation and labelling (tagging). Using the time-energy waveform, spectrogram and sound playback, we extract only the section from the recording which we require. This process is demonstrated in Figure 3.2.

Before we began processing the data we listened to all the sounds in each of the categories listed in Table 3.4 and then removed the sounds which differed too excessively or were not consistent within a vowel grouping or did not "sound" correct to the data labeller. It was found that certain of the speakers misunderstood some of the instructions or misread certain of the words. This can partially be attributed to the lengthy process of recording almost 300 words, but, alternatively, it can be argued that without such an exhaustive sampling session, if a few mistakes were made, simple errors would be a far more significant percentage of the database. A few of the words had to be discarded due to excessive clipping of the signal due to a change in volume of the speaker and some of the recordings were of lip smacks or coughs. The amount of data thrown out in this initial stage was about 10% of the initial data.

The second stage of data checking took place after the formants and pitch trajectories had been extracted. This, in the case of formants, involved superimposing the extracted formants onto the same section of speech's spectrogram. We then manually observed each of the almost five thousand words and where necessary and possible, manually corrected the trajectories. This usually occurred when a few of the points could be seen to have been incorrectly extracted. This is demonstrated in Figure 3.3. This step is essential as a few large misplaced values can have drastic effects on the means and variances of the data and this carries over and biases the statistical significance of differences between the two language groups.

It is important to note that we decided to exclude all the normal word-in-context data, i.e. the vowels and diphthongs which were recorded as complete words (not the h-vowel-t or h-diphthong-t structure). It was found after exhaustive plotting in formant space that the influence of the consonants was far too excessive to make for any useful comparison or pooling with the isolated and h-structure-t data.

When checking the pitch trajectories we simply discarded any utterance in which we observed unrealistic fundamental frequencies. De Villiers[40] states that under normal relaxed speech conditions, males speak with a fundamental frequency of 109 to 163 Hz. To be safe we plotted pitch trajectories from 75 to 250 Hz and then discarded those where sudden large (25 Hz) jumps occurred. The remaining trajectories are then used to determine mean intonations for the vowels and diphthongs for each of the language groups.

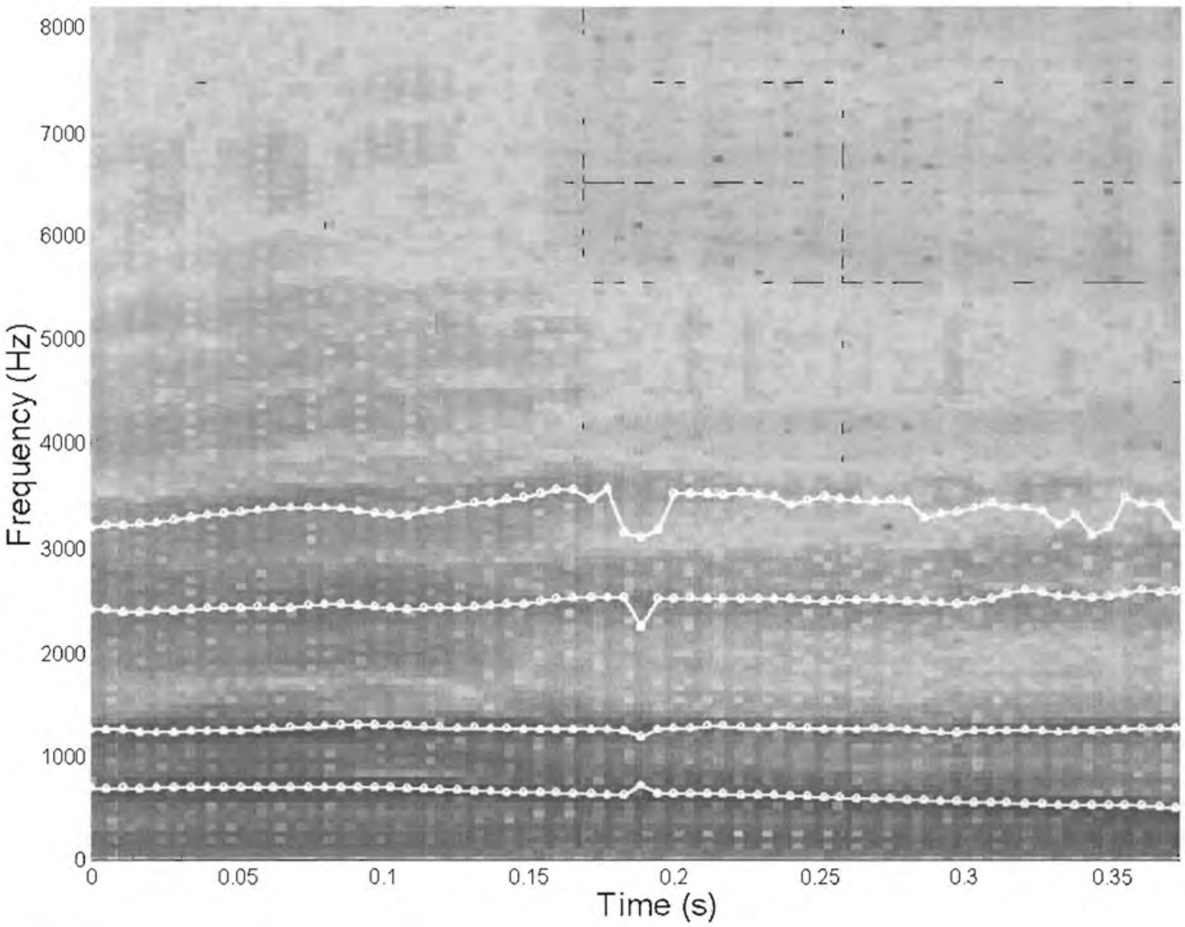


Figure 3.3: Miscalculated formant values at time 0.19s superimposed on a spectrogram and requiring correction.

3.3.2 Formant extraction

The formants were extracted using the techniques described in Section 2.4 on page 26 and using the functions found in the program Praat[1] with the following parameters:

- LPC Prediction order : 17
- Analysis width : 25ms
- Time step : 5ms
- Pre-emphasis from : 50Hz

We have used a rather high LPC order. This was determined experimentally, using Praat and iteratively trying various orders with various recordings. This was the value which tracked the formants most accurately.

Each segmented file was processed individually and the extracted formants then saved to separate files. This allowed for editing of the formant trajectories if required. This process was described in Section 3.3.1.

Although the formant extraction process is relatively quick, the checking and correcting of poorly extracted formants is an arduous task requiring many days to complete.

3.3.3 Pitch extraction

The pitch trajectories were extracted using one of Praat's pitch extraction techniques using a modified autocorrelation technique developed by Boersma[33]. The algorithm corrects many of the problems associated with standard autocorrelation techniques discussed in Sections 2.4.1 and 2.5. This method is more accurate, noise-resistant

---

[1] Praat by Paul Boersma. A system for doing phonetics by computer, IFA, University of Amsterdam

and robust than methods based on cepstrum, combs or the original autocorrelation methods. The reason why other methods were invented was the failure to recognise the fact that if you want to estimate a signal's short-term autocorrelation function, you should divide the autocorrelation function of the windowed signal ($r_{xw}(t)$) by the autocorrelation function of the window ($r_w(t)$). This is represented by:

$$r_x(\tau) \approx \frac{r_{xw}(\tau)}{r_w(\tau)} \tag{3.1}$$

A summary of the complete 9-parameter algorithm, as it is implemented in the speech analysis and synthesis program Praat, is given in Appendix A.2.

With reference to the algorithm described, we have used the following parameter values:

To find pitch candidates:

- Time step : 10ms
- Minimum pitch : 75Hz
- Maximum number of candidates : 5

And to find a path:

- Silence threshold : 3% (0.03)
- Voicing threshold : 45% (0.45)
- Octave cost : 0.01
- Cost of octave jump : 0.35
- Voiced/Unvoiced cost : 0.14

- Ceiling : 600Hz

The *Silence threshold* is the point at which all frames with amplitudes less than this normalised (relative to the global maximum amplitude) value are considered to be silence.

The *Voicing threshold* is the strength of the unvoiced candidate, relative to the maximum possible autocorrelation.

The *Octave cost* is the degree of favouring of high-frequency candidates, relative to the maximum possible autocorrelation. This is necessary because even (or, especially) in the case of a perfectly periodic signal, all under-tones of F_0 are equally strong candidates as F_0 itself.

The *Octave jump cost* is the degree of disfavouring of pitch changes, relative to the maximum possible autocorrelation.

The *Voiced/unvoiced cost* is the degree of disfavouring of voiced/unvoiced transitions, relative to the maximum possible autocorrelation.

3.3.4 Data visualisation and comparison

Due to the graphical nature of data plotting, and the desire to make a single application which could be used to view and analyse the data, it was decided to write a program to run on the Windows operating system using the Borland C++ Builder development platform. The resulting program allows for the following:

- Formant plotting

    1) Each extracted formant point for each utterance - useful for spotting rogue data or poorly extracted formants.

    2) The mean location of the formants for each utterance - used to determine the mean formant frequencies for each of the accent groups' vowels.

  3) Individual formant trajectories for each utterance - useful for spotting rogue data or poorly extracted formant trajectories for diphthongs.

  4) Mean formant trajectories for a number of utterances - used to determine the mean trajectories of the diphthongs for each of the accent groups.

- Pitch plotting

  1) Individual pitch trajectories - useful for spotting poorly extracted pitch contours.

  2) Mean pitch trajectories - used to plot the mean pitch trajectories (for a number of utterances) for the different accent groups.

When plotting the mean formant positions (as in Figure 3.4 on page 70) of each utterance we also plot a mean/variance cluster bubble around the data, so orientated to indicate the direction of maximum variance. The centre of the bubble lies at the mean of the formant values (these values can be seen in Table 3.5 on page 69). The border of the bubble indicates the mean variance of the data set. The variance ($\sigma^2 = \frac{[(x_1-\mu)^2+(x_2-\mu)^2+\cdots+(x_n-\mu)^2}{n}$) is a measure of the dispersion or scatter of the local mean formant values around the global mean formant value. If the values tend to be concentrated near the mean, the variance is small and the bubble will be small. So as to include most of the points within the variance bubble we actually plot the border at twice the variance. We also calculate the direction of greatest variance and rotate the oval bubble to reflect this direction.

When plotting diphthong trajectories (as in Figure 3.6 on page 79) we plot a small circle around the originating point to show the direction of articulation i.e. where the diphthong starts and by implication, where it ends.

The program also allows for swapping the axes and inverting them. The orientation used in the plots given in this dissertation was chosen so that the data always fits in with the IPA vowel chart. The locations of the Peterson and Barney vowels are also plotted for reference purposes.

The plotting software also allows us to perform analysis of variance comparisons between any two batches of data (section 2.9.1 on page 46). For the vowel formants the independent groups of samples are simply the mean formant frequencies for each utterance. For the diphthong formants and pitch trajectories we make our comparison between the cubic spline coefficients as determined and explained in Section 2.7 on page 42. For cubic spline comparisons we end up with multiple indicators of difference (12 per formant or pitch trajectory) which is not an efficient means of indicating trajectory differences. To this effect we have utilised simple "or" logic, if any one of the coefficients differs significantly, then we consider the entire trajectory to differ. We indicate this in the tables (for example Table 3.7 on page 78) with black blocks indicating a significant difference. We have also used gray blocks to indicate significant differences in the third formant. F3 is more prone to tracking errors and we have therefore just indicated this difference to remind readers of this possibility.

3.4 Results

3.4.1 Long vowel results

We present here a discussion of the long vowels analysed and try to explain the trends visible in the figures and tables offered in this section.

We have decided to work at the 99% level of significance. All F ratios which exceed the 99% significance level are indicated in the tables by grayed boxes. The degrees of freedom are also indicated and the tables can therefore be used to determine the significance level at 95% if required by checking which F ratios exceed those indicated on a F distribution table at a 95% significance level.

We have summarised the analysis of variance results for the long vowels in Table 3.5. The table is structured as a number of sets of rows. Each of these sets represents a specific long vowel (indicated in a black block) and consists of a number of rows where each row is the mean result for a specific group of utterances. Note that the words indicated with the vowels in the table do not imply that these are the results of instances of vowels in context. The words are merely given as example or context. For example in Table 3.5 at the top left hand corner we see "aa" in a black block which indicates the set of results for the long vowel <a:>. Under this we see "*a klaar*" which indicated to us that these results pertain to the long vowel extracted when the Afrikaans ("a") mother-tongue speakers were told to utter the long vowel in the word "*klaar* (finished)". The next three columns contain the mean formant values for F1,F2 and F3 in hertz. The fifth column (labelled "Num") indicates the number of utterances which were used to determine the mean. We then have a number of "blocks" of 4 columns which indicate the ANOVA F ratio results (as described in Section 2.9). For example, referring to our previous example for <a:> we see that the F ratio value for a comparison of the means of F1 for the Afrikaans first language <a:> from "*klaar*" and the Afrikaans second language <a:> from "*klaar*" is 0.05. There were 38 utterances used for the Afrikaans L1 mean and 26 utterances used for the Afrikaans L2 mean. This leads us to 62 degrees of freedom(indicated in the ninth column). At a significance level of .99 we require a F ratio in excess of 7.08 (according to the Fisher tables[19]) and can therefore safely state that the means are statistically equal. In cases where the F ratio has exceeded the F distribution values we highlight the value with a dark block.

The graphs of the individual utterance means and their cumulative means are shown in a number of sub-figures in Figures 3.4 and 3.5.

We did not show <e:> and <o:> with the long vowels but rather plotted them in Figures 3.6 and 3.7, with the diphthongs. During the recording session we indicated to the speakers that these were vowels (as many phonologists state), however, after analysis and confirmation from various references such as Taylor and Uys[12] we concluded that these "vowels" are in fact diphthongs. We have therefore plotted them as diphthong trajectories and we will discuss and analyse them as such. Further justification for this decision is provided by an experiment and the results are given in

Section 3.4.5.

<i:> and <y:>

<i:> (unrounded) and <y:> (rounded) are high front vowels found in words like ["*dier*" and "heat"] or ["*uur*"] respectively.

Rousseau[39] claims that English has had such a large influence on the development of Afrikaans that <i:> has pushed aside many "traditional" or "correct" ways of saying words, for example [ji:səs] ("*Jesus*") instead of [je:səs]. This intense replacement may even have had a large effect on the pronunciation of <i:>. There is no way for us to say where <i:> may have lain historically, but as we state later, it appears that Afrikaans and English mother-tongue speakers now appear to use statistically similar versions of <i:>.

De Villiers[40] states that the unrounding of <y:> to <i:> is quite common (for example in "*askies*" in stead of "*ekskuus*"(excuse me)). He goes further to say that <y> is seldom still found in general speech, except amongst older people and in careful and cultured speech. As a result of this, when this vowel is expected to be produced it is quite often hyper-corrected. We can therefore expect that our data may not entirely correctly reflect the position of <y:>. It is found that when <y> is used, it is often in a stressed position, such as in words like "*luuks*" and "*muur*". In unstressed positions it is often replaced by its unrounded companion <i:> as in words like "*murasie*".

Ward[20] states that in some types of South African speech <ɪ> is a close variety, often approaching <i>.

Looking at Table 3.5 (row group "ii") we see the following:

- Although <y:> does not occur in English, we can not at a 99% level of surety state that the first and second language speakers generated a different sound (first line,

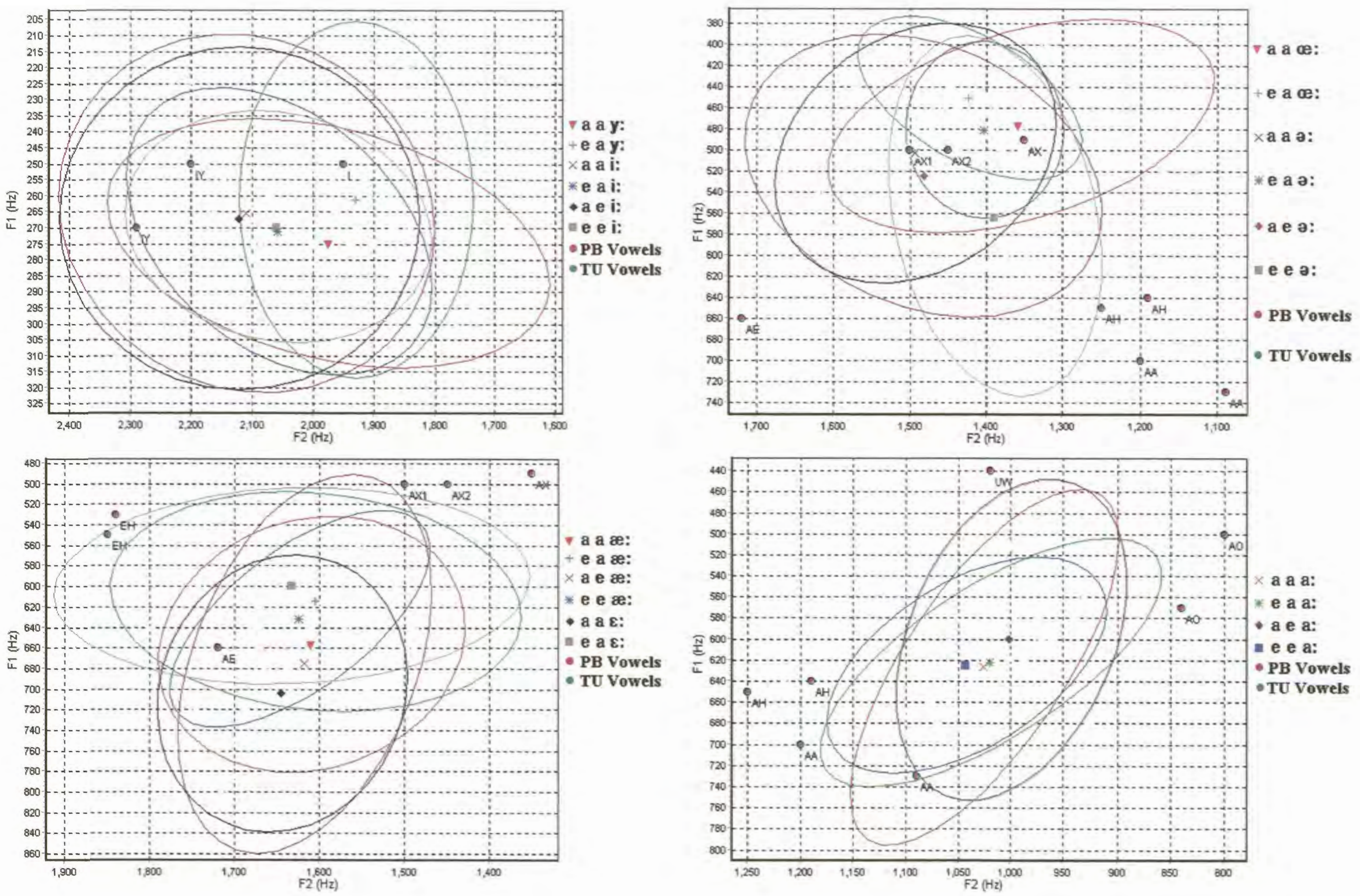


Figure 3.4: Vowel formant clusters: [<y:> in "*uur*" and also <i:> in "*dier*" and "heat"], [<œ:> in "*brûe*" and also <ə:> in "*wîe*" and "about"], [<æ:> in "*werk*" and "hat" and also the incorrectly used <ɛ:> in "*êrens*"] and [<a:> in "*klaar*" and "father"]. The first *a*/*e* indicates the mother-tongue of the speakers and the second *a*/*e* indicates from which language the vowel was indicted as coming from. *PB Vowels* are the Peterson and Barney vowels and the *TU Vowels* the Taylor and Uys vowels.

