Representations Of Mandarin Syllables
And
Computation Of Their Frequency Distributions

Wei Rosa Lee

A Thesis
in
The Department
of
Computer Science

Presented in Partial Fulfillment of the Requirements
for the degree of Master of Computer Science at
Concordia University
Montreal, Quebec, Canada

April 1984

ABSTRACT

Representations Of Mandarin Syllables

And

Computation Of Their Frequency Distributions

Wei Rosa Lee

Mandarin syllables represented by three phonetic systems: Pinyin, Chan's and Suen's; have been studied and analyzed. A segmentation algorithm has been developed and implemented to decompose Pinyin words into syllables and convert them into Chan's Chinese Phonetic Codes and Suen's phonetic symbols. This algorithm has been tested with a large database consisting of 24,271 spoken words. The frequency distributions of syllables and phonetic symbols have been tabulated. N-gram statistics of Mandarin syllables have also been analyzed and presented.

ACKNOWLEDGEMENTS

My first acknowledgement is to my thesis advisor, Professor C. Y. Suen, to whom I owe my first interest in the subject, and who has given me constructive advice and comments, moral support and encouragement, throughout the course of this research. I also thank Dr. E. Regener for his comments and suggestions.

I am greatly indebted to Mr. Peter K. L. Chan, for his generous financial support and helpful personal discussions, for this project.

Finally, but no means least, I must express a word of appreciations to my parents and brothers, as well as to my family, Eric and Augustin, whose encouragement has made it possible for me to complete this thesis.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

Chapter I

Introduction

Since written Chinese characters are very complicated to be processed by computer, speech can be an important mode of man-machine communication. In the past ten years, many researchers have tried to analyze and synthesize the Chinese official language, Mandarin, with the aid of a computer. A good representation of Mandarin in input/output of computer processing can be an important tool for research in this area. In this thesis, a segmentation algorithm has been developed and implemented to decompose Pinyin words into syllables and convert them into Chan's Chinese Phonetic Codes and Suen's phonetic symbols. The properties of the three phonetic systems: Pinyin, Chan's, and Suen's; are analyzed and compared. The algorithm has been tested with three sets of data consisting of 24,271 spoken words. Also, the frequency distributions and n-gram statistics of syllables and phonemes have been computed. It is hoped that the results will be useful to computer input/output of spoken and written Mandarin.

1.1 Speech Analysis

The most common mode of human communication is speech. Human speech is produced through the specialized movements of vocal organs, such as vocal cords, lips, etc. Once produced, the speech signal might be received by the listener's ears and understood. A listener does not only understand the meaning of speech, but also can identify the speaker's accent without seeing the speaker.

In man-machine communication by voice, a lot of research and studies are in progress with the aim to develop systems, that can speak, hear, identify and understand human voice. There are two major categories of man-machine communication systems[12]. The first one is from man to machine. In this system, it is composed of two parts[19]; (1) Speaker recognition, which is to verify a speaker identity or to identify the speaker from some known ensemble. (2) Speech recognition and understanding, which is to recognize and understand the entire spoken utterance. To recognize, verify and understand speech signal, a set of pre-determined references from the analysis stage must be stored in the computer first. In the stage of analysis, some speech signal will be extracted and processed according to intensity, duration, pitch and intonation, as well as frequency distribution of the signal. Based on the analysis of the acoustic features described above, the phonetic features of the input speech signal are determined, and some quantized information is generated. All quantized

information is stored in computer memory as reference patterns, if it is in the learning mode. Otherwise, the recognition mode is assumed, all information will be matched with the pre-stored references in the memory. Once the sounds are matched, the identity of the speech sound can be sent to the next stage for further processing, such as speech understanding[18].

The second mode of communication is from machine to man, which is also called the "synthesis system". It is simple and economical to synthesize speech, by concatenation of phonemes. It requires a simulation of the vocal cords for sound generation, and a set of phonetic rules to interpret the input text into a phoneme string. To make synthesized speech sound natural, it is important to take into considerations the various characteristics of phonemes and their combinations of prosodic features, stress and intonation[2],[24]. For example, English pronunciation is based on spelling, but the letters may be pronounced differently in different cases, such as the letter 's' which is pronounced differently /was/ (as /z/), /sum/ (as /s/) or /sure/ (as /sh/) etc.

Therefore, speech analysis is a significant and fundamental ingredient of all the important technical problems of speech communication by machine. It is also involved in many aids to the handicapped such as adjusting speech speed for the blind and visual training for teaching

the deaf to speak. In addition, it can also enhance the quality of speech signals that have been degraded by noise, reverberation, or sound produced in an unusual atmosphere, for instance, driver's speech.

In general, speech analysis can be classified into four different areas as follows:

(A) detection and processing of the speech signal

(B) extraction of acoustic and phonetic features

(C) display and examination of the acoustic waveform

(D) text analysis and phonetic transcription.

In this thesis, the text analysis and phonetic transcription is considered in detail on Mandarin.

1.2 Chinese Language And Mandarin

Chinese, a hieroglyphic language, is very different from a Western language. The pronunciation of Western language is based on the alphabetic phonetic system. So the spoken word can be spelled accordingly. But this is not true in Chinese. The written part of Chinese, namely character, does not always relate to its pronunciation. And all the Chinese characters were invented according to the following criteria[26],[27]:

(1) the shapes of objects

(2) phonetic combinations

(3) logical combinations

(4) indicating to situation or indirect symbols

(5) mutually interpretative symbols

(6) false borrowing.

Chinese characters were unified about two thousand years ago. After many years of improvement and modification, all characters could be derived from about 220 basic radicals. But, there are more than fifty dialects in China, none of them have a complete phonetic system to indicate the pronunciation of characters before this century. After the 1911 revolution, although the Mandarin dialect has been chosen as the Chinese official language, yet the shapes of characters and their pronunciations are still independent of each other. Therefore, it is very difficult to establish phonetic rules between the character and its sound. Hence to implement a Mandarin speech system in computer, it would be easier to find a set of phonetic symbols as input/output, rather than using characters directly. In other words, a set of phonetic symbols is required by computer to perform speech processing. However, several phonetic systems have been developed in this century. They will be introduced in the following chapter.

1.3 Criteria For The Ideal Phonetic System

In order to have an ideal phonetic system to serve as input/output in speech processing by computer, there are several essential criteria that we should look into[8]:

(1) Consistency

The phonetic representations must be unique. There should not be any confusion in usage in the representations.

(2) Easy recognition

The representations should be simple enough to be recognized not only by human beings, but also by the computer.

(3) Accuracy

Every phonetic representation can be accurately and easily pronounced by human beings.

(4) Easy for computer input/output

From a computer implementation point of view, easy to enter through key-board and display from a computer are the essential criteria.

(5) Flexibility

The set of phonetic representations can be easily expanded to build a new syllable.

(6) Length of syllable

From the economic point of view, the shorter the average syllable length is, the less the cost is in the printing process.

CHAPTER II

Characteristics Of Mandarin And Its Phonetic Systems

2.1 Characteristics Of Mandarin

In order to set up a system which can process Mandarin speech directly, a detailed study of the following characteristics of Mandarin is necessary[7],[16].

(A) The sounds of all Chinese characters are monosyllables. Each Chinese character can only be pronounced by one syllable.

(B) Most consonants appear at the beginning of the syllable except the nasals 'n' and 'ng'.

(C) Chinese is a tonal language. Every syllable is associated with a tone. In Mandarin, there are four tones due to the shift of the fundamental frequency of vowel or diphthong. A neutral tone is used while the syllable is unstressed. Those unstressed syllables appear mostly at the end of a word or a sentence. The tone specifies the pitch contour of the syllable[22]:

Tone/Notation		Description	Pitch
1	—	high-level	55
2	ˊ	high-rising	35
3	ˇ	low-dipping	214
4	ˋ	high-falling	51
5	·	neutral	5

For example, the syllable /m-a/ in tone one, means mother. It has a high level but flat tone; the same syllable, in tone two, means hemp, it has a high level and rises at the end; in tone three, it means horse, the tone starts at a low level, and rises at the end of the syllable; in tone four, it means scold, the tone starts at a high level, then goes down to a low level rapidly. The neutral tone of this syllable indicates the ending of a question, the syllable will be pronounced very short without stress.

(D) Chinese words may be represented by more than one character.

A meaningful Chinese word may contain more than one character. In other words, a word may be pronounced by more than one syllable. And the different combinations and permutations of characters give different meanings. For example the character (樓) means building, is spelled with /l-oŭ/ in the second tone. Another character (上) means up, is spelled with /sh-a-ng/ in

the fourth tone. The combination of these two characters can be (樓上) and (上樓), the first word, which is a noun, means upstairs, and the second word, which is a verb, means to go upstairs.

(E) Mandarin speech can be defined as the following:

(1) <CLAUSE> ::= <WORD> {<WORD>} <punctuation>

(2) <WORD> ::= <SYLLABLE> {<SYLLABLE>}

(3) <SYLLABLE> ::= <INITIAL> <FINAL> <TONE>

(4) <INITIAL> ::= <CONSONANT> | <empty>

(5) <FINAL> ::= {<SEMI-VOWEL>} <VOWEL> {<VOWEL>} {<NASAL>}
| {<SEMI-VOWEL>} <DIPHTHONG>

(6) <TONE> ::= <1> | <2> | <3> | <4> | <NEUTRAL>

2.2 Mandarin Phonetic Systems

So far many phonetic systems have been invented. The corresponding phonetic representations used in the main systems are shown in Table 2.1.

(A) JIFH(1918)[7],[14]

This is the first phonetic system of Mandarin. It consists of a set of thirty seven phonetic symbols to represent Mandarin sounds. Any of the syllables can be

spelled by one, two or three JIFH symbols with a tonal mark.

In this system, each Mandarin sound has a unique representation. Yet, the symbols are difficult to familiarize with, and not easy to enter through an ordinary key-board.

(B) IPA and Yale systems[10]

IPA converts those thirty seven JIFH phonetic symbols into International Phonetic Letters. Yale system converts them into an English alphabet. These two systems try to use Latin transcriptions to help a person in Mandarin pronunciation. They are easy to enter through an ordinary key-board, but the representations of both systems are not unique. The same representation may be pronounced differently, for example, in Yale, the letter [j] represents two different consonants.

(C) Chan system(1933)[3],[4]

Mr. Shui Ki Chan(陳瑞棋) of Hong Kong invented two sets of symbols, called Chinese Phonetic Characters. Recently, it has been modified by his son Mr. Peter K. L. Chan(陳經綸). The symbols used in this system are different from JIFH. One set is called 'Initial', which contains twenty five symbols. Among them, twenty

one symbols represent all the consonant starting initials, two symbols represent two semi-vowels starting initials and two symbols represent eight vowels starting initials. The other set is called 'Final', which contains thirty four symbols, all of them are the combination or permutation of vowels, diphthongs semi-vowels or nasal consonants that could happen in Mandarin. The tone is indicated by the position of the final symbol.

In this system, not only the representation of each syllable is unique, but also the length is fixed and short. However, it takes time to learn to recognize the symbols used and the structure of the representation complicates the process of recognition or entering into a computer. Also, the phonetic characters are not general enough to represent all Mandarin syllables, for instance the syllable /eh/(诶) cannot be represented by this system[17]. Although the occurrences of this syllable are very low, yet it exists.

(D) Pinyin system(1957)[20]

The Pinyin system was developed in China, it made use of the Latin alphabet to present the Mandarin sound. Unlike IPA and Yale systems, this system tries to represent all the phonemes by the English alphabet, but they may not be pronounced the same way as they occur in

a Western language. As a result, some pronunciations are totally different from Western languages, or do not occur in some Western languages, such as [c], [q], [x], [z], [zh], etc.

The majority of the representations are unique in this system[7], but there are some exceptions, such as the letter 'u', which represents either vowel 'u:' or 'oo(u)' depending on the starting consonant. More examples can be found in [23]. These exceptions have to be interpreted by rules. And the spelling is easy to enter through an ordinary keyboard except the tonal mark and one German letter um-laut [ü]. Nevertheless, the spelling of this system is according to the word, not the character, so that syllable segmentation of each word is required.

(E) Suen system(1979)[22]

Professor C. Y. Suen of our university developed this system to provide a unique transcription of Mandarin phonetic symbols for input/output of computer. All of the symbols can be represented by the English alphabet.

This system is different from Pinyin. The pronunciations of phonetic symbols are closer to other Latin languages. The representations of this system are unique, and also easy to enter through an ordinary

key-board. This system has been modified further and perfected by Dr. Suen himself recently[25].

2.3 Review Of Studies On Mandarin Speech Analysis

Several studies on Mandarin speech analysis have been made in the past ten years.

(A) In 1973, T. Y. Chou and K. C. Huang[8] studied the vowels and consonants of the Mandarin according to JIFH. They also classified the thirty-seven JIFH symbols into different classes based on the acoustic parameters. The synthesis of several vowels has been tried using an EAI hybrid computer.

(B) In 1973, K. P. Li[15], analyzed Mandarin speech according to its monosyllabic structure. Each Mandarin syllable was decomposed into three parts: Initial, Tone and Final. Each part in turn was classified into several elements. Elements in the Initial and Final were classified according to the phonemic feature of Mandarin. He also provided a list of phonemes which were represented by International Phonetic Symbols that could happen in each part. He claimed that such a syllable constraint might help Mandarin recognition system at the syllable level.

(C) John M. Howie[13] classified Mandarin syllables into nine different types, according to the combination of

the features of the phonemes involved. He performed measurements on the four distinctive tones of those types of syllables. The results showed that the different combinations of phonemes did affect the pitch pattern of the tone. For example, a syllabic vowel has pitch pattern of the tone different from a syllable with an initial voiced consonant or non-syllabic vowel.

(D) In 1976, based on JIFH phonetic system, Professor C. Y. Suen of our university developed a Mandarin synthesizer on a PDP-10 computer[21]. In 1979, he provided a set of rules to convert Mandarin syllables into phonemic syllables based on thirty-eight basic phonemic symbols[22]. These symbols are represented by the English alphabet.

In addition, he made a study on Pinyin system and compared it with his own system by using four criteria for computer speech processing[23]. He also published some statistical analysis results on a large database composed of 753,000 Mandarin syllables[22].

Table 2.1 Phonetic symbols used in some systems.

	Suen	Pinyin	Chan	JIPH	IPA	Yale
1	b	b	ㄅ	ㄅ	b	b
2	p	p	ㄆ	ㄆ	p'	p
3	m	m	ㄇ	ㄇ	m	m
4	f	f	ㄈ	ㄈ	f	f
5	d	d	ㄉ	ㄉ	d	d
6	t	t	ㄊ	ㄊ	t'	t
7	n	n	ㄋ	ㄋ	n	n
8	l	l	ㄌ	ㄌ	l	l
9	g	g	ㄍ	ㄍ	g	g
10	k	k	ㄎ	ㄎ	k'	k
11	h	h	ㄏ	ㄏ	x	h
12	j	j	ㄐ	ㄐ	ʤ	j
13	ch	q	ㄑ	ㄑ	tɕ	ch
14	x	x	ㄒ	ㄒ	ɕ	s
15	rj	zh	ㄓ	ㄓ	ʤ	j
16	rc	ch	ㄔ	ㄔ	tʂ'	ch
17	sh	sh	ㄕ	ㄕ	ʂ	sh
18	r	r	ㄖ	ㄖ	ʐ	
19	ds	z	ㄗ	ㄗ	ʣ	dz
20	ts	c	ㄘ	ㄘ	ʦ'	ts
21	s	s	ㄙ	ㄙ	s	s
22	i	i	ㄧ	ㄧ	i	i
23	oo(u)	u	ㄨ	ㄨ	u	u
24	ü	u(u)	ㄩ	ㄩ	y	yu
25	a	a	ㄚ	ㄚ	ɑ	a
26	o	o	ㄛ	ㄛ	ɔ	o
27	u(uh)	e	ㄜ	ㄜ	ɤ	e
28	e(eh)	eh		ㄝ	ɛ	e
29	ai	ai	ㄞ	ㄞ	ai	ai
30	ei	ei	ㄟ	ㄟ	ei	ei
31	au	ao	ㄠ	ㄠ	au	au
32	ou	ou	ㄡ	ㄡ	ou	ou
33	an	an	ㄢ	ㄢ	an	an
34	un	en	ㄣ	ㄣ	ən	en
35	ang	ang	ㄤ	ㄤ	ɑŋ	ang
36	ung	eng	ㄥ	ㄥ	ʌŋ	eng
37	er	er	ㄦ	ㄦ	ɜ	er

Chapter III

Mandarin Syllable Structures In Different Phonetic Systems

In order to set up an ideal way of processing Mandarin by computer, it is necessary to analyse Mandarin in different aspects. In this study, Mandarin is analysed by computer in terms of Suen's phonetic representations and Chan's initial and final representations using input texts which are represented by Pinyin. Therefore, in this chapter, a detailed discussion of Mandarin syllable structures of these three phonetic systems is made. In the following discussions, the following notations are used:

[ ] - to denote Mandarin sounds in Pinyin,

( ) - to denote a Pinyin syllable or parts of it,

/ / - to denote a phoneme string,

' ' - to denote Mandarin sounds in Suen's system.

3.1 Suen's System

This system analyses Mandarin sounds according to the basic phoneme features by using the combinations of the twenty-four English letters, with the exception of letters "q" and "v", and the German letter um-laut "ü" to represent all the phonemes in Mandarin speech. There are twenty-two consonants, two semi-vowels, eight vowels and five

diphthongs in this system.

3.1.1 Mandarin Phoneme

Using Suen's phonetic system as the reference, Table 3.1 lists all the consonants according to the manner and place in which they are pronounced. Among them, the nasal consonant 'n' may appear at either the beginning or the end of a syllable, and 'ng' appears only at the end of the syllable. All other consonants are only used to initialize a syllable. But, none of the consonants can form a syllable alone. Therefore, if a consonant is involved in a syllable it must be associated with other phonemes such as a vowel, or diphthong.

The two semi-vowels defined in this system are 'w' and 'y'. When a syllable starts with a vowel in a high-tongue position, a semi-vowel instead of a consonant is used to initialize the pronunciation of this syllable. However, there are several cases, in which a semi-vowel may appear between a starting consonant and a vowel or a diphthong such as /hwo/, /hwai/, ...etc.. Similar to the consonants, neither one of these two semi-vowels appears at the end of any syllable nor can be pronounced independently.

Table 3.2 shows all the vowels according to the tongue positions. Vowel 'er' is an independent syllable, it will never follow or be followed by any other phonemes. All

other vowels can form syllables by themselves or be grouped together with any other phonemes, but not more than two vowels to form a Mandarin syllable. There are three vowels, 'oo(u)', 'e(eh)' and 'u(uh)', each of them can have two kinds of representations in this system. Anyone of these three vowels can be either the ending of a syllable or followed by a nasal consonant. If they are the ending of a syllable, the one in parentheses is used; otherwise the first one is used. For example, in the cases of /toong/ and /tu/, 'oo' and 'u' represent the same vowel; similarly, 'u' and 'uh' of /dung/ and /duh/ represent the same vowel. Therefore, the letter "u" in these cases has different meanings depending on the type of phonemes it is associated with in the same syllable. For instance, letter "u" represents different vowels in syllables /du/ and /dung/, which are quite obvious to English speakers.

In Suen's system, five diphthongs have been defined. Each of them is represented by two letters, they are 'ai', 'ei', 'au', 'ou' and 'iu'. None of them is followed by a nasal consonant, or a vowel. In other words, they are the endings of a syllable. So, if there are any other phonemes involved in the syllable, they must be in front of the diphthong.

