CHAPTER 3

RESEARCH METHODOLOGY

This chapter describes the method of data selection, data collection and analysis in this study. This chapter is divided into four sections. The first section describes subjects of this study, the second section describes the data, third and fourth sections explain how the data were collected and analysed.

3.1 Subjects

The subjects in this study consisted of 13[1] Iranian students, six males and seven females, from the Iranian High school (in Kuala Lumpur, Malaysia). Both male and female students were selected for analysis in order to ascertain if there are gender variables in the production of English vowels. Since Iranian students begin learning English when they are 12 years old, this group of students would have had approximately four to five years of exposure to learning English. After high school their exposure to English is likely to be limited since English is not used extensively in Iran. This is why high school students were chosen as subjects of this study as they reflect a cohort which have already started to learn English and are still learning and using English in the classroom. Thus, the extent of exposure to English is relatively controlled with this group of speakers. The subjects were all in the second semester of the 2010/2011 academic year, and the first language for all the subjects is Persian. The data were collected in April 2010 from the subjects who were between the ages of 16 to 17 years. We are aware that there may be differences in the vocal apparatus between males and females in this age group as they are in what Beck (1999, p. 256) considered as the second phase of the life cycle. At this stage there will be significant changes in the male

[1] Initially there was 14 subjects, but one of the male subjects was removed from the analysis as he was from mixed parentage

vocal apparatus for males, typically noticed as a change in voice quality among young males. Keeping this in mind, it was ensured that for all the male subjects, their voices had broken at the time of the study. The acoustic data from both males and females were also analysed separately in view of the physical differences in the vocal tract between them (see 4.1.1- 4.2).

Discussions with teachers indicated that the samples are representative of Iranian learners aged from 15-17 years and are from the same area of Iran. At the time of the study, the subjects had been in Malaysia for an average of 26 months. Based on auditory impression and verification by their English teachers, it was felt that the duration that the subjects had been in Malaysia did not influence their English pronunciation. This could be due to the fact that most of their communication is among fellow Iranians, where Persian is used. Conversations with the subjects showed that they did not use English very much outside of the classroom, or have much contact with Malaysian speakers of English.

All of them had studied English as a subject for five to six years. Seven students were in the Second Grade and six students were in the Third Grade of high school (O level equivalent). Upon successful completion of Grade Three at High School, Iranian students may choose to do a pre-university course (A level equivalent). The subjects were selected based on their English language proficiency. To ensure similar levels proficiency in English, 13 students were selected based on their previous semester's final English examination results and by verifying the selection with their current English language teachers. The school- based English examination included both written and oral components. The written test comprises sections on vocabulary (choosing the correct word in a cloze passage), grammar (e.g. using the correct tense

and aspect of the given verb) and sound recognition (e.g. choosing a word with a different sound among a list given) and reading comprehension. For the oral component, the students are evaluated more on their content knowledge (for example providing synonyms) and less on their pronunciation ability. Thus, although the subjects have done well in their exam, the facts are not reflective of their productive skills. The final grade for each component is 20 marks, and the description for the range of marks are provided in Table 3.1, while the Grades obtained by the subjects are provided in Table 3.2. The names of the students are not used in this study so as not to reveal their identities. Instead, codes are used to identify them: SF1-SF7 refers to the female subjects while SM1-SM6 refers to the male subjects.

No other description is provided apart from the general grade descriptions presented in Table 3.1. However, a discussion with the English language teachers indicated that an 'excellent' grade meant that the students are able to use English appropriately and accurately in both spoken and written contexts, while a 'very good' grade meant that the students can use English appropriately and accurately in both spoken and written contexts but occasionally with occasional grammar mistakes.