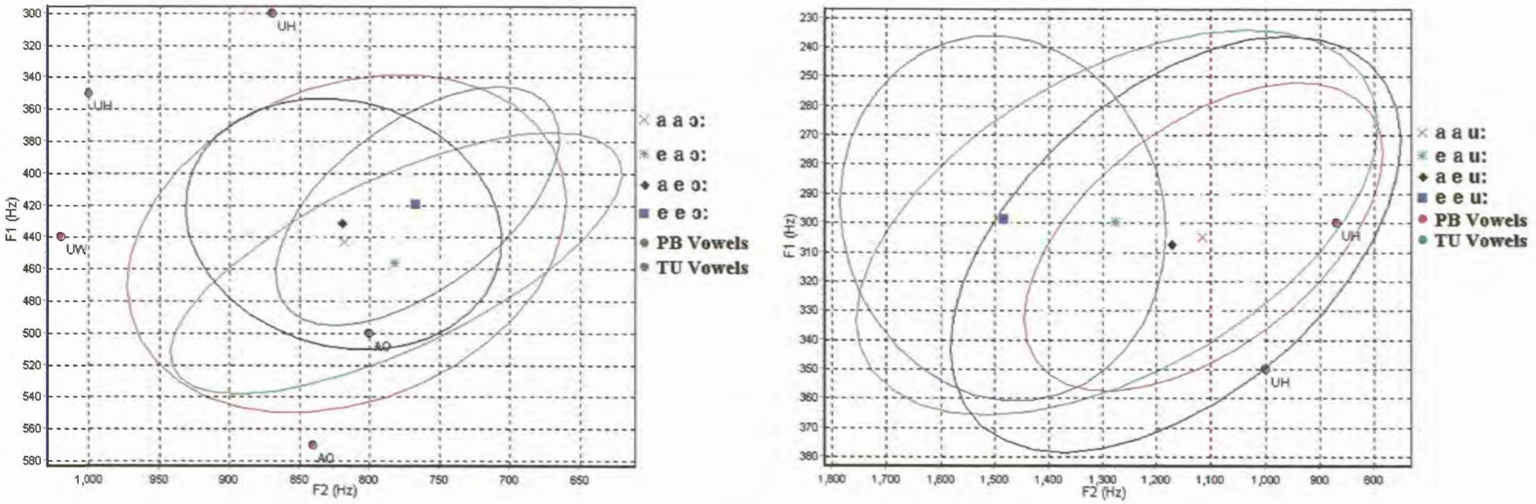


Figure 3.5: Vowel formant clusters: [<ɔ:> in "*dom*" and in "bought"] and [<u:> in "*boer*" and "soon"]. The first *a*/*e* indicates the mother-tongue of the speakers and the second *a*/*e* indicates from which language the vowel was indicted as coming from. *PB Vowels* are the Peterson and Barney vowels and the *TU Vowels* the Taylor and Uys vowels.

Vowel Formant Analysis of Variance

Vwls	mF1	mF2	mF3	Num	anovaF1	anovaF2	anovaF3	DOF	anovaF1	anovaF2	anovaF3	DOF	anovaF1	anovaF2	anovaF3	DOF	anovaF1	anovaF2	anovaF3	DOF	anovaF1	anovaF2	anovaF3	DOF
ii					a uur				e uur				a dier				e dier				a heat			
a uur	275.07	1973.32	2441.90	26																				
e uur	261.46	1927.86	2576.43	27	3.89	1.26	5.93	51																
a dier	265.78	2108.04	2959.45	31	1.85	8.81	122.29	55	0.33	26.56	52.97	56												
e dier	271.30	2054.40	2858.13	26	0.36	3.36	72.57	50	1.85	16.40	25.69	51	0.60	1.95	4.62	55								
a heat	267.06	2120.98	2940.84	34	1.54	11.57	121.82	58	0.62	33.20	51.70	59	0.03	0.12	0.18	63	0.40	3.28	3.30	58				
e heat	269.87	2056.36	2872.16	31	0.97	3.95	103.69	55	1.83	17.97	36.75	56	0.44	2.02	4.53	60	0.06	0.00	0.11	55	0.23	3.45	3.00	63
oe					a brûe				e brûe				a wie				e wie				a about			
a brûe	477.66	1355.69	2272.02	31																				
e brûe	451.18	1421.00	2357.21	23	4.25	4.75	6.61	52																
a wie	502.97	1492.98	2389.88	17	2.29	14.89	7.22	46	10.06	7.15	0.75	38												
e wie	481.41	1401.34	2427.22	15	0.06	1.74	16.37	44	4.83	0.77	7.58	36	1.23	10.75	0.78	30								
a about	525.15	1481.58	2380.37	19	7.75	11.93	6.79	48	18.64	4.00	0.42	40	1.00	0.10	0.04	34	4.53	5.78	1.35	32				
e about	563.39	1388.35	2447.83	14	16.84	0.79	17.95	43	27.68	1.70	9.21	35	4.87	11.38	1.53	29	10.14	0.30	0.48	27	1.92	6.65	2.30	31
ae					a werk				e werk				a hat				e hat				a êrens			
a werk	656.97	1608.55	2475.03	21																				
e werk	615.20	1603.68	2469.95	27	5.99	0.02	0.03	46																
a hat	675.76	1616.59	2466.57	25	0.61	0.10	0.06	44	8.25	0.20	0.01	50												
e hat	632.06	1622.29	2503.01	28	2.24	0.32	0.85	47	1.35	0.45	1.43	53	4.48	0.07	1.31	51								
a êrens	703.92	1643.67	2503.22	17	4.72	1.59	0.53	36	22.29	1.43	0.89	42	1.11	1.28	0.81	40	15.35	0.81	0.00	43				
e êrens	599.94	1631.33	2488.42	13	9.93	0.36	0.15	38	0.95	0.49	0.34	44	10.24	0.19	0.35	42	4.41	0.08	0.20	45	27.35	0.10	0.12	34
aa					a klaar				e klaar				a father											
a klaar	627.06	1025.16	2416.62	38																				
e klaar	622.65	1019.34	2472.76	26	0.05	0.10	1.99	62																
a father	600.59	999.95	2391.09	31	1.79	2.89	0.45	67	1.41	1.10	3.95	55												
e father	625.21	1041.80	2468.83	23	0.01	0.90	1.58	62	0.03	1.14	0.01	50	1.91	6.49	3.24	55								
openo					a dom				e dom				a bought											
a dom	444.51	816.22	2329.01	24																				
e dom	456.53	780.91	2404.74	8	0.33	1.12	0.70	30																
a bough	431.48	819.02	2372.11	30	1.04	0.02	0.63	52	2.38	2.24	0.20	36												
e bough	420.83	765.86	2426.15	14	2.09	4.39	2.01	36	4.03	0.26	0.07	20	0.69	8.85	0.96	42								
uu					a boer				e boer				a soon											
a boer	305.09	1114.15	2136.03	27																				
e boer	300.06	1273.04	2151.85	21	0.37	7.03	0.15	46																
a soon	307.27	1167.23	2185.32	29	0.07	1.07	1.68	54	0.57	2.65	1.29	48												
e soon	298.71	1482.48	2113.45	25	0.64	67.46	0.39	50	0.02	12.37	2.95	44	0.90	38.12	7.36	52								

Table 3.5: An analysis of variance table of the long-vowel formants.

Vowel	Strong Form	Replaced Form
ε	pence [pεns]	sixpence [sikspəns]
æ	valid [ˈvælɪd]	validity [vəˈlɪdɪtɪ]
ɑ	particle [pɑtikl]	particular [pəˈtɪkjʊlə]
ɔ	ward [wɔ:d]	backward [ˈbækwəd]
u	to [tu:]	today [təˈdeɪ]
ʌ	some [sʌm]	handsome [ˈhænsəm]
ɜ	Bert [bɜt]	Herbert [ˈhɜbət]

Table 3.6: Examples of the neutral vowel <ə> replacing the strong forms.

second column). We can with safety state though that both groups' <y:> differ significantly from both the English and Afrikaans utterances of <i:> (columns two and three). This would seem to contradict the research of Flege[18] on "equivalence classification", but it may however be attributed to the bilinguality of the two groups.

- The <i:> uttered by both groups in both first and second languages were found to be statistically identical (fourth, fifth and sixth columns). Therefore we can say that South African English and Afrikaans speakers use the same <i:>, even when speaking their second language.

<œ:> and <ə:>

<œ:>(rounded) and <ə>(unrounded) are central vowels. The results are in row group "oe" of Table 3.5.

<ə> is an interesting vowel in that it tends to replace all vowels that are in unstressed positions[43]. There are two exceptions: <i> unstressed becomes <ɪ> and <ɪ> unstressed remains <ɪ>[20]. Besides these exceptions though we can generalise by saying the neutral vowel replaces the strong forms as in Table 3.6. This is especially noticeable in continuous, fluent speech.

- <œ>, although lying in the general vicinity of <ə> is seen to be statistically separate from it. It is seen to be an identical sound for the two language groups though (first row, second column).
- The <ə:> spoken by Afrikaans speakers when saying "*wîe*" is seen to be the same sound as when they say the <ə:> in "about" (third row, fourth column), and this sound is statistically distinct from the <ə:> uttered by the English speakers (first and third rows, fourth column). So the Afrikaans speakers use the same <ə:> when speaking their mother-tongue or a second language, and this <ə:> is significantly different from the <ə:> used by mother-tongue English speakers.
- For reasons which we can not explain the <ə:> produced by the English speakers saying "*wîe*" is greatly different from that which they used when saying "about" yet it is statistically similar to the <ə:> produced by the Afrikaans speakers saying "about". This may be due to a labelling error caused by the diphthongization some speakers enforce, especially the English speakers pronouncing "*wîe*", saying something more in the line of [w-œ-ə]. This is observable in Figure 3.4 (top right). This may also be as a result of the unfamiliarity of the word *wîe* as it not a common word and possibly new to many of the SA English speakers.
- The <ə:> used by Afrikaans and English speakers to pronounce "about" appears to be identical (sixth column). This ability of the Afrikaans speakers to produce authentic Afrikaans and English sounds seems to support our suspicions that Afrikaans mother-tongue speakers are more bilingual than English mother-tongue speakers. We do not have enough data to confirm this though.

<æ:> and <ɛ:>

<æ> is traditionally considered to be a short vowel but there seems to be a recent tendency for people to lengthen it as in "bad" [bæ:d]. It was therefore recorded and analysed as a long vowel. The results are in row group "ae" of Table 3.5.

It is important to note (as Ward[20] states) that many people find it difficult to pronounce an isolated <ɛ> and quite often end up saying something which approaches <ɜ> as in "bird"[2].

We have no useful data for <ɛ> as we made the error of requesting that the speakers utter <ɛ> as in "*êrens*". This would have worked in certain regions of the Cape (where we would have got [ɛrəns], but in the region where the data was recorded the acceptable pronunciation tends toward [ærəns]. This problem could have been rectified by using an Afrikaans source word such as "*hê*" and an English source word like "bet". This use of <æ> instead of <ɛ> is quite clear in the clustering observable in Figure 3.4 on page 70.

Interestingly enough, this phenomenon of the loose definition(the same symbol being used to indicate 2 different phonemes) of <ɛ> does not only occur in Afrikaans but also in British English. Ward[20] observes that <ɛ> can be as close as in "bet" and as open as in "bell" and goes on to note that many areas in the United Kingdom may pronounce an <æ> verging on <ɛ>.

Referring to Table 3.5 we see that the <æ> spoken by the Afrikaans speakers in "*werk*" only differs significantly from the unfortunate <æ> in "*êrens*" spoken by the English speakers (fifth row, first column). After this we see the general trend that the Afrikaans speakers' utterances differ from the English speakers' utterances, whether in first or second language words. In other words the <æ> spoken by Afrikaans speakers is always the same, whether it is spoken in the mother-tongue or not, and the same is true for the English speakers. More importantly, there is a noticeable difference between the Afrikaans and English speakers' versions of the vowel. This is also observable in Figure 3.4 where the English mother-tongue speakers consistently utter their versions of <æ> with lower F1's (as is demonstrated by the black blocks showing significant differences in the "anovaF1" columns).

---

[2] This may however be peculiar to British English

<ɑ:>

<ɑ:> is a back, open, unrounded long-vowel. The results are in row group "aa" of Table 3.5.

According to Ward[20], South African English uses an <ɑ> which is very near to the cardinal <ɑ>. According to De Villiers[40], <ɑ> and <ɑ:> differ from each other phonemically e.g.: "*dan*" versus "*Daan*" and "*man*" versus "*maan*" although Taylor[12] claims that length is dependent on syllable structure i.e. length is not phonemic in Afrikaans.

If we look at Table 3.5 on page 69 we see that the F ratios calculated for the analysis of variance of the <ɑ> in "*klaar*" and "father" indicate that there is no significant difference between the four possibilities i.e. Afrikaans word spoken by Afrikaans speaker, English word spoken by Afrikaans speaker and the same for an English Speaker. The F ratio for F1 between the Afrikaans first language "*klaar*" and Afrikaans second language "*klaar*" is 0.05. The F ratio for F1 between the Afrikaans first language "*klaar*" and English second language "father" is 1.79. The F ratio for F1 between the Afrikaans first language "*klaar*" and English first language "father" is 0.01. As none of these values exceed the 99% percentile values for the respective degrees of freedom it follows that Afrikaans and English speakers use the same <ɑ>.

<ɔ:>

<ɔ:> is classified as a back, open-mid, rounded vowel. The results are in row group "openo" of Table 3.5.

Rousseau[39] contends that the use of <ɔ> instead of many other vowels is as a result of the influence of English. For example, <o> is replaced by <ɔ> in "*dokument*" and replaces <o:> in "*horisontaal*" and also <u> in "*moderniseer*".

It is possible that this overwhelming usage of the English <ɔ> may have resulted in Afrikaans speakers using the same vowel sound as mother-tongue English speakers. Whatever the cause, we find that Afrikaans speakers generally appear to use the same sound as their English speaking counterparts, except that the English speakers demonstrated a significantly lower F2 when saying "bought" than the Afrikaans speakers. This can be observed by the mean F2 values shown in the third column of Table 3.5 on page 69 in the subsection labelled "openo"(<ɔ>) where the Afrikaans speakers spoke with a mean F2 of 819 Hz as opposed to the English speakers who spoke with a mean F2 of 766 Hz. This proves to be statistically significant as we have indicated in the last block of the <ɔ> section where the second language "bought" is compared with the first language "bought". We see that for F2 in 42 degrees of freedom we have a F ratio of 8.65 which is greater then 7.31 (the value determined from the F distribution) and therefore significant.

<u:>

Ward[20] identifies two varieties of this vowel. The first occurs in words like "fool" where the vowel is followed by a dark l. The second lies in a slightly more forward position and is found in words like "rude". The <u:> associated with the source words "*boer*" and "soon" would most likely be the first and second kinds respectively.

Ward claims that many people diphthongise this vowel considerably, however we only found this (see Section 3.4.5) to a small extent with the Afrikaans mother-tongue speakers, as Ward suggests, moving from <ʊ> to <u>. A measure of the diphthongization of vowels and diphthongs is given in Section 3.4.5.

From the F ratios of row group "uu" in Table 3.5 we can easily deduce (and confirm in Figure 3.5) that the English speakers have generated an <u:> with significantly higher F2 in their first-language than in their second language (which is statistically similar to the utterances by the Afrikaans speakers). It would appear therefore that if we take

Ward's reasoning further, that the Afrikaans speakers produce a single version of <u:> which is like the <u:> Ward identifies with words like "rude" and lies in a slightly more forward position. The English speakers on the other hand appear to demonstrate both types of <u:>. The one used in their utterance of "soon" is of the type which Ward identifies as usually being followed by a dark l. This would place it closer to a cardinal <o> which has a lower F1 but a higher F2. This is indeed the case as can be deduced from Figure 3.5 where the right hand diagram clearly portrays the Afrikaans utterances of first(×) and second(♦) language <u:> as lying in the same location and the English second(∗) language cluster of utterances lying in a similar position. The English first(■) language <u:> is clearly seen to have a higher F2 though.