The abbreviations, C for consonant, V for vowel, D for diphthong, NC for nasal consonant, and if there are two vowels in one syllable, V1 for the first vowel, V2 for the

second vowel, are used in the summary of the syllabic structure of Mandarin as follows:

(A) Starting with consonant

	Syllabic Structure	example
(1)	C+V	/ba/
(2)	C+V+NC	/bang/
(3)	C+D	/bau/
(4)	C+V+D	/biau/
(5)	C+V1+V2	/bieh/
(6)	C+SV+V	/bwo/
(7)	C+SV+V+NC	/bwan/
(8)	C+SV+D	/twei/

(B) Starting with semi-vowel

	Syllabic Structure	example
(1)	SV+V	/yi/
(2)	SV+D	/yai/
(3)	SV+V+NC	/yen/
(4)	SV+V1+V2	/yu:eh/
(5)	SV+V1+V2+NC	/yu:oong/

(C) Starting with vowel or diphthong

	Syllabic Structure	example
(1)	V	/a/
(2)	V+NC	/an/
(3)	D	/ai/

3.2 Chinese Phonetic Character Set(Chan's System)

As mentioned previously, the representations of Chan's system may not be easily recognized and accepted by people. However, it represents Mandarin syllables in a unique length of only two symbols. This uniqueness facilitates the computer to handle information processing and input/output. In this system, each Mandarin syllable is represented by a so-called Chinese Phonetic Character. And each character contains two parts. The first part is the initial symbol and the second part is the final symbol. Each final symbol is contained in a square box, and each initial symbol is contained in an oblong box which is about double the size of the final symbols.

3.2.1 Initials

There are twenty-five initial symbols. They have been classified into five groups according to the shape of the symbols. These symbols, and their corresponding code numbers used in this research are shown in the Table 3.3.

However, these initials can also be classified according to the phonemic features into three groups as follows:

(A) Vowel representations

Code I-01 and Code I-07 are the initials in this group. The Code I-01 represents the initial of the syllable which contains only vowel 'er'. And the Code I-07

represents the initial of a syllable which starts with either vowel or diphthong, for example /a/, /ai/, or /ang/, etc.

(B) Semi-vowel representations

Code I-11 and Code I-21 are the initials in this group. The Code I-11 represents the initial of a syllable which starts with semi-vowel 'w'. The Code I-21 represents the initial of a syllable which starts with semi-vowel 'y'.

(C) Consonant representation

All other twenty-one symbols represent twenty-one different consonant initials .

3.2.2 Finals

There are all together thirty-four final symbols. They are also classified according to the pattern of the symbols into nine groups by Chan as shown in Table 3.4.

These finals represent all the possible combinations of vowel, diphthong, semi-vowel or nasal consonant. However, all these finals in Mandarin can stand alone as a syllable without consonant initials. Although most of the phonemic structures of these syllables with consonant initial are different from those syllables without consonant initial, in this system, the symbols for finals with or without consonant initial have no difference at all in terms of representation. Therefore, the finals may represent

different phonemic strings depending on their respective initials. Table 3.4 shows the possible phonemic string of every final.

The representation of tone in this system requires no other special marks or symbol. It is indicated by the position of the final symbol related to the initial symbol of the syllable. The first and second tones are placed correspondingly on the right and left hand side of the upper part of the initial symbol, while the third and fourth tones are placed correspondingly on the right and left hand side of the lower part of the initial symbol. The neutral tone is indicated on top of the right hand side, but further apart from the initial symbol. Fig. 3.1 shows five different tones of the syllable /ma/ in Chinese Phonetic Characters.

Fig. 3.1 Chinese Phonetic Characters of syllable /ma/ in five different tones

3.3 Pinyin System

Pinyin is represented by twenty-five English lower-case letters "a" to "z" excluding letter "v", one German letter

um-laut "ü", and four different tonal marks for stressing. The four different tonal marks represent the four distinct tones of Mandarin denoted by ('-','ˊ','ˇ','ˋ'). If a syllable is stressed, there should be a tonal mark on top of the vowel or diphthong. No special tonal mark is used to indicate the neutral tone in Pinyin.

3.3.1 Representations Of Phonemes In Pinyin

As mentioned earlier, the non-unique use of letters in the Pinyin representation of Mandarin sounds entails the necessity of establing a large number of phonetic rules to translate Pinyin into phonemes. Again, with reference to Suen's phonetic symbols (sometimes also called phonemes, see Table 3.1), these rules are summarized below:

(A) Vowels

(1) 'a' is represented by the letter [a].

(2) 'er', if it appears at the end of the word without stress, it is represented by the letter [r]; otherwise, it is represented by [er].

(3) 'i' is represented by the letter [i].

(4) 'o' is represented by the letter [o].

(5) 'uh' is represented by the letter [e].

(6) 'e(eh)' is represented by the letter [a], if the nasal consonant 'n' follows; otherwise, this vowel is represented by letter [e]. For example, (bian) represents /bien/, (xie) represents /xieh/.

(7) 'oo(u)' is represented by the letter [o] if nasal consonant 'ng' follows; otherwise, letter [u] represents this vowel. For example, (dong) represents /doong/, (dun) represents /doon/, and (du) represents /du/ etc.

(8) 'u:' has the following cases:

i) represented by German letter um-laut [ü] if the initial consonant is 'n' or 'l'.

ii) represented by [i] if it is followed by phonemes 'oong'.

iii) represented by [y] in the syllable /yu:oong/ as (yong).
In this case, the letter [y] represents both the semi-vowel 'y' and vowel 'u:'.

iv) it is represented by [u] in other cases.

(B) Diphthongs

(1) 'ai' is represented by [ai].

(2) 'ei' is represented by [i] when it is preceded by the semi-vowel 'w'; otherwise, it is represented by [ei].

(3) 'au' is represented by [ao].

(4) 'ou' is represented by [ou].

(5) 'iu' is represented by [iu].

(C) Semi-vowels

(1) 'y' is represented by [y].

(2) 'w' is represented by [w].

(D) Consonants

(1) double-letter representation

'rj' is represented by [zh].

'rc' is represented by [ch].

'sh' is represented by [sh].

'ng' is represented by [ng].

(2) single-letter representation

'b' is represented by [b].

'p' is represented by [p].

'm' is represented by [m].

'f' is represented by [f].

'd' is represented by [d].

't' is represented by [t].

'n' is represented by [n].

'l' is represented by [l].

'g' is represented by [g].

'k' is represented by [k].

'h' is represented by [h].

'j' is represented by [j].

'ch' is represented by [q].

'x' is represented by [x].

'r' is represented by [r].

'ds' is represented by [z].

'ts' is represented by [c].

's' is represented by [s].

3.3.2 Syllable Structures Of Pinyin

According to the Table of Speech Sounds of Peking Dialect(Appendix I) [11] published by the government of The People's Republic of China in 1976, each Mandarin syllable in Pinyin can also be decomposed into three parts - the initial(Shung-mu聲母), the final(Yün-mu韻母), and tone. This table lists twenty-two initials and thirty-five finals. Among those twenty-two initials, there are twenty-one consonant initials and one initial which is used for those syllables starting with non-consonant. All thirty-five finals are the combinations and permutations of vowel, diphthong, semi-vowel, and one of the nasal consonants 'n' and 'ng'. The combinations of these twenty-two initials and thirty-five finals give more than four hundred syllables which can be pronounced without tone consideration in Mandarin(Appendix I). All these syllables can also be classified into two categories: syllables without consonant initial and syllables with consonant initial.

3.3.3 Syllables Without Consonant Initial

There are thirty-five syllables in Mandarin which do not have initials. They normally start with one of the following letters [a], [e], [o], [y], and [w], and can be classified into two categories depending on the types of

finals associated with: (1) non-semi-vowel starting and (2) semi-vowel starting.

In the cases of non-semi-vowel starting, the syllable begins with either [a], [e], or [o]. All the possibilities are summarized as follows:

(A) Syllable starting with letter [a].

(1) if letter [i] follows, [ai] represents diphthong 'ai'.

(2) if letter [o] follows, [ao] represents diphthong 'au'.

(3) if [n] or [ng] follows, [a] represents vowel 'a'.

(4) if [a] stands alone, it represents vowel 'a'.

(B) Syllable starting with letter [e].

(1) if letter [i] follows, [ei] represents diphthong 'ei'.

(2) if [r] follows, [er] represents vowel 'er'.

(3) if [n] or [ng] follows, [e] represents vowel 'u(uh)'.

(4) if [e] stands alone, it represents vowel 'e(eh)'.

(C) Syllable starting with [o].

(1) if [u] follows, [ou] represents diphthong 'ou'.

(2) if letter [o] stands alone, it represents vowel 'o'.

The tonal mark is always placed above the first letter of

the syllable representation.

In case of the semi-vowel starting, the syllable begins with either 'w' or 'y'. There is always a semi-vowel to initiate the pronunciation. These syllables can be derived from the following three types:

(A) [yu-]

letter [y] represents semi-vowel 'y' with [u] representing vowel 'u:'. If the syllable ends with a nasal sound, the tonal mark is placed on the letter in front of the nasal consonant, 'n'; otherwise, the tonal mark is placed on the last letter of the syllable.

(B) letter [y] followed by one of the following letters: [a], [e], [i], and [o].

Only in the case of (yong), phonemes 'yu:' are represented by letter [y]. In all other cases, letter [y] stands for semi-vowel 'y'. The tonal mark is placed on the second letter of the syllable.

(C) Syllable staring with [w]

Semi-vowel 'w' is represented by [w]. The tonal mark is always placed on the second letter of the syllable.

Table 3.5 shows all those thirty-five syllables without consonant, together with the corresponding phonetic symbols of Suen's system and Chinese Phonetic Characters of Chan's system.

3.3.4 Syllables With Consonant Initial

There are twenty-one syllables starting with a consonant. They include three double-letter initials, like [ch], [sh] and [zh], and eighteen single letter initials, such as [b], [c], [d], [f], [g], [h], [j], [k], [l], [m], [n], [p], [q], [r], [s], [t], [x] and [z]. These twenty-one consonant initials are listed in Table 3.6 together with the corresponding phonetic symbols of Suen's system and initial codes of the Chinese Phonetic Characters of Chan's system.

According to Pinyin, there are thirty-five finals which follow the consonant initial letters. These finals can be classified into three groups as follows:

(A) Finals starting with a vowel.

Five different categories of this type of finals are classified as follows:

(1) Single vowel only.

For example, [-a], [-o], [-e], [-i], [-u] and um-laut [-ü] representing vowels 'a', 'o', 'e(eh)', 'i', 'oo(u)' and 'u:' respectively. Every final of this type is represented by only one letter. The tonal mark is placed on top of the final in Pinyin representation.

(2) Vowel followed by the nasal consonant 'n' or 'ng'.

For example, (-an), (-en), (-ang), (-eng), (-ong),

(-ing), (-in) and (-un). The letter [u] in the final (-un) represents the vowel 'y', if the initial consonant is either 'j','q' or 'x'; otherwise it represents the vowel 'u'. Letters [a], [i], [e], and [o] represent vowels 'a', 'i', 'u(uh)' and 'oo(u)' respectively. The letters [-n] and [-ng] are pronounced as in English. The tonal mark of this type is placed on the first letter of the final.

(3) Two vowels.

for example, [-ia], [-ie] and [-ue]

[-ia] is the combination of vowels 'i' and 'a'.

[-ie] is the combination of vowels 'i' and 'e(eh)'.

[-ue] is the combination of vowels 'u:' and 'e(eh)'.

The tonal mark is placed on top of the second final letter.

(4) Two vowels followed by nasal consonant 'n' or 'ng'. For example,(-ian), (-iong), (-iang), and (-uan). They follow one of these initial consonants: 'j','q','x' and are listed as follows:

(-ian) - 'i-eh-n'

(-uan) - 'u:-eh-n'

(-iong) - 'y-u:-oo-ng'

(-iang) - 'i-a-ng'

The tonal mark is placed on top of the letter in front of the nasal letter [n] or [ng].

(5) Vowel followed by a diphthong.

For example, in (-iao), [i] represents the vowel 'i' and [ao] represents the diphthong 'au'. The tonal mark is placed above 'a'.

(B) Finals starting with diphthongs only.

For example, [-ai], [-ei], [-ao], [-ou] and [-iu]. The initial consonant is followed by one of the following diphthongs: 'ai', 'ei', 'au', 'ou' and 'iu'. Every final is represented by two letters in this category. The tonal mark is placed on top of the first letter of the final.

(C) Final starting with semi-vowel.

For example, [-ua], [-uo], [-uai], [-ui], [-uang], and [-uan].

Every syllable of this type has one semi-vowel which is followed by the initial consonant other than 'j', 'q' and 'x'. They are listed below:

(-ua) - 'w-a'

(-uo) - 'w-o'

(-uai) - 'w-ai'

(-ui) - 'w-ei'

(-uan) - 'w-a-n'

(-uang) - 'w-a-ng'

The tonal mark is placed on the second final letter.

These thirty-five finals are listed in Table 3.7 together with the corresponding phonemes of Suen's system

and final codes of the Chinese Phonetic Characters of Chan's system.

Table 3.1 Listing Of Consonants In Suen's System.

Manner / Place	Plosive		Nasal	Lateral	Fricative
	Un-aspirated	Aspirated			
Labial	b	p	m		f
Dental Alveolar	d	t	n	l	
Gutteral	g	k	ng		h
Palatal	j	ch			x
Retroflex	rj	rc			r,sh
Dental Sibilant	ds	ts			s

Table 3.2 Listing Of Vowels In Suen's System.

Tongue Position	Front	Central	Back
High	i,u:		oo(u)
Mid	e(eh)	er,u(uh)	o
Low		a	

Table 3.3 The Listing Of Phonetic Symbols And Their Corresponding Symbols/Codes Of Chan's And Pinyin Systems

(A) Initial of syllable starting vowel

Chan's Code/Symbol		phonetic Symbol	Pinyin Letter
I-01	ʕ	er	er
I-07	ʔ	-	(a-,e-,o-)

(B) Initial of syllable starting semi-vowel

Chan's Code/Symbol		phonetic Symbol	Pinyin Letter
I-11	ʷ	w	w
I-21	ʲ	y	y

(C) Initial of syllable starting consonant

Chan's Code/Symbol		phonetic Symbol	Pinyin Letter
I-02	ɾ	r	r
I-03	ʃ	sh	sh
I-04	ʧ	rc	ch
I-05	ʒ	rj	zh
I-06	h	h	h
I-08	s	s	s
I-09	ʦ	ts	c
I-10	ʣ	ds	z
I-12	f	f	f
I-13	m	m	m
I-14	p	p	p
I-15	b	b	b
I-16	t	t	t
I-17	d	d	d
I-18	l	l	l
I-19	k	k	k
I-20	g	g	g
I-22	x	x	x
I-23	n	n	n
I-24	ʧʰ	ch	q
I-25	ʤ	j	j

Table 3.4 The Listing Of Chan's Final Code With Corresponding Phonetic String And Pinyin Representation.

Code/Symbol		With Consonant Initial		Without Consonant Initial	
		phonetic String	Pinyin Letter	phonetic String	Pinyin Letter
F-11	ɱ	a-ng	ang	a-ng	ang
F-12	ɛ	i-ng	ing	y-i-ng	ying
F-13	ɜ	oo-ng	ong	w-u-ng	weng
F-14	ʷ	i-a-ng	iang	y-a-ng	yang
F-21	∩	a-n	an	a-n	an
F-22	ɕ	i-n	in	y-i-n	yin
F-23	ɔ	oo-n	un	w-u-n	wen
F-24	ᵛ	u-n	en	u-n	en
F-31	ʭ	i-a	ia	y-a	ya
F-32	ɐ	i-au	iao	y-au	yao
F-33	ə	i-e-n	ian	y-e-n	yan
F-34	ʊ	u-ng	eng	u-ng	eng
F-41	ʌ	uh	e	uh	e
F-42	ʕ	u:-e-n	uan	y-u:-e-n	yuan
F-43	ʔ	u:-n	un	y-u:-n	yun
F-44	ʏ	u:-oo-ng	iong	y-u:-oo-ng	yong
F-51	⌐	ai	ai	ai	ai
F-52	¬	ei	ei	ei	ei
F-53	ʟ	au	ao	au	ao
F-54	ɹ	ou	ou	ou	ou
F-61	ʌ	w-a	ua	w-a	wa
F-62	‹	w-ai	uai	w-ai	wai
F-63	›	w-a-ng	uang	w-a-ng	wang
F-64	ᵛ	w-a-n	uan	w-a-n	wan
F-71	ʁ	u:	u/u:	y-u:	yu
F-72	ɑ	u	u	w-u	wu
F-73	ɷ	iu	iu	y-ou	you
F-74	ɤ	w-ei	ui	w-ei	wei
F-81	ɐ	a	a	a	a
F-82	ɒ	o	o	o	o
F-83	ɑ	u:-eh	ue/u:e	y-u:-eh	yue
F-84	ɑ	i-eh	ie	y-eh	ye
F-91	ʌ	w-o	uo	w-o	wo
F-92	ɪ	i	i	y-i	yi

Table 3.5 Listing Of Non-consonant Starting Pinyin Syllables With Corresponding Suen's Phonetic Symbols And The Codes Of Chinese Phonetic Characters.

Pinyin Syllables	Suen's Phonemes	Chinese Phonetic Characters Initial and Final Codes	
a	a	I-07	F-81
o	o	I-07	F-82
e	e	I-07	F-41
er	er	I-01	F-92
ai	ai	I-07	F-51
ei	ei	I-07	F-52
ao	au	I-07	F-53
ou	ou	I-07	F-54
an	an	I-07	F-21
en	un	I-07	F-24
ang	ang	I-07	F-11
eng	ung	I-07	F-34
yu	yu:	I-21	F-71
yue	yu:eh	I-21	F-83
yuan	yu:en	I-21	F-42
yun	yu:n	I-21	F-43
yi	yi	I-21	F-92
ya	ya	I-21	F-31
yao	yau	I-21	F-32
ye	yeh	I-21	F-84
you	you	I-21	F-73
yan	yen	I-21	F-33
yin	yin	I-21	F-22
yang	yang	I-21	F-14
ying	ying	I-21	F-12
yong	yu:oong	I-21	F-44
wu	wu	I-13	F-72
wa	wa	I-13	F-61
wo	wo	I-13	F-91
wai	wai	I-13	F-62
wei	wei	I-13	F-74
wan	wan	I-13	F-64
wen	wun	I-13	F-23
wang	wang	I-13	F-63
weng	wung	I-13	F-13

Table 3.6 List Of Consonant Initial Of Pinyin System With Corresponding Suen's Phonetic Symbol And The Code Of Chinese Phonetic Characters.

Pinyin Initial	Suen's Phoneme	Chinese Phonetic Characters Initial Code
b	b	I-15
p	p	I-14
m	m	I-13
f	f	I-12
d	d	I-17
t	t	I-16
n	n	I-23
l	l	I-18
z	ds	I-10
c	ts	I-09
s	s	I-08
zh	rj	I-05
ch	rc	I-04
sh	sh	I-03
r	r	I-02
j	j	I-25
q	ch	I-24
x	x	I-22
g	g	I-20
k	k	I-19
h	h	I-06

Table 3.7 List Of The Finals Of Pinyin System Which Can Be Followed By Consonant Initial With Suen's Phonetic Symbol And The Code Of Chinese Phonetic Character

Pinyin Final	Suen's Phoneme	Chinese Phonetic Characters Final Code
a	a	F-81
o	o	F-82
e	eh	F-41
ai	ai	F-51
ei	ei	F-52
ao	au	F-53
ou	ou	F-54
an	an	F-21
en	un	F-24
ang	ang	F-11
eng	ung	F-34
u:/u	u:	F-71
u:e/ue	u:eh	F-83
uan	u:en	F-42
un	u:n	F-43
i	i	F-92
ia	ia	F-31
iao	iau	F-32
ie	ieh	F-84
iu	iu	F-73
ian	ien	F-33
in	in	F-22
iang	iang	F-14
ing	ing	F-12
iong	u:oong	F-44
u	u	F-72
ua	wa	F-61
uo	wo	F-91
uai	wai	F-62
ui	wei	F-74
uan	wan	F-64
un	oon	F-23
uang	wang	F-63
ong	oong	F-13

Chapter IV

Implementation

The major concern of this project is to analyse Mandarin speech by computer using Pinyin, Chan's and Suen's systems. However, Pinyin is expressed in words rather than syllables. Therefore, before converting Pinyin syllables into Chan's and Suen's representations, pre-processing is required to decompose Pinyin words into syllables.

In this chapter, the database used in this study is introduced first, followed by the modifications of Pinyin representations, then the segmentation algorithm which decomposes every Pinyin word into a group of syllables. Each of these syllables is represented by Chan's initial and final codes with a tone. At the last stage, rules are applied to convert Pinyin syllables and Chan's Codes into Suen's phonetic symbols.

