Chapter 3

Methodology

3.1 Subjects

A total of six male subjects were invited to participate in the study. The main reason for using the data from only male subjects was that the data produced by male speakers have clearer formant structure than that of female speakers. It is found that the longer the vocal tract, the smaller separations between the resonance frequencies that together make up the formants (Kent & Read, 1992). Given that generally the length of vocal tract is longer in male speaker than in female speakers because vocal the length of vocal tract correlates positively with the body height, speech sounds produced by male speakers would therefore have concentrate resonance frequencies in formants and consequently a clear formant structure. Clear formant structure facilitates the identification of formants and the measuring of formant frequencies. Besides, it is also because the easy identification of formants that many studies in the acoustic literature (Howie, 1976; Y. Chang, 1998; Catford, 2001) used only data from male subjects; thus, this study also used data from male subjects for the ease of cross-linguistic comparison. The subjects were paid for participating in the recording[5]. The subjects age from nineteen to twenty-two years old. All of them are collage students but none of them is language-related or linguistic majors. Their birthplaces and living places are within the Taipei city; three of the subjects live in the Wenshan district, two of them in the Shilin district, and one in the Zhongshan district. As for the

[5] Thank Prof. Wan for providing funds for the subjects of this study.

language background of the subjects, all of the subjects are fluent in using Mandarin. Most of them can speak Taiwanese Southern Min and English; only one subject can speak neither Taiwanese Southern Min nor English. All of the subjects have learned some English in school but most are considered not fluent by self-evaluation[6].

3.2 Materials

3.2.1 Design of testing words

There were 70 test items in total in this study[7]. The testing words were produced in two forms. One is in the citation form, in which the testing words were produced in individual syllables with several seconds of interval between syllables. The other is in the sentence form, in which the testing words were produced in the short sentence *wo shuo* ___ “I say ___” with intervals between the short sentences. The motivation for the design to put the testing words was to reduce the influence from other syllables. In addition, three tokens for each testing words were required. Among the testing words, there are syllables with initial consonants (IC) and the corresponding syllables without initial consonants (if possible). Consequently, there were totally 420 tokens for each subject and 2520 tokens for the whole study. Almost all of the testing words have the tone value (51). Exceptions are the testing item *en* 恩 [ən] ‘favor’, which have the tone value (55), and two heteronyms such as *ou* 漚 [ow], and *a* 阿 [a], which can have both the tone value (51) and (55) with tone value (55) being the frequent one in ordinary speech. The rationale behind the selection of the tone value is simply because in syllables with the tone value (51), there are more possible syllables that include a greater variety of combinations of segments in the syllable final[8] and

6 For the detailed data and language background of the subjects, please see Appendix A.
7 Please see Appendix B for the full chart of testing items.
8 For the combinations of segments in the syllable final of the testing items without and with IC, please

could also be written down in Chinese characters[9]. The interaction between tones and vowels is beyond the scope of this study. Studies such as Howie (1976), Tsao & Yang (1984), and Hoole & Hu (2004) have already devoted on this topic. The results of these study generally found that tones do have influences on the vowel quality. For example, Howie (1976) found syllables with high tones have higher vowel quality; Hoole & Hu (2004) found tones with the tone value (214) showed both articulatory and head-positions differences from syllables with other tone values.

3.2.2 Equipments

3.2.2.1 Stimuli

Testing word cards were used as the stimuli for this study. The testing word cards were made of papers in 6cm × 11cm size. Characters of the testing words and carrier sentences that contain the testing words sized 72pts were printed in black New Ximing Font in the center of word cards. Since there were five practice cards for the citation section and five for the sentence section and 420 tokens for each subject, a total of 430 word cards were employed in this study.

3.2.2.2 Recording apparatus

The recording apparatus used in the present study was KAY Electronics CSL (Computerised Speech Lab) Model 4100 in the phonetic and psycholinguistic laboratory of the Graduate Institute of Linguistics, National Chengchi University. The advantage of using KAY CSL 4100 to record the speech sound is in that it could convert the analog signals into digital signals with little distortion. This advantage makes it more convenient to use the computer to edit, store, and analyze the speech

---

see Appendix C and Appendix D for reference.