3.4.2 Diphthong results

As we mentioned with the discussion of the long vowels, in continuous, fluent speech, a neutralisation of the strong forms of the vowels takes place, moving toward the location of <ə>. In the same way, many of the diphthongs are prone to neutralisation. Examples of this are given in Table 3.8, for example, face([feɪs]) as compared to preface([ˈprɛfəs]). It is important to record the data in context, but it is more important to record the diphthongs in isolation or pseudo-word context where the chosen consonants for the pseudo-word must have minimal effect on the diphthong and not warp the inherent diphthong structure.

We will now discuss each of the diphthongs in turn. We have drawn up a summary shown as Table 3.7. Similarly to Table 3.5 we have clusters of rows where the clusters consist of a specific diphthong's analysis. For example, the first cluster of rows in Table 3.7 represent the ANOVA results for the diphthong <ei>. The first column shows the word that was given to the speaker to indicate (in some context) which diphthong to utter. The "a" and "e" indicate whether the utterances are those of Afrikaans mother-tongue or English mother-tongue speakers respectively. The second column indicates the number of utterances which were used in the processing. We have analysed each

Diphthong Formant Analysis of Variance

Formant: 1 2 3 DOF (repeated for each of five panel columns)

Section: 1 2 3 1 2 3 1 2 3

Diphthong	Num	a ryk	e ryk	a play	e play	a bly
ei						
a ryk	42					
e ryk	38	78				
a play	45	85	70			
e play	38	78	74	81		
a bly	38	78	74	81	74	
e bly	34	74	70	77	70	70

oey	Num	a trui
a trui	41	
e trui	33	72

ou	Num	a home	e home	a blou	e blou	a gou
a home	40					
e home	34	72				
a blou	47	85	79			
e blou	33	71	65	78		
a gou	45	83	77	90	76	
e gou	31	69	63	76	62	74

ooi	Num	a mooi	e mooi	a boy	e boy	a hondjie
a mooi	41					
e mooi	26	65				
a boy	54	93	78			
e boy	31	70	55	83		
a hondjie	39	78	63	91	68	
e hondjie	28	67	52	80	57	65

aai	Num	a haai	e haai	a time
a haai	40			
e haai	36	74		
a time	40	78	74	
e time	36	74	70	74

oe	Num	a bees
a bees	23	
e bees	25	46

oo	Num	a kool
a kool	27	
e kool	15	40

Table 3.7: An analysis of variance table of the diphthong formants. (See text on page 77 for explanation.)

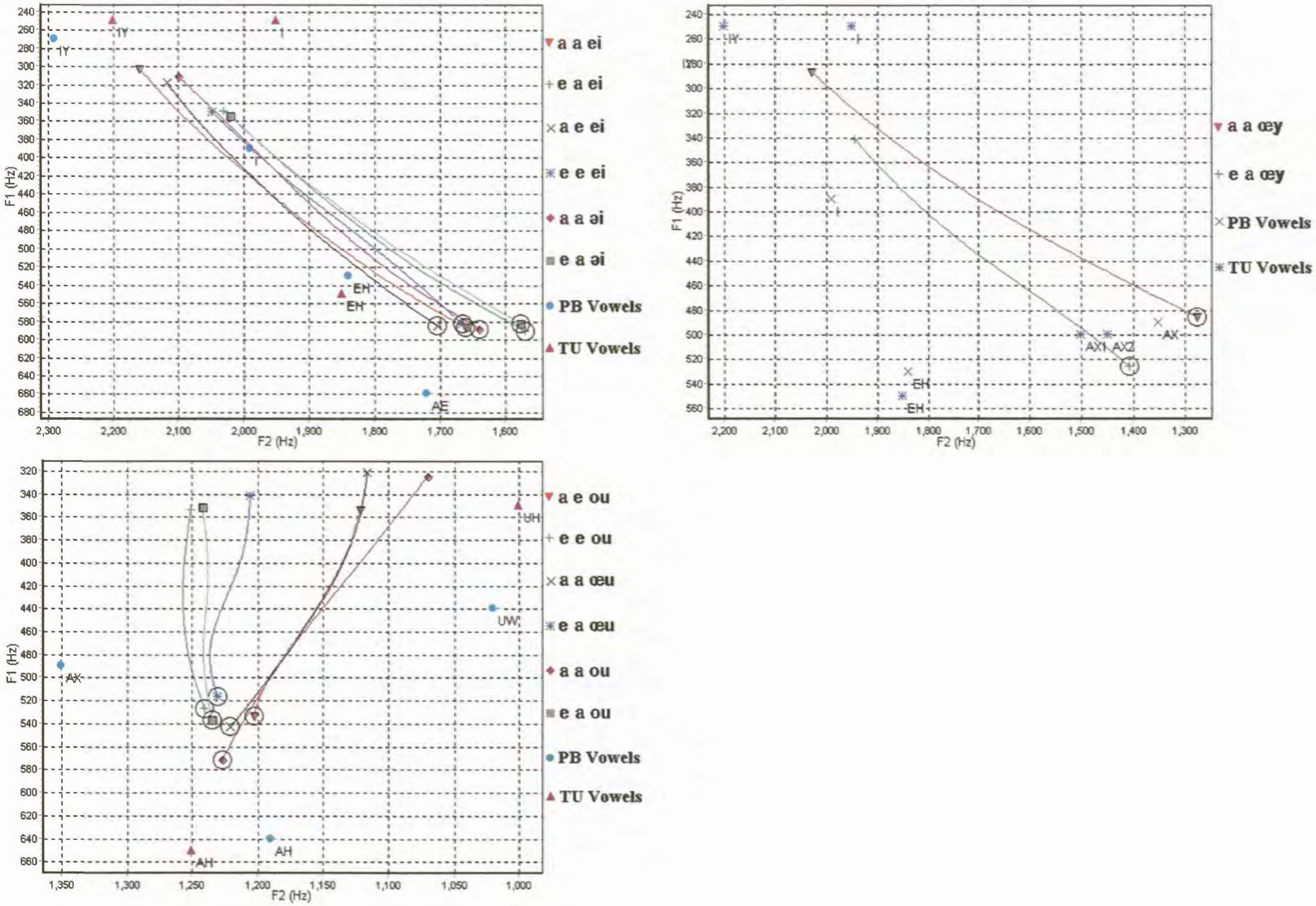


Figure 3.6: Diphthong formant trajectories: [<ei> in “*ryk*” and “play” and also <əi> in “*bly*”], [<œy> in “*trui*”] and [<ou> in “*gou*” and “home” and also <œu> in “*blou*”]. The first *a*/*e* indicates the mother-tongue of the speakers and the second *a*/*e* indicates from which language the vowel was indicted as coming from. *PB Vowels* are the Peterson and Barney vowels and the *TU Vowels* the Taylor and Uys vowels.

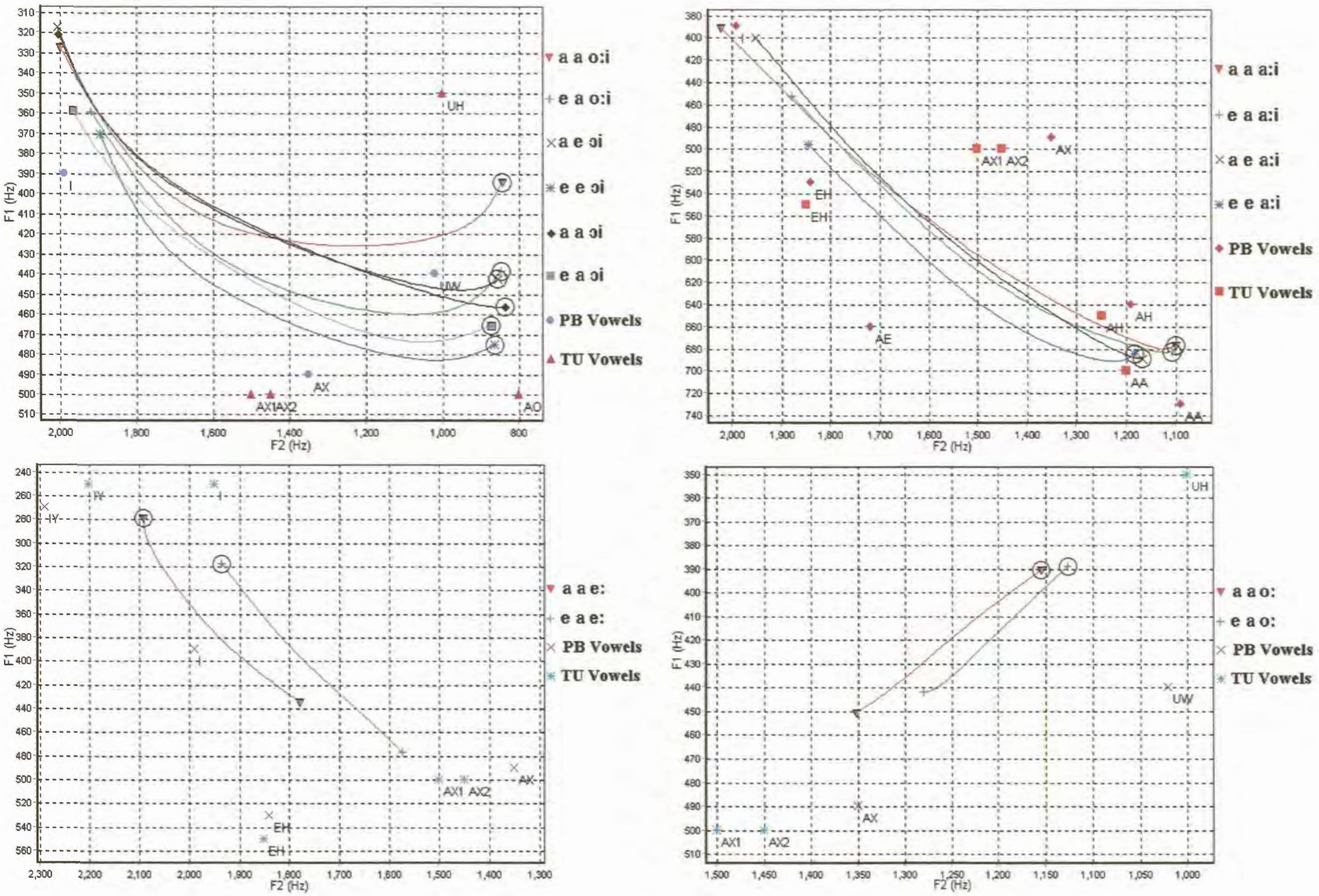


Figure 3.7: Diphthong formant trajectories: [<o:i> in "*mooi*" and also <ɔi> in "*hondjie*" and "boy"], [<a:i> in "*haai*" and "time"], [<e:> in "*bees*"] and [<o:> in "*kool*"]. The first *a/e* indicates the mother-tongue of the speakers and the second *a/e* indicates from which language the vowel was indicted as coming from. *PB Vowels* are the Peterson and Barney vowels and the *TU Vowels* the Taylor and Uys vowels.

Diphthong	Strong Form	Replaced Form
eɪ	f**a**ce [feɪs]	pref**a**ce [ˈprɛfəs]
oʊ	m**o**st [moʊst]	topm**o**st [ˈtɒpməst]
aʊ	m**ou**th [maʊΘ]	Plym**ou**th [plɪməΘ]
aɪə	sh**i**re [ʃaɪə]	Yorksh**i**re [ˈjɔkʃə]

Table 3.8: Examples of the neutral vowel <ə> replacing the strong forms.

diphthong in 3 sections and therefore, seeing as we are observing 3 formants, the next 9 columns of each set in Table 3.7 consist of a number of blocks indicating the presence of significant differences. Unlike Table 3.5 we do not indicate the actual F ratio values as they have no intuitive meaning(comparing curve coefficients may have significant meaning, but the values are not meaningful if simply observed). Returning to our example in Table 3.7 we see that in an ANOVA comparison between the L1 utterances of "*ryk*"(42 utterances) and the L2 utterances of "*ryk*"(38 utterances) we find a significant difference in the final section for F2(this is duly indicated by a black block). We also find significant differences in all three sections of F3, but due to the seemingly "unstable"(due to difficult extraction) nature of F3 we consider this to be less important and therefore indicate all significant differences in F3 by gray boxes in stead of black ones.

Reminder: We have placed a circle around the originating point in the figures.

<ei>

We notice in Table 3.7 that the differences observable within the diphthong <ei> are only in the second and third formants. We find it difficult to find any consistent differences between the two accent groups. From the graphs in Figure 3.6 (top left) we can see that it appears as if the Afrikaans speakers' formant trajectories (▼×◆) occur at a slightly higher F2 than the English trajectories (+∗■). The statistical significance of this seems to be reflected in Table 3.7 except for the English speakers saying <ei> as in "play" which begins in the vicinity of the Afrikaans speakers' beginning of articulation. Interestingly enough it still terminates in the vicinity of the Peterson and Barney <ɪ> where the other English spoken <ei> and <əi> terminate. The Afrikaans speakers on the other hand tend to terminate the diphthong closer to the Peterson and Barney <i>.

<œy>

The diphthong <œy> shows clear separation between the two groups in Figure 3.6 (top right). This diphthong does not occur in English, and this explains the difference. The English speakers pronounce a formant trajectory which begins more or less at their mother-tongue <ə> and then stops well short of <y>. The Afrikaans speakers on the other hand begin in the vicinity of their <œ> and finish fairly close to their location of <y>. This can be seen quite clearly from Figure 3.6, especially if we substitute the vowel locations from Table 3.5, namely the mean formant values calculated and shown in columns 2,3 and 4 (row group "oey"). The analysis of variance results in Table 3.7 echo this large difference with the only region not being significantly different being the first formant about half way along the trajectory.

We can argue that the substitution of <ə> for <œ> by the English speakers is due to the phenomenon of equivalence classification. The closest "vowel" to <œ>, the rounded version of <ə> is naturally <ə>. <ə> is used frequently in English, but <œ> is not. The fact that the English speakers do not reach <y> is once again support of the concept of equivalence classification. The interesting paradox, however, is the fact that in our vowel analysis we found that the English speakers managed to articulate <y> to the point that it was statistically equivalent to that of the Afrikaans speakers. Perhaps this can be explained by the dynamic nature of diphthongs and a difference in articulation of diphthongs and vowels between the two groups (i.e. "laziness" or neutralisation). We continue this argument when discussing <e:>. It is also important to note that in the long vowel analysis the emphasis was only on the single vowel whereas in diphthong analysis this is only partially true.

<ou>

The <ou> cluster consists of diphthongs with two distinct transcriptions. The first is <ou> and is used by De Villiers[40] in words like "*oud*" and "*gou*" and the second is <œu> and is used by Coetzee[23] in words like "*blou*" and "*troue*". We suspect that this was merely a difference in transcription labels, but as we wanted an acoustic model for this diphthong anyway, we decided to include both in the recording session. As surmised, the two transcriptions do represent the same diphthong. For the two accent groups this diphthong has significantly different trajectories.

The diphthongs begin in more or less the same vicinity (as can be seen in Figure 3.6 (bottom right) F1=540 Hz, F2=1225 Hz, which has no vowel associated with this location) and then the English speakers articulate in the direction of the location of the vowel they pronounced when saying the vowel <u:> from the Afrikaans word "*boer*".[3] The Afrikaans speakers on the other hand move towards the location of their <u:> as formed in the word "*boer*". We may have expected the English <ou> to migrate to a much higher F2. We attribute the lack of this to the effect of hypercorrection, in other words, the English speakers may have forced an unnaturally long diphthong in the (wrong) assumption that they were helping to create better data. This problem has been noticed throughout the data and especially prominent with certain speakers and can be heard when playing back the data. Although this effect appears to have had some influence on the results, we can not easily determine how much, and must therefore carry on regardless. Nevertheless, what is noticeable is that there is a significant observable difference(Figure 3.6 (bottom right)) between the English (+ ∗ ■) and Afrikaans (▼ × ◆) pronunciation of <ou>, and it may be slightly larger than we have measured.

---

[3] Notice that the location of the <u:> spoken by the English speakers when using the English word "soon" has an even greater F2.

The mean diphthong tracks plotted in Figure 3.6 (bottom right) and the staggering of the black blocks in Table 3.7 (the third row group "ou") clearly demonstrate the separation between the two pronunciations. Remember that the presence of a black block indicates that there is a significant difference. Therefore the staggered black blocks are as a result of the Afrikaans versus Afrikaans diphthong comparisons (which are similar in this case and therefore do not give rise to black blocks) and the Afrikaans versus English comparisons (which are dissimilar and therefore give rise to black blocks). For example, the first language English utterance of <ou> as in "home" is dissimilar to the Afrikaans first language <ou> as in "*blou*" (and therefore gives rise to black blocks in F2), but it is similar to the <ou> as in second language Afrikaans "*blou*" and therefore does not give rise to black blocks.

<o:i>

The <o:i> diphthongs have proven difficult to define as consistently different between the two groups. Although they can be proven significantly different (see row group "ooi" in Table 3.7) they are also different within the two groups, and this gives rise to a complex "black block" structure in the table.

The English and Afrikaans groups terminate in clusters which we can associate with those languages i.e. the Afrikaans speakers terminate in a cluster which has a higher F2 and lower F1 than the English group(as can be seen in Figure 3.7 (top left)). This follows the same trend as was observed with the vowel <i:> discussed in Section 3.4.1.