4.1 Data Base

The input texts used in the current study are taken from three different books. The first one is "Chinese For Beginner"[6] published in China. This book is intended to help a beginner to learn to speak Mandarin. The second one

is "Chinese Conversation For Tourist "[5] published in Hong Kong. These two books include most common conversations. The last one is part of a text book[9] on speeches and poems used in elementary schools in China.

4.2 Pinyin System Modification

Although Pinyin represents Mandarin in Latin transcriptions, there are problems which prevent the direct entry of all those phonetic letters into the computer through an ordinary keyboard. As explained, the tonal mark is associated with one of the finals, hence, one of the final letters must carry a tonal mark on top, if the syllable is stressed. Even though the entry of these letters through an ordinary keyboard presents a problem, yet Mandarin is a tonal language, the tone is so important for pronunciation that it must be included. Therefore, in order to minimize the cost and effort for computer input, a modification on the tonal mark has been proposed. Since there are only four distinct tonal marks and one neutral tone, it is simple to substitute the tonal marks by digits placed on the right hand side of the stressed letter, for instance [mǎ] ---> [ma3]. For those syllables with neutral tone, they have no tonal mark on any of the letter, and require no adjustment.

However, an ordinary English keyboard does not have the German um-laut [ü]. To solve this problem, this letter has

been broken up into two characters, an ordinary [u] followed by a colon [:].

4.3 Segmentation Algorithm

This algorithm attempts to separate a Pinyin word into syllables. Then every syllable will be represented by initial and final Chinese Phonetic Character codes with a corresponding tone. Finally, all the Chinese Phonetic Characters will be decoded into Suen's phoneme string representation and its tone.

4.3.1 The Segmentation of Pinyin Word

First, a Mandarin word is read in. A word being defined as a string of letters, delimited by any non-letter, except [:] for German letter um-laut [ü] and digits [1] to [4] for tone indication.

Once the character string is decided, a scanner is employed to decode all Pinyin syllables into two Chinese Phonetic Character Codes, Two sets of rules are used in the scanner. The first rule decodes the initial part of the syllable, while the second rule decides where the syllable should end, the final code and the tonal mark. The scanner is used repeatedly to determine all the syllables within the word string until the end of the string or any error is encountered.

4.3.2 Syllable scanner

As mentioned, there are two steps performed by the scanner to extract the syllables from a word. The first step is to scan the initial of the syllable; the second step is to distinguish one syllable from the other, and to determine the tone as well as the final of the syllable. The rules used in these two processes are listed below.

(A) Rules to determine the initial

In Pinyin representation, a syllable cannot start with any digits or one of the following letters [i], [u], [u:], or [v]. Once the invalid entry is found, an error is encountered, the word will be dropped by the scanner, and the next character string will be read in.

All the valid entries can be classified into three types:

(1) Vowel initials

The entry is either [a] or [e] or [o], the syllable starts with either a vowel or a diphthong. For example (a), (a1n), or (a3i) etc.

In Chan's system, the initial code of this type of syllables is represented by I-07, except string (e#r) or (er). Therefore, whenever the scanner encounters the letters [a], [e], or [o]; the code I-07 is determined. However, these entry letters

can also form part of the final, the scanner shall not move until the final is identified.

(2) Semi-vowel initials

The entry is either [y] or [w], the syllable starts with one of the following semi-vowels 'y', or 'w'. For example, (yue), (yin) or (wei) etc. In Chan's system, the code I-11 represents the semi-vowel initial 'w' while the code I-21 represents the semi-vowel initial 'y'.

In order to simplify the rules of final code determination, several adjustments are required once a semi-vowel initial code is determined.

i) code I-13 is determined (semi-vowel 'w' is found),

(a) in [wa-] or [wo-], [w] is converted to [u], the scanner will remain in the same position.

(b) in case of [we-], but not [weng-], [e] is converted to [u], the scanner will move to the letter [u].

(c) in case of [weng], the [e] is converted to [o], the scanner will move to the letter [o], to apply the second rule.

ii) code I-21 is determined (semi-vowel 'y' is found).

(a) in case of [yu-], the letter [u] is

converted to the German letter um-laut [u], the scanner will move to the letter [u].

(b) in case of [yo-], the letter [o] is converted to letter [i], the scanner will move to the letter [i] too.

(c) in cases of [ya-] or [ye-], the letter [y] is converted to letter [i], the scanner will remain at the same position.

(d) there is no adjustment for [yi-], the scanner will move to the letter [i].

(3) Consonant initials

All the other valid entries can be mapped into Chan's initial code without any adjustments. However, there is a special case on letter [r]. If this letter is the last letter in the word, a syllable which contains only vowel 'er' with neutral tone is determined. For instance, the word (hular), the first syllable is ended at letter [a], the second syllable contains only one letter [r] at the end of the word, so that a vowel 'er' with neutral tone is found, and the end of the word is reached. All the other cases of [r] are considered as a consonant 'r'.

Furthermore, there are three double-letter consonants in Pinyin, viz. ch, sh, and zh. Whenever

one of these letters [c], [s] or [z] is found, the scanner has to check the next character to decide the initial code.

Once the initial code is decided, the scanner will move to the next one or two characters depending on the initial code. In this type of initial code, only I-03, I-04, and I-05 are decided by two letters, therefore, the scanner will move two positions further, while in other cases it moves only one position.

(B) Rules to determine the end of the syllable.

The tone of a syllable is always associated with the final so that, once the initial is determined, the tone and final part of the syllable must be identified in order to break the syllables. As discussed previously, all the finals in Mandarin must start with either a vowel, or a diphthong, or a semi-vowel. Therefore, in Pinyin representation the final should start with one of the following letters [a], [e], [i], [o] or [u]. All other characters are considered as invalid entries. Every final ends with either one of these letters [a], [e], [g], [i], [n], [o], [u] or digits [1] to [4] or the special character [:]. If the syllable is stressed, the digit which represents the tonal mark normally appears in the second or the third position of the final character string. Every final can be represented by one

to five characters. All the valid entries can be divided into four types listed as follows:

Note : [#] stands for digits.
[!] stands for letter [x], [q], [j], or [y].
[%] stands for letter [i] or [u].
[*] stands for letter [a], [e] or [o].

(1) Letters [a], [e], [i], [o], [u] followed by a digit. Since the tone is placed on the second position of the final string, the syllable can be a single vowel, a diphthong or a vowel plus a nasal consonant. However, the scanner has to check several letters further in order to determine the final code. For example, a vowel plus a nasal consonant followed by a digit, the scanner must make sure that the letter which represents the nasal consonant is not the initial consonant of the next syllable. All the final codes and endings of syllables can be determined according to the following rules:

i) [a] followed by a digit.

a#ng - F-11 except (a#ng%) and (a#ng*#)
a#n - F-21 except (a#n%) and (a#n*#)
a#i - F-51
a#o - F-53 except case (a#o#)
a# - F-81

ii) [e] followed by a digit.

e#ng - F-34 except (e#ng%) and (e#ng*#)

e#n - F-24 except (e#n%) and (e#n*#)

e#r - I-01, F-92 except (e#r%)

e#i - F-52

e# - F-41

iii) [i] followed by a digit.

i#ng - F-12 except (i#ng%) and (i#ng*#)

i#n - F-22 except (i#n%) and (i#n*#)

i#u - F-73

i# - F-92

iv) [o] followed by a digit.

o#ng - F-13 except (o#ng%) and (o#ng*#)

o#u - F-54

o# - F-82

vi) [u] followed by a digit.

(!)u#n - F-43 except (u#n%) and (u#n*#)

(!)u# - F-71

u#ng - F-13 except (u#ng%) and (u#ng*#)

u#n - F-23 except (u#n%) and (u#n*#)

(2) letter [u] followed by the special character [:]. The German um-laut [u] is found, for four valid possibilities which are classified as follows:

u:#e - F-83 except (u:#e#)

u:# - F-71

u:e - F-83 except (u:e#)

u: - F-71

(3) Letter [i] or [u] followed by one of these letters [a], [e], [i], [o], [u] and a digit.

If the second letter is either [a], or [e], or [o], it is possible that the current syllable contains only one single vowel with neutral tone which ends at the first letter of the final string. However, in Mandarin, a neutral tone frequently appears at the end of a word. The representation of this kind of final is considered as a vowel followed by either another vowel, or another vowel plus a nasal consonant, or a diphthong, and the tone is represented by the digit. It is also required to check several characters further in order to break the syllable at the proper place. The following listing shows the corresponding final string and its final code.

i) [ia] followed by a digit.

ia#ng - F-14 except (ia#ng%) and (ia#ng*#)

ia#n - F-33 except (ia#n%) and (ia#ng*#)

ia#o - F-32 except (ia#o#)

ia# - F-31

ii) [ie] followed by a digit.

ie# - F-84

iii) [io] followed by a digit.

(1)io#ng - F-44

i - F-92

iv) [iu] followed by a digit

iu# - F-73

v) [ua] followed by a digit

(!)ua#n - F-42 except (!ua#n%) and (!ua#n*#)

ua#ng - F-63 except (ua#ng%) and (ua#ng*#)

ua#n - F-64 except (ua#n%) and (uan*#)

ua#i - F-62

ua# - F-61

vi) [ue] followed by a digit.

(!)ue# - F-83

u - F-71

vii) [ui] followed by a digit.

ui# - F-74

viii) [uo] followed by a digit.

u(o#u) - F-71

uo# - F-91

(4) There is no digit found in the first three places. In this case, the syllable is unstressed, the final code can be determined in the following manner:

i) string starting with [a].

ang - F-11 except (ang%) and (ang*#)

an - F-21 except (an%) and (an*#)

ai - F-51

ao - F-53

a - F-81

ii) string starting with [e].

er - I-01, F-92 except (er%)

eng - F-34 except (eng%) and (eng*#)

en - F-24 except (en%) and (en*#)

ei - F-52

e - F-41

iii) string starting with [i].

ing - F-12 except (ing%) and (ing*#)

in - F-22 except (in%) and (in*#)

iang - F-14 except (iang%) and (iang*#)

ian - F-33 except (ian%) and (ian*#)

iao - F-32 except (iao#)

ia - F-31

iu - F-73

ie - F-84

(!)iong - F-44 except (iong%) and (iong*#)

i - F-92

iv) string starting with [o].

ong - F-13 except (ong%) and (ong*#)

ou - F-54

o - F-82

v) string starting with [u].

(!)uan - F-42 except (!uan%) and (!uan*#)

(!)un - F-43 except (!un%) and (!un*#)

(!)ue - F-83

(!)u - F-71

un - F-23 except (un%) and (un*#)

uang - F-63 except (uang%) and (uang*#)

uan - F-64 except (uan%) and (uan*#)

uai - F-62

ua - F-61

ui - F-74

uo - F-91

u - F-72

4.4 Converting Pinyin Syllable Into Suen's Phonemes

Once each Pinyin word has been decomposed into syllable(s), then each of these syllables can be converted to Suen's phoneme representation according to the following rules.

(A) Rules converting initial letter(s) of Pinyin syllables.

i) The letter [z] or [c] followed by letter [h].

[zh] ---> 'rj'

[ch] ---> 'rc'

ii) The letter [y] followed by [u] or [ong].

[yu] ---> 'yu:'

[y(ong)] ---> 'yu:'

iii) The syllable starts with [c] or [z].

[z] ---> 'ds'

[c] ---> 'ts'

(B) Rules converting the final part of Pinyin syllable.

i) the final part of the syllable consisting only letter [e].

[e] ---> 'uh'

ii) syllable ends with [en], [eng].

[e] ---> 'u'

iii) syllable ends with letter [e].

[e] ---> 'eh'

iv) Letter [i], [u:] or [y] in front of [an].

[a] ---> 'e'

v) Letter [o] followed by [ng].

[o] ---> 'oo'

vi) Letter [u] followed by [n] or [i].

[ui] ---> 'w' + 'ei'

[u(n)] ---> 'oo'

vii) Letter[a] followed by [o].

[ao] ---> 'au'

4.5 Converting Chinese Phonetic Characters Into Phonemes

Every Chinese Phonetic Character contains one tone, one initial code and one final code. The pair of initial and final codes can be easily converted into Suen's phoneme representation as follows:

(A) decode initial code.

i) If code I-01 or I-07 is found, the syllable starts with a vowel or diphthong, the initial code will not be converted.

ii) There are twenty-three recognizable initial codes, which can be decoded according to Table 4.1.

iii) Other cases are invalid entries.

(B) decode final code.

i) If the initial code is found to be I-21, the semi-vowel 'y' initializes the syllable. There are fourteen valid final codes which can be decoded according to Table 4.2.

ii) If the initial code is I-11, the syllable starts with the semi-vowel 'w'. There are nine recognizable final codes which are listed in Table 4.3 with phonemes represented in Suen's system.

iii) The syllable starts with a consonant, a vowel or a diphthong, the final code can be decoded according to Table 4.4.

Table 4.1 Listing Of Chan's Initial Code vs. Suen's Phoneme Representation

Chan's Initial Code	Suen's Phoneme Representation
I-02	r
I-03	sh
I-04	rc
I-05	rj
I-06	h
I-08	s
I-09	ts
I-10	ds
I-11	w
I-12	f
I-13	m
I-14	p
I-15	b
I-16	t
I-17	d
I-18	l
I-19	k
I-20	g
I-21	y
I-22	x
I-23	n
I-24	ch
I-25	j

Table 4.2 Listing Of Chan's Final Code With Initial Code I-21 vs. Suen's Phoneme Representation

Chan's Final Code	Suen's Phoneme Representation
F-12	ing
F-14	ang
F-22	in
F-31	a
F-32	au
F-33	en
F-42	u:en
F-43	u:n
F-44	u:oong
F-71	u:
F-73	ou
F-83	u:eh
F-84	eh
F-92	i

Table 4.3 Listing Of Chan's Final Code With Initial Code I-11 vs. Suen's Phoneme Representation

Chan's Final Code	Suen's Phoneme Representation
F-13	ung
F-23	un
F-61	a
F-62	ai
F-63	ang
F-64	an
F-72	u
F-74	ei
F-91	o

Table 4.4 : Listing Of Chan's Final Code With A Consonant, A Vowel Or A Diphthong Initial vs. Suen's Phoneme Representation

Chan's Final Code	Suen's Phoneme Representation
F-11	ang
F-12	ing
F-13	oong
F-14	iang
F-21	an
F-22	in
F-23	oon
F-24	un
F-31	ia
F-32	iau
F-33	ien
F-34	ung
F-41	uh
F-42	u:en
F-43	u:n
F-44	u:oong
F-51	ai
F-52	ei
F-53	au
F-54	ou
F-61	wa
F-62	wai
F-63	wang
F-64	wan
F-71	u:
F-72	u
F-73	iu
F-74	wei
F-81	a
F-82	o
F-83	u:uh
F-84	ieh
F-91	wo
F-92	er(with initial code I-01)
F-92	i(other cases)

Chapter V

Results

5.1 Distribution Of Initials And Finals Of Chan's System

Table 5.1 presents the frequency and percent frequency distribution of the twenty-five Chan's initial codes in the entire data. Code I-17(9.71%) represents the most frequently used consonant 'd' and code I-07(0.48%) represents the least frequently used syllable among those which start with a vowel or a diphthong. The rank order of each Chan's initial with the corresponding tone is listed in Table 5.2. The percent proportion of Chan's initials in the data is illustrated in Fig. 5.1.

Table 5.3 presents the frequency and percent frequency distribution of the thirty-four Chan's final codes in the entire set of data. Code F-92(17.04%) represents the most frequently used vowel 'i' and code F-82(0.80%) represents the least frequently used vowel 'o'. Table 5.4 lists the rank order of all Chan's finals. The percent proportion of Chan's finals in the overall data is illustrated in Fig. 5.2.

