NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

A FEASIBILITY STUDY USING CHINESE SPEECH
AS
A COMMAND/CONTROL TOOL FOR COMPUTER
SYSTEMS

By

Liu, I Kang

March 1987

Thesis Advisor: Gary K. Poock

Approved for public release; distribution is unlimited

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; Distribution is Unlimited
2b DECLASSIFICATION / DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S)	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION Naval Postgraduate School 6b OFFICE SYMBOL (If applicable) Code 55	7a NAME OF MONITORING ORGANIZATION Naval Postgraduate School
6c ADDRESS (City, State, and ZIP Code) Monterey, California 93943-5000	7b ADDRESS (City, State, and ZIP Code) Monterey, California 93943-5000
8a NAME OF FUNDING / SPONSORING ORGANIZATION 8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c ADDRESS (City, State, and ZIP Code)	10 SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO / PROJECT NO / TASK NO / WORK UNIT ACCESSION NO

11 TITLE (Include Security Classification)
A FEASIBILITY STUDY USING CHINESE SPEECH AS A COMMAND/CONTROL TOOL FOR COMPUTER SYSTEMS (u)

12 PERSONAL AUTHOR(S)
Liu, I Kang

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Master's Thesis	FROM _____ TO _____	1987 March	64

16 SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Voice recognition, Chinese phonetic system, phoneme

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

This thesis examined whether American English speech recognition technology can be used by Chinese speakers, in their native tongue, to achieve a reasonable degree of recognition accuracy. Three experiments were completed. The first showed that 88.25% of 4305 trials of Chinese phoneme recognition was correctly recognized. The second showed that 74.67% of 900 trials of simulated speaker independent mode Chinese utterance recognition was correctly recognized. The third showed that 12.44% of 900 trials of speaker dependent mode Chinese utterance recognition was incorrectly recognized on the first attempt. Only 16 utterances required a retraining to eventually obtain a correct recognition.

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT XX UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a NAME OF RESPONSIBLE INDIVIDUAL Gary K. Poock	22b TELEPHONE (Include Area Code) 408-646-2636 22c OFFICE SYMBOL 55Pk

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

Approved for public release; distribution is unlimited.

A Feasibility Study Using Chinese Speech As A Command/Control Tool For Computer Systems

by

Liu, I Kang
Commander, Republic of China Navy
B.S., Chinese Naval Academy, 1975

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN INFORMATION SYSTEMS

from the

NAVAL POSTGRADUATE SCHOOL
March 1987

ABSTRACT

This thesis examined whether American English speech recognition technology can be used by Chinese speakers, in their native tongue, to achieve a reasonable degree of recognition accuracy. Three experiments were completed. The first showed that 88.25% of 4305 trials of Chinese phoneme recognition was correctly recognized. The second showed that 74.67% of 900 trials of simulated speaker independent mode Chinese utterance recognition was correctly recognized. The third showed that 12.44% of 900 trials of speaker dependent mode Chinese utterance recognition was incorrectly recognized on the first attempt. Only 16 utterances required a retraining to eventually obtain a correct recognition.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENT

I would like to take this opportunity to thank the United States Navy and the Navy of the Republic of China for making my graduate education possible. I would also like to thank my thesis adviser, Professor Gary K. Poock for his valuable guidance and the use of his laboratory. Additionally, I appreciate the dedicated reviews that Professor Richard A. McGonigal of NPS and Doctor Sarah W. Blackstone of American Speech-Language-Hearing Association conducted of this thesis. This thesis would not have been possible without the untiring support of the test subjects who particpated in my experiments. Last, but not least, I dedicate this thesis to my wife who has supported me on all of my professional endeavors.

I. BACKGROUND

A. INTRODUCTION

There were about forty speech recognition/input studies conducted at United States Naval Postgraduate School (NPS) during the past six years. The conclusion suggested by these studies is quite significant: that speech input, compared to conventional manual input, is much more accurate and faster. And, since hands are free from typing on the keyboard, users may be capable of performing a secondary assignment. From an early experiment conducted by Prof. Gary Poock in 1980, he concluded three results. (1) Manual input had 183.2% more entry errors. (2) Speech input was 17.5% faster. (3) Speech input allowed subjects to concurrently perform 25% more on a secondary job. See [Ref. 1] for detailed information.

Another highly valued finding is that speech input needs only a small amount of time to acquaint brand-new users with this input device, and results in a better performance than that of a well-trained operator who uses a keyboard as an input device. From the same experiment mentioned above, Prof. Poock found that the average time for the subjects to practice with the voice recognition equipment and feel ready to conduct the experiment was only 3.26 hours. This is much less than the time needed for familiarizing an individual with a keyboard device.

The usage of English speech to input