Of interest and harder to explain, is that the initial F1 points are distributed over a range of just less than 100 Hz. The greatest culprit causing this large range is the Afrikaans utterance of <o:i>[▼] (compare this to the trajectory of the Afrikaans utterance of the Afrikaans <ɔi>[♦] which is clustered closer to the <o:i> and <ɔi> utterances from the English speakers[■∗]). The extreme curvature observable in this trajectory (and that of the Afrikaans utterance of the English <ɔi>) is most likely due to a segmentation and labelling error where F1 was regularly incorrectly estimated due to poor formant definition (flat peaks) in the initial stages of the segment. The sharpness of the curve would lead us to suspect that this is a justifiable explanation of this phenomenon. This type of error could occur quite easily if we consider that we are speaking of an F1 deviation of about 50 Hz which would be quite unobservable on a spectrogram with a range of 4 kHz or more (i.e. about a 1% shift in the scale or a couple of pixels on a computer screen).

<a:i>

The <a:i> diphthong has an interesting trajectory structure. We find that the initiating articulator or diphthong half vowel is fairly similar for the two language groups (as is expected from the statistical similarity of the vowel <a:> in the two groups as discussed in Section 3.4.1), but what is interesting is a contrary language clustering visible in Figure 3.7 (top right). In other words, the English utterance for the Afrikaans <a:i> is similar to the Afrikaans utterance for the Afrikaans <a:i> and likewise for the English utterance <a:i>. This is reflected in Table 3.7 (the fifth row group "aai") where we can see that the first part of F1 does not differ significantly between the English "*haai*" and Afrikaans "time" or between the Afrikaans "time" and English "time". On the graph this is visible as [$*$ and $\times$] and [■ and ♦] starting together in clusters, but ending clustered as [$*$ and ■] and [$\times$ and ♦].

In spite of this initial "cluster swapping" the diphthong trajectories *terminate* in language clusters as we would have expected. Even though they form language clusters, there are still significant differences between the two groups of utterances by the individual language groups i.e. the <a:i> spoken by the Afrikaans speakers for both English and Afrikaans source words were different and is also evident for the English speakers (third columns).

3.4.3 Diphthongization of <e:> and <o:>

<e:>

<e:> is regarded as a high middle vowel or potential diphthong to some phoneticians but a definite diphthong to others.

Like Taylor[12], De Villiers[40] recognises that <e:> is only found as a vowel in areas of the Cape Province's rural districts, but otherwise it is in fact a diphthong <iə>. This is clearly seen in Figure 3.7. <e:> is prone to becoming <i:> in diminutives as for example in "*seun*"[siən]→"*seuntjie*"[<si:ŋki>].

The Afrikaans speakers are found to produce a diphthong which begins in the location of <y> (close to the point where their articulation of <œy> ends) and ends before reaching their <ə>. The English speakers produce an interesting phenomenon. Like the Afrikaans speakers they begin the diphthong articulation at a point very close to where their <œy> ended giving credence to our previous statement that this may be a articulatory characteristic of English speakers in fluent speech.

The large displacement between the two groups' diphthongs <e:> is quite clearly seen in Figure 3.7 (bottom left). Once again, this is also directly echoed by the overwhelming presence of black blocks (which indicate significant differences) in Table 3.7 (second last row cluster).

<o:>

Like <e:>, <o:> is recognised as being a potential diphthong or by some as a centring diphthong transcribed as <uə>. We confirm this with the plot in Figure 3.7. English does not appear to have a vowel or diphthong similar to <o:>. This may have had an influence on the diphthongization which <o> has undergone in Afrikaans.

The Afrikaans and SA English utterances of this diphthong are proven to be statistically similar. This can be see by the absence of black blocks in Table 3.7 (row group "oo").

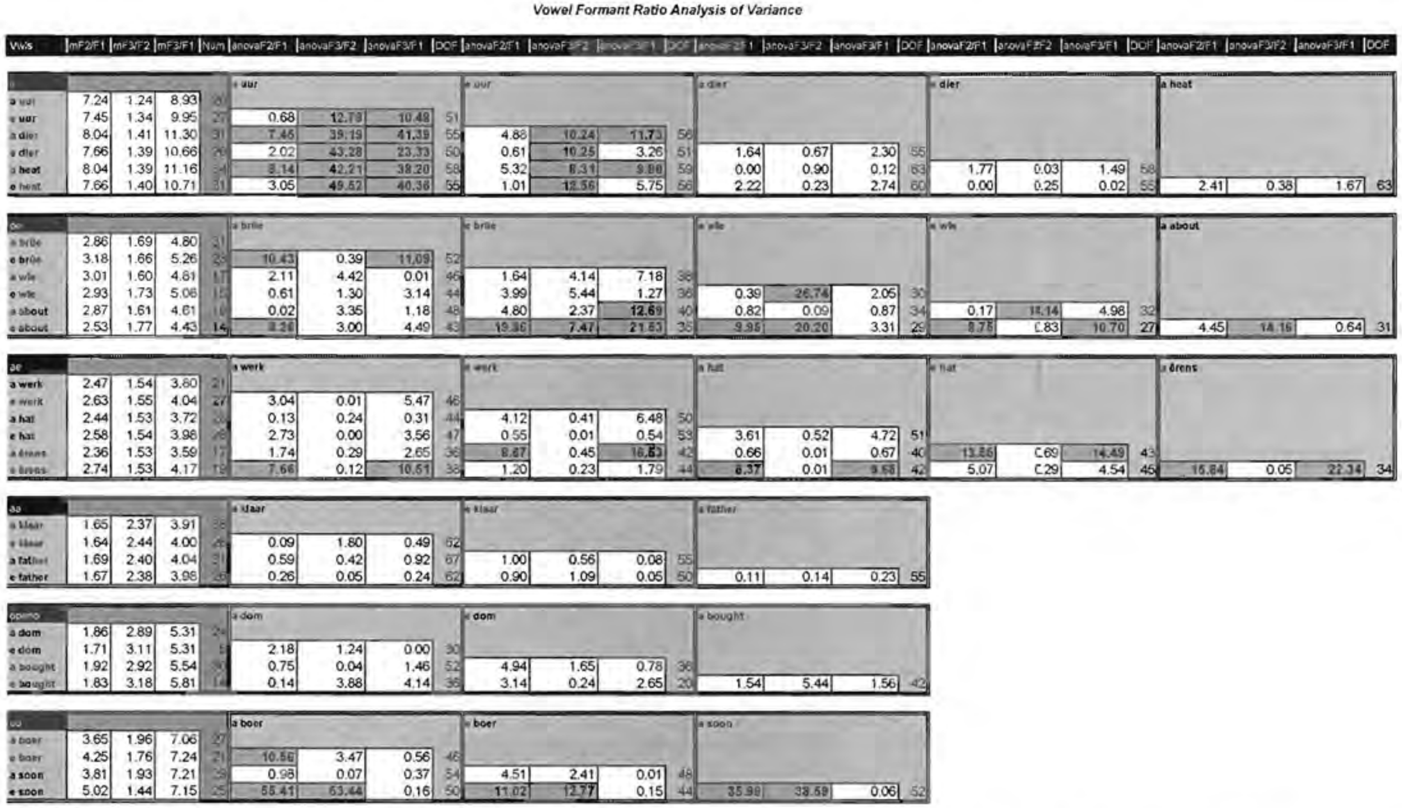
3.4.4 Long vowel and diphthong results - ratios

The vowel formant ratios largely reflect the results discussed above for the direct formant values. Tables 3.9 and 3.10 summarise the results of the formant ratios for the long vowels and diphthongs respectively. We note that some of the accent clustering trends visible in the normal formant data as discussed above becomes even more noticeable in the formant ratio data.

The way the tables are layed out and the use of black squares and F ratio values (in the long vowel ratio chart) are the same as in Tables 3.5 and 3.7.

Generally the normal formant long vowel and formant ratio tables (Tables 3.5 and 3.9) correlate quite well. We do see some differences though. These are:

- Row group "ii", the comparison between "e uur" and "a uur". The difference lies in the second and third columns of the ratio table. Most likely this is due to an F3 extraction problem.
- Row group "oe", all fields. The fairly low F ratio values indicate that <œ:> and $< e :>$ are similar, yet noticeably different. We can not account for the exact differences between the normal formant and ratio formant ANOVAs, but we would prefer to go with the normal analysis and commentary due to the proven value of the standard formant principles.
- Row group "ae". Only a single marginal difference is noted between "a hat" and "e werk".
- Row group "openo", the comparison between "e bought" and "a bought". Once again, a marginal difference which if probably as a result of an F3 extraction



Vowel Formant Ratio Analysis of Variance

Vwls	mF2/F1	mF3/F2	mF3/F1	Num	anovaF2/F1	anovaF3/F2	anovaF3/F1	DOF	anovaF2/F1	anovaF3/F2	anovaF3/F1	DOF	anovaF2/F1	anovaF3/F2	anovaF3/F1	DOF	anovaF2/F1	anovaF3/F2	anovaF3/F1	DOF	anovaF2/F1	anovaF3/F2	anovaF3/F1	DOF
i					a uur				e uur				a dier				e dier				a heat			
a uur	7.24	1.24	8.93	20																				
e uur	7.45	1.34	9.95	27	0.68	12.79	10.48	51																
a dier	8.04	1.41	11.30	31	7.46	39.19	41.39	55	4.88	10.24	11.73	56												
e dier	7.66	1.39	10.66	26	2.02	43.28	23.33	50	0.61	10.25	3.26	51	1.64	0.67	2.30	55								
a heat	8.04	1.39	11.16	34	8.14	42.21	38.20	58	5.32	8.31	9.88	59	0.00	0.90	0.12	63	1.77	0.03	1.49	58				
e heat	7.66	1.40	10.71	31	3.05	49.52	40.36	55	1.01	12.56	5.75	56	2.22	0.23	2.74	60	0.00	0.25	0.02	55	2.41	0.38	1.67	63
oe					a brüe				e brüe				a wie				e wie				a about			
a brüe	2.86	1.69	4.80	31																				
e brüe	3.18	1.66	5.26	23	10.43	0.39	11.09	52																
a wie	3.01	1.60	4.81	17	2.11	4.42	0.01	46	1.64	4.14	7.18	38												
e wie	2.93	1.73	5.06	15	0.61	1.30	3.14	44	3.99	5.44	1.27	36	0.39	26.74	2.05	30								
a about	2.87	1.61	4.61	16	0.02	3.35	1.18	48	4.80	2.37	12.69	40	0.82	0.09	0.87	34	0.17	14.14	4.98	32				
e about	2.53	1.77	4.43	14	8.26	3.00	4.49	43	19.36	7.47	21.63	35	9.95	20.20	3.31	29	8.75	0.83	10.70	27	4.45	14.16	0.64	31
ae					a werk				e werk				a hat				e hat				a ôrens			
a werk	2.47	1.54	3.60	21																				
e werk	2.63	1.55	4.04	27	3.04	0.01	5.47	46																
a hat	2.44	1.53	3.72	[illegible]	0.13	0.24	0.31	44	4.12	0.41	6.48	50												
e hat	2.58	1.54	3.98	28	2.73	0.00	3.56	47	0.55	0.01	0.54	53	3.61	0.52	4.72	51								
a ôrens	2.36	1.53	3.59	17	1.74	0.29	2.65	36	8.67	0.45	16.83	42	0.66	0.01	0.67	40	13.86	0.69	14.49	43				
e ôrens	2.74	1.53	4.17	19	7.66	0.12	10.61	38	1.20	0.23	1.79	44	8.37	0.01	9.58	42	5.07	0.29	4.54	45	15.84	0.05	22.34	34
aa					a klaar				e klaar				a father											
a klaar	1.65	2.37	3.91	38																				
e klaar	1.64	2.44	4.00	26	0.09	1.80	0.49	62																
a father	1.69	2.40	4.04	31	0.59	0.42	0.92	67	1.00	0.56	0.08	55												
e father	1.67	2.38	3.98	26	0.26	0.05	0.24	62	0.90	1.09	0.05	50	0.11	0.14	0.23	55								
opmo					a dom				e dom				a bought											
a dom	1.86	2.89	5.31	24																				
e dom	1.71	3.11	5.31	5	2.18	1.24	0.00	30																
a bought	1.92	2.92	5.54	30	0.75	0.04	1.46	52	4.94	1.65	0.78	36												
e bought	1.83	3.18	5.81	14	0.14	3.88	4.14	36	3.14	0.24	2.65	20	1.54	5.44	1.56	42								
uu					a boer				e boer				a soon											
a boer	3.65	1.96	7.06	27																				
e boer	4.25	1.76	7.24	21	10.56	3.47	0.56	46																
a soon	3.81	1.93	7.21	29	0.98	0.07	0.37	54	4.51	2.41	0.01	48												
e soon	5.02	1.44	7.15	25	55.41	63.44	0.16	50	11.02	12.77	0.15	44	35.98	38.59	0.06	52								

Table 3.9: An analysis of variance table of the vowel formant ratios. See text on page 87 for explanation.

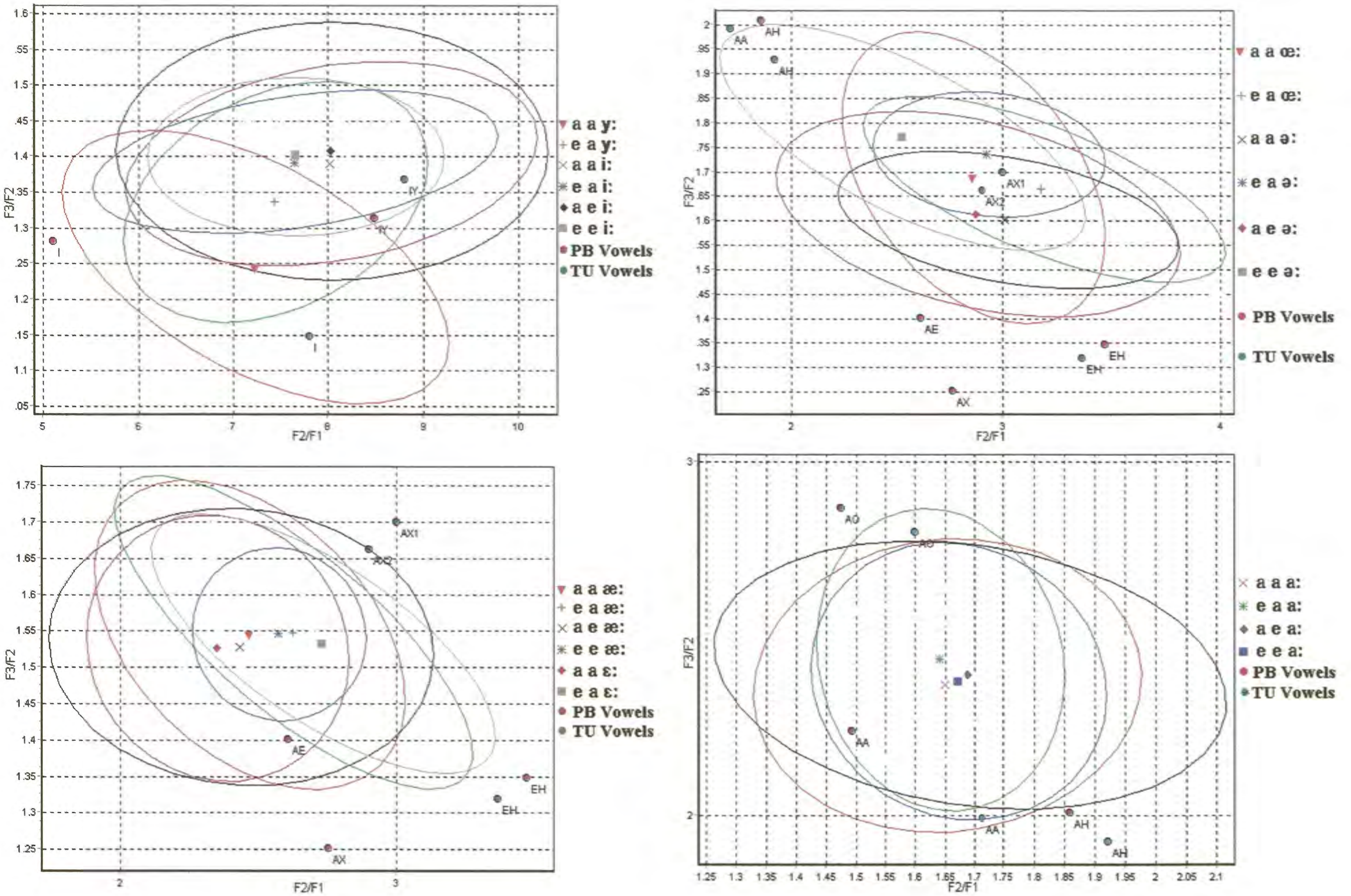


Figure 3.8: Vowel formant ratio clusters: [<y:> in "*uur*" and also <i:> in "*dier*" and "heat"], [<œ:> in "*brûe*" and also <ə:> in "*wîe*" and "about"], [<æ:> in "*werk*" and "hat" and also the incorrectly used <ɛ:> in "*êrens*"] and [<a:> in "*klaar*" and "father"]. The first *a*/*e* indicates the mother-tongue of the speakers and the second *a*/*e* indicates from which language the vowel was indicted as coming from. *PB Vowels* are the Peterson and Barney vowels and the *TU Vowels* the Taylor and Uys vowels.