Table 3.1: Description of Grades

Marks	Grades	Description
18-20	A+	Excellent
16-17.90	A	Very Good
14-15.90	B	Good
12-13.90	C	Satisfactory
10-11.90	D	Marginal
below 10	F	Fail

Table 3.2: First Semester English Grades

Female subjects	**Grade**	**Writing Score (20-20)**	**Speaking Score (20-20)**
SF 1	Two	20	20
SF 2	Two	20	20
SF 3	Two	20	20
SF 4	Three	17.5	20
SF 5	Three	20	20
SF 6	Three	20	20
SF 7	Three	20	20
Male Subjects	**Grade**	**Writing Score(20-20)**	**Speaking Score(20-20)**
SM 1	Two	20	20
SM 2	Two	20	20
SM 3	Two	20	20
SM 4	Two	20	20
SM 5	Three	20	20
SM 6	Three	20	20

3.2 Data

The subjects were recorded in two speaking contexts: reading a word list and an informal speaking context.

(i) Word List Context

The target vowels were embedded in hVd words (except for *hoar* to avoid confusion with *horde)* which was placed in a carrier frame, *Please say hVd again*. The carrier frame is an attempt at providing a more naturalistic context for the target word as opposed to eliciting words in isolation. The use of the frame also keeps the speaking rate more constant: "If the words "say" and "again" are about the same on each occasion, then we can assume that the speaker is reading at a constant rate" (Ladefoged, 1993, p. 188-189). An hVd context was selected as this is a frequently used context in the acoustic study of vowels (e.g. Cox, 2006; Cox & Palethorpe, 2005; de Jong, McDougall, Hudson & Nolan, 2007; Ferragne & Pellegrino, 2010; Hillenbrand et al., 1995). The rationale for using an hVd context was to ensure easy identification of the

target vowel as well as to minimize the effects of co-articulatory features on vowel. Huang (1991, p. 19), for example, points out that:

> Previous acoustic studies have shown that consonantal context, lexical stress, and speech style exert a strong influence on vowel formant trajectories. Several investigators have found consonantal context, the consonant immediately before the vowel and the consonant immediately after the vowel, to affect the vowel formant frequencies systematically.

However, we acknowledge that an hVd context has its limitations as it contains unfamiliar words especially for second and foreign language learners which is why in cases where the subjects were unsure of how to pronounce a word, they were given a rhyming word or a supporting word (see Table 3.3 and 3.4) to help them to pronounce the target word correctly. For example, *beard* for *hered*.

The list of words containing English monophthongs and diphthongs which were used in the present study is shown in Tables 3.3 and 3.4.

Table 3.3: Word List for Monophthongs

Vowel	Word	Supporting Word	Vowel	Word	Supporting Word
/iː/	**heed**	**need**	/ɒ/	**hod**	**dog**
/ɪ/	**hid**	**sit**	/ɔː/	**horde**	**horse**
/e/	**head**	**bed**	/ʊ/	**hood**	**book**
/æ/	**had**	**bad**	/uː/	**who'd**	**soon**
/ɑː/	**hard**	**card**	/ɜː/	**heard**	**bird**
/ʌ/	**hud**	**but**			

Table 3.4: Word List for Diphthongs:

Vowel	Word	Supporting Word	Vowel	Word	Supporting Word
/ɪə/	**hered**	**beard**	/eɪ/	**hayed**	**pay**
/əʊ/	**hoed**	**go**	/aʊ/	**how'd**	**cloud**
/aɪ/	**hide**	**ride**	/eə/	**haired**	**beared**
/ɔɪ/	**hoyed**	**boy**	/ʊə/	**hoar**	**poor**

The subjects were presented with the carrier sentences containing the words with the target vowels in random order. The first and last sentence in the list of sentence were not used in the study (*happy*/*house*) but instead acted as fillers to counter for "beginning- and end-of lists effects" in reading" (Hawkins & Midgley, 2005, p. 108) such as a sudden drop or rise in pitch or speaking rate. Carrier frames rather than isolated words are commonly used by researchers studying target sounds (Pillai, Mohd. Don, Knowles & Tang, 2010). The recordings yielded two sets of 19 tokens of vowels (eleven English monophthongs and eight English diphthongs) per speaker giving a total of 494 tokens of vowels for analysis.