[9] This generalization was based on the observation of Prof. I-Ping Wan (personal communication).

sounds, which is a kind of analog signals. The microphone was placed near the subjects' mouth in a distance that they felt comfortable. The whole recording section was recorded in PCM, 11.025 kHz, 16 bits, monaural WAV sound files.

3.2.2.3 Acoustic analysis apparatus

KAY Electronics CSL (Computer Speech Lab) Model 4100 was also utilized to analyze the recorded files. Besides its convenience in edit, store, and analyze, the reason for selecting KAY as the acoustic analysis apparatus of this study is that KAY has been used in numerous studies of acoustic analysis in the fields of acoustics, audiology, speech pathology, and acoustic phonetics.

3.2.2.4 Statistic software and vowel space plotting software

The statistic data will be processed with the *SPSS 11.0* (Statistics Package for Social Science, version 11.0) released by SPSS Inc. *SPSS* is one of the most widely used programs for statistical analysis in social sciences. The analysis method used in this study was one-way ANOVA (analysis of variance), which compares the means of two groups of numbers, and the post hoc method of Bonferroni, which compares the means of several groups of numbers.

Most of the vowel spaces or formant plots were plotted with the freeware *PlotFormant 4.0*, developed and distributed by SCICONRD.com based on the program written by Dr. Peter Ladefoged at UCLA phonetic lab. Its strength is in that it is a professional but handy software. It helps the user to create a professional formant plot in just one or two steps.

Some of the figures were created with *Microcal Origin 8.0*, a software that allows the users to customize their graph or figure. With its flexibility in the design of the graph, it can meet the needs for creating graph or figure in this study.

3.3 Procedures

3.3.1 Recording

The recordings were held in the phonetic and psycholinguistic laboratory of the Graduate Institute of Linguistics, National Chengchi University. Before the recording begins, there were a section of personal data filling[10] and an instruction section. During the instruction section, the subjects were informed of the requirements and things to be noticed in the recording. The subjects were asked to speak the sentences on the word cards in ordinary pace loudly and clearly. They were informed of the possible repetition of testing words.

After the instruction section, the recoding was further divided into two sections: a citation form section and a sentence form section. Before each section, five practice word cards in the form of that section were given to simulate the process in the recording. During each section, three piles of word cards were given since three tokens for each testing word were required and consequently three rounds of the reading of word cards were needed. The word cards were presented to the subjects one by one by the experimenter rather than handled by the subjects. This is to ensure that the reading speed was roughly controlled at similar pace. Otherwise, the reading speed might vary in a great degree if the word cards were handled by the subjects because subjects might have different habitual speed in reading. Besides, the subjects were asked to read the testing words in the speed of ordinary conversation. The order

[10] For the personal data form, please see Appendix E.

of word cards was randomized in each round with the purpose to prevent any possible patterns created by the order of the testing words. The random list of numbers was created by a small java program.

There were short breaks between sections for the subjects to take a rest and for the experimenter to set up the recording instrument. The whole recording section for each subject, including the breaks, lasts about half an hour in general.

3.3.2 Acoustic analysis

In this study, the first two formants (F1 and F2) of each main vowel in the testing words are measured. The formant frequencies were attained with the help of the spectrogram function and the formant history function of KAY CSL 4100. The formant history function located the formants and indicated the position of formants with series of dots in the spectrogram (as the small white dots in Figure 3.1). Besides, once the point of measuring was selected by the researcher with curser, the frequency of the point of measuring was displayed in the bottom of the window.

The following paragraphs of this section show the point of measuring of vowel quality in monophthongs, diphthongs, and triphthongs. To begin with, in the case of monophthongs, the formant frequencies in the center of the steady section of the first and second formant were attained as the vowel quality.