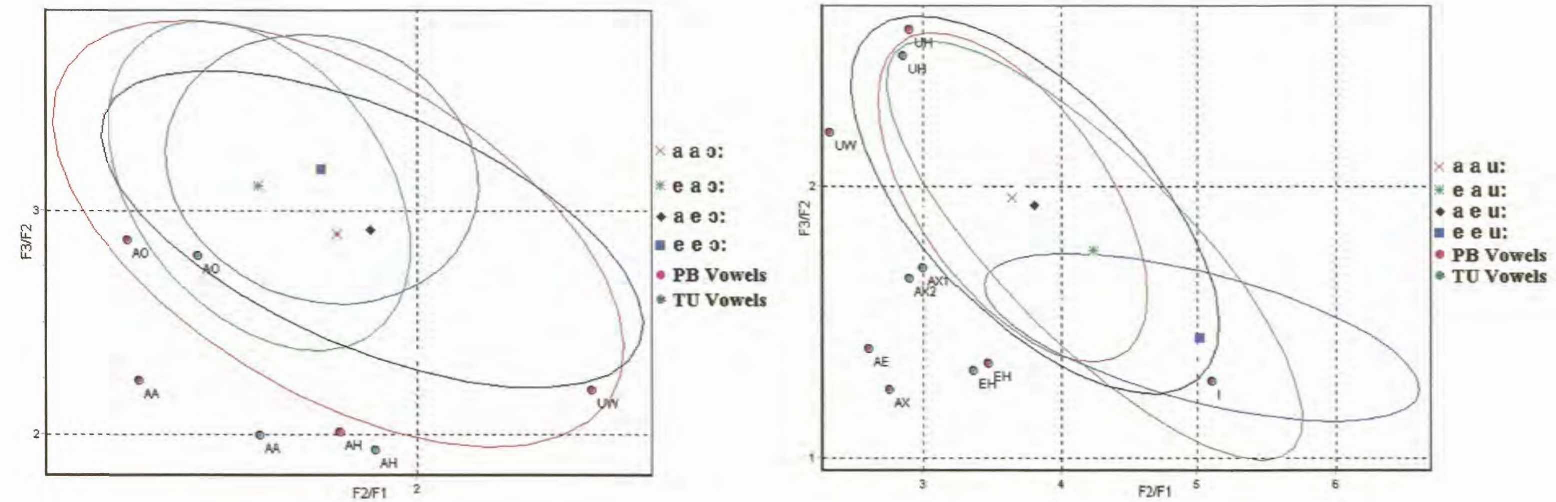


Figure 3.9: Vowel formant ratio clusters: [<ɔ:> in "*dom*" and in "bought"] and [<u:> in "*boer*" and "soon"]. The first *a*/*e* indicates the mother-tongue of the speakers and the second *a*/*e* indicates from which language the vowel was indicted as coming from. *PB Vowels* are the Peterson and Barney vowels and the *TU Vowels* the Taylor and Uys vowels.

Diphthong Formant Ratio Analysis of Variance

Diphthong	Num	a ryk	e ryk	a play	e play	a bly
ei						
a ryk	42					
e ryk	38	78				
a play	45	85	81			
e play	38	78	74	81		
a bly	38	78	74	81	74	
e bly	34	74	70	77	70	70

Diphthong	Num	a trui
oey		
a trui	41	
e trui	33	72

Diphthong	Num	a home	e home	a blou	e blou	a gou
ou						
a home	40					
e home	34	72				
a blou	47	85	79			
e blou	33	71	65	78		
a gou	45	83	77	90	76	
e gou	31	69	63	76	62	74

Diphthong	Num	a mooi	e mooi	a boy	e boy	a hondjie
ooi						
a mooi	41					
e mooi	26	65				
a boy	54	93	78			
e boy	31	70	55	83		
a hondjie	39	78	63	91	68	
e hondjie	28	67	52	80	57	65

Diphthong	Num	a haai	e haai	a time
aai				
a haai	40			
e haai	36	74		
a time	40	78	74	
e time	36	74	70	74

Diphthong	Num	a bees
eu		
a bees	23	
e bees	25	46

Diphthong	Num	a kool
oo		
a kool	27	
e kool	15	40

Table 3.10: An analysis of variance table of the diphthong formant ratios. See text on page 87 for explanation.

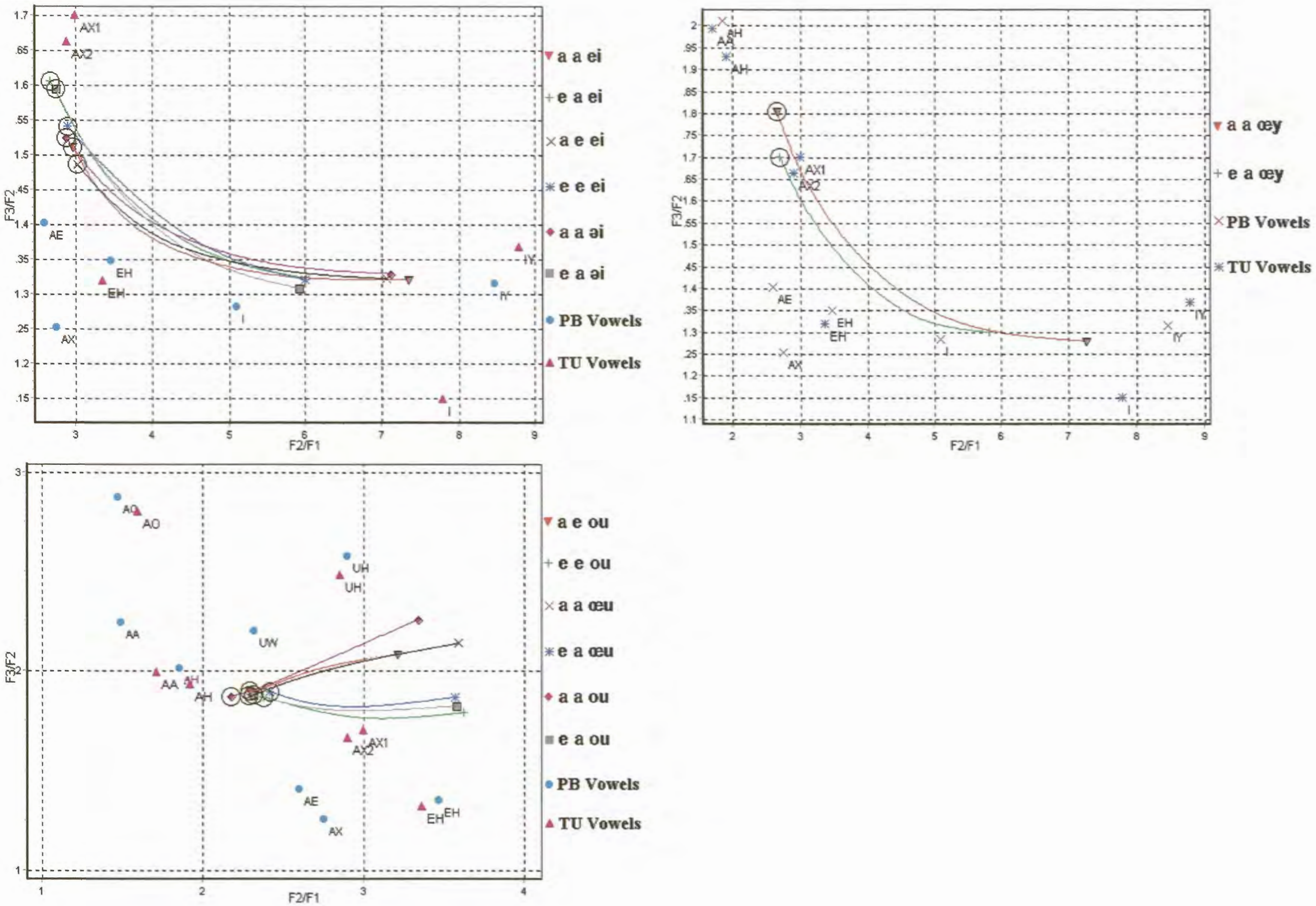


Figure 3.10: Diphthong ratio formant trajectories: [<ei> in "*ryk*" and "play" and also <əi> in "*bly*"], [<œy> in "*trui*"] and [<ou> in "*gou*" and "home" and also <œu> in "*blou*"]. The first *a*/*e* indicates the mother-tongue of the speakers and the second *a*/*e* indicates from which language the vowel was indicted as coming from. *PB Vowels* are the Peterson and Barney vowels and the *TU Vowels* the Taylor and Uys vowels.

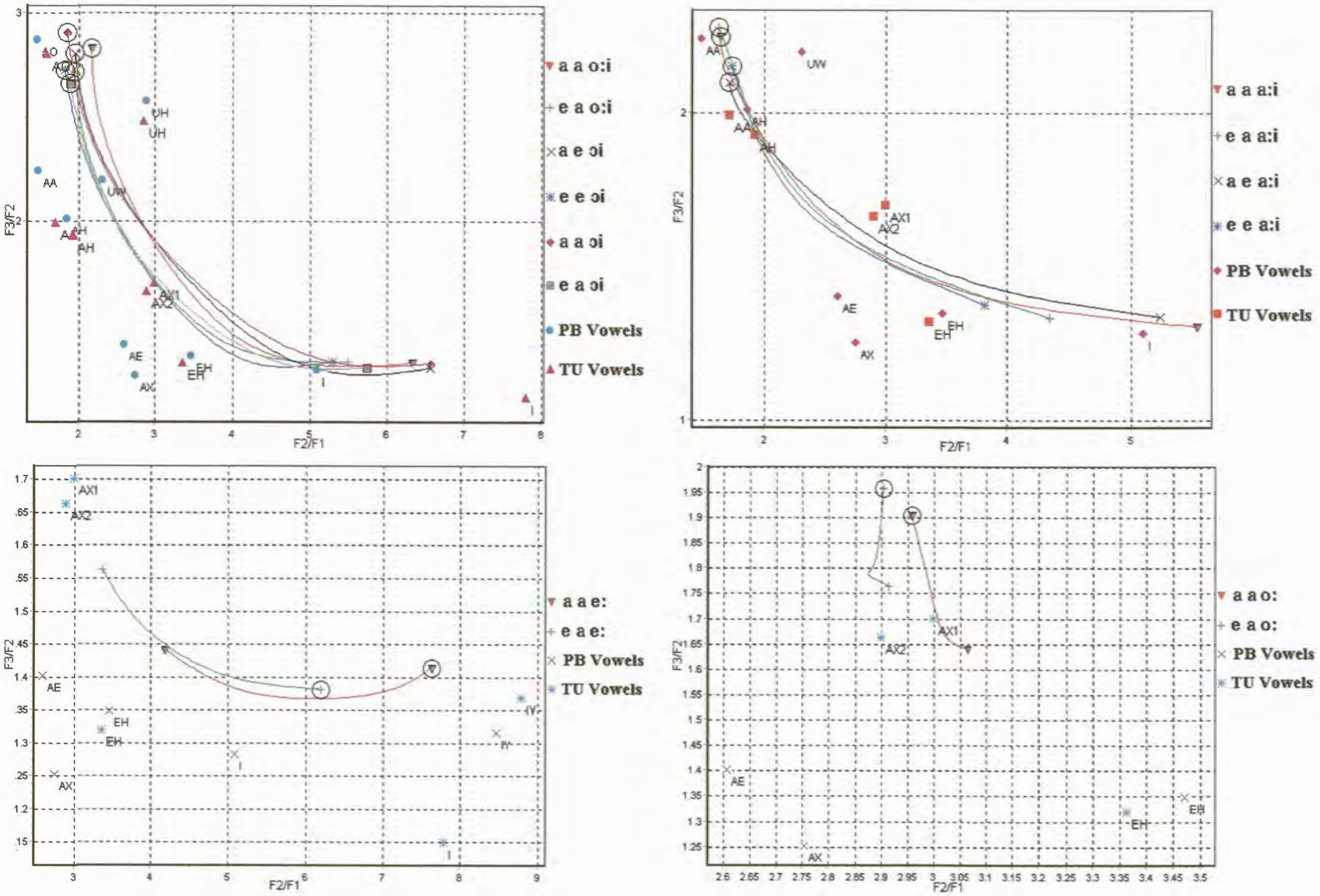


Figure 3.11: Diphthong formant ratio trajectories: [<o:i> in "*mooi*" and also <ɔi> in "*hondjie*" and "boy"], [<a:i> in "*haai*" and "time"], [<e:> in "*bees*"] and [<o:> in "*kool*"].

difficulty.

The diphthong tables correlate excellently. Tables 3.7 and 3.10 only differ in one row group, namely "ei". The differences can probably largely be attributed to F3 extraction problems. The normal formant analysis results should rather be used as their analysis techniques have been accepted and are understood.

We have previously noted the controversy surrounding the "formant ratio theory". As stated, it is not entirely clear whether the formant "ratios" i.e. $\frac{F3}{F2}$ and $\frac{F2}{F1}$ have any true significance. This is primarily because we do not fully understand how the brain processes speech, and although the formant ratios do appear to have significance to the brain as pattern recogniser, there is still much scientific debate as to the validity of the theory. Formants are clearly the resonance peaks of voiced speech, and obviously have linguistic importance, but science has been unable to prove whether it is the actual locations of these formants, or merely the spacing between them which is important.

We do not attempt to speculate here as to the correctness of this theory, but for completeness and out of scientific curiosity we have calculated the formant ratios, both for the static vowels and for the dynamic diphthongs, and graphed them.

3.4.5 Long vowel and diphthong diphthongization results

The long vowels are considered to be quasi-stationary. It is impossible for anything but synthetic speech to have truly stationary formants. We would however like to have some measure of the level of diphthongization of voiced speech and use this as a metric for whether to label a sound as a vowel or a diphthong.

Using the cubic spline formant trajectories we have calculated for the long vowels and diphthongs, we calculate the difference between initial and terminal points of F1, F2 and F3 namely ΔF1, ΔF2 and ΔF3, and also the cumulative absolute shift in

Vowels

	$\Delta F1$	$\Delta F2$	$\Delta F3$	$\Sigma \Delta F1$	$\Sigma \Delta F2$	$\Sigma \Delta F3$	$\lvert\Delta F1\rvert+\lvert\Delta F2\rvert$	$\lvert\Delta F1\rvert+\lvert\Delta F2\rvert+\lvert\Delta F3\rvert$
i								
a uur	-12.70	27.22	34.87	12.70	40.41	35.56	39.92	74.79
e uur	-7.50	5.48	-6.30	15.48	16.75	52.45	12.98	19.28
a dier	-5.21	13.39	-20.31	5.21	32.84	23.71	18.60	38.91
e dier	-7.51	27.76	10.44	9.06	27.76	25.28	35.27	45.71
a heat	-10.60	-0.11	-22.43	10.69	10.00	28.55	10.71	33.14
e heat	-8.18	23.72	-25.32	8.18	23.72	38.10	31.90	57.22
oe								
a brue	-2.95	-1.99	-19.19	4.45	3.29	24.58	4.94	24.13
e brue	6.36	1.03	-24.82	6.98	7.71	25.53	7.39	32.21
a wie	-1.37	-5.42	21.98	11.37	17.09	28.64	6.79	28.77
e wie	8.66	11.36	-4.52	14.50	11.36	14.67	20.02	24.54
a about	-10.43	1.67	-12.72	12.01	7.95	17.00	12.10	24.82
e about	-0.15	33.61	-6.21	10.02	33.78	11.29	33.76	39.97
ae								
a werk	-12.92	5.57	-4.91	13.63	16.21	50.28	18.49	23.40
e werk	8.87	5.86	-18.33	13.28	7.89	38.77	14.73	33.06
a hat	-2.34	21.39	-22.34	11.18	21.39	57.26	23.73	46.07
e hat	-4.74	-2.31	-15.30	13.17	22.54	24.46	7.05	22.35
a erans	-0.91	25.01	7.79	11.44	25.01	32.46	25.92	33.71
e erens	-2.95	-17.66	4.42	8.88	20.67	28.57	20.61	25.03
a:								
a klaar	-8.89	-25.77	-15.57	20.28	25.87	28.49	34.66	50.23
e klaar	-11.66	-18.49	52.85	11.71	19.98	52.85	30.15	83.00
a father	-22.29	-16.79	8.66	22.29	20.57	14.67	39.08	47.74
e father	-3.25	-1.94	-7.20	5.56	6.50	24.54	5.19	12.39
openo								
a dom	-8.24	-19.97	13.20	8.66	19.97	13.30	28.21	41.41
e dom	14.52	-27.50	40.99	32.82	27.51	51.69	42.02	83.01
a bought	-14.80	-13.06	21.90	14.80	18.83	37.05	27.86	49.76
e bought	-14.14	-13.35	-9.38	14.14	13.35	30.47	27.49	36.87
u:								
a boer	-4.91	-9.99	31.57	8.31	10.68	31.57	14.90	46.47
e boer	-4.94	-7.61	-5.64	6.64	27.11	33.23	12.55	18.19
a soon	-16.81	11.15	19.71	16.81	13.40	50.95	27.96	47.67
e soon	-4.36	-10.13	-29.34	4.36	17.07	29.34	14.49	43.83

Diphthongs

	$\Delta F1$	$\Delta F2$	$\Delta F3$	$\Sigma \Delta F1$	$\Sigma \Delta F2$	$\Sigma \Delta F3$	$\lvert\Delta F1\rvert+\lvert\Delta F2\rvert$	$\lvert\Delta F1\rvert+\lvert\Delta F2\rvert+\lvert\Delta F3\rvert$
ei								
a ryk	-283.67	498.08	355.56	283.67	498.08	355.56	781.75	1137.31
e ryk	-241.67	461.64	174.15	241.67	461.64	174.15	703.31	877.46
a play	-266.93	413.98	281.54	266.93	413.98	281.54	680.91	962.45
e play	-232.79	382.23	146.19	232.79	382.23	146.19	615.02	761.21
a bly	-277.19	459.08	299.89	277.19	459.08	299.89	736.27	1036.16
e bly	-227.82	444.96	137.49	227.82	444.96	137.49	672.78	810.27
oey								
a trui	-199.18	754.05	307.28	199.18	754.05	328.76	953.23	1260.51
e trui	-184.32	536.24	138.45	184.32	536.24	187.11	720.56	859.01
ou								
a home	-179.17	-80.93	5.99	179.17	80.93	14.71	260.10	266.09
e home	-173.65	10.08	-100.30	173.65	22.73	124.85	183.73	284.03
a blou	-221.82	-104.27	17.88	221.82	104.27	72.79	326.09	343.97
e blou	-175.06	-25.19	-120.32	175.06	35.78	137.27	200.25	320.57
a gou	-246.20	-156.74	-7.77	246.20	156.74	36.31	402.94	410.71
e gou	-184.88	7.23	-102.55	184.88	14.79	116.39	192.11	294.66
ooi								
a mooi	-66.98	1157.96	313.55	130.88	1157.96	428.25	1224.94	1538.49
e mooi	-79.03	1071.53	257.18	122.82	1071.53	375.30	1150.56	1407.74
a boy	-125.46	1152.61	228.86	135.53	1152.61	399.95	1278.07	1506.93
e boy	-104.67	1029.09	186.53	120.03	1029.09	281.17	1133.76	1320.29
a hondjie	-135.59	1167.89	252.15	135.59	1167.89	470.64	1303.48	1555.63
e hondjie	-107.42	1092.70	231.83	123.32	1092.70	376.76	1200.12	1431.95
aai								
a haai	-285.45	921.28	153.48	292.09	921.28	289.17	1206.73	1360.21
e haai	-229.64	769.91	-17.61	231.40	769.91	125.05	999.55	1017.16
a time	-289.05	783.90	148.10	289.05	783.90	187.55	1072.95	1221.05
e time	-188.12	661.14	-9.12	202.42	661.14	111.90	849.26	858.38
[illegible]								
a bees	155.37	-311.14	-396.41	155.37	311.14	396.41	466.51	862.92
e bees	159.44	-361.62	-224.89	159.44	361.62	230.44	521.06	745.95
[illegible]								
a kool	60.26	197.07	23.85	60.26	197.07	40.49	257.33	281.18
e kool	53.24	153.71	69.66	53.24	153.71	109.17	206.95	276.61

Table 3.11: A table indicating the level of diphthongization of the long vowels and diphthongs.