Fig. 5.1 Histogram of Chan's Initial symbols

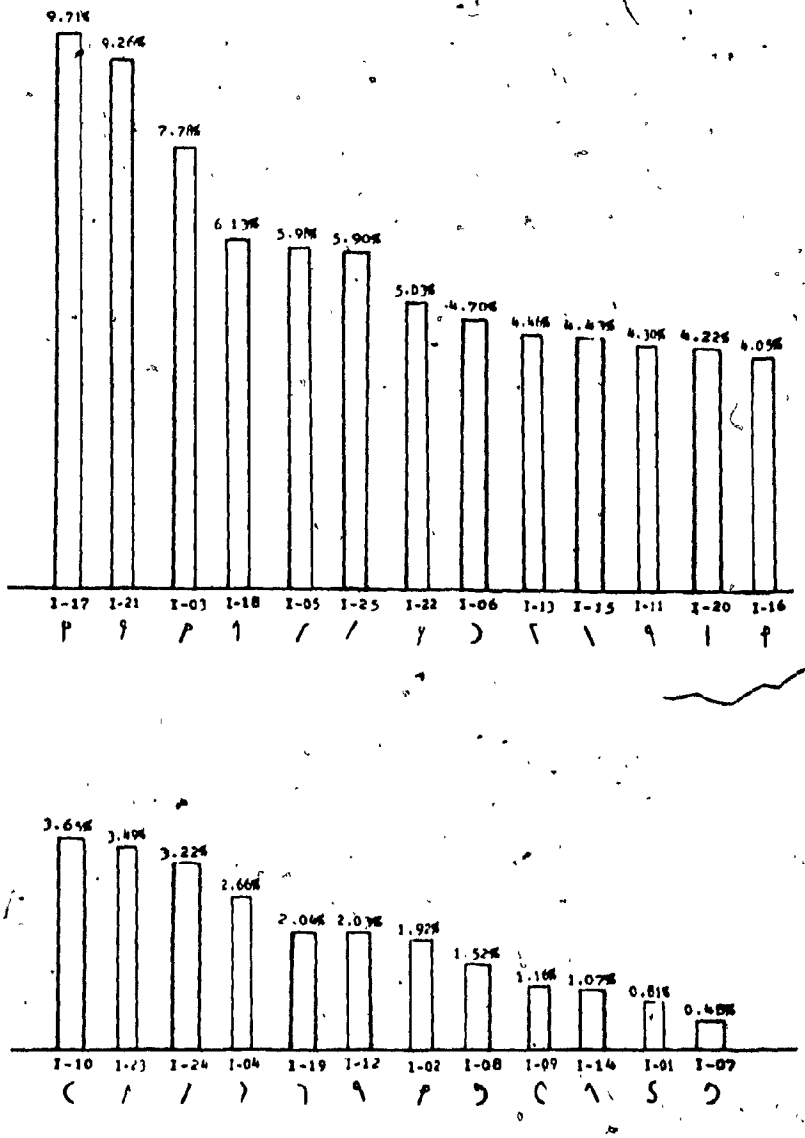


Fig. 5.2 Histogram of Chan's Final symbols

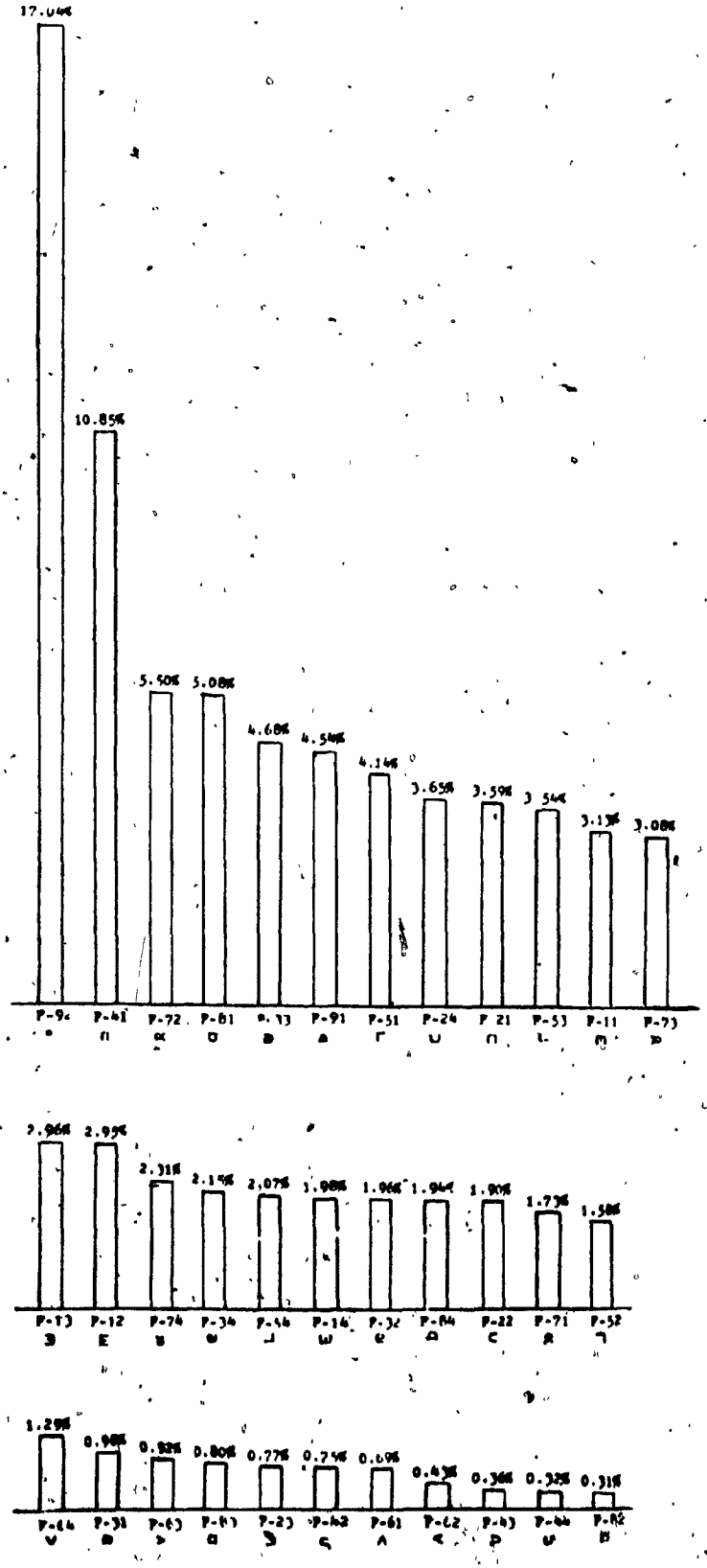
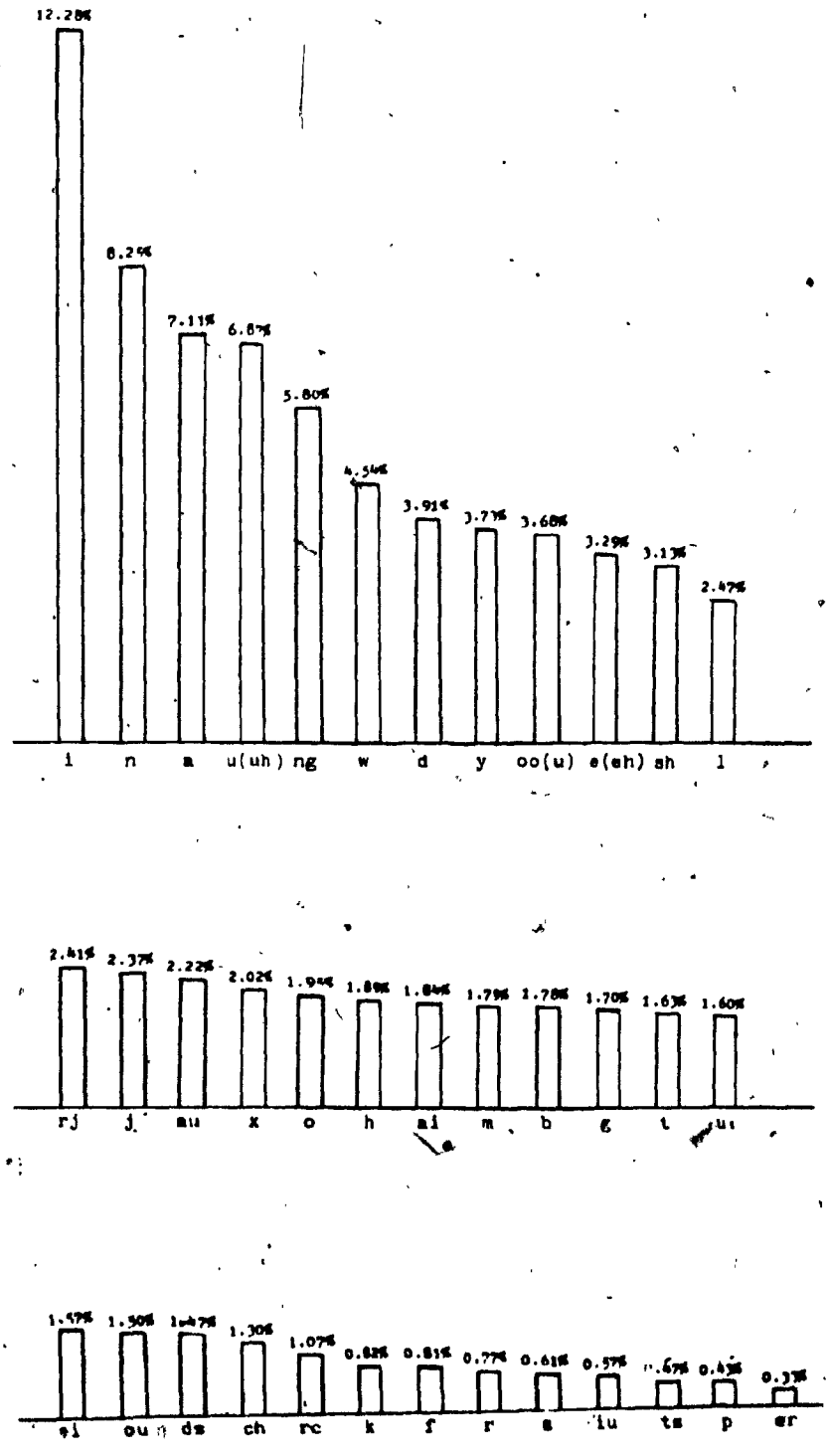


Fig. 5.3 Histogram of Suen's phonetic symbols



5.2 Distribution Of Suen's Phonemes

Table 5.5 presents the frequency and percent frequency distribution of Suen's phonemes in all three sets of data. The rank order of the thirty-seven phonemes is listed in Table 5.6. The percent distributions of the various types of phonemes are as follows:

Consonants : 46.93%

Semi-vowels : 8.26%

Vowels : 37.11%

Diphthongs : 7.70%

Table 5.7 shows the relative percent proportion of phonemes in the data which have been classified into consonants, semi-vowels, vowels and diphthongs. The histogram of all the thirty-seven phonemes is illustrated in Fig 5.3.

Among the consonants, the nasal 'n'(17.58%) is the most frequently used consonant. However, only 17.05% of this consonant is used to initialize a syllable. The nasal 'ng'(12.36%) is the second most frequently used consonant, followed by the unaspirated plosive 'd'(8.33%), fricative 'sh'(6.67%) and lateral 'l'(5.26%). The least frequently used consonants are listed in ascending order as follows: aspirated plosive 'ts'(1.00%), fricative 's'(1.31%),

'r'(1.65%), 'f'(1.74%), and aspirated plosive 'k'(1.75%).

Among the vowels, the front-high vowel 'i'(33.10%) is the one most frequently used, followd by the central-low vowel 'a'(19.17%), and central-mid vowel 'u(uh)'(18.52%). The central-mid 'er'(0.88%) is the least frequently used vowel.

Among the diphtongs, the diphthong 'iu'(7.43%) is the one least frequently used while the diphthong 'au'(28.77%) is the most frequently used one.

5.3 Distribution Of Tones

Table 5.8 lists the frequency and percent frequency distribution of every tone in all three sets of data. It shows that the fourth tone is used most frequently, and the neutral tone is used least frequently.

5.4 Distribution Of Syllables

There are totally 37,639 syllables in the data, among them, 4,989 syllables come from set one, 6,151 syllables from set two, and 26,499 syllables from set three. All these syllables can be classified into 396 identical syllables without taking tone into consideration and 1,194 syllables when the tone is taken into consideration. Therefore, the average appearance of each identical syllable is 66.92 times without tone and 22.19 for syllables with

tone.

In all these syllables, 32,503(85.16%) syllables begin with a consonant, 5,104(13.56%) syllables begin with a semi-vowel, 442(1.17%) syllables begin with a vowel, and only 40(0.11%) syllables begin with a diphthong. Also, there are 11,824(31.41%) syllables which end with a nasal consonant, and among them, 6,399(17%) syllables end with the nasal consonant 'n' and 5,425(14.41%) syllables end with 'ng'. There are 18,616(49.46%) syllables ending with a vowel, and 7,199(19.13%) syllables with a diphthong. Concerning the number of phonemes in a syllable, there are 429(1.14%) syllables which contain only one phoneme, 21,805(57.93%) syllables which contain two phonemes, 12,162(32.31%) syllables which contain three phonemes, and 3,243(8.62%) syllables which contain four phonemes. The average number of phonemes per syllable is 2.48.

The ten most frequently used syllables excluding and including tone in the data have been listed in Tables 5.9 and 5.10 respectively.

5.5 N-gram Analysis Of Phonemes In Database

There are 5,251 clauses in the database. Among them 752 clauses come from set one, 1,109 from set two and 3,390 from set three. Every clause can be considered as a phoneme string which is used as an entry in n-gram analysis.

The n-gram analysis is used to find the occurrence frequency and percent distribution of a substring which contains from one up to five phonemes, since n is not greater than five in this analysis. For example, the clause 'ni3 hau3 ma?' which means how are you, contains six phonemes. Every phoneme can be the entry of a substring. Entry 'n' can form five substrings 'n', 'n-i','n-i-h', 'n-i-h-au', and 'n-i-h-au-m'. Table 5.11 shows the frequency of occurrence of a substring with the starting phoneme and the percentage of frequency of occurrence. Table 5.12 lists the most frequent occurring phoneme string together with their percent frequency distribution.

5.6 Summary

The total computing time required for the process of segmenting all the data is 27.65 seconds, in which set one takes 3.76 seconds, set two takes 4.54 seconds and set three takes 19.35 seconds. All the results, which include the number of clauses, words, syllables, and letters in each set of data, are summarized in Table 5.13. Table 5.14 lists the distribution frequency and the percent frequency of consonants, semi-vowels, vowels and diphthongs found in each set of data.

Table 5.1 Distribution Of Chan's Initial Phonetic Character vs. Tone

Code	Phoneme	Frequency						% (w.r.t. Total Initial Characters)					
		1th	2th	3th	4th	Neutral	Sum	1th	2th	3th	4th	Neutral	Sum
I- 1	er	0	90	11	77	126	304	0.00	0.24	0.03	0.20	0.33	0.81
I- 2	r	2	496	12	203	9	722	0.01	1.32	0.03	0.54	0.02	1.92
I- 3	sh	715	477	248	1196	292	2928	1.90	1.27	0.66	3.18	0.78	7.78
I- 4	rc	382	400	115	67	38	1002	1.01	1.06	0.31	0.18	0.10	2.66
I- 5	rj	764	83	319	863	220	2249	2.03	0.22	0.85	2.29	0.58	5.98
I- 6	h	207	633	343	502	85	1770	0.55	1.68	0.91	1.33	0.23	4.70
I- 7	(a-,e-,o-,u-)	60	5	6	43	66	180	0.16	0.01	0.02	0.11	0.18	0.48
I- 8	a	187	13	126	231	16	573	0.50	0.03	0.33	0.61	0.04	1.52
I- 9	ts	72	194	83	87	1	437	0.19	0.52	0.22	0.23	0.00	1.16
I-10	ds	47	86	289	764	188	1374	0.12	0.23	0.77	2.03	0.50	3.65
I-11	(w-)	63	263	829	439	26	1620	0.17	0.70	2.20	1.17	0.07	4.30
I-12	f	284	132	125	144	79	764	0.75	0.35	0.33	0.38	0.21	2.03
I-13	m	28	504	256	245	647	1680	0.07	1.34	0.68	0.65	1.72	4.46
I-14	p	47	158	47	147	3	402	0.12	0.42	0.12	0.39	0.01	1.07
I-15	b	210	303	377	638	138	1666	0.56	0.81	1.00	1.70	0.37	4.43
I-16	t	920	348	98	124	34	1524	2.44	0.92	0.26	0.33	0.09	4.05
I-17	d	471	134	312	1189	1550	3656	1.25	0.36	0.83	3.16	4.12	9.71
I-18	l	30	558	469	377	872	2306	0.08	1.48	1.25	1.00	2.32	6.13
I-19	k	143	5	225	367	27	767	0.38	0.01	0.60	0.98	0.07	2.04
I-20	g	490	128	340	344	288	1590	1.30	0.34	0.90	0.91	0.77	4.22
I-21	(y-)	482	777	1026	1132	67	3484	1.28	2.06	2.73	3.01	0.18	9.26
I-22	x	534	254	455	534	116	1893	1.42	0.67	1.21	1.42	0.31	5.03
I-23	n	11	429	540	244	91	1315	0.03	1.14	1.43	0.65	0.24	3.49
I-24	ch	240	382	222	250	119	1213	0.64	1.01	0.59	0.66	0.32	3.22
I-25	j	750	178	366	840	86	2220	1.99	0.47	0.97	2.23	0.23	5.90
	Total	7139	7030	7239	11047	5184	37639	18.97	18.68	19.23	29.35	13.77	100.00

Table 5.2 Rank Order Of Chan's Initials

Rank	Tone 1	Tone 2	Tone 3	Tone 4	Neutral	Overall
1	I-16	I-21	I-21	I-03	I-17	I-17
2	I-05	I-06	I-11	I-17	I-18	I-21
3	I-25	I-18	I-23	I-21	I-13	I-03
4	I-03	I-13	I-18	I-05	I-03	I-18
5	I-22	I-02	I-22	I-25	I-20	I-05
6	I-20	I-03	I-15	I-10	I-05	I-25
7	I-21	I-23	I-25	I-15	I-10	I-22
8	I-17	I-04	I-06	I-22	I-15	I-06
9	I-04	I-24	I-20	I-06	I-01	I-13
10	I-12	I-16	I-05	I-11	I-24	I-15
11	I-24	I-15	I-17	I-18	I-22	I-11
12	I-15	I-11	I-10	I-19	I-23	I-20
13	I-06	I-22	I-13	I-20	I-25	I-16
14	I-08	I-09	I-03	I-24	I-06	I-10
15	I-19	I-25	I-19	I-13	I-12	I-23
16	I-09	I-14	I-24	I-23	I-21	I-24
17	I-11	I-17	I-08	I-08	I-07	I-04
18	I-07	I-12	I-12	I-02	I-04	I-19
19	I-10	I-20	I-04	I-14	I-16	I-12
20	I-14	I-01	I-16	I-12	I-19	I-02
21	I-18	I-10	I-09	I-16	I-11	I-08
22	I-13	I-05	I-14	I-09	I-08	I-09
23	I-23	I-08	I-02	I-01	I-02	I-14
24	I-02	I-07	I-01	I-04	I-14	I-01
25		I-19	I-07	I-07	I-09	I-07

Table 5.3 Distribution Of Chan's Final Phonetic Character vs. Tone

Code	Phoneme		Frequency						$ (w.r.t. Total Initial Characters)					
			1th	2th	3th	4th	Neutral	Sum	1th	2th	3th	4th	Neutral	Sum
F-11	ang		232	282	228	248	187	1177	0.62	0.75	0.61	0.66	0.50	3.13
F-12	ing	/ying	309	342	192	229	38	1110	0.82	0.91	0.51	0.61	0.10	2.95
F-13	ong	/wung	494	303	101	215	1	1114	1.31	0.81	0.27	0.57	0.00	2.96
F-14	iang	/yang	104	84	310	225	24	747	0.28	0.22	0.82	0.60	0.06	1.98
F-21	an		376	228	220	496	32	1352	1.00	0.61	0.58	1.32	0.09	3.59
F-22	in	/yin	267	166	71	203	8	715	0.71	0.44	0.19	0.54	0.02	1.90
F-23	oon	/wun	51	96	36	101	7	291	0.14	0.26	0.10	0.27	0.02	0.77
F-24	un		214	571	201	88	301	1375	0.57	1.52	0.53	0.23	0.80	3.65
F-31	ia	/ya	146	21	12	142	47	368	0.39	0.06	0.03	0.38	0.12	0.98
F-32	iau	/yau	88	91	166	390	2	737	0.23	0.24	0.44	1.04	0.01	1.96
F-33	ian	/yen	583	339	227	529	82	1760	1.55	0.90	0.60	1.41	0.22	4.68
F-34	ung		269	267	53	146	74	809	0.71	0.71	0.14	0.39	0.20	2.15
F-41	uh		187	301	189	696	2710	4083	0.50	0.80	0.50	1.85	7.20	10.85
F-42	u:en	/yu:en	29	163	51	40	0	283	0.08	0.43	0.14	0.11	0.00	0.75
F-43	u:n	/yu:n	42	52	1	41	0	136	0.11	0.14	0.00	0.11	0.00	0.36
F-44	u:oong	/yu:oong	12	14	20	74	1	121	0.03	0.04	0.05	0.20	0.00	0.32
F-51	ai		140	430	169	670	149	1558	0.37	1.14	0.45	1.78	0.40	4.14
F-52	ei		96	170	228	98	4	596	0.26	0.45	0.61	0.26	0.01	1.58
F-53	au		118	137	483	539	57	1334	0.31	0.36	1.28	1.43	0.15	3.54
F-54	ou		202	100	202	199	78	781	0.54	0.27	0.54	0.53	0.21	2.07
F-61	wa		84	28	11	121	16	260	0.22	0.07	0.03	0.32	0.04	0.69
F-62	wai		5	19	2	135	0	161	0.01	0.05	0.01	0.36	0.00	0.43
F-63	wang		95	127	35	89	1	347	0.25	0.34	0.09	0.24	0.00	0.92
F-64	wan		179	75	108	113	12	487	0.48	0.20	0.29	0.30	0.03	1.29
F-71	u:	/yu:	52	127	176	229	67	651	0.14	0.34	0.47	0.61	0.18	1.73
F-72	u	/wu	327	486	334	775	149	2071	0.87	1.29	0.89	2.06	0.40	5.50
F-73	iu	/you	50	227	536	337	11	1161	0.13	0.60	1.42	0.90	0.03	3.08
F-74	wei		86	125	128	530	2	871	0.23	0.33	0.34	1.41	0.01	2.31
F-81	a		694	131	287	467	333	1912	1.84	0.35	0.76	1.24	0.88	5.08
F-82	o		20	44	5	38	11	118	0.05	0.12	0.01	0.10	0.03	0.31
F-83	u:eh	/yu:eh	44	110	15	132	0	301	0.12	0.29	0.04	0.35	0.00	0.80
F-84	ieh	/yeh	122	114	321	148	26	731	0.32	0.30	0.85	0.39	0.07	1.94
F-91	wo		415	162	772	294	65	1708	1.10	0.43	2.05	0.78	0.17	4.54
F-92	i	/yi	1007	1098	1349	2270	689	6413	2.68	2.92	3.58	6.03	1.83	17.04
	Total		7139	7030	7239	11047	5184	37639	18.97	18.68	19.23	29.35	13.77	100.00