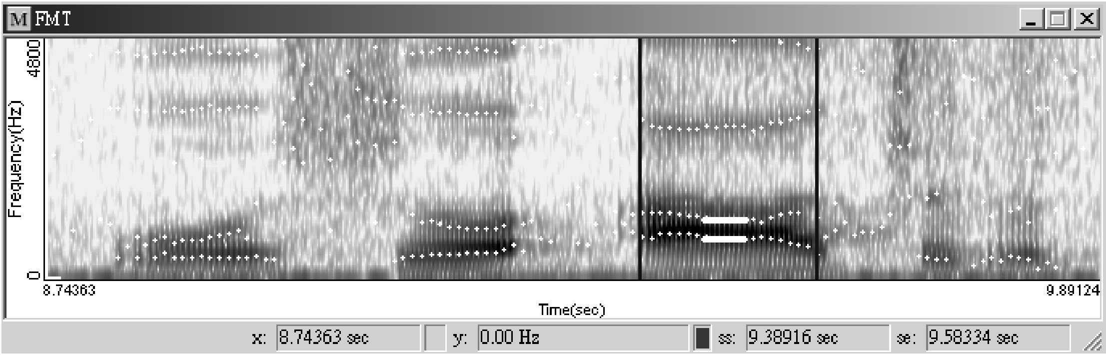

Figure 3.1 The measuring of [a]

The measuring of the F1 and F2 of the monophthong [a] is illustrated in Figure 3.1. The two bold vertical lines indicate the syllable boundary of the testing item. The white bars at the center of formants indicate the position of measuring. The unsteady sections in the beginning and near the end of the syllable were ignored.

Unlike the measurement of monophthongs, in the case of diphthongs, both the values at the steady section near the starting and ending point of the vocalic section were recorded for the two components respectively; however, only the vowel quality of the main vowel was reported in this study. Figure 3.2 and 3.3 demonstrates the measuring of the diphthongs [ja] and [aj].

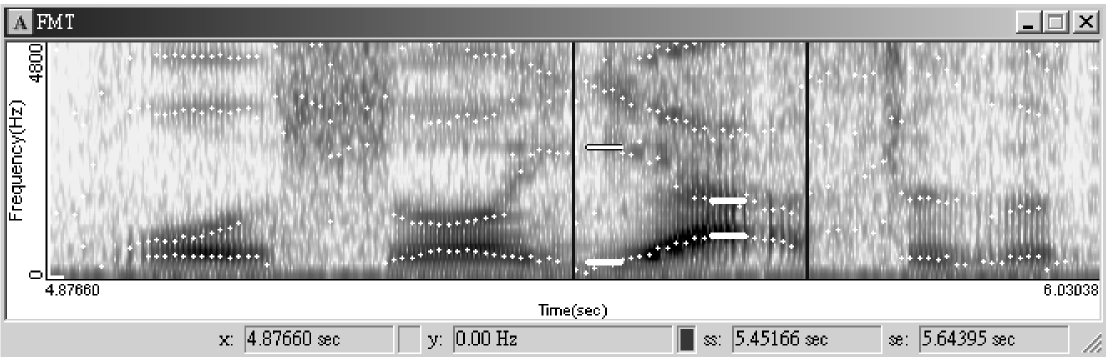

Figure 3.2 The measuring of [ja]

The strong energy section is represented with darker area in the spectrogram while the weak energy section is represented with lighter area as the center of the syllable and the beginning of the syllable in Figure 3.2 has shown. Since the main vowel in the syllable should be the section that has the strongest energy in the syllable, the darkest section with comparatively steady formants was measured for the vowel quality for the main vowel as Figure 3.2 showed. The dots in the spectrogram provided by the formant history function of KAY CSL 4100 facilitate the identification of the formants. The formants at the beginning of the syllable with weaker energy were measured for the formant frequencies of glide [j].