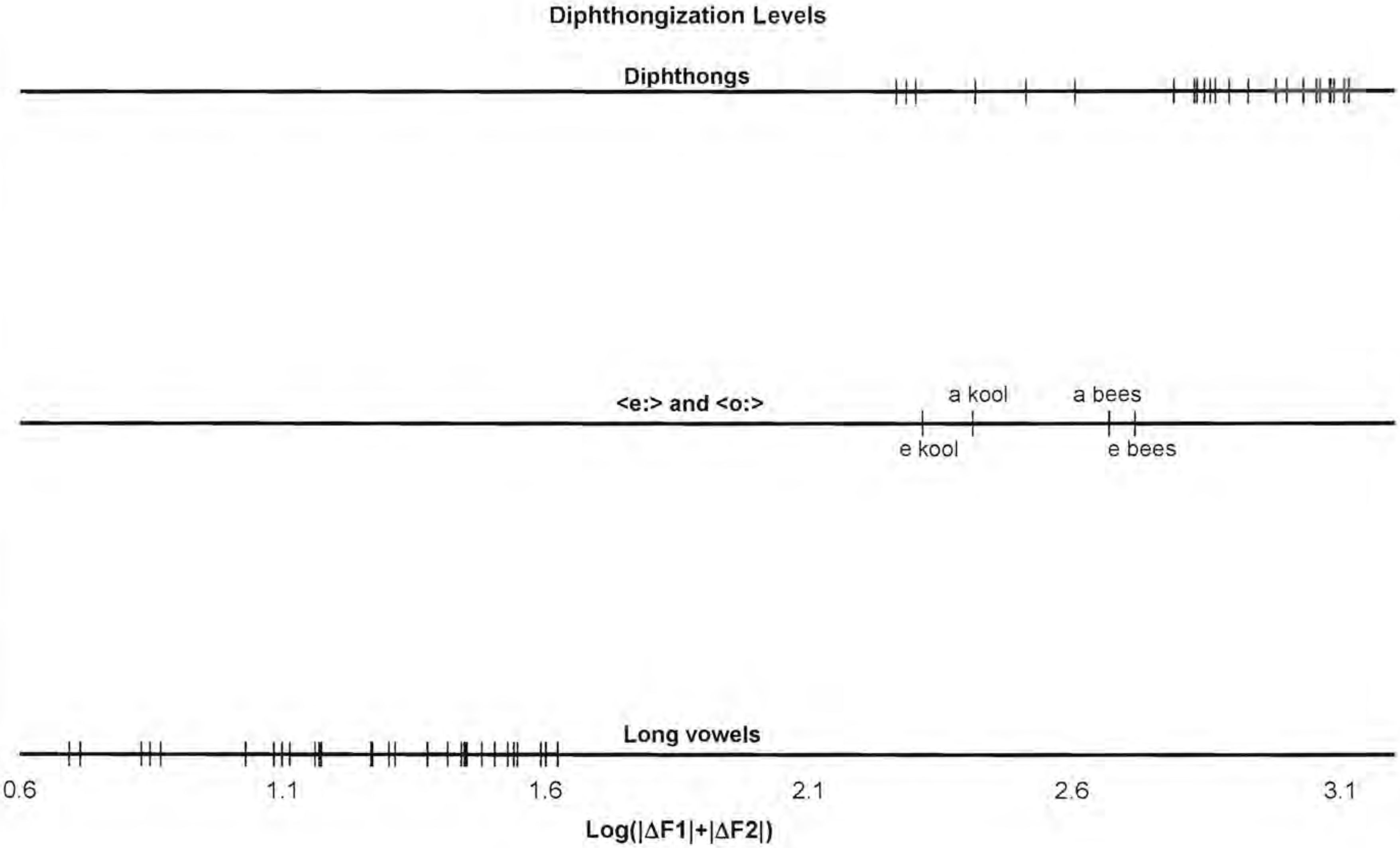


Figure 3.12: A figure giving a graphical representation of the data in Table 3.11 with the level of diphthongization of the long vowels and diphthongs indicated as well as <e:> and <o:> which is clearly seen to lie in the range of the diphthongs.

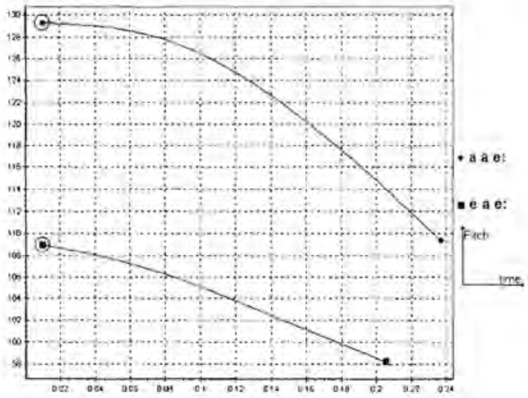


Figure 3.13: Vowel pitch trajectory: [<e:> in "*bees*"].

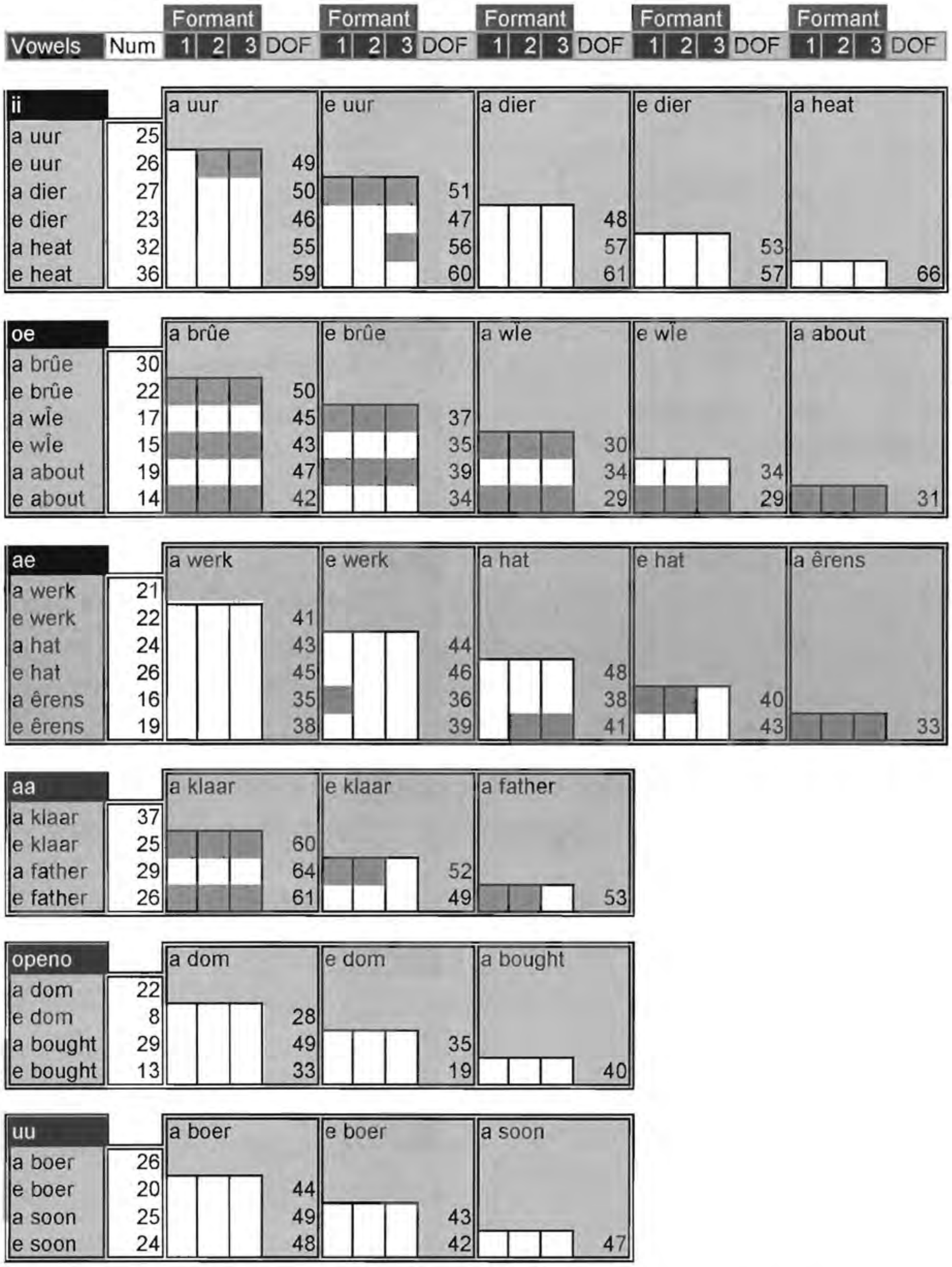
formant values ΣF1, ΣF2 and ΣF3, as demonstrated in Figure 2.11. The results are given in Table 3.11 where we have included a measure of diphthongization, namely, $|\Delta F1| + |\Delta F2|$ (and also $|\Delta F1| + |\Delta F2| + |\Delta F3|$). The results are also represented graphically in Figure 3.12 where we have plotted the logarithm of $|\Delta F1| + |\Delta F2|$ to scale the data for plotting clarity. We can see from Table 3.11 that by simply stating the condition that if the measure $|\Delta F1| + |\Delta F2|$ exceeds about 200Hz then we can consider the sound to be a diphthong or conversely, if the measure is less than 40Hz then it is a vowel. It is of special importance to note that this is also true of the much argued over "potential" diphthongs <e:> and <o:>. This adds further credibility to the notion that these sounds should be classified as diphthongs and not vowels, in spite of their short duration. This has been emphasised in Figure 3.12.

3.4.6 Pitch results

Tables 3.12 and 3.13 summarise the analysis of variance comparisons between the pitch trajectories of the vowels and diphthongs respectively. We have also included a single plot of pitch trajectories for the vowel/diphthong <e:> in Figure 3.13. The rest of the plots are available in Appendix A.3.

It is important to note that, while we determine significant differences between the means of the English and Afrikaans pitch trajectories, this appears to be more a speaker

Vowel Pitch Analysis of Variance



Vowels	Num	Formant 1 2 3 DOF	Formant 1 2 3 DOF	Formant 1 2 3 DOF	Formant 1 2 3 DOF	Formant 1 2 3 DOF

ii		a uur	e uur	a dier	e dier	a heat
a uur	25					
e uur	26	☐ ▨ ▨ 49				
a dier	27	☐ ☐ ☐ 50	▨ ▨ ▨ 51			
e dier	23	☐ ☐ ☐ 46	☐ ☐ ☐ 47	☐ ☐ ☐ 48		
a heat	32	☐ ☐ ☐ 55	☐ ☐ ▨ 56	☐ ☐ ☐ 57	☐ ☐ ☐ 53	
e heat	36	☐ ☐ ☐ 59	☐ ☐ ☐ 60	☐ ☐ ☐ 61	☐ ☐ ☐ 57	☐ ☐ ☐ 66

oe		a brûe	e brûe	a wîe	e wîe	a about
a brûe	30					
e brûe	22	▨ ▨ ▨ 50				
a wîe	17	☐ ☐ ☐ 45	▨ ▨ ▨ 37			
e wîe	15	▨ ▨ ▨ 43	☐ ☐ ☐ 35	▨ ▨ ▨ 30		
a about	19	☐ ☐ ☐ 47	▨ ▨ ▨ 39	☐ ☐ ☐ 34	☐ ☐ ☐ 34	
e about	14	▨ ▨ ▨ 42	☐ ☐ ☐ 34	▨ ▨ ▨ 29	▨ ▨ ▨ 29	▨ ▨ ▨ 31

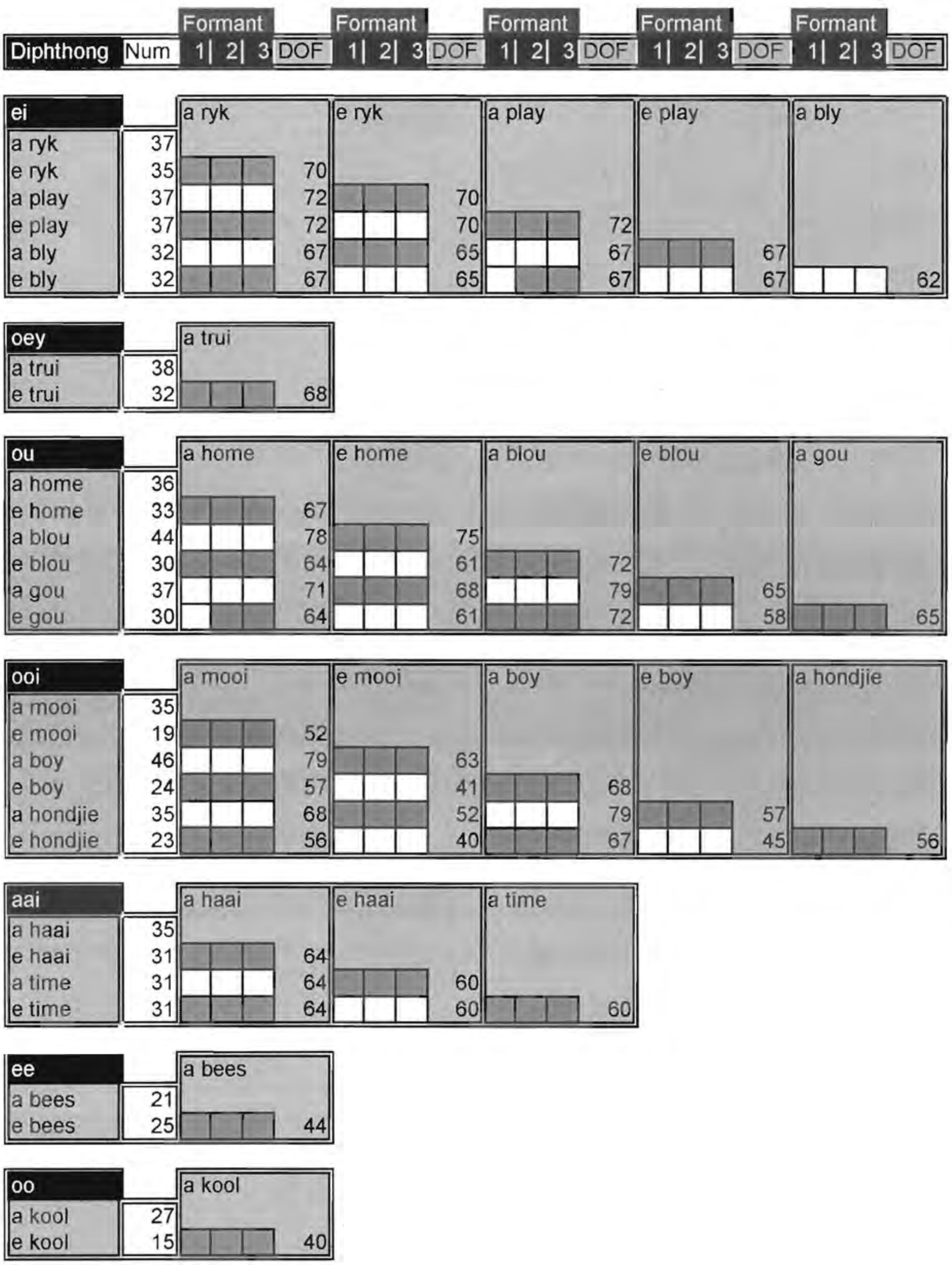
ae		a werk	e werk	a hat	e hat	a êrens
a werk	21					
e werk	22	☐ ☐ ☐ 41				
a hat	24	☐ ☐ ☐ 43	☐ ☐ ☐ 44			
e hat	26	☐ ☐ ☐ 45	☐ ☐ ☐ 46	☐ ☐ ☐ 48		
a êrens	16	☐ ☐ ☐ 35	▨ ☐ ☐ 36	☐ ☐ ☐ 38	▨ ▨ ☐ 40	
e êrens	19	☐ ☐ ☐ 38	☐ ☐ ☐ 39	☐ ▨ ▨ 41	☐ ☐ ☐ 43	▨ ▨ ▨ 33

aa		a klaar	e klaar	a father
a klaar	37			
e klaar	25	▨ ▨ ▨ 60		
a father	29	☐ ☐ ☐ 64	▨ ▨ ☐ 52	
e father	26	▨ ▨ ▨ 61	☐ ☐ ☐ 49	▨ ▨ ☐ 53

openo		a dom	e dom	a bought
a dom	22			
e dom	8	☐ ☐ ☐ 28		
a bought	29	☐ ☐ ☐ 49	☐ ☐ ☐ 35	
e bought	13	☐ ☐ ☐ 33	☐ ☐ ☐ 19	☐ ☐ ☐ 40

uu		a boer	e boer	a soon
a boer	26			
e boer	20	☐ ☐ ☐ 44		
a soon	25	☐ ☐ ☐ 49	☐ ☐ ☐ 43	
e soon	24	☐ ☐ ☐ 48	☐ ☐ ☐ 42	☐ ☐ ☐ 47

Table 3.12: An analysis of variance table of the long-vowel pitches.