Table 5.4 Rank Order Of Chan's Finals

Rank	Tone 1	Tone 2	Tone 3	Tone 4	Neutral	Overall
1	F-92	F-92	F-92	F-92	F-41	F-92
2	F-81	F-24	F-91	F-72	F-92	F-41
3	F-33	F-72	F-73	F-41	F-81	F-72
4	F-13	F-51	F-53	F-51	F-24	F-81
5	F-91	F-12	F-72	F-53	F-11	F-33
6	F-21	F-33	F-84	F-74	F-51	F-91
7	F-72	F-13	F-14	F-33	F-72	F-51
8	F-12	F-41	F-81	F-21	F-33	F-24
9	F-34	F-11	F-11	F-81	F-54	F-21
10	F-22	F-34	F-52	F-32	F-34	F-53
11	F-11	F-21	F-33	F-73	F-71	F-11
12	F-24	F-73	F-21	F-91	F-91	F-73
13	F-54	F-52	F-54	F-11	F-53	F-13
14	F-41	F-22	F-24	F-12	F-31	F-12
15	F-64	F-42	F-12	F-71	F-12	F-74
16	F-31	F-91	F-41	F-14	F-21	F-34
17	F-51	F-53	F-71	F-13	F-84	F-54
18	F-84	F-81	F-51	F-22	F-14	F-14
19	F-53	F-63	F-32	F-54	F-61	F-32
20	F-14	F-71	F-74	F-84	F-64	F-84
21	F-52	F-74	F-64	F-34	F-73	F-22
22	F-63	F-84	F-13	F-31	F-82	F-71
23	F-32	F-83	F-22	F-62	F-22	F-52
24	F-74	F-54	F-34	F-83	F-23	F-64
25	F-61	F-23	F-42	F-61	F-52	F-31
26	F-71	F-32	F-23	F-64	F-32	F-63
27	F-23	F-14	F-63	F-23	F-74	F-83
28	F-73	F-64	F-44	F-52	F-13	F-23
29	F-83	F-43	F-83	F-63	F-44	F-42
30	F-43	F-82	F-31	F-24	F-63	F-61
31	F-42	F-61	F-61	F-44		F-62
32	F-82	F-31	F-82	F-43		F-43
33	F-44	F-62	F-62	F-42		F-44
34	F-62	F-44	F-43			F-62

Table 5.5 The Distribution Of Suen's Phonemes

Phonemes	Frequency Of Occurrence						Percentage Of Frequency Of Occurrence					
	Tone 1	Tone 2	Tone 3	Tone 4	Neutral	Sum	Tone 1	Tone 2	Tone 3	Tone 4	Neutral	Sum
b	210	303	377	638	138	1666	0.2246	0.3241	0.4032	0.6824	0.1476	1.7819
p	47	158	47	147	3	402	0.0503	0.1690	0.0503	0.1572	0.0032	0.4300
m	28	504	256	245	647	1680	0.0299	0.5391	0.2738	0.2620	0.6920	1.7969
f	284	132	125	144	79	764	0.3038	0.1412	0.1337	0.1540	0.0845	0.8172
d	471	134	312	1189	1550	3656	0.5038	0.1433	0.3337	1.2717	1.6578	3.9104
t	920	348	98	124	34	1524	0.9840	0.3722	0.1048	0.1326	0.0364	1.6300
n	1752	2119	1455	1855	533	7714	1.8739	2.2664	1.5562	1.9841	0.5701	8.2507
l	30	558	469	377	872	2306	0.0321	0.5968	0.5016	0.4032	0.9327	2.4664
g	490	128	340	344	288	1590	0.5241	0.1369	0.3637	0.3679	0.3080	1.7006
k	143	5	225	367	27	767	0.1529	0.0053	0.2407	0.3925	0.0289	0.8204
h	207	633	343	502	85	1770	0.2214	0.6770	0.3669	0.5369	0.0909	1.8931
j	750	178	366	840	86	2220	0.8022	0.1904	0.3915	0.8984	0.0920	2.3745
ch	240	382	222	250	119	1213	0.2567	0.4086	0.2374	0.2674	0.1273	1.2974
x	534	254	455	534	116	1893	0.5712	0.2717	0.4867	0.5712	0.1241	2.0247
rj	764	83	319	863	220	2249	0.8172	0.0888	0.3412	0.9230	0.2353	2.4055
rc	382	400	115	67	38	1002	0.4086	0.4278	0.1230	0.0717	0.0406	1.0717
sh	715	477	248	1196	292	2928	0.7647	0.5102	0.2653	1.2792	0.3123	3.1317
r	2	496	12	203	9	722	0.0021	0.5305	0.0128	0.2171	0.0096	0.7722
ds	47	86	289	764	188	1374	0.0503	0.0920	0.3091	0.8172	0.2011	1.4696
ts	72	194	83	87	1	437	0.0770	0.2075	0.0888	0.0931	0.0011	0.4674
s	187	13	126	231	16	573	0.2000	0.0139	0.1348	0.2471	0.0171	0.6129
ng	1515	1419	939	1226	326	5425	1.6204	1.5177	1.0043	1.3113	0.3487	5.8024
w	894	679	1157	1400	111	4241	0.9562	0.7262	1.2375	1.4974	0.1187	4.5361
y	482	777	1026	1132	67	3484	0.5155	0.8311	1.0974	1.2108	0.0717	3.7264
a	1910	976	1211	1901	652	6650	2.0429	1.0439	1.2953	2.0333	0.6974	7.1127
e(eh)	778	726	614	849	108	3075	0.8321	0.7765	0.6567	0.9081	0.1155	3.2889
i	2571	2058	2364	3711	781	11485	2.7499	2.2012	2.5285	3.9692	0.8353	12.2841
o	435	206	777	332	76	1826	0.4653	0.2203	0.8311	0.3551	0.0813	1.9530
oo(u)	871	826	487	1103	152	3439	0.9316	0.8835	0.5209	1.1797	0.1626	3.6783
u:	179	466	263	516	68	1492	0.1915	0.4984	0.2813	0.5519	0.0727	1.5958
u(uh)	683	1212	447	992	3091	6425	0.7305	1.2963	0.4781	1.0610	3.3061	6.8720
er	0	90	11	77	126	304	0.0000	0.0963	0.0118	0.0824	0.1348	0.3252
ai	145	449	171	805	149	1719	0.1551	0.4802	0.1829	0.8610	0.1594	1.8386
ei	182	295	356	628	6	1467	0.1947	0.3155	0.3808	0.6717	0.0064	1.5691
au	206	228	649	929	59	2071	0.2203	0.2439	0.6942	0.9936	0.0631	2.2151
ou	208	176	652	286	85	1407	0.2225	0.1882	0.6974	0.3059	0.0909	1.5049
iu	44	151	86	250	4	535	0.0471	0.1615	0.0920	0.2674	0.0043	0.5722
Total	19378	18319	17492	27104	11202	93495	20.7262	19.5936	18.7090	28.9898	11.9814	100.0000

Table 5.6 Rank Order Of Suen's Phonemes

RANK	Tone 1	Tone 2	Tone 3	Tone 4	Neutral	Overall
1	i	n	i	i	u(uh)	i
2	a	i	n	a	d	n
3	n	ng	a	n	l	a
4	ng	u(uh)	w	w	i	u(uh)
5	t	a	y	ng	a	ng
6	w	oo(u)	ng	sh	m	w
7	oo(u)	y	o	d	n	d
8	rj	e(eh)	ou	y	ng	y
9	j	w	au	oo(u)	sh	oo(u)
10	sh	h	e(eh)	u(uh)	g	e(eh)
11	e(eh)	l	oo(u)	e(eh)	rj	sh
12	u(uh)	m	l	au	ds	l
13	x	r	x	rj	oo(u)	rj
14	g	sh	u(uh)	j	ai	j
15	Y	u:	b	ai	b	au
16	d	ai	j	ds	er	x
17	o	rc	h	b	ch	h
18	rc	ch	g	x	x	ai
19	f	t	ei	u:	W	o
20	ch	b	rj	h	e(eh)	m
21	b	x	d	ei	j	b
22	ou	ei	ds	l	h	g
23	h	au	u:	k	ou	t
24	au	ts	m	g	f	u:
25	s	j	sh	o	u:	ou
26	u:	ou	k	ou	Y	ds
27	ei	o	ch	ch	o	ch
28	ai	p	ai	iu	au	ei
29	k	iu	s	m	rc	rc
30	ts	d	f	s	t	k
31	p	f	rc	r	k	f
32	ds	g	t	p	s	r
33	iu	er	iu	f	r	s
34	l	ds	ts	t	ei	iu
35	m	rj	p	ts	iu	ts
36	r	s	r	er	p	p
37	er	k	er	rc	ts	er

Table 5.7 Relative Percent Proportion Of Phonemes In Database: Classified into Consonants, Semi-vowels, Vowels and Diphthongs

Consonants

MANNER / PLACE	PLOSIVE		NASAL	LATERAL	FRICATIVE	TOTAL
	UNASPIRATED	ASPIRATED				
LABIAL	b 1666 (1.78)	p 402 (0.43)	m 1680 (1.80)		f 764 (0.82)	4512 (4.83)
DEN. ALVEOLAR	d 3656 (3.91)	t 1524 (1.63)	n 7714 (8.25)	l 2306 (2.47)		15200 (16.26)
GUTTERAL	g 1590 (1.70)	k 767 (0.82)	ng 5425 (5.80)		h 1770 (1.89)	9552 (10.22)
PALATAL	j 2220 (2.37)	ch 1213 (1.30)			x 1893 (2.02)	5326 (5.70)
RETROFLEX	rj 2249 (2.41)	rc 1002 (1.07)			sh 2928 (3.13) r 722 (0.77)	6901 (7.38)
DEN. SIBILANT	ds 1374 (1.47)	ts 437 (0.47)			s 573 (0.61)	2384 (2.55)
TOTAL	12755 (13.64)	5345 (5.72)	14819 (15.85)	2306 (2.47)	8650 (9.25)	43875 (46.93)

Vowels

TONGUE POSITION	FRONT	CENTRAL	BACK	TOTAL
HIGH	i 11485 (12.28) u: 1492 (1.60)		oo(u) 3439 (3.68)	16416 (17.56)
MID	e(eh) 3075 (3.29)	er 304 (0.33) u(uh) 6425 (6.87)	o 1826 (1.95)	11630 (12.44)
LOW		a 6650 (7.11)		6650 (7.11)
TOTAL	16052 (17.17)	13379 (14.31)	5265 (5.63)	34696 (37.11)

Semi-vowels

w	4241 (4.54)	y	3484 (3.73)	TOTAL	7725 (8.26)	

Diphthongs

ai	1719 (1.84)	ei	1467 (1.57)	au	2071 (2.22)	ou	1407 (1.50)	iu	535 (0.57)	TOTAL 7199 (7.70)

Table 5.8 The Distribution Of Tones

TONE	Set 1	Set 2	Set 3	Overall
1st	964(19.32%)	1245(20.24%)	4930(18.60%)	7139(18.97%)
2nd	905(18.14%)	915(14.88%)	5210(19.66%)	7030(18.68%)
3rd	938(18.80%)	1497(24.34%)	4804(18.13%)	7239(19.23%)
4th	1398(28.02%)	1848(30.04%)	7801(29.44%)	11047(29.35%)
neutral	784(15.71%)	646(10.50%)	3754(14.17%)	5184(13.77%)

Table 5.9 Listing of the first hundred frequently used syllables without taking tone into consideration

Rank	Set I		Set II		Set III		Overall	
1	duh	171(3.43%)	shí	237(3.85%)	duh	1191(4.49%)	duh	1529(4.06%)
2	yi	168(3.37%)	wo	227(3.69%)	yi	893(3.37%)	shi	1263(3.36%)
3	shi	150(3.01%)	yi	201(3.27%)	shi	876(3.31%)	yi	1262(3.35%)
4	luh	127(2.55%)	duh	167(2.72%)	rjuh	490(1.85%)	rjuh	713(1.89%)
5	wo	101(2.02%)	ni	145(2.36%)	rji	462(1.74%)	wo	649(1.72%)
6	you	84(1.68%)	rjuh	141(2.29%)	bu	449(1.69%)	you	626(1.66%)
7	rjuh	82(1.64%)	you	136(2.21%)	ta	437(1.65%)	luh	578(1.54%)
8	mun	79(1.58%)	ma	110(1.79%)	you	406(1.53%)	bu	550(1.46%)
9	guh	71(1.42%)	li	91(1.48%)	luh	406(1.53%)	ta	539(1.43%)
10	run	70(1.40%)	ching	89(1.45%)	li	333(1.26%)	rji	525(1.39%)
11	aa	68(1.36%)	mun	87(1.41%)	wo	321(1.21%)	li	462(1.23%)
12	ta	65(1.30%)	dsai	82(1.33%)	dsi	306(1.15%)	ji	417(1.11%)
13	ji	56(1.12%)	ji	80(1.30%)	ji	281(1.06%)	dsai	413(1.10%)
14	gwo	56(1.12%)	xiang	70(1.14%)	run	279(1.05%)	ni	381(1.01%)
15	ba	55(1.10%)	chi	68(1.11%)	dsai	278(1.05%)	dsi	379(1.01%)
16	er	54(1.08%)	jien	66(1.07%)	dau	275(1.04%)	run	371(0.99%)
17	dsai	53(1.06%)	dien	64(1.04%)	jien	242(0.91%)	dau	367(0.98%)
18	hai	52(1.04%)	wu	57(0.93%)	shang	241(0.91%)	guh	356(0.95%)
19	jiu	49(0.98%)	bu	57(0.93%)	guh	241(0.91%)	jien	355(0.94%)
20	jien	47(0.94%)	tien	56(0.91%)	chi	235(0.89%)	mun	347(0.92%)
21	shang	45(0.90%)	xi	55(0.89%)	yeh	235(0.89%)	chi	342(0.91%)
22	nau	45(0.90%)	na	55(0.89%)	wei	229(0.86%)	shang	317(0.84%)
23	bu	44(0.88%)	dau	54(0.88%)	jiu	221(0.83%)	er	304(0.81%)
24	ni	41(0.82%)	yau	51(0.83%)	xiang	213(0.80%)	xiang	303(0.81%)
25	rjoong	40(0.80%)	muh	50(0.81%)	na	212(0.80%)	jiu	302(0.80%)
26	goong	40(0.80%)	xieh	50(0.81%)	lai	212(0.80%)	da	302(0.80%)
27	dwo	40(0.80%)	xing	50(0.81%)	er	207(0.78%)	lai	299(0.79%)
28	chi	39(0.78%)	lai	49(0.80%)	ni	195(0.74%)	na	287(0.76%)
29	tien	39(0.78%)	shun	48(0.78%)	da	194(0.73%)	wei	274(0.73%)
30	xi	38(0.76%)	luh	45(0.73%)	mei	192(0.72%)	rjoong	270(0.72%)
31	li	38(0.76%)	guh	44(0.72%)	rjoong	191(0.72%)	xi	264(0.70%)
32	chu:	38(0.76%)	shau	44(0.72%)	shwo	186(0.70%)	yeh	260(0.69%)
33	dau	38(0.76%)	lfang	44(0.72%)	xien	181(0.68%)	ba	259(0.69%)
34	lai	38(0.76%)	er	43(0.70%)	mun	181(0.68%)	gwo	256(0.68%)
35	shwo	37(0.74%)	ba	43(0.70%)	gwo	173(0.65%)	mei	251(0.67%)
36	rji	37(0.74%)	jin	41(0.67%)	shung	172(0.65%)	wu	249(0.66%)
37	dsi	36(0.72%)	goong	41(0.67%)	jin	171(0.65%)	xien	242(0.64%)
38	hun	36(0.72%)	fang	40(0.65%)	xi	171(0.65%)	tien	240(0.64%)
39	aoong	35(0.70%)	da	40(0.65%)	wu	167(0.63%)	jin	235(0.62%)
40	yau	33(0.66%)	jiau	39(0.63%)	hwa	163(0.62%)	shung	235(0.62%)
41	kan	33(0.66%)	rjoong	39(0.63%)	fu	162(0.61%)	yau	231(0.61%)
42	jia	33(0.66%)	yu:en	38(0.62%)	ba	161(0.61%)	shwo	228(0.61%)
43	xiau	33(0.66%)	rcuh	38(0.62%)	kan	153(0.58%)	hau	215(0.57%)
44	xien	33(0.66%)	dsi	37(0.60%)	rou	150(0.57%)	kan	209(0.56%)
45	rcang	32(0.64%)	ta	37(0.60%)	huh	148(0.56%)	fu	209(0.56%)
46	jing	32(0.64%)	gwan	36(0.59%)	yau	147(0.55%)	hwa	209(0.56%)

47	dai	32(0.64%)	dswo	36(0.59%)	di	147(0.55%)	ma	207(0.55%)
48	shung	32(0.64%)	jiau	35(0.57%)	tien	145(0.55%)	shun	206(0.55%)
49	chien	31(0.62%)	dwo	35(0.57%)	kuh	141(0.53%)	fang	200(0.53%)
50	fu	30(0.60%)	chien	33(0.54%)	yu:	140(0.53%)	ching	197(0.52%)
51	nin	30(0.60%)	rcang	33(0.54%)	fang	139(0.52%)	di	195(0.52%)
52	mei	30(0.60%)	fei	33(0.54%)	shun	137(0.52%)	chu:	195(0.52%)
53	yu:en	29(0.58%)	jiu	32(0.52%)	jing	132(0.50%)	huh	194(0.52%)
54	ma	29(0.58%)	hai	32(0.52%)	hau	131(0.49%)	jing	194(0.52%)
55	yu:	29(0.58%)	jieh	32(0.52%)	hwei	129(0.49%)	kuh	192(0.51%)
56	dswo	29(0.58%)	shang	31(0.50%)	chu:	128(0.48%)	xieh	191(0.51%)
57	jieh	28(0.56%)	xia	31(0.50%)	yu:en	122(0.46%)	yu:	191(0.51%)
58	hou	28(0.56%)	shung	31(0.50%)	muh	116(0.44%)	yu:en	189(0.50%)
59	rcung	27(0.54%)	jing	30(0.49%)	yang	116(0.44%)	muh	185(0.49%)
60	xu:eh	26(0.52%)	huh	30(0.49%)	xieh	115(0.43%)	hwei	183(0.49%)
61	dou	26(0.52%)	kuh	30(0.49%)	xin	115(0.43%)	hai	182(0.48%)
62	xieh	26(0.52%)	di	29(0.47%)	si	114(0.43%)	goong	181(0.48%)
63	hwei	26(0.52%)	chu:	29(0.47%)	xia	113(0.43%)	rcu	180(0.48%)
64	gwan	25(0.50%)	hwa	29(0.47%)	wang	113(0.43%)	liang	169(0.45%)
65	wu	25(0.50%)	mei	29(0.47%)	liang	108(0.41%)	dien	167(0.44%)
66	toong	24(0.48%)	xien	28(0.46%)	wun	106(0.40%)	xia	165(0.44%)
67	dien	24(0.48%)	san	28(0.46%)	jia	106(0.40%)	chien	162(0.43%)
68	jin	23(0.46%)	wun	28(0.46%)	dou	105(0.40%)	dwo	162(0.43%)
69	rju	23(0.46%)	hwei	28(0.46%)	hou	104(0.39%)	xing	162(0.43%)
70	xia	21(0.42%)	gwo	27(0.44%)	bien	104(0.39%)	jia	158(0.42%)
71	bien	21(0.42%)	ming	26(0.42%)	xing	103(0.39%)	hou	157(0.42%)
72	jiau	21(0.42%)	rji	26(0.42%)	shan	103(0.39%)	rcang	157(0.42%)
73	shun	21(0.42%)	mai	26(0.42%)	yen	102(0.38%)	si	154(0.41%)
74	kuh	21(0.42%)	hou	25(0.41%)	ming	102(0.38%)	jiau	153(0.41%)
75	fang	21(0.42%)	wei	25(0.41%)	xiau	102(0.38%)	dswo	152(0.40%)
76	na	20(0.40%)	gei	24(0.39%)	toong	100(0.38%)	xin	152(0.40%)
77	rjang	20(0.40%)	rju	24(0.39%)	goong	100(0.38%)	wun	151(0.40%)
78	xiang	20(0.40%)	kan	23(0.37%)	hai	98(0.37%)	xiau	151(0.40%)
79	lau	20(0.40%)	hwan	23(0.37%)	chien	98(0.37%)	doong	146(0.39%)
80	wei	20(0.40%)	jiang	22(0.36%)	shu	98(0.37%)	dou	144(0.38%)
81	lieh	20(0.40%)	run	22(0.36%)	yin	97(0.37%)	bien	142(0.38%)
82	wan	20(0.40%)	yu:	22(0.36%)	jiau	97(0.37%)	ming	142(0.38%)
83	yeh	19(0.38%)	si	22(0.36%)	rcung	93(0.35%)	yang	141(0.37%)
84	dan	19(0.38%)	xin	22(0.36%)	ching	93(0.35%)	jieh	138(0.37%)
85	nien	19(0.38%)	wan	21(0.34%)	rcang	92(0.35%)	wang	135(0.36%)
86	bai	19(0.38%)	tsoong	21(0.34%)	doong	92(0.35%)	toong	135(0.36%)
87	di	19(0.38%)	yu:oong	20(0.33%)	ran	90(0.34%)	gwan	134(0.36%)
88	muh	19(0.38%)	bau	20(0.33%)	rjung	90(0.34%)	rcung	131(0.35%)
89	bei	19(0.38%)	rjang	20(0.33%)	bai	89(0.34%)	shan	130(0.35%)
90	shan	18(0.36%)	piau	20(0.33%)	shou	89(0.34%)	shu	129(0.34%)
91	kwai	18(0.36%)	tsan	20(0.33%)	lau	88(0.33%)	bai	126(0.33%)
92	si	18(0.36%)	dai	20(0.33%)	bau	88(0.33%)	yen	123(0.33%)
93	lang	18(0.36%)	ju:	19(0.31%)	hu	88(0.33%)	rjang	121(0.32%)
94	tsan	18(0.36%)	doong	19(0.31%)	dwo	87(0.33%)	hun	120(0.32%)
95	gu	17(0.34%)	yu:eh	19(0.31%)	dswo	87(0.33%)	rju	120(0.32%)
96	wun	17(0.34%)	bei	19(0.31%)	dan	87(0.33%)	bau	119(0.32%)
97	yu:eh	17(0.34%)	jia	19(0.31%)	ting	82(0.31%)	shou	119(0.32%)
98	shu	17(0.34%)	shou	18(0.29%)	nien	81(0.31%)	yu:eh	117(0.31%)
99	ting	17(0.34%)	kai	18(0.29%)	rjang	81(0.31%)	rjung	117(0.31%)
100	hwa	17(0.34%)	wang	18(0.29%)	dang	81(0.31%)	yin	117(0.31%)
Total		3854(77.12%)		4764(77.52%)		19133(72.21%)		27354(72.71%)