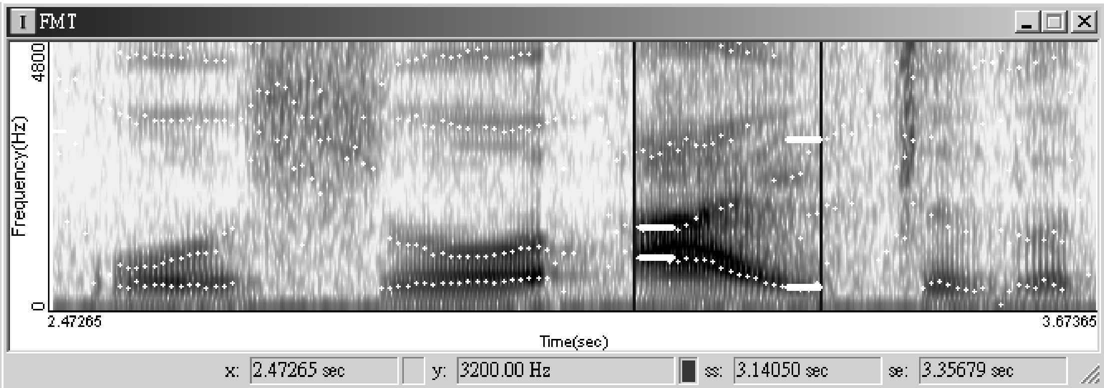

Figure 3.3 The measuring of [aj]

Similarly, in the measuring of the vowel quality in the syllable [aj] in Figure 3.3, the darkest section with comparatively steady formants in the beginning of the syllable was measured for the vowel quality of [a]. And in this case, the formants at the end of the syllable with weaker energy were measured for the formant frequencies of glide [j].

Finally, in the case of triphthongs, in addition to the F1 and F2 values at the starting and ending point of the vocalic section, the F1 and F2 values in the turning point of F1 were measured. The rationale behind this method is based on the characteristic of Mandarin co-occurrence restriction that it only allows a high-low-high sequence of vocalic sounds, which has been accounted for extensively

(Lin (1989), Wiese (1997)). What is more, low vowels are more prominent acoustically and are thus the main vowels. As a consequence, the time point at the turning point of F1 represents the vowel quality of the main vowel. Figure 3.4 illustrates the measuring of the formant frequency values in the triphthongs [waj].

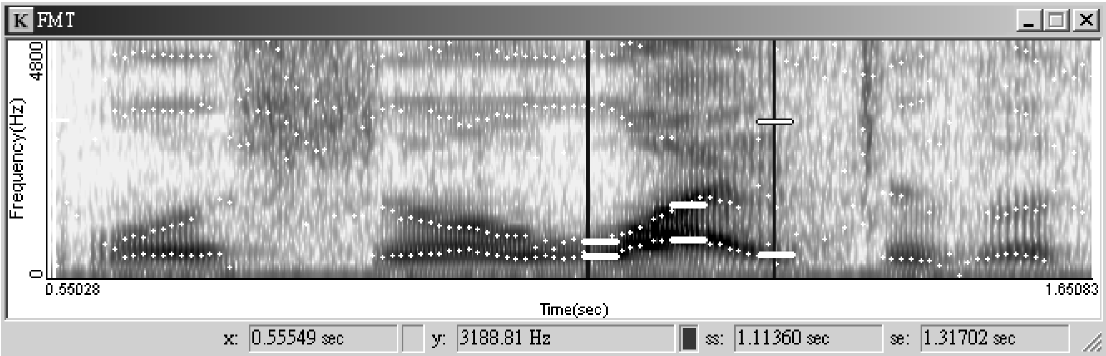

Figure 3.4 the measuring of [waj]

Figure 3.4 exemplifies the measuring of the vowel quality of [a] in the syllable [waj]. Based on the rationale explained above, since there is no clue for locating the vowel [a] in the second formant, the first formant of the vowel must be located first. The following steps demonstrates the procedures of measuring the vowel quality of [a] in the triphthong [waj] in Figure 3.4. Firstly, the highest point of the F1 in the syllable is located as the vowel [a]. Next, the frequency in this point is measured as the F1 of the vowel [a]. Then, the frequency of the second formant at the same time point is measured as the F2 of the vowel [a].

3.3.3 Statistical treatments

The mean value of the F1 and F2 of each vowel were calculated. Standard deviations among the values were also computed. The value of standard deviation shows the difference among the values. The smaller the standard deviation, the more reliable the mean value is. The one-way ANOVA was used to compare the mean formant frequency values between different vowel groups.

Further detail comparison of the means between groups more than two were performed with the post hoc method of Bonferroni. The statistical analysis was processed with SPSS 11.0. The confidence interval used was set at .95 in this study. In other words, the significant level of p-value in one-way ANOVA was set at .05, which means the chance level of refuting the null hypothesis of having no difference between two mean values is above 95%.