(ii) Informal Speech

For the informal speaking context, the subjects were asked to talk about their future plans for about five minutes. The recording of the informal speech began with the subjects being asked to talk about themselves as a warm up and then to start talking about their plans for the future in English. Informal speech was obtained with the aim of examining the vowels in a natural setting as those produced in a word list context may have different qualities due a more careful pronunciation of the words. However, as anticipated, not all the target vowels were produced and the numbers were also not constant, especially as only vowels in lexical words in stressed syllables that were not preceded or followed by approximants and nasals were used. This was done as these sounds are likely to have coarticulatory effects on the vowels (Deterding, 1997; Pillai,

Mohd. Don, Knowles & Tang, 2010). For example, all the subjects pronounced the post vocalic /r/ in the word *university*, and thus, the vowel preceding /r/ was not considered in this study.

Informal speech was also used because read speech can yield different results due to a more careful pronunciation of words. However, the spontaneous nature of the speech also means that the quality and duration of the vowels will tend to vary according to its phonological environment (Leemann, 2007). The frequency of vowels extracted in this context is shown in Table 3.5.

Table 3.5: Frequency of Vowels in Informal Speech Context

Monophthongs	**Female Frequency**	**Male Frequency**
iː	26	23
ɪ	24	25
e	25	16
æ	39	31
ɑː	5	4
ʌ	35	20
ɒ	24	10
ɔː	5	1
ʊ	19	8
uː	11	14
ɜː	7	3
Diphthongs	**Female Frequency**	**Male Frequency**
ɪə	3	1
əʊ	42	26
aɪ	15	11
ɔɪ	1	3
eɪ	34	18
aʊ	13	14
eə	1	5
ʊə	2	3

The subjects talked about their plans for the future, like: where they planned to study and visit; what they wanted to buy in the future; and what they wanted to own. Thus, words like *study*, *people*, *good*, *go*, *house* were frequently produced. As can be seen in

Table 3.5, the most frequently occurring vowels were: /iː/, /ɪ/, /e/, and /eɪ/ while, vowels like /ɜː/, /eə/, /ʊə/ occurred less frequently consistent with lists of frequently occurring English vowels (Cruttenden, 1994; Knowles, 1987). According to Knowles (1987, p. 224), "In any text, we can expect an inflated frequency of the phonemes of the key words". For instance, the occurrence of phoneme frequency for /ɜː/, /eə/ and /ʊə/ is (0.62%), (0.31%) and (0.04%) while for /iː/, /ɪ/, /e/ and /eɪ/ is (1.80%), (8.26%), (2.57%) and (1.54%). Fry (1947), cited in Cruttenden, 1994, p. 136) also presented similar results: (0.52%) for /ɜː/, (0.34%) for /eə/ and (0.06%) for /ʊə/ and for /iː/, /ɪ/, /e/ and /eɪ/ is (1.65%), (8.33%), (2.97%) and (1.71%). The list of words extracted from the informal speech context is shown in Appendix A.

3.3 Procedure

All the recordings were carried out in a quiet section of the library at the subjects' school which is an Iranian school located in Kuala Lumpur. Written permission was sought from the school authorities by way of letters of support for the study from the Faculty of Languages and Linguistics and from the Iranian Embassy. The subjects were asked to fill a questionnaire to obtain information about their age, how long they have been in Malaysia, their first language, how long they have studied English, and their English proficiency level (later verified by their English teacher) (see Appendix B). The subjects were briefed about the study and the nature of their participation and were assured that their identity would be kept confidential, and that their participation was purely voluntary.

Before the actual recording the speakers were given five minutes to read through the word list on their own and were told to read in a natural manner. This was to minimize

any anxiety among the speakers and to warm up their voice. Then the speakers were placed in a quiet area in the school library to record the data. They were asked to read the word list once and followed by a short break to minimize fatigue and subsequent effects on their voice quality. They then read the word list for the second time and were recorded once again. After another short break the speakers were asked to talk about their plans for future. For the word list the recording took approximately 20 minutes per subject while recording of the informal speech took about five minutes. At the end of the recording, the students were given a token of appreciation for their cooperation.