Diphthong Pitch Analysis of Variance

Diphthong	Num	Formant 1 2 3 DOF	Formant 1 2 3 DOF	Formant 1 2 3 DOF	Formant 1 2 3 DOF	Formant 1 2 3 DOF
ei		a ryk	e ryk	a play	e play	a bly
a ryk	37					
e ryk	35	▨▨▨ 70				
a play	37	☐☐☐ 72	▨▨▨ 70			
e play	37	▨▨▨ 72	☐☐☐ 70	▨▨▨ 72		
a bly	32	☐☐☐ 67	▨▨▨ 65	☐☐☐ 67	▨▨▨ 67	
e bly	32	▨▨▨ 67	☐☐☐ 65	☐▨▨ 67	☐☐☐ 67	☐☐☐ 62
oey		a trui				
a trui	38					
e trui	32	▨▨▨ 68				
ou		a home	e home	a blou	e blou	a gou
a home	36					
e home	33	▨▨▨ 67				
a blou	44	☐☐☐ 78	▨▨▨ 75			
e blou	30	▨▨▨ 64	☐☐☐ 61	▨▨▨ 72		
a gou	37	☐☐☐ 71	▨▨▨ 68	☐☐☐ 79	▨▨▨ 65	
e gou	30	☐▨▨ 64	☐☐☐ 61	▨▨▨ 72	☐☐☐ 58	▨▨▨ 65
ooi		a mooi	e mooi	a boy	e boy	a hondjie
a mooi	35					
e mooi	19	▨▨▨ 52				
a boy	46	☐☐☐ 79	▨▨▨ 63			
e boy	24	▨▨▨ 57	☐☐☐ 41	▨▨▨ 68		
a hondjie	35	☐☐☐ 68	▨▨▨ 52	☐☐☐ 79	▨▨▨ 57	
e hondjie	23	▨▨▨ 56	☐☐☐ 40	▨▨▨ 67	☐☐☐ 45	▨▨▨ 56
aai		a haai	e haai	a time		
a haai	35					
e haai	31	▨▨▨ 64				
a time	31	☐☐☐ 64	▨▨▨ 60			
e time	31	▨▨▨ 64	☐☐☐ 60	▨▨▨ 60		
ee		a bees				
a bees	21					
e bees	25	▨▨▨ 44				
oo		a kool				
a kool	27					
e kool	15	▨▨▨ 40				

Table 3.13: An analysis of variance table of the diphthong pitches.

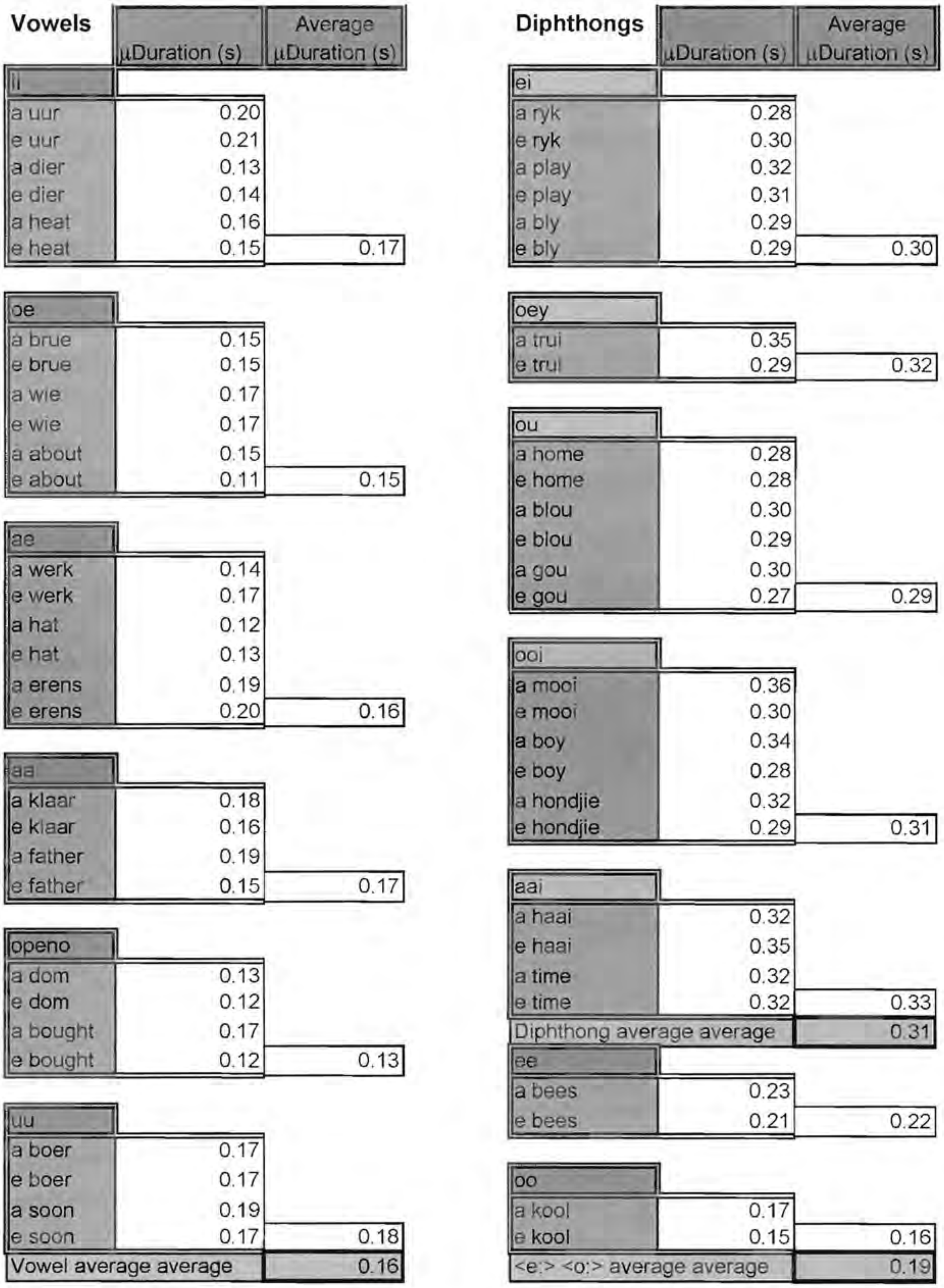
dependent influence than a language or accent group influence. This means that, because we have relatively few speakers, it is quite probable that the difference in pitch trajectories is purely due to the fact that a few of the speakers in the Afrikaans group spoke with a slightly higher pitched voice. We doubt that this is a general trend observable across the language groups.

It is also important to note that if we assume that this is true, and then adjust all the pitch trajectories to begin at the same pitch (i.e. introduce an offset), then we find no significantly observable differences between the two groups. In other words, there appears to be no difference in the intonation (change in pitch over time) of the two groups even though there are observable differences in pitch. It is unlikely therefore that we can make use of intonation as a means of distinguishing between the two language/accent groups.

There are a few interesting features which we can deduce from the pitch trajectory plots in Appendix A.3 and from Table 3.14 which shows mean durations.

Firstly, the diphthongs have noticeably longer duration than the vowels. This is true for all the diphthongs, except the potential diphthongs or centring diphthongs <e:> and <o:> which are called vowels by some phonologists. Perhaps it is this relatively short duration that has led to them being labelled as vowels in the past. Although the centring diphthongs have short durations similar to vowels (as can be seen at the bottom right of Table 3.14) they are clearly diphthongized (as is reflected in Table 3.11 and Figure 3.12). If we perform an analysis of variance between the long vowels, diphthongs and potential diphthongs we find that there is no statistically measurable difference between the mean durations of the long vowels and <e:> and <o:>. Between the long vowels and diphthongs there is though (as expected), and between the diphthongs and <e:> and <o:> there is too (which is also expected). This, with F ratio values is summarised in Table 3.15.

Another interesting feature is that the initial pitch of the diphthongs appears to be held



Vowels	µDuration (s)	Average µDuration (s)
ii		
a uur	0.20	
e uur	0.21	
a dier	0.13	
e dier	0.14	
a heat	0.16	
e heat	0.15	0.17
oe		
a brue	0.15	
e brue	0.15	
a wie	0.17	
e wie	0.17	
a about	0.15	
e about	0.11	0.15
ae		
a werk	0.14	
e werk	0.17	
a hat	0.12	
e hat	0.13	
a erens	0.19	
e erens	0.20	0.16
aa		
a klaar	0.18	
e klaar	0.16	
a father	0.19	
e father	0.15	0.17
openo		
a dom	0.13	
e dom	0.12	
a bought	0.17	
e bought	0.12	0.13
uu		
a boer	0.17	
e boer	0.17	
a soon	0.19	
e soon	0.17	0.18
Vowel average average		0.16

Diphthongs	µDuration (s)	Average µDuration (s)
ei		
a ryk	0.28	
e ryk	0.30	
a play	0.32	
e play	0.31	
a bly	0.29	
e bly	0.29	0.30
oey		
a trui	0.35	
e trui	0.29	0.32
ou		
a home	0.28	
e home	0.28	
a blou	0.30	
e blou	0.29	
a gou	0.30	
e gou	0.27	0.29
ooi		
a mooi	0.36	
e mooi	0.30	
a boy	0.34	
e boy	0.28	
a hondjie	0.32	
e hondjie	0.29	0.31
aai		
a haai	0.32	
e haai	0.35	
a time	0.32	
e time	0.32	0.33
Diphthong average average		0.31
ee		
a bees	0.23	
e bees	0.21	0.22
oo		
a kool	0.17	
e kool	0.15	0.16
<e:> <o:> average average		0.19

Table 3.14: A table showing the mean duration of the long-vowels and diphthongs.

First Set	Second Set	F ratio
Vowels	<e:> and <o:>	1.00
Diphthongs	<e:> and <o:>	**61.05**
Vowels	Diphthongs	**84.54**

Table 3.15: Analysis of variance of the long vowels, diphthongs and potential diphthongs' mean durations as given in Table 3.14.

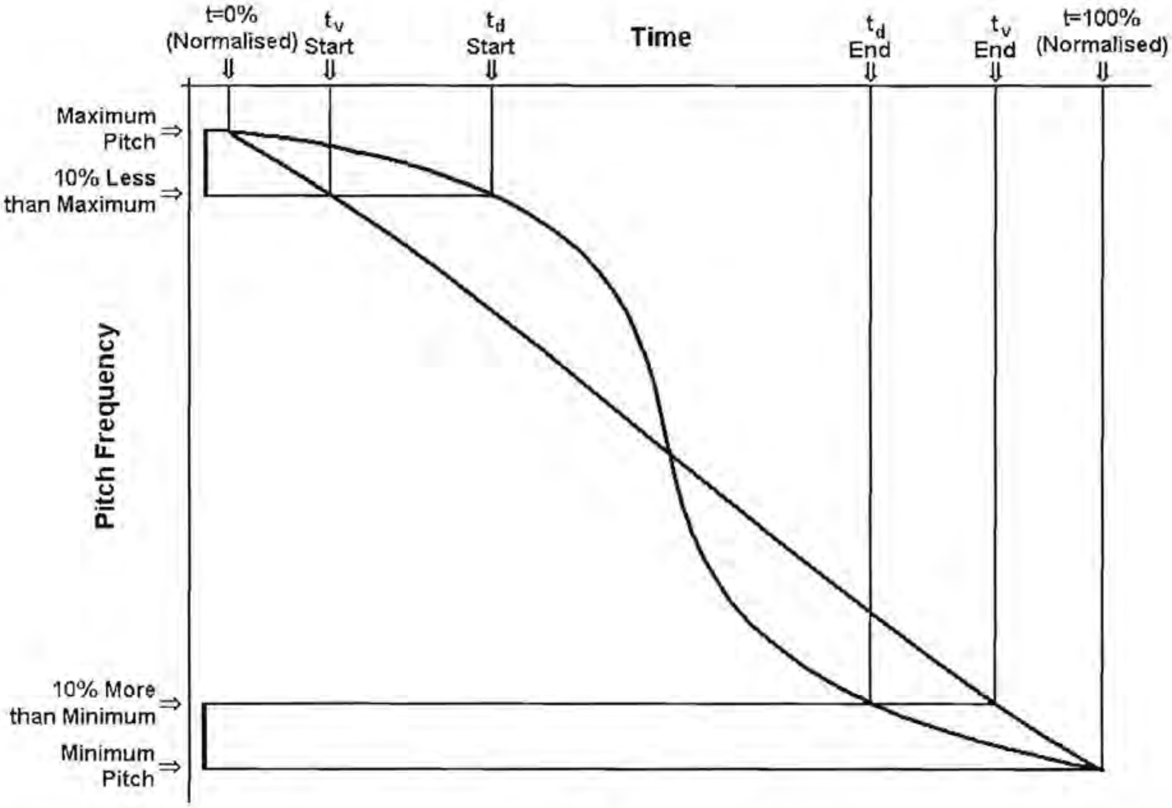


Figure 3.14: This graph demonstrates how we measured the change of pitch from the initial point to final point for the long vowels and diphthongs.

Vowels

	tv Start	tv End	μtv Start	μtv End
ii				
a uur	0.38	0.95		
e uur	0.29	0.92		
a dier	0.14	0.93		
e dier	0.13	0.94		
a heat	0.55	0.95		
e heat	0.13	0.93	0.27	0.94
oe				
a brue	0.36	0.94		
e brue	0.18	0.95		
a wie	0.45	0.94		
e wie	0.24	0.94		
a about	0.24	0.94		
e about	0.19	0.91	0.28	0.94
ae				
a werk	0.15	0.92		
e werk	0.40	0.96		
a hat	0.11	0.96		
e hat	0.29	0.94		
a erens	0.27	0.93		
e erens	0.38	0.95	0.27	0.94
aa				
a klaar	0.23	0.94		
e klaar	0.20	0.93		
a father	0.44	0.91		
e father	0.27	0.94	0.28	0.93
openo				
a dom	0.20	0.90		
e dom	0.18	0.83		
a bought	0.21	0.94		
e bought	0.22	0.87	0.20	0.88
uu				
a boer	0.25	0.95		
e boer	0.57	0.95		
a soon	0.28	0.94		
e soon	0.27	0.94	0.34	0.94
Average average vowel pitch onset time			0.27	0.93

Diphthongs

	td Start	td End	μtd Start	μtd End
ei				
a ryk	0.20	0.88		
e ryk	0.39	0.92		
a play	0.34	0.89		
e play	0.36	0.89		
a bly	0.34	0.92		
e bly	0.35	0.91	0.33	0.90
oey				
a trui	0.25	0.88		
e trui	0.24	0.80	0.24	0.84
ou				
a home	0.36	0.91		
e home	0.31	0.89		
a blou	0.16	0.84		
e blou	0.18	0.88		
a gou	0.40	0.94		
e gou	0.18	0.86	0.27	0.89
ooi				
a mooi	0.36	0.78		
e mooi	0.22	0.92		
a boy	0.23	0.60		
e boy	0.24	0.95		
a hondjie	0.41	0.90		
e hondjie	0.32	0.95	0.30	0.85
aai				
a haai	0.19	0.76		
e haai	0.35	0.93		
a time	0.41	0.91		
e time	0.43	0.92	0.35	0.88
Average average diphthong pitch onset time			0.30	0.87
ee				
a bees	0.27	0.93		
e bees	0.21	0.92	0.24	0.92
oo				
a kool	0.18	0.90		
e kool	0.20	0.93	0.19	0.91
Average average <e:> <o:> pitch onset time			0.22	0.92

Table 3.16: The fraction of the normalised length of the long vowels and diphthongs at which the pitch has dropped by 10% from the maximum pitch and still has 10% to drop to the minimum pitch. The last line in the table reflects the fact that for diphthongs, the pitch "spends more time" at the initial and final values.

First Set	Second Set	F ratio Start	F ratio End
Vowels	<e:> and <o:>	0.88	0.88
Diphthongs	<e:> and <o:>	3.78	1.13
Vowels	Diphthongs	0.98	**13.27**

Table 3.17: Analysis of variance of the long vowels, diphthongs and potential diphthongs mean pitch onset times, for $t_{v/d}Start$ and $t_{v/d}End$ as shown in Figure 3.14 and Table 3.16.

for a while before shifting to the lower pitch, and reaches the lower pitch sooner than the long vowels do (This concept is demonstrated in Figure 3.14). In the case of the long vowels the pitch shift is a relatively smooth transition. A table with measurements that confirm this are given in Table 3.16. This may be a side-effect of the hyper-correction induced by the recording of non-continuous speech. This warrants further research as it may have a substantial influence on the realism of synthetic speech generated with prosodic influences added to improve naturalness.

If we submit these starting and ending threshold times to an analysis of variance (ANOVA) test between the long vowels, diphthongs and potential diphthongs we yield Table 3.17. The results are a bit inconclusive due to the low degree of freedom introduced by <e:> and <o:>, but clearly, by looking at the means and the ANOVA results, we can see that the vowels and diphthongs have similar initial pitch slopes (intonations), but have dissimilar final slopes, with the diphthongs reaching their minimum pitch sooner than the vowels. This can be seen by the bold **13.27** in Table 3.17 which indicates a large F ratio and hence a statistically measurable difference.