Table 5.10 Listing of the first hundred frequently used syllables taking tone into consideration

Rank	Set I			Set II			Set III			Overall		
1	duh	0	167(3.35%)	wo	3	223(3.63%)	duh	0	1151(4.34%)	duh	0	1485(3.95%)
2	luh	0	126(2.53%)	shi	4	171(2.78%)	shi	4	583(2.20%)	shi	4	841(2.23%)
3	wo	3	96(1.92%)	duh	0	167(2.72%)	ta	1	433(1.63%)	wo	3	635(1.69%)
4	shi	4	87(1.74%)	yi	1	148(2.41%)	luh	0	393(1.48%)	luh	0	561(1.49%)
5	mun	0	74(1.48%)	ni	3	144(2.34%)	yi	4	342(1.29%)	ta	1	534(1.41%)
6	ta	1	65(1.30%)	rjuh	4	135(2.19%)	wo	3	316(1.19%)	you	3	450(1.20%)
7	run	2	65(1.30%)	you	3	102(1.66%)	you	3	293(1.11%)	rjuh	4	450(1.20%)
8	da	4	62(1.24%)	ma	0	102(1.66%)	yi	2	280(1.06%)	yi	4	413(1.10%)
9	yi	4	59(1.18%)	ching	3	84(1.37%)	dsai	4	269(1.02%)	dsai	4	403(1.07%)
10	guh	0	57(1.14%)	dsai	4	81(1.32%)	rjuh	4	261(0.98%)	ni	3	373(0.99%)
11	you	3	55(1.10%)	mun	0	74(1.20%)	bu	4	255(0.96%)	yi	2	334(0.89%)
12	rjuh	4	54(1.08%)	li	3	67(1.09%)	rji	1	252(0.95%)	run	2	333(0.88%)
13	dsai	4	53(1.06%)	tien	1	54(0.88%)	run	2	248(0.94%)	bu	4	323(0.86%)
14	yi	2	49(0.98%)	xiang	3	51(0.83%)	dau	4	211(0.80%)	dau	4	296(0.79%)
15	shi	2	44(0.88%)	muh	0	50(0.81%)	shi	2	197(0.74%)	shi	2	291(0.77%)
16	hai	2	44(0.88%)	shi	2	50(0.81%)	ni	3	188(0.71%)	mun	0	288(0.77%)
17	ni	3	41(0.82%)	yau	4	49(0.80%)	yeh	3	183(0.69%)	yi	1	279(0.74%)
18	jiu	4	40(0.80%)	lai	2	48(0.78%)	shwo	1	179(0.68%)	rji	1	263(0.70%)
19	hau	3	40(0.80%)	dau	4	48(0.78%)	guh	0	177(0.67%)	guh	0	235(0.62%)
20	dau	4	37(0.74%)	jien	1	45(0.73%)	jiu	4	170(0.64%)	da	4	232(0.62%)
21	tien	1	37(0.74%)	bu	4	45(0.73%)	rjuh	0	168(0.63%)	tien	1	225(0.60%)
22	shang	4	36(0.72%)	shun	2	44(0.72%)	shang	0	165(0.62%)	shwo	1	218(0.58%)
23	hun	3	36(0.72%)	luh	0	42(0.68%)	bu	2	158(0.60%)	jiu	4	217(0.58%)
24	shwo	1	34(0.68%)	ji	1	41(0.67%)	wei	4	154(0.58%)	rjuh	0	200(0.53%)
25	dwo	1	34(0.68%)	guh	4	39(0.63%)	li	0	153(0.58%)	yeh	3	198(0.53%)
26	jien	4	33(0.66%)	na	3	38(0.62%)	na	4	150(0.57%)	yau	4	195(0.52%)
27	goong	1	33(0.66%)	dien	4	37(0.60%)	da	4	147(0.55%)	rjoong	1	191(0.51%)
28	yau	4	32(0.64%)	chi	1	37(0.60%)	mun	0	140(0.53%)	yi	3	190(0.50%)
29	jia	1	32(0.64%)	rcuh	1	37(0.60%)	yi	3	137(0.52%)	dsi	0	187(0.50%)
30	ji	1	32(0.64%)	ww	3	36(0.59%)	kan	4	136(0.51%)	bu	2	185(0.49%)
31	jing	1	30(0.60%)	ta	1	34(0.55%)	tien	1	134(0.51%)	wei	4	182(0.48%)
32	nin	2	29(0.58%)	dwo	1	34(0.55%)	dsi	0	134(0.51%)	kan	4	181(0.48%)
33	rjoong	1	29(0.58%)	liang	3	33(0.54%)	huh	2	132(0.50%)	muh	0	176(0.47%)
34	yi	3	28(0.56%)	goong	1	32(0.52%)	rjoong	1	132(0.50%)	jien	4	175(0.46%)
35	rjuh	0	28(0.56%)	rjoong	1	30(0.49%)	mei	2	126(0.48%)	na	4	173(0.46%)
36	dsi	0	28(0.56%)	shau	3	29(0.47%)	jien	4	123(0.46%)	shang	0	170(0.45%)
37	rji	4	27(0.54%)	dswo	4	29(0.47%)	yau	4	114(0.43%)	huh	2	169(0.45%)
38	kan	4	27(0.54%)	you	2	29(0.47%)	shung	1	111(0.42%)	lai	2	169(0.45%)
39	rcung	2	27(0.54%)	shang	4	28(0.46%)	dsi	4	110(0.42%)	li	3	164(0.44%)
40	er	0	27(0.54%)	jin	1	27(0.44%)	di	4	110(0.42%)	li	0	162(0.43%)
41	yi	1	27(0.54%)	dien	3	27(0.44%)	muh	0	107(0.40%)	xiang	3	159(0.42%)
42	chien	2	25(0.50%)	fei	1	27(0.44%)	lai	0	105(0.40%)	ma	0	154(0.41%)
43	dswo	4	25(0.50%)	fang	2	26(0.42%)	yi	1	104(0.39%)	mei	2	152(0.40%)
44	li	3	24(0.48%)	di	4	26(0.42%)	lai	2	103(0.39%)	ji	1	151(0.40%)
45	chu:	4	24(0.48%)	yuien	2	26(0.42%)	xiang	3	102(0.38%)	di	4	150(0.40%)
46	yuien	2	24(0.48%)	san	1	26(0.42%)	kuh	3	99(0.37%)	hau	3	146(0.39%)

47	dai	4	24(0.48%)	ming	2	26(0.42%)	jin	4	97(0.37%)	dwo	1	141(0.37%)
48	xiau	3	24(0.48%)	dai	0	25(0.41%)	ba	3	95(0.36%)	hai	2	139(0.37%)
49	dou	1	24(0.48%)	yi	3	25(0.41%)	rou	1	90(0.34%)	goong	1	139(0.37%)
50	toong	2	23(0.46%)	chien	2	25(0.41%)	chi	2	90(0.34%)	dawo	4	137(0.36%)
51	bu	4	23(0.46%)	jiu	3	25(0.41%)	rji	4	89(0.34%)	jia	1	130(0.35%)
52	gwo	2	23(0.46%)	chu:	4	24(0.39%)	ran	2	88(0.33%)	shang	4	129(0.34%)
53	ba	0	23(0.46%)	hau	3	24(0.39%)	dou	1	88(0.33%)	dai	4	128(0.34%)
54	xuieh	2	23(0.46%)	gei	3	24(0.39%)	li	4	87(0.33%)	kuh	3	128(0.34%)
55	rcang	2	23(0.46%)	xi	3	24(0.39%)	shan	1	87(0.33%)	shung	1	127(0.34%)
56	doong	1	23(0.46%)	hwa	4	23(0.37%)	jia	1	84(0.32%)	ba	3	127(0.34%)
57	ba	3	22(0.44%)	xia	4	23(0.37%)	dawo	4	83(0.31%)	shun	2	127(0.34%)
58	xien	1	22(0.44%)	da	4	23(0.37%)	rji	3	83(0.31%)	er	0	126(0.33%)
59	lieh	4	20(0.40%)	xing	2	22(0.36%)	rcung	2	82(0.31%)	lai	0	126(0.33%)
60	lai	0	20(0.40%)	xing	1	22(0.36%)	hau	3	82(0.31%)	rji	4	125(0.33%)
61	nien	2	19(0.38%)	hou	4	22(0.36%)	er	2	80(0.30%)	jin	4	123(0.33%)
62	muh	0	19(0.38%)	huh	2	22(0.36%)	yu:	2	80(0.30%)	yuien	2	122(0.32%)
63	hou	4	19(0.38%)	tsoong	2	21(0.34%)	ji	3	79(0.30%)	dou	1	120(0.32%)
64	wu	3	19(0.38%)	hwei	4	21(0.34%)	ming	2	79(0.30%)	xia	4	119(0.32%)
65	kwai	4	18(0.36%)	shung	0	21(0.34%)	xia	4	79(0.30%)	rcung	2	118(0.31%)
66	lang	2	18(0.36%)	jing	1	21(0.34%)	hwei	4	79(0.30%)	hun	3	115(0.31%)
67	tsan	1	18(0.36%)	er	0	21(0.34%)	er	0	78(0.29%)	rcu	1	113(0.30%)
68	er	4	18(0.36%)	hai	2	21(0.34%)	xin	1	78(0.29%)	hwei	4	113(0.30%)
69	shan	1	18(0.36%)	gwan	1	21(0.34%)	ji	1	78(0.29%)	ming	2	112(0.30%)
70	lái	2	18(0.36%)	piau	4	20(0.33%)	xiang	4	77(0.29%)	shan	1	111(0.29%)
71	ma	0	18(0.36%)	ba	0	20(0.33%)	hwa	4	75(0.28%)	ji	3	110(0.29%)
72	xia	4	17(0.34%)	tsan	1	20(0.33%)	niu	2	74(0.28%)	jien	1	110(0.29%)
73	jiau	4	17(0.34%)	run	2	20(0.33%)	goong	1	74(0.28%)	chu:	4	109(0.29%)
74	fu	0	17(0.34%)	jien	4	19(0.31%)	hai	2	74(0.28%)	rcuh	1	109(0.29%)
75	lau	3	16(0.32%)	wu	4	19(0.31%)	xien	4	73(0.28%)	ching	3	106(0.28%)
76	gwan	1	16(0.32%)	er	4	19(0.31%)	dwo	1	73(0.28%)	li	4	106(0.28%)
77	bai	2	16(0.32%)	xieh	1	19(0.31%)	li	3	73(0.28%)	hwa	4	106(0.28%)
78	ting	1	16(0.32%)	dai	4	19(0.31%)	shun	2	73(0.28%)	chien	2	104(0.28%)
79	xi	2	15(0.30%)	xi	1	19(0.31%)	yuien	2	72(0.27%)	liang	3	104(0.28%)
80	dan	4	15(0.30%)	rcang	3	18(0.29%)	toong	2	71(0.27%)	jing	1	102(0.27%)
81	bu	2	15(0.30%)	rju	4	18(0.29%)	you	4	70(0.26%)	yu:	2	99(0.26%)
82	shung	0	15(0.30%)	mai	3	18(0.29%)	hun	3	69(0.26%)	xien	1	99(0.26%)
83	huh	2	15(0.30%)	kai	1	18(0.29%)	fang	3	69(0.26%)	toong	2	99(0.26%)
84	bien	1	15(0.30%)	kuh	3	18(0.29%)	tsoong	2	68(0.26%)	jiau	4	98(0.26%)
85	gwo	1	15(0.30%)	kan	4	18(0.29%)	rcuh	1	68(0.26%)	xin	1	98(0.26%)
86	chi	3	14(0.28%)	ji	3	18(0.29%)	dwei	4	67(0.25%)	hou	4	98(0.26%)
87	di	4	14(0.28%)	ji	4	18(0.29%)	wu	2	67(0.25%)	chi	2	98(0.26%)
88	mei	2	14(0.28%)	wun	4	18(0.29%)	yang	4	67(0.25%)	dien	3	98(0.26%)
89	rjang	1	14(0.28%)	rjang	1	17(0.28%)	hwa	1	66(0.25%)	xiau	3	97(0.26%)
90	dsou	3	14(0.28%)	gwo	2	16(0.26%)	shang	4	65(0.25%)	wu	3	97(0.26%)
91	yuieh	4	14(0.28%)	kwai	4	16(0.26%)	jiau	4	65(0.25%)	tsoong	2	96(0.26%)
92	san	1	14(0.28%)	xien	1	16(0.26%)	ri	4	63(0.24%)	xiang	4	96(0.26%)
93	li	4	14(0.28%)	yuioong	4	16(0.26%)	ji	4	63(0.24%)	ran	2	96(0.26%)
94	jin	4	14(0.28%)	jiau	4	16(0.26%)	hwang	2	62(0.23%)	gwo	2	95(0.25%)
95	kai	1	14(0.28%)	wei	4	16(0.26%)	xiau	3	62(0.23%)	gwan	1	95(0.25%)
96	chiu	2	13(0.26%)	chi	4	15(0.24%)	nien	2	61(0.23%)	xien	4	94(0.25%)
97	hwei	4	13(0.26%)	rjan	4	15(0.24%)	nu:	3	61(0.23%)	dwei	4	92(0.24%)
98	ji	3	13(0.26%)	xieh	4	15(0.24%)	xien	1	61(0.23%)	rji	3	92(0.24%)
99	min	2	13(0.26%)	dsau	3	14(0.23%)	chu:	4	61(0.23%)	nien	2	90(0.24%)
100	bei	3	13(0.26%)	dwei	4	14(0.23%)	mei	3	61(0.23%)	rcang	2	90(0.24%)
Total			3027(60.56%)			3820(62.11%)			13775(52.01%)			19634(52.17%)

Table 5.11 The total frequency of occurrences and percentage of frequency of occurrences of each entry phoneme in n-gram analysis

Starting	Frequency Of Occurrence						Percentage Of Frequency Of Occurrence					
Phoneme	Tone 1	Tone 2	Tone 3	Tone 4	Neutral	Sum	Tone 1	Tone 2	Tone 3	Tone 4	Neutral	
b	3678	4150	4674	8319	1384	22205	0.3087	0.3483	0.3923	0.6983	0.1162	1.8638
p	761	1821	606	1793	404	5385	0.0639	0.1528	0.0509	0.1505	0.0339	0.4520
m	1418	6054	3841	3969	6414	21696	0.1190	0.5081	0.3224	0.3331	0.5384	1.8211
f	3413	1713	1626	2231	883	9866	0.2865	0.1438	0.1365	0.1873	0.0741	0.8281
d	8682	4676	6226	15875	14485	49944	0.7287	0.3925	0.5226	1.3325	1.2158	4.1921
t	10120	4423	1900	2974	1436	20853	0.8494	0.3712	0.1595	0.2496	0.1205	1.7503
n	19182	23034	18283	26014	12294	98807	1.6100	1.9334	1.5346	2.1835	1.0319	8.2934
l	1706	6748	5771	5843	6323	26391	0.1432	0.5664	0.4844	0.4904	0.5307	2.2151
g	6456	2638	4554	5404	3413	22465	0.5419	0.2214	0.3822	0.4536	0.2865	1.8856
k	1657	474	2690	4716	812	10349	0.1391	0.0398	0.2258	0.3958	0.0682	0.8686
h	3209	7514	4315	6471	1586	23095	0.2693	0.6307	0.3622	0.5431	0.1331	1.9385
j	9540	2866	4896	11516	1983	30801	0.8007	0.2406	0.4109	0.9666	0.1664	2.5853
ch	3319	4639	2855	3402	1271	15486	0.2786	0.3894	0.2396	0.2855	0.1067	1.2998
x	7166	3542	5884	7220	1982	25794	0.6015	0.2973	0.4939	0.6060	0.1664	2.1650
rj	9338	2554	4745	11134	2978	30749	0.7838	0.2144	0.3983	0.9345	0.2500	2.5809
rc	4228	5012	1741	1580	870	13431	0.3549	0.4207	0.1461	0.1326	0.0730	1.1273
sh	8989	6478	4018	13631	4766	37882	0.7545	0.5437	0.3373	1.1441	0.4000	3.1796
r	460	5708	518	2460	314	9460	0.0386	0.4791	0.0435	0.2065	0.0264	0.7940
da	1505	1985	3991	8948	1938	18367	0.1263	0.1666	0.3350	0.7511	0.1627	1.5416
ta	1228	2411	944	1061	133	5777	0.1031	0.2024	0.0792	0.0891	0.0112	0.4849
s	2347	537	1682	2607	422	7595	0.1970	0.0451	0.1412	0.2188	0.0354	0.6375
ng	15530	15192	11362	18535	6982	67601	1.3035	1.2751	0.9537	1.5557	0.5860	5.6741
w	10709	10078	13351	16993	4084	55215	0.8989	0.8459	1.1206	1.4263	0.3428	4.6345
y	7490	10278	11734	15609	2131	47242	0.6287	0.8627	0.9849	1.3101	0.1789	3.9653
a	20151	14308	14919	23452	8972	81802	1.6914	1.2009	1.2522	1.9685	0.7531	6.8661
e(eh)	8239	8102	6156	11098	3364	36959	0.6915	0.6800	0.5167	0.9315	0.2824	3.1022
i	32686	25802	27728	44351	14478	145045	2.7435	2.1657	2.3274	3.7226	1.2152	12.1744
o	3424	3189	6198	4922	2955	20688	0.2874	0.2677	0.5202	0.4131	0.2480	1.7365
oo(u)	10363	10233	6664	13307	3255	43822	0.8698	0.8589	0.5593	1.1169	0.2732	3.6782
u:	3007	5205	2960	5856	1138	18166	0.2524	0.4369	0.2484	0.4915	0.0955	1.5248
u(uh)	13979	17802	10765	19500	18697	80743	1.1733	1.4942	0.9036	1.6367	1.5693	6.7772
er	455	897	475	857	578	3262	0.0382	0.0753	0.0399	0.0719	0.0485	0.2738
ai	3297	4319	3375	7371	2323	20685	0.2767	0.3625	0.2833	0.6187	0.1950	1.7362
ei	2755	2867	3695	4285	1028	14630	0.2312	0.2406	0.3101	0.3597	0.0863	1.2280
au	4032	4003	5658	8681	2445	24819	0.3384	0.3360	0.4749	0.7286	0.2052	2.0832
ou	3138	2954	5101	4645	1559	17397	0.2634	0.2479	0.4282	0.3899	0.1309	1.4602
iu	1077	1555	1172	2744	371	6919	0.0904	0.1305	0.0984	0.2303	0.0311	0.5807
Total	248734	235761	217073	349374	140451	1191393	20.8776	19.7887	18.2201	29.3248	11.7888	100.0000