3.3.1 Instrumentation

The recordings were carried out using the Marantz PMD661 Professional Solid State Recorder and an Audio Technica ATM73a cardiod condenser head worn microphones placed at the recommended 3 cm from the speakers' mouth (Plichta, 2010). A sampling rate of 44.1 kHz at a 16-bit rate was used to ensure high quality recording that will be suitable for acoustic analysis.

3.3.2 Data Transcription and Analysis

The data were transcribed and annotated using Praat Version 5.1.32 (Boersma and Weenink, 2010). In the first pass, the data was orthographically transcribed using the TextGrid function of Praat. Then, the target vowels were isolated and measured for all the speaking contexts.

Visual inspection of the waveforms and spectrograms together with auditory examination of the data were used to determine and measure the first and second formant and the duration of the monophthong vowels and also first and second formant for diphthong vowels in the target words. The analysis of the data was based on the

formant frequency model. However, for the informal speech, the orthographic transcription was done manually and the vowels to be measured were identified, avoiding vowels, which precede and follow approximants and nasals, as there may be coarticulatory effects on the vowel quality. All the measurements were annotated in subsequent tiers created in the TextGrid (see Figure 3.1).

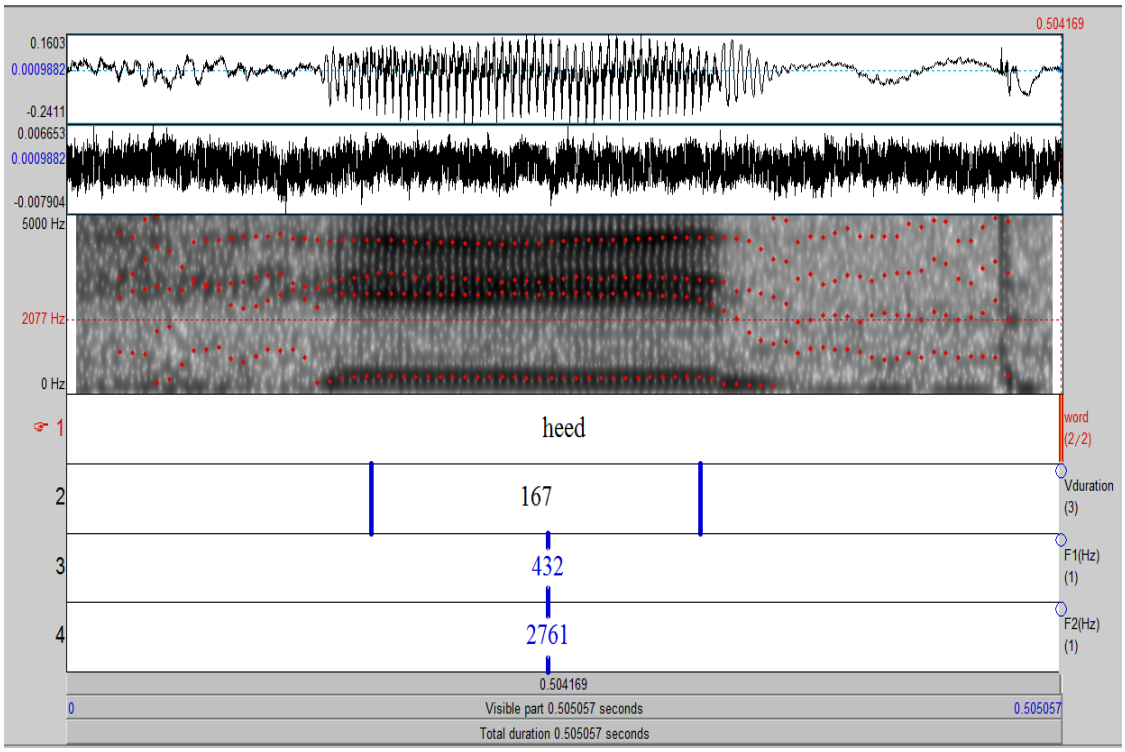

Figure 3.1: Screenshot of the word *heed*

As mentioned in chapter two, the first two frequencies of vowels have been found to be the most important cues for vowel perception: F1 for vowel high/low and F2 for front/back dimension (Fry, 1979; Kent and Read, 2002; Watt & Tillotson, 2001). For vowel length, the duration of the vowel was measured from the onset to the offset of the vowel. For the hVd context, the onset of the vowel determined from the onset of voicing for the vowel while the vowel offset preceded the absence of acoustic signal for the following stop consonant *d* (see Figure 3.1).