Chapter 4

Summary and conclusion

This dissertation has presented the motivation, background theory, technique and results on an acoustical modelling of the long vowels and diphthongs of Afrikaans and South African English.

We believe such a study was justifiably motivated by:

- A need for a further study of the acoustic phonetics of Afrikaans and South African English.
- With the use of text to speech(TTS) systems, a thorough understanding of the pronunciation of phonemes and their uniqueness in an accent is required for natural and realistic sounding speech.
- With the likely large-scale roll-out of Automatic Speech Recognition technologies in the near future, a need for compensating for "foreign accents".

4.1 Summary of results

Using a multiple stage recording technique we collected long-vowels and diphthongs from Afrikaans and South African English first language speakers and also recorded them speaking their second languages. We recorded the long vowels and diphthongs in three ways:

- Isolated form: For example, just the <aː> in f**a**ther.
- Contextual form: For example, the complete word "f**a**ther".
- Pseudo-word form: Using consonants which have minimal influence on the vowels and diphthongs, for example, "h-**a**-t" ([haːt]).

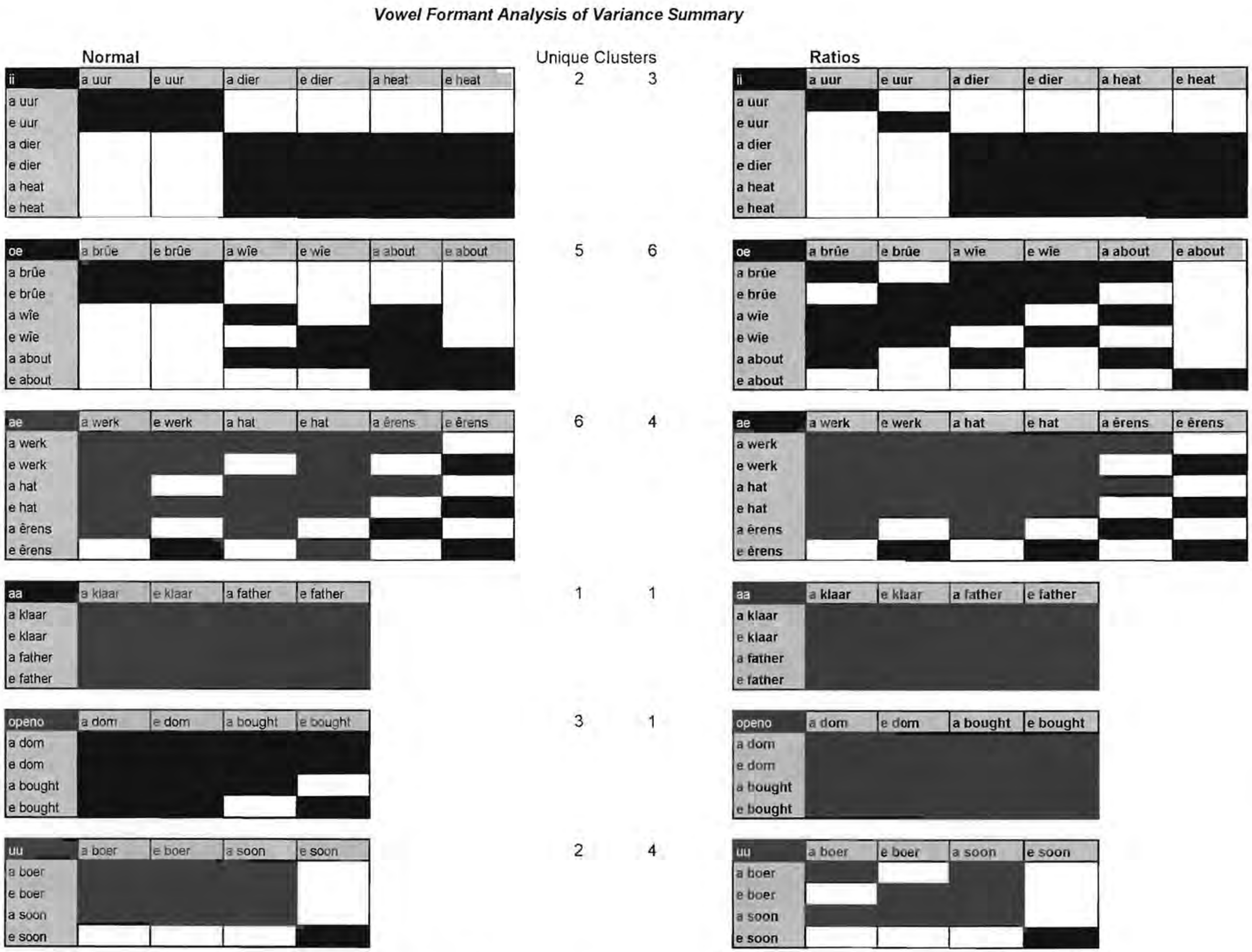
The data was verified (by listening), segmented and labelled (using time and spectrogram representations). The primary analysis/modelling technique consisted of formant plotting. The formants are the resonant peaks of voiced speech. We utilised linear prediction techniques to extract the formants. This extraction was then verified by superimposing the extracted formants on the spectrograms of the speech segments from which they came and visual inspection. Any obvious extraction mistakes were manually corrected.

Each of the vowels and diphthongs have been discussed in detail in Section 3.4. It is important however not to "miss the forest for the trees". We have not noticed any general trends between the two language groups' formant structures in terms of global features. For example, some researchers have noted that Afrikaans vowels appear to be more centred around <ə> than British English vowels (for example, compare Ward[20] and Coetzee[23]). This may be true, but no such trend between Afrikaans and South African English has been observed by us. This would suggest that South African English has been noticeably influenced by Afrikaans and vice-versa. Some of the vowels demonstrate distinct differences and other not. We can therefore not apply a single linear transformation to all vowels or diphthongs to adapt the one formant space to the other, however, knowing which vowels differ, and in which way, is an important result, especially in context of the justification for this study(as mentioned at the beginning of this chapter).

Table 4.1 and table 4.2 summarise the results that were discussed in Chapter 3 "Experiments". The black blocks represent that particular set of vowels or diphthongs (from a particular language group) that has been found to be statistically similar (at the 0.99 significance level) to other vowel and diphthong sets from the same or the other accent group. So, for example, we see in Table 4.1 that for the group of vowels usually transcribed as <u:> ("uu"), the set of English utterances of the vowel <u:> as in the word "soon" form their own cluster (i.e. are dissimilar to the other sets) and the sets of Afrikaans utterances of <u:> as in "*boer*" (farmer) and "soon" and the set of English utterances of <u:> as in "*boer*" form their own cluster. We define a cluster as being a unique row in a row group. Discussions of this clustering and other phenomena were discussed in detail in Chapter 3.

We have developed a new measure of diphthongization and used it to cluster voiced speech segments as either vowels (having low diphthongization) or diphthongs (having high diphthongization). Using this metric we were able to justify the claim by some phoneticians that <e:> and <o:> are in fact diphthongs and not long vowels.

The pitch trajectories were also analysed for the vowels and diphthongs, and even though the mean trajectories were found to be different, a simple logical transformation (i.e. mean normalisation) results in the pitch trajectories generally being statistically identical. We therefore suggest that the differences are purely a speaker dependent phenomenon and not accent dependent.



Vowel Formant Analysis of Variance Summary

Normal

ii	a uur	e uur	a dier	e dier	a heat	e heat
a uur	■	■				
e uur	■	■				
a dier			■	■	■	■
e dier			■	■	■	■
a heat			■	■	■	■
e heat			■	■	■	■

oe	a brûe	e brûe	a wîe	e wîe	a about	e about
a brûe	■	■				
e brûe	■	■				
a wîe			■		■	
e wîe				■	■	
a about			■	■	■	■
e about					■	■

ae	a werk	e werk	a hat	e hat	a êrens	e êrens
a werk	▨	▨	▨	▨	▨	
e werk	▨	▨		▨		■
a hat	▨		▨	▨	▨	
e hat	▨	▨	▨	▨		■
a êrens	▨		▨		■	
e êrens		■		▨		■

aa	a klaar	e klaar	a father	e father
a klaar	▨	▨	▨	▨
e klaar	▨	▨	▨	▨
a father	▨	▨	▨	▨
e father	▨	▨	▨	▨

openo	a dom	e dom	a bought	e bought
a dom	■	■	■	■
e dom	■	■	■	■
a bought	■	■	■	
e bought	■	■		■

uu	a boer	e boer	a soon	e soon
a boer	▨	▨	▨	
e boer	▨	▨	▨	
a soon	▨	▨	▨	
e soon				■

Unique Clusters

2	3
5	6
6	4
1	1
3	1
2	4

Ratios

ii	a uur	e uur	a dier	e dier	a heat	e heat
a uur	■					
e uur		■				
a dier			■	■	■	■
e dier			■	■	■	■
a heat			■	■	■	■
e heat			■	■	■	■

oe	a brûe	e brûe	a wie	e wie	a about	e about
a brûe	■		■	■	■	
e brûe		■	■	■		
a wie	■	■	■		■	
e wie	■	■		■		
a about	■		■		■	
e about						■

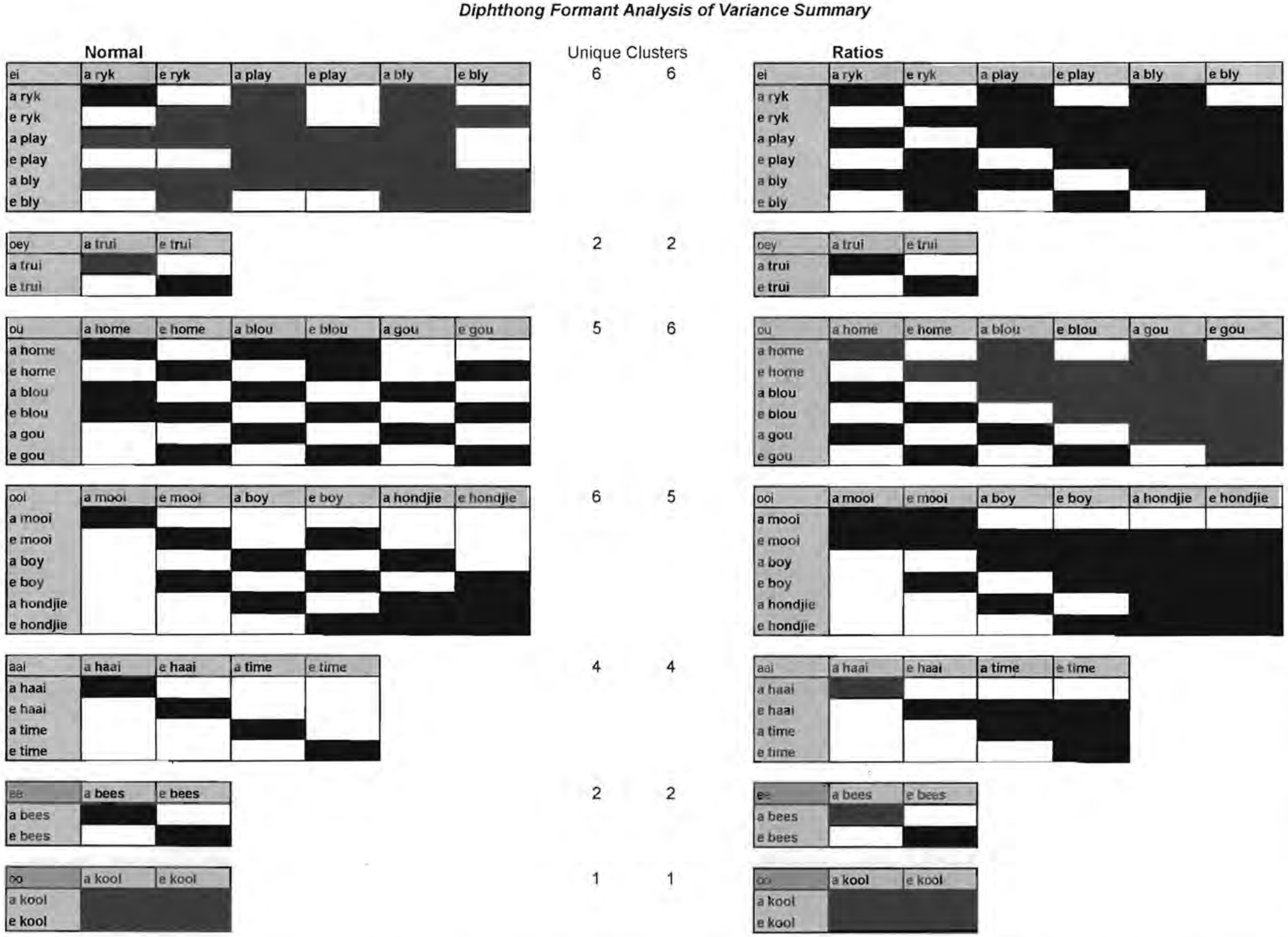
ae	a werk	e werk	a hat	e hat	a êrens	e êrens
a werk	▨	▨	▨	▨	▨	
e werk	▨	▨	▨	▨		■
a hat	▨	▨	▨	▨	▨	
e hat	▨	▨	▨	▨		■
a êrens	▨		▨		■	
e êrens		■		■		■

aa	a klaar	e klaar	a father	e father
a klaar	▨	▨	▨	▨
e klaar	▨	▨	▨	▨
a father	▨	▨	▨	▨
e father	▨	▨	▨	▨

openo	a dom	e dom	a bought	e bought
a dom	▨	▨	▨	▨
e dom	▨	▨	▨	▨
a bought	▨	▨	▨	▨
e bought	▨	▨	▨	▨

uu	a boer	e boer	a soon	e soon
a boer	▨		▨	
e boer		▨	▨	
a soon	▨	▨	▨	
e soon				■

Table 4.1: A statistical similarity cluster table for the long-vowel formants (normal[left] and ratios[right]).



Diphthong Formant Analysis of Variance Summary

Normal — Unique Clusters — **Ratios**

Normal

ei	a ryk	e ryk	a play	e play	a bly	e bly
a ryk	■		▨		▨	
e ryk		▨	▨		▨	▨
a play	▨	▨	▨	▨	▨	
e play			▨	▨	▨	
a bly	▨	▨	▨	▨	▨	▨
e bly		▨			▨	▨

Unique Clusters: 6 6

Ratios

ei	a ryk	e ryk	a play	e play	a bly	e bly
a ryk	■		■		■	
e ryk		■	■	■	■	■
a play	■		■	■	■	■
e play		■		■	■	■
a bly	■	■	■		■	■
e bly		■		■		■

Normal

oey	a trui	e trui
a trui	▨	
e trui		■

Unique Clusters: 2 2

Ratios

oey	a trui	e trui
a trui	■	
e trui		■

Normal

ou	a home	e home	a blou	e blou	a gou	e gou
a home	■		■	■		
e home		■		■		■
a blou	■		■		■	
e blou	■	■		■		■
a gou			■		■	
e gou		■		■		■

Unique Clusters: 5 6

Ratios

ou	a home	e home	a blou	e blou	a gou	e gou
a home	▨		▨		▨	
e home		▨	▨	▨	▨	▨
a blou	■		▨	▨	▨	▨
e blou		■		▨	▨	▨
a gou	■		■		▨	▨
e gou		■		■		▨

Normal

ooi	a mooi	e mooi	a boy	e boy	a hondjie	e hondjie
a mooi	■					
e mooi		■		■		
a boy			■		■	
e boy		■		■		■
a hondjie			■		■	■
e hondjie				■	■	■

Unique Clusters: 6 5

Ratios

ooi	a mooi	e mooi	a boy	e boy	a hondjie	e hondjie
a mooi	■	■				
e mooi	■	■	■	■	■	■
a boy			■	■	■	■
e boy		■		■	■	■
a hondjie			■		■	■
e hondjie				■	■	■

Normal

aai	a haai	e haai	a time	e time
a haai	■			
e haai		■		
a time			■	
e time				■

Unique Clusters: 4 4

Ratios

aai	a haai	e haai	a time	e time
a haai	▨			
e haai		■	■	■
a time			■	■
e time				■

Normal

ee	a bees	e bees
a bees	■	
e bees		■

Unique Clusters: 2 2

Ratios

ee	a bees	e bees
a bees	▨	
e bees		■

Normal

oo	a kool	e kool
a kool	▨	▨
e kool	▨	▨

Unique Clusters: 1 1

Ratios

oo	a kool	e kool
a kool	▨	▨
e kool	▨	▨

Table 4.2: A statistical similarity cluster table for the diphthong formant trajectories (normal[left] and ratios[right]).

4.2 Shortcomings and future work

Due to a mistake in the initial data collection planning the long vowel <ɛ:> was incorrectly replaced with the long vowel <æ:>. This was as a result of an alternative pronunciation of the word "*êrens*" (somewhere) which is used in certain parts of South Africa, namely [ɛ:rəns] as opposed to [æ:rəns]. As a result, to complete the study a thorough analysis of <ɛ:> would have to be undertaken to complete this study. This problem clearly demonstrates the need for an accurate pronunciation dictionary for the South African languages, one of the reasons this study was undertaken in the first place!

We reiterate that we are aware the the research only holds true for a particular group of speakers in South Africa. There are a couple of L1 accents for both Afrikaans and South African English, and this study focuses on L1 and L2 common to well educated white males on the Gauteng Province.

The temporal nature of the diphthongs has largely been down-played in this study. The diphthong formant graphs clearly display the mean path followed by the formants during articulation of the diphthongs, but the rate at which they do this is not visible. We suggest that there may be importance in the temporal shift from one "vowel" to another during diphthong articulation. This should be studied in future work.

Important future work includes the use of the models developed here in ASR and TTS systems. Although fairly basic, the models have the potential to increase recognition rates of "foreign" accents in automatic speech recognition systems and make more natural and familiar sounding text-to-speech systems.