Table 5.12 The First Hundred Frequently Appearing Phoneme Strings In N-gram Analysis

List Of Phonemes	Frequency Of Occurrence						Percentage Of Frequency Of Occurrence					
	Tone 1	Tone 2	Tone 3	Tone 4	Neutral	Sum	Tone 1	Tone 2	Tone 3	Tone 4	Neutral	Sum
b	210	303	377	638	138	1666	0.0176	0.0254	0.0316	0.0536	0.0116	0.1398
b u	0	185	12	323	30	550	0.0000	0.0155	0.0010	0.0271	0.0025	0.0462
	0	185	12	323	30	550	0.0000	0.0155	0.0010	0.0271	0.0025	0.0462
p	47	158	47	147	3	402	0.0039	0.0133	0.0039	0.0123	0.0003	0.0337
m	28	504	256	245	647	1680	0.0024	0.0423	0.0215	0.0206	0.0543	0.1410
m i	0	187	36	121	20	364	0.0000	0.0157	0.0030	0.0102	0.0017	0.0306
	0	187	36	121	20	364	0.0000	0.0157	0.0030	0.0102	0.0017	0.0306
m uh	11	60	3	1	464	539	0.0009	0.0050	0.0003	0.0001	0.0389	0.0452
	11	60	3	1	464	539	0.0009	0.0050	0.0003	0.0001	0.0389	0.0452
f	284	132	125	144	79	764	0.0238	0.0111	0.0105	0.0121	0.0066	0.0641
d	471	134	312	1189	1550	3656	0.0395	0.0112	0.0262	0.0998	0.1301	0.3069
d a	68	17	104	318	0	507	0.0057	0.0014	0.0087	0.0267	0.0000	0.0426
	68	17	104	318	0	507	0.0057	0.0014	0.0087	0.0267	0.0000	0.0426
d i	18	37	118	275	8	456	0.0015	0.0031	0.0099	0.0231	0.0007	0.0383
	18	37	118	275	8	456	0.0015	0.0031	0.0099	0.0231	0.0007	0.0383
d uh	18	41	25	4	1485	1573	0.0015	0.0034	0.0021	0.0003	0.1246	0.1320
	18	41	25	4	1485	1573	0.0015	0.0034	0.0021	0.0003	0.1246	0.1320
d au	10	1	16	296	44	367	0.0008	0.0001	0.0013	0.0248	0.0037	0.0308
	10	1	16	296	44	367	0.0008	0.0001	0.0013	0.0248	0.0037	0.0308
t	917	348	98	124	37	1524	0.0770	0.0292	0.0082	0.0104	0.0031	0.1279
t a	544	39	15	9	3	610	0.0457	0.0033	0.0013	0.0008	0.0003	0.0512
	544	39	15	9	3	610	0.0457	0.0033	0.0013	0.0008	0.0003	0.0512
t i	311	81	61	16	9	478	0.0261	0.0068	0.0051	0.0013	0.0008	0.0401
	311	81	61	16	9	478	0.0261	0.0068	0.0051	0.0013	0.0008	0.0401

List Of Phonemes		Frequency Of Occurrence						Percentage Of Frequency Of Occurrence					
		Tone 1	Tone 2	Tone 3	Tone 4	Neutral	Sum	Tone 1	Tone 2	Tone 3	Tone 4	Neutral	Sum
n		1752	2119	1455	1855	533	7714	0.1471	0.1779	0.1221	0.1557	0.0447	0.6475
n	m	53	212	46	38	16	365	0.0044	0.0178	0.0039	0.0032	0.0013	0.0306
		3	88	35	37	202	365	0.0003	0.0074	0.0029	0.0031	0.0170	0.0306
n	d	142	141	129	144	52	608	0.0119	0.0118	0.0108	0.0121	0.0044	0.0510
		100	14	50	155	289	608	0.0084	0.0012	0.0042	0.0130	0.0243	0.0510
n	sh	187	76	72	132	30	497	0.0157	0.0064	0.0060	0.0111	0.0025	0.0417
		117	51	37	180	112	497	0.0098	0.0043	0.0031	0.0151	0.0094	0.0417
n	y	98	112	64	117	49	440	0.0082	0.0094	0.0054	0.0098	0.0041	0.0369
		53	83	142	159	3	440	0.0044	0.0070	0.0119	0.0133	0.0003	0.0369
n	i	0	177	381	28	16	602	0.0000	0.0149	0.0320	0.0024	0.0013	0.0505
		0	177	381	28	16	602	0.0000	0.0149	0.0320	0.0024	0.0013	0.0505
l		30	558	469	377	872	2306	0.0025	0.0468	0.0394	0.0316	0.0732	0.1936
l	i	0	180	307	206	171	864	0.0000	0.0151	0.0258	0.0173	0.0144	0.0725
		0	180	307	206	171	864	0.0000	0.0151	0.0258	0.0173	0.0144	0.0725
l	uh	1	0	6	19	561	587	0.0001	0.0000	0.0005	0.0016	0.0471	0.0493
		1	0	6	19	561	587	0.0001	0.0000	0.0005	0.0016	0.0471	0.0493
g		490	128	340	344	288	1590	0.0411	0.0107	0.0285	0.0289	0.0242	0.1335
g	w	151	95	107	119	52	524	0.0127	0.0080	0.0090	0.0100	0.0044	0.0440
		151	95	107	119	52	524	0.0127	0.0080	0.0090	0.0100	0.0044	0.0440
g	uh	82	32	0	82	235	431	0.0069	0.0027	0.0000	0.0069	0.0197	0.0362
		82	32	0	82	235	431	0.0069	0.0027	0.0000	0.0069	0.0197	0.0362
k		143	5	225	367	27	767	0.0120	0.0004	0.0189	0.0308	0.0023	0.0644
h		207	633	343	502	85	1770	0.0174	0.0531	0.0288	0.0421	0.0071	0.1486
h	w	108	205	36	296	24	669	0.0091	0.0172	0.0030	0.0248	0.0020	0.0562
		108	205	36	296	24	669	0.0091	0.0172	0.0030	0.0248	0.0020	0.0562

List Of Phonemes	Frequency Of Occurrence						Percentage Of Frequency Of Occurrence					
	Tone 1	Tone 2	Tone 3	Tone 4	Neutral	Sum	Tone 1	Tone 2	Tone 3	Tone 4	Neutral	Sum
ɔ	750	178	366	840	86	2220	0.0630	0.0149	0.0307	0.0705	0.0072	0.1863
ɔ i	684	114	268	581	83	1730	0.0574	0.0096	0.0225	0.0488	0.0070	0.1452
	684	114	268	581	83	1730	0.0574	0.0096	0.0225	0.0488	0.0070	0.1452
ɔ i eh	149	47	51	201	45	493	0.0125	0.0039	0.0043	0.0169	0.0038	0.0414
	149	47	51	201	45	493	0.0125	0.0039	0.0043	0.0169	0.0038	0.0414
	149	47	51	201	45	493	0.0125	0.0039	0.0043	0.0169	0.0038	0.0414
ch	240	382	222	250	119	1213	0.0201	0.0321	0.0186	0.0210	0.0100	0.1018
ch i	213	285	210	86	54	848	0.0179	0.0239	0.0176	0.0072	0.0045	0.0712
	213	285	210	86	54	848	0.0179	0.0239	0.0176	0.0072	0.0045	0.0712
x	534	254	455	534	116	1893	0.0448	0.0213	0.0382	0.0448	0.0097	0.1589
x i	460	166	399	489	116	1630	0.0386	0.0139	0.0335	0.0410	0.0097	0.1368
	460	166	399	489	116	1630	0.0386	0.0139	0.0335	0.0410	0.0097	0.1368
x i a	48	7	161	215	39	470	0.0040	0.0006	0.0135	0.0180	0.0033	0.0394
	48	7	161	215	39	470	0.0040	0.0006	0.0135	0.0180	0.0033	0.0394
	48	7	161	215	39	470	0.0040	0.0006	0.0135	0.0180	0.0033	0.0394
x i eh	173	42	79	123	16	433	0.0145	0.0035	0.0066	0.0103	0.0013	0.0363
	173	42	79	123	16	433	0.0145	0.0035	0.0066	0.0103	0.0013	0.0363
	173	42	79	123	16	433	0.0145	0.0035	0.0066	0.0103	0.0013	0.0363
rɔ	764	83	319	863	220	2249	0.0641	0.0070	0.0268	0.0724	0.0185	0.1888
rɔ i	263	43	92	125	2	525	0.0221	0.0036	0.0077	0.0105	0.0002	0.0441
	263	43	92	125	2	525	0.0221	0.0036	0.0077	0.0105	0.0002	0.0441
rɔ u	205	12	80	99	11	407	0.0172	0.0010	0.0067	0.0083	0.0009	0.0342
	205	12	80	99	11	407	0.0172	0.0010	0.0067	0.0083	0.0009	0.0342
rɔ uh	85	10	71	529	204	899	0.0071	0.0008	0.0060	0.0444	0.0171	0.0755
	85	10	71	529	204	899	0.0071	0.0008	0.0060	0.0444	0.0171	0.0755

List Of Phonemes	Frequency Of Occurrence						Percentage Of Frequency Of Occurrence					
	Tone 1	Tone 2	Tone 3	Tone 4	Neutral	Sum	Tone 1	Tone 2	Tone 3	Tone 4	Neutral	Sum
rc	382	400	115	67	38	1002	0.0321	0.0336	0.0097	0.0056	0.0032	0.0841
sh	715	477	248	1196	292	2928	0.0600	0.0400	0.0208	0.1004	0.0245	0.2458
sh a	158	0	9	146	170	483	0.0133	0.0000	0.0008	0.0123	0.0143	0.0405
	158	0	9	146	170	483	0.0133	0.0000	0.0008	0.0123	0.0143	0.0405
sh i	39	291	52	841	40	1263	0 0033	0.0244	0.0044	0.0706	0.0034	0.1060
	39	291	52	841	40	1263	0.0033	0.0244	0.0044	0.0706	0.0034	0.1060
sh uh	193	152	8	72	70	495	0.0162	0.0128	0.0007	0.0060	0.0059	0.0415
	193	152	8	72	70	495	0.0162	0.0128	0.0007	0.0060	0.0059	0.0415
r	2	496	12	203	9	722	0.0002	0.0416	0.0010	0.0170	0.0008	0.0606
r uh	2	340	2	48	6	398	0.0002	0.0285	0.0002	0.0040	0.0005	0.0334
	2	340	2	48	6	398	0.0002	0.0285	0.0002	0.0040	0.0005	0.0334
r u n	0	333	1	32	6	372	0.0000	0.0280	0.0001	0.0027	0.0005	0.0312
	0	333	1	32	6	372	0.0000	0.0280	0.0001	0.0027	0.0005	0.0312
	0	333	1	32	6	372	0.0000	0.0280	0.0001	0.0027	0.0005	0.0312
ds	47	86	289	764	188	1374	0.0039	0.0072	0.0243	0.0641	0.0158	0.1153
ds i	15	1	48	128	187	379	0.0013	0.0001	0.0040	0.0107	0.0157	0.0318
	15	1	48	128	187	379	0.0013	0.0001	0.0040	0.0107	0.0157	0.0318
as ai	8	0	2	403	0	413	0.0007	0.0000	0.0002	0.0338	0.0000	0.0347
	8	0	2	403	0	413	0.0007	0.0000	0.0002	0.0338	0.0000	0.0347
ts	72	194	83	87	1	437	0.0060	0.0163	0.0070	0.0073	0.0001	0.0367
s	187	13	126	231	16	573	0.0157	0.0011	0.0106	0.0194	0.0013	0.0481
ng	1515	1419	939	1226	326	5425	0.1272	0.1191	0.0788	0.1029	0.0274	0.4553
ng a	118	134	95	134	38	519	0.0099	0.0112	0.0080	0.0112	0.0032	0.0436
	43	14	26	164	272	519	0.0036	0.0012	0.0022	0.0138	0.0228	0.0436
ng y	124	106	69	72	24	395	0.0104	0.0089	0.0058	0.0060	0.0020	0.0332
	64	109	95	115	12	395	0.0054	0.0091	0.0080	0.0097	0.0010	0.0332

List Of Phonemes	Frequency Of Occurrence						Percentage Of Frequency Of Occurrence					
	Tone 1	Tone 2	Tone 3	Tone 4	Neutral	Sum	Tone 1	Tone 2	Tone 3	Tone 4	Neutral	Sum
w	894	679	1157	1400	111	4241	0.0750	0.0570	0.0971	0.1175	0.0093	0.3560
w a	358	230	154	323	29	1094	0.0300	0.0193	0.0129	0.0271	0.0024	0.0918
	358	230	154	323	29	1094	0.0300	0.0193	0.0129	0.0271	0.0024	0.0918
w a n	180	75	108	118	12	493	0.0151	0.0063	0.0091	0.0099	0.0010	0.0414
	180	75	108	118	12	493	0.0151	0.0063	0.0091	0.0099	0.0010	0.0414
	179	76	110	114	14	493	0.0150	0.0064	0.0092	0.0096	0.0012	0.0414
w o	415	162	772	294	65	1708	0.0348	0.0136	0.0648	0.0247	0.0055	0.1434
	415	162	772	294	65	1708	0.0348	0.0136	0.0648	0.0247	0.0055	0.1434
w ei	86	125	128	530	2	871	0.0072	0.0105	0.0107	0.0445	0.0002	0.0731
	86	125	128	530	2	871	0.0072	0.0105	0.0107	0.0445	0.0002	0.0731
y	482	777	1026	1132	67	3484	0.0405	0.0652	0.0861	0.0950	0.0056	0.2924
y eh	20	43	252	60	8	383	0.0017	0.0036	0.0212	0.0050	0.0007	0.0321
	20	43	252	60	8	383	0.0017	0.0036	0.0212	0.0050	0.0007	0.0321
y i	373	356	221	442	50	1442	0.0313	0.0299	0.0185	0.0371	0.0042	0.1210
	373	356	221	442	50	1442	0.0313	0.0299	0.0185	0.0371	0.0042	0.1210
y u:	48	238	82	255	1	624	0.0040	0.0200	0.0069	0.0214	0.0001	0.0524
	48	238	82	255	1	624	0.0040	0.0200	0.0069	0.0214	0.0001	0.0524
y ou	6	76	450	87	7	626	0.0005	0.0064	0.0378	0.0073	0.0006	0.0525
	6	76	450	87	7	626	0.0005	0.0064	0.0378	0.0073	0.0006	0.0525
a	1910	976	1211	1901	652	6650	0.1603	0.0819	0.1016	0.1596	0.0547	0.5582
a n	577	309	343	628	44	1901	0.0484	0.0259	0.0288	0.0527	0.0037	0.1596
	565	321	352	615	48	1901	0.0474	0.0269	0.0295	0.0516	0.0040	0.1596
a ng	431	493	573	562	212	2271	0.0362	0.0414	0.0481	0.0472	0.0178	0.1906
	431	493	573	562	212	2271	0.0362	0.0414	0.0481	0.0472	0.0178	0.1906
eh	778	726	614	849	108	3075	0.0653	0.0609	0.0515	0.0713	0.0091	0.2581
eh n	614	511	286	583	91	2085	0.0515	0.0429	0.0240	0.0489	0.0076	0.1750
	612	527	288	524	84	2085	0.0514	0.0442	0.0242	0.0482	0.0071	0.1750

List Of Phonemes	Frequency Of Occurrence						Percentage Of Frequency Of Occurrence					
	Tone 1	Tone 2	Tone 3	Tone 4	Neutral	Sum	Tone 1	Tone 2	Tone 3	Tone 4	Neutral	Sum
i	2568	2058	2364	3711	784	11485	0.2155	0.1727	0.1984	0.3115	0.0658	0.9640
i d	127	112	134	190	41	604	0.0107	0.0094	0.0112	0.0159	0.0034	0.0507
	44	21	76	163	300	604	0.0037	0.0018	0.0064	0.0137	0.0252	0.0507
i n	317	176	95	255	15	858	0.0266	0.0148	0.0080	0.0214	0.0013	0.0720
	268	220	116	238	16	858	0.0225	0.0185	0.0097	0.0200	0.0013	0.0720
i j	52	71	101	142	19	385	0.0044	0.0060	0.0085	0.0119	0.0016	0.0323
	98	22	111	134	20	385	0.0082	0.0018	0.0093	0.0112	0.0017	0.0323
i sh	83	40	69	146	26	364	0.0070	0.0034	0.0058	0.0123	0.0022	0.0306
	94	64	32	127	47	364	0.0079	0.0054	0.0027	0.0107	0.0039	0.0306
i ng	309	342	192	229	38	1110	0.0259	0.0287	0.0161	0.0192	0.0032	0.0932
	309	342	192	229	38	1110	0.0259	0.0287	0.0161	0.0192	0.0032	0.0932
i y	59	44	127	155	35	420	0.0050	0.0037	0.0107	0.0130	0.0029	0.0353
	47	88	128	148	9	420	0.0039	0.0074	0.0107	0.0124	0.0008	0.0353
i a	231	60	305	283	110	989	0.0194	0.0050	0.0256	0.0238	0.0092	0.0830
	238	59	305	276	111	989	0.0200	0.0050	0.0256	0.0232	0.0093	0.0830
i a ng	97	54	294	138	25	608	0.0081	0.0045	0.0247	0.0116	0.0021	0.0510
	97	53	294	138	26	608	0.0081	0.0044	0.0247	0.0116	0.0022	0.0510
	97	53	294	138	26	608	0.0081	0.0044	0.0247	0.0116	0.0022	0.0510
i eh	685	410	296	617	103	2111	0.0575	0.0344	0.0248	0.0518	0.0086	0.1772
	687	410	296	617	101	2111	0.0577	0.0344	0.0248	0.0518	0.0085	0.1772
i eh n	565	308	174	526	91	1664	0.0474	0.0259	0.0146	0.0441	0.0076	0.1397
	565	308	174	526	91	1664	0.0474	0.0259	0.0146	0.0441	0.0076	0.1397
	563	319	183	516	83	1664	0.0473	0.0268	0.0154	0.0433	0.0070	0.1397
i au	74	74	163	195	1	507	0.0062	0.0062	0.0137	0.0164	0.0001	0.0426
	74	74	163	195	1	507	0.0062	0.0062	0.0137	0.0164	0.0001	0.0426
o	435	206	777	332	76	1826	0.0365	0.0173	0.0652	0.0279	0.0064	0.1533

List Of Phonemes	Frequency Of Occurrence						Percentage Of Frequency Of Occurrence					
	Tone 1	Tone 2	Tone 3	Tone 4	Neutral	Sum	Tone 1	Tone 2	Tone 3	Tone 4	Neutral	Sum
u	871	826	487	1103	152	3439	0.0731	0.0693	0.0409	0.0926	0.0128	0.2887
oo ng	502	314	121	289	2	1228	0.0421	0.0264	0.0102	0.0243	0.0002	0.1031
	502	314	121	289	2	1228	0.0421	0.0264	0.0102	0.0243	0.0002	0.1031
u:	179	466	263	516	68	1492	0.0150	0.0391	0.0221	0.0433	0.0057	0.1252
u: eh :	73	273	66	172	0	584	0.0061	0.0229	0.0055	0.0144	0.0000	0.0490
	73	273	66	172	0	584	0.0061	0.0229	0.0055	0.0144	0.0000	0.0490
uh	683	1212	447	992	3091	6425	0.0573	0.1017	0.0375	0.0833	0.2594	0.5393
u n	225	647	207	159	384	1622	0.0189	0.0543	0.0174	0.0133	0.0322	0.1361
	223	671	226	181	321	1622	0.0187	0.0563	0.0190	0.0152	0.0269	0.1361
uh sh	16	25	37	99	230	407	0.0013	0.0021	0.0031	0.0083	0.0193	0.0342
	88	93	31	186	9	407	0.0074	0.0078	0.0026	0.0156	0.0008	0.0342
u ng	273	270	53	146	74	816	0.0229	0.0227	0.0044	0.0123	0.0062	0.0685
	273	270	53	146	74	816	0.0229	0.0227	0.0044	0.0123	0.0062	0.0685
uh y	11	20	51	74	241	397	0.0009	0.0017	0.0043	0.0062	0.0202	0.0333
	53	95	95	148	6	397	0.0044	0.0080	0.0080	0.0124	0.0005	0.0333
ai	145	449	171	805	149	1719	0.0122	0.0377	0.0144	0.0676	0.0125	0.1443
ei	182	295	356	628	6	1467	0.0153	0.0248	0.0299	0.0527	0.0005	0.1231
au	206	228	649	929	59	2071	0.0173	0.0191	0.0545	0.0780	0.0050	0.1738
ou	208	176	652	266	85	1407	0.0175	0.0148	0.0547	0.0240	0.0071	0.1181
iu	44	151	86	250	4	535	0.0037	0.0127	0.0072	0.0210	0.0003	0.0449
Total	43351	38487	36453	57745	24398	200434	3.6387	3.2304	3.0597	4.8468	2.0479	16.8235

Table 5.13 The occurrence frequency of clauses, words, syllables, phonemes, and letters of the database

	Set 1	Set 2	Set 3	Total
Total No. Of Clauses	752	1109	3390	5251
Total No. Of Words	3235	3998	17038	24271
Total No. Of Syllables	4989	6151	26499	37639
Total No. Of Phonemes	12457	15297	65741	93495
Total No. Of Pinyin Letters (counting tone as a letter)	19046	23863	101407	144316
Total No. Of Suen's Phonetic Letters (without counting tone)	16400	20125	87076	123601
Syllables per Word	1.54	1.54	1.56	1.55
Phonemes per Syllable	2.50	2.49	2.48	2.48
Pinyin Letters per Syllable	3.82	3.88	3.83	3.83
Suen's Phonetic Letters per Syllable (without counting tone)	3.29	3.27	3.29	3.28
Suen's Phonetic Letters per Syllable (counting tone as a letter)	4.29	4.27	4.29	4.28

Table 5.14 The occurrence frequency of consonants, semi-vowels, vowels, and diphthongs.