After an auditory examination of the target words and a visual examination of the formant tracks, the F1 and F2 frequencies were measured using the automatic linear predictive coding (LPC) tracker overlaid on a wide-band spectrogram in Praat. For monophthong vowels, the midpoint of the vowel where the vowel is at its most steady state were measured (see Adank, Smits and Hout, 2004; Hawkins and Midgley, 2005; Ladefoged, 2003; Watt and Tillotson, 2001). The midpoint of the vowel is the point which a vowel is considered to be at its most steady state and is least influenced by preceding and following sounds (in this study the fricative /h/ and the stop consonant /d/). Based on Deterding's study (1997), the distance of each vowel was calculated from the centroid using the Euclidean distance (ED) between the F1 and F2 of each vowel according to the average F1 and F2 values of all vowels. In cases where the computer generated formants in Praat were not clear, especially for the back vowels measurements were carried out manually.

The F1 and F2 measurements for the monophthongs were annotated into the second and third tiers, which were created as point tiers in the TextGrid. The measurements were cross-checked by another rater. A statistically significant positive degree of agreement was found for the measurements between Rater 1 and Rater 2 for the monophthongs in WLC and ISC (F1 WLC: Pearson's r (284) = 0.99, $p < 0.001$; F2 WLC: $r(284) = 0.99$, $p < 0.001$; F1 ISC: $r(373) = 0.99$, $p < 0.001$; F2 ISC: $r(373) = 0.99$, $p < 0.001$). Any discrepancies in the measurements were corrected upon checking by the second rater. The measurements of the vowel durations were also annotated in an interval tier in the TextGrid. There was a significant positive correlation between the measurements of the duration of the monophthong vowels in both speaking contexts between the two raters (WLC: $r(284) = 0.99$, $p < 0.001$; ISC: $r(373) = 0.99$, $p < 0.001$).

This enabled the sound file and spectrogram to be accessed simultaneously with the relevant information (see Figure 3.1). The measurements were also entered into a spread sheet to enable means and standard deviations to be calculated, and for scatter plots and vowel charts to be generated, as well as to enable comparisons with vowels in other varieties of English, where relevant. Once the measurements were taken and checked for accuracy, the average F1 and F2 values for the monophthongs were converted into Bark scale to enable the vowels to be plotted on a F1 vs. F2 vowel chart (Zwicker & Terhardt, 1980). The Bark scale was used because it "is thought to be a good approximation of the actual frequency analysis performed by the ear" (Kent and Read, 2002, p. 115). Zwicker and Terhardst's (1980, cited in Deterding 1997, p. 50) formula is as follows:

$$Z = 13\text{Arctan}\,(0.00076F) + 3.5\text{Arctan}\,(F/7500)^2$$

For diphthongs, the F1 frequency of the vowel was measured at the beginning and the end of the vowel. The duration of the diphthong was also measured in seconds. All these measurements were entered into the TextGrid. These measurements were used to determine the *rate of change* (ROC). As discussed in Chapter 2, to measure ROC, the difference in F1 is calculated and then divided by the duration in seconds. Thus, for measuring diphthongs the following formula was used (Deterding, 2000, p. 94-95):

$$F1_{end} - F1_{start}/\text{Duration (seconds)} = \text{ROC (Hz/second)}$$

The ROC can be expected to be a negative value for closing diphthongs like /eɪ/ and /əʊ/ since the F1 decreases toward the offset due to the trajectory towards a high vowel. Figure 3.2 shows the annotations for the measurements for diphthongs. There were significant positive correlations in the ROC measurements between the two raters in

both speaking contexts (F1 ROC WLC: $r(206) = 0.96$, $p < 0.001$; F2 ROC WLC: $r(206) = 0.99$, $p < 0.001$; F1 ROC ISC: $r(190) = 0.91$, $p < 0.001$).

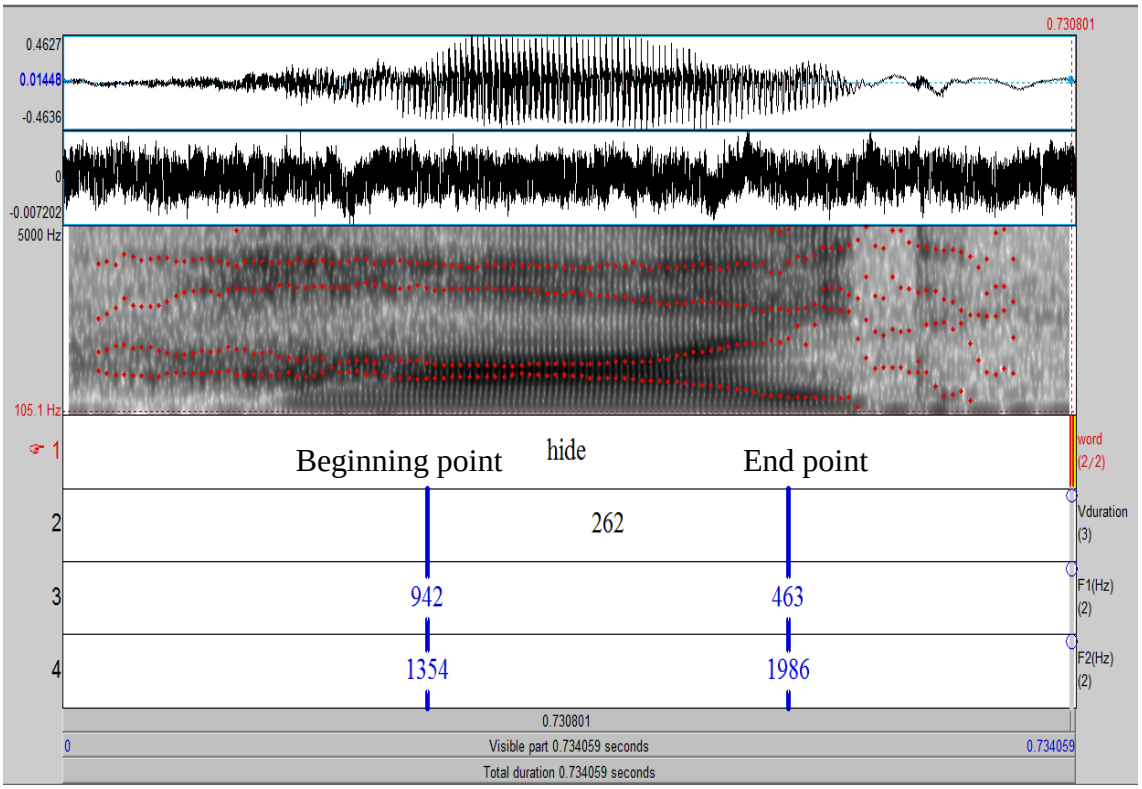

Figure 3.2: Spectrogram of a Persian Female Saying *hide*

The data was then compared with published findings of similar vowels in British English (Deterding, 1997; 2000; 2003) and American vowels (Hillenbrand et al., 1995). The comparison was aimed at determining the extent to which the vowels produced by the subjects, who have been learning English as a Foreign Language, compare to vowels produced by native speakers in terms of vowel distribution, vowel quality and vowel contrast.

3.4 Conclusion

This chapter presented the methodology used to examine the English monophthongs and diphthongs produced by a group of Iranian speakers. The following chapter will present and discuss the results of the study.