	Set 1	Set 2	Set 3	Total
Consonants	5880 47.20%	7057 46.13%	30938 47.06%	43875 46.93%
Semi-vowels	1001 8.04%	1304 8.52%	5420 8.24%	7725 8.26%
Vowels	4554 36.56%	5679 37.12%	24463 37.21%	34696 37.11%
Diphthongs	1022 8.20%	1257 8.22%	4920 7.48%	7199 7.70%

Chapter VI

Conclusions

The Pinyin, Chan's and Suen's phonetic systems have been studied in detail in this thesis. An algorithm has been implemented to segment each Pinyin word into syllables and convert each syllable into Chan's codes. Some analytical results have been obtained and listed in Chapter Five.

In this Chapter, the problems encountered during the segmentation process are discussed. Later, Chan's system is compared with Suen's based on the criteria for an ideal phonetic system and Suen's results published in 1979. Finally, possible improvements and future studies on Mandarin are discussed in the end.

6.1 The Problems In Segmentation

The algorithms presented in Chapter Five attempts to segment each Pinyin word into syllables at the proper position. However, several difficult cases have been encountered in determining the proper position for segmentation of syllables in Pinyin system and they are listed below.

(1) The letter [n] is preceded by letter [a] or [e] or [i]

or [o] or [u]. This [n] may be or may not be the initial consonant of the following syllable, if the following letter is either [a], or [e] or [o]. For example, /he3ne4/, where [n] can be considered as the ending consonant of the first syllable or the initial consonant of the next syllable. If this word is divided into two syllables, based on the fact that the vowel starting syllables have lower frequency, it becomes /he3/ and /ne4/, which has no meaning at all. This word should be interpreted as /he3n/(很) and /e4/(餓), it means very hungry.

(2) The letter [g] is preceded by the letter [n].
This letter [g] may be part of the nasal consonant 'ng' or the initial consonant of the following syllable. For instance, the word /xialngaln/, can be divided as /xialn/ and /galn/(光軋) or /xialng/ and /aln/(相安).

(3) The letter [a], or [e], or [o] is preceded by either [a], or [i], or [u] or [u:], such as /ao/, /ia/, /ie/, /ua/, /uo/, /ue/, /u:e/ and /u:a/.
These letters can also form another syllable which starts with a vowel or diphthong. For example, /lia4o/ can contain three syllables /li/, /a4/ and /o/ or two syllables /lia4/ and /o/, or /li/ and /a4o/, or only one syllable /lia4o/. As a matter of fact, this character string is treated as one syllable in the current study because the neutral tone usually occurs at the end

instead of the beginning of the syllable.

6.2 Comparison Of Suen's And Chan's Representations

Based on the criteria for an ideal phonetic system as mentioned in Chapter One, a comparison of both Suen's and Chan's system has been made as follows:

(1) Consistency In Suen's system, only letter [u] represents two different vowels. However, in Chan's system, some of the finals have two pronunciations depending on the initial. For example the final code F-13(ㆥ) can be pronounced as 'oo-ng', if the initial is a consonant or pronounced as 'u-ng', if the initial is a semi-vowel. However, these inconsistencies can be detected by using the rules provided in Chapter Three for both systems.

(2) Easy Recognition Suen's representations using Latin letters are easier for the computer and human to recognize, but Chan's Chinese Phonetic Characters look more like Chinese characters, the images are hard for the computer to recognize, and for human beings to learn.

(3) Accuracy Suen's representations are easy for those people who know a Western language, especially English, to pronounce the syllables accurately, because this system uses the International Phonetic Symbols to

represent most of the Mandarin phonemes. However, Chan's representations are difficult to learn not only for a foreigner, but also for Chinese people, because all the symbols are totally new. And many shapes of Chan's symbols look very similar, that is, very easy to be confused.

(4) Easy For Computer Input/Output Suen's representations consist of letter strings, any ordinary key-board or printer can handle input/output without any problem. Chan's symbols present difficulties in entry through an ordinary computer keyboard. Since a pair of Chan's symbols, one initial and one final, looks like a Chinese character, which must be displayed in graphics mode.

(5) Flexibility Since Suen's representations are created according to phoneme features, it is easy to create a new syllable. However, in Chan's system, if the syllable has a new initial or final, a new symbol must be invented for representation.

(6) Length Of Syllable Chan's representation has the advantage that it requires only two symbols to represent one syllabe with tone, but in Suen's representation, it requires one or two letters to represent one phoneme; hence a syllable in Suen's system may have up to four phonemes plus the tone. According to the data used in this study, the average length of each syllable in

Suen's representations is 3.28 letters without tone; with tone, 4.15 letters not including the neutral tone, and 4.28 letters including the neutral tone.

6.3 Results of Comparison

The data used in this study are collected from three different books. Therefore, the number of words, syllables and phonemes involved in these three sets of data are different. However, as discussed in Chapter Five, the average number of syllables in each word and the average number of phonemes in each syllable are more or less the same in these three sets of data. Moreover, the percent frequency of consonants, semi-vowels, vowels and diphthongs are very similar in these three sets of data.

With respect to tone distribution, the results of set one and set three are very close to each other. The percent frequency of tone four is close to 30%, neutral tone around 15%, and the rest of the three tones between 18 and 19%. The reusults of set two show that the percent frequency of tone four is 30.04%, but, the neutral tone only occupies 10.50%, tone two 14.88%, tone one 20.24%, and tone three 24.34%. Nevertheless, tone four is the most frequently used and the neutral tone is the least frequently used in these three sets of data.

In April 1979, Dr. Suen published a book, called "Computational Analysis Of Mandarin"[22], with results obtained from a database of 753,941 syllables from a book called "A Study On The High Frequency Words Used In Chinese Elementary School Reading Materials"[1]. In that book, there are 1,883,462 phonemes in total. All the results are listed as follows:

	My results	Suen's results
Total syllables	37,639	753,941
Total phonemes	93,495	1,883,462
Average phonemes per syllable	2.50	2.48
Consonants	46.78%	46.93%
Semi-vowels	8.38%	8.26%
Vowels	37.13%	37.11%
Diphthongs	7.70%	7.70%
Tone 1	21.39%	18.97%
Tone 2	20.40%	18.68%
Tone 3	17.75%	19.23%
Tone 4	34.46%	29.35%
Neutral tone	13.77%	6.01%

As can be seen from the results, they are very similar except the differences in percent frequency in the tones ranging from 2.5% to 7.0%. The main reason is that Suen's data are taken from syllables of independent words mainly from written texts and my syllables are segmented from words

in meaningful clauses or sentences used in conversation. In Mandarin, the syllables at the end of the clauses or sentences are sometimes spoken in a neutral tone. Therefore, Suen's results have a lower percent frequency of the neutral tone, this gives rise to a slightly higher percentage in the other four tones.

6.4 Further Improvements And Research

In order to eliminate those problems mentioned in 6.1 in the Pinyin system, some ambiguous phoneme strings need to be analysed further by the computer and improved.

To make Chan's system more convenient and efficient for computer input/output, it is suggested to change the symbols and revise the sequence of the initial and final codes to facilitate computer storage and processing.

Among all the existing Mandarin phonetic systems, Suen's system is the most suitable one for computer input/output. Right now, the tonal marks of syllables are the same as Pinyin. For computer processing, it would be more convenient if these tonal marks are denoted by digits and placed at the end of the syllables.

REFERENCES

[1] "A study on the High Frequency Words Used in Chinese Elementary School Reading Materials", Chung Hwa Book Co., Taipei, 1967.

[2] R. M. Brend(ed.), "Studies in Tone and Intonation", S. Karger, Basel, 1975.

[3] Peter K. L., Chan, "Chinese Phonetic Characters", Hong Kong, 1981.

[4] S. K., Chan, "Dau-Han Phonetic System", Hong Kong, 1939.

[5] "Chinese Conversation For Tourist", (中国旅遊会话), (香港萬里書店), Hong Kong, 1976.

[6] "Chinese For Beginners", Foreign Language Press, Peking, China, 1976.

[7] Y. R. Chao, "Mandarin Primer", Harvard University Press, Cambridge, P. 9, 1974.

[8] T. Y. Chou and K. C. Huang, "Chinese Phonemes Analysis and Synthesis", Proceedings of The First International Symposium on Computers and Chinese Input/Output Systems, 1227-1241, Aug. 14-16, 1973.

[9] "Chuzhong Yuwen Langdu Jiaoxue Cankao" (初中語文朗读教学参攷), 上海教育出版社, Shanghai, 1981.

[10] "Dictionary of Spoken Chinese", Compiled by Staff of The Institute of Far Eastern Language, Yale Univeristy Press, P. 1066-1071, 1966.

[11] "Elementary Chinese", part I, Shangwuyinshukwan

(商務印書館) Peking, China, 1975.

[12] J. L. Flanagan, "Computer that Talk and Listen: Man-Machine Communication by Voice", Proc. IEEE, Vol. 64, No. 4, 405-415 April 1976.

[13] J. M. Howie, "On The Domain of Tone in Mandarin", Phonetica, Vol 30, 129-148, 1974.

[14] R. Huang, "Mandarin Pronunciation", Hong Kong University Press, 1969.

[15] K. P. Li, "Speech Recognition And Chinese Voice Input For Computer", Proceedings of The First International Symposium on Computers and Chinese Input/Output Systems, 211-223, Aug. 1973.

[16] T. S. Lin and H. C. Wang, "Segmentation Method for Isolated Mandarin Word Speech Recognition", Proceedings of International Computer Symposium, Vol. II, 919-931, Dec. 1980.

[17] "New Edition of Mandarin Dictionary", Taiwan Shan-Wu-Yin-Shu-Guan(商務印書館), re-edited by the Education Department of Taiwan, 1981.

[18] A. Newell et al, "Speech Understanding Systems", North-Holland/American Elsevier, 1973.

[19] L. R. Rabiner and R. W. Schafer, "Digital Processing of Speech Signals", Bell Lab., 1978.

[20] P. J. Seybot and G. K. K. Chiang(eds), "Language Reform in China", M. E. Sharpe, Inc. White Plains New York, 1979.

[21] C. Y. Suen, "Computer Synthesis of Mandarin", Proc.

IEEE Int. Conf. Acoustics, Speech and Signal Processing 698-700, April 1976.

[22] C. Y. Suen, "Computational Analysis of Mandarin", Birkhauser Boston, Cambridge, 1979.

[23] C. Y. Suen, "A Comparative Study of Mandarin Phonetic Systems by Computer", Proc. Int. Computer Conference, 7.3.1 - 7.3.15, Oct. 1980.

[24] C. Y. Suen and R. De Mori(eds.), "Computer Analysis and Perception of Visual and Auditory Signals", CRC Press, Boca Raton, 1982.

[25] C. Y. Suen, "Computer Aided Design of Mandarin Phonetic System", Proc. Chinese-American Academic and Professional Association 8th Annual Convention, 37-38 Nov. 1983.

[26] W. S. -Y. Wang, "The Chinese Language", Scientific American, Vol. 228, 51-60, Feb. 1973.

[27] G. D. Wilder and J. H. Ingram, "Analysis of Chinese Characters", Dover Publications Inc., New York, vi-viii, 1974.

Appendix 1. Table of Mandarin Sounds in Pinyin System

北 京 语 音 表

Table of the Speech Sounds of Peking Dialect

韵母 / 声母	1															2											3									4			
	a	o	e	-i	er	ai	ei	ao	ou	an	en	ang	eng	ong	i	ia	iao	ie	iou	ian	in	iang	ing	iong	u	ua	uo	uai	uei	uan	uen	uang	ueng	ü	üe	üan	ün
b	ba	bo				bai	bei	bao		ban	ben	bang	beng		bi		biao	bie		bian	bin		bing		bu												
p	pa	po				pai	pei	pao	pou	pan	pen	pang	peng		pi		piao	pie		pian	pin		ping		pu												
m	ma	mo				mai	mei	mao	mou	man	men	mang	meng		mi		miao	mie	miu	mian	min		ming		mu												
f	fa	fo					fei		fou	fan	fen	fang	feng												fu												
d	da		de			dai	dei	dao	dou	dan		dang	deng	dong	di		diao	die	diu	dian			ding		du		duo		dui	duan	dun						
t	ta		te			tai		tao	tou	tan		tang	teng	tong	ti		tiao	tie		tian			ting		tu		tuo		tui	tuan	tun						
n	na		ne			nai	nei	nao	nou	nan	nen	nang	neng	nong	ni		niao	nie	niu	nian	nin	niang	ning		nu		nuo			nuan				nü	nüe		
l	la		le			lai	lei	lao	lou	lan		lang	leng	long	li	lia	liao	lie	liu	lian	lin	liang	ling		lu		luo			luan	lun			lü	lüe		
z	za		ze	zi		zai	zei	zao	zou	zan	zen	zang	zeng	zong											zu		zuo		zui	zuan	zun						
c	ca		ce	ci		cai		cao	cou	can	cen	cang	ceng	cong											cu		cuo		cui	cuan	cun						
s	sa		se	si		sai		sao	sou	san	sen	sang	seng	song											su		suo		sui	suan	sun						
zh	zha		zhe	zhi		zhai	zhei	zhao	zhou	zhan	zhen	zhang	zheng	zhong											zhu	zhua	zhuo	zhuai	zhui	zhuan	zhun	zhuang					
ch	cha		che	chi		chai		chao	chou	chan	chen	chang	cheng	chong											chu	chua	chuo	chuai	chui	chuan	chun	chuang					
sh	sha		she	shi		shai	shei	shao	shou	shan	shen	shang	sheng												shu	shua	shuo	shuai	shui	shuan	shun	shuang					
r			re	ri				rao	rou	ran	ren	rang	reng	rong											ru	rua	ruo		rui	ruan	run						
j															ji	jia	jiao	jie	jiu	jian	jin	jiang	jing	jiong										ju	jue	juan	jun
q															qi	qia	qiao	qie	qiu	qian	qin	qiang	qing	qiong										qu	que	quan	qun
x															xi	xia	xiao	xie	xiu	xian	xin	xiang	xing	xiong										xu	xue	xuan	xun
g	ga		ge			gai	gei	gao	gou	gan	gen	gang	geng	gong											gu	gua	guo	guai	gui	guan	gun	guang					
k	ka		ke			kai	kei	kao	kou	kan	ken	kang	keng	kong											ku	kua	kuo	kuai	kui	kuan	kun	kuang					
h	ha		he			hai	hei	hao	hou	han	hen	hang	heng	hong											hu	hua	huo	huai	hui	huan	hun	huang					
	a	o	e		er	ai	ei	ao	ou	an	en	ang	eng		yi	ya	yao	ye	you	yan	yin	yang	ying	yong	wu	wa	wo	wai	wei	wan	wen	wang	weng	yu	yue	yuan	yun

Appendix 2. Table of Mandarin Sounds in Suen's System

	a	o	uh er	ai	ei	au	ou	an	un	ang	ung	i	ia	ie	iau	ieh	iu	ien	in	iang	ing	u	ua	uo	uai	uei	uan	uang	uon	ueng	ui	ueh	uen	uin	uaong
b	ba	bo		bai	bei	bau		ban	bun	bang	bung	bi	bia		biau	bieh		bien	bin		bing	bu													
p	pa	po		pai	pei	pau	pou	pan	pun	pang	pung	pi			piau	pieh		pien	pin		ping	pu													
m	ma	mo	muh	mai	mei	mau	mou	man	mun	mang	mung	mi			miau	mieh		mien	min		ming	mu													
f	fa	fo			fei		fou	fan	fun	fang	fung											fu													
d	da		deh	dai	dei	dau	dou	dan		dang	dung	di			diau	dieh	diu	dien			ding	du		duo		dwei	duan		doon	doong					
t	ta		tuh	tai		tau	tou	tan		tang	tung	ti			tiau	tieh		tien			ting	tu		two		twei	twan		toon	toong					
n	na		nuh	nai	nei	nau		nan	nun	nang	nung	ni			niau	nieh		nien		niang	ning	nu		nwo			nwan		noon	nong	nui	nueh			
l	la	lo	luh	lai	lei	lau	lou	lan		lang	lung	li	lia		liau	lieh	liu	lien	lin	liang	ling	lu		lwo			lwan		loon	loong	lui	lueh			
g	ga		guh	gai	gei	gau	gou	gan	gun	gang	gung											gu	gua	guo	guai	gwei	gwan	gwang	goon	goong					
k	ka		kuh	kai		kau	kou	kan	kun	kang	kung											ku	kua	kuo	kuai	kwei	kwan	kwang	koon	koong					
h	ha		huh	hai	hei	hau	hou	han	hun	hang	hung											hu	hua	huo	huai	hwei	hwan	hwang	hoon	hoong					
j												ji	jia		jiau	jieh	jiu	jien	jin	jiang	jing										jui	jueh	juen	juin	juaong
ch												chi	chia		chiau	chieh	chiu	chien	chin	chiang	ching										chui	chueh	chuen	chuin	chuaong
x												xi	xia		xiau	xieh	xiu	xien	xin	xiang	xing										xui	xueh	xuen	xuin	xuaong
rj	rja		rjuh	rjai	rjei	rjau	rjou	rjan	rjun	rjang	rjung	rji										rju	rjua	rjuo	rjuai	rjuei	rjuan	rjuang	rjoon	rjoong					
rc	rca		rcuh	rcai		rcau	rcou	rcan	rcun	rcang	rcung	rci										rcu	rcua	rcuo	rcuai	rcuei	rcuan	rcuang	rcoon	rcoong					
sh	sha		shuh	shai	shei	shau	shou	shan	shun	shang	shung	shi										shu	shua	shuo	shuai	shuei	shuan	shuang	shoon						
r			ruh			rau	rou	ran	run	rang	rung	ri										ru		ruo		ruei	ruan		roon	roong					
dz	dza		dzuh	dzai	dzei	dzau	dzou	dzan	dzun	dzang	dzung	dzi										dzu		dzuo		dzuei	dzuan		dzoon	dzoong					
ts	tza		tsuh	tsai		tsau	tsou	tsan	tsun	tsang	tsung	tsi										tsu		tsuo		tsuei	tsuan		tsoon	tsoong					
s	sa		suh	sai		sau	sou	san	sun	sang	sung	si										su		suo		suei	suan		soon	soong					
	a	o	uh er	ai	ei	au	ou	an	un	ang	ung	yi	ya	yo	yau	yeh	you	yen	yin	yang	ying	wu	wa	wo	wai	wei	wan	wang	wun	wung	yu	yueh	yuen	yuin	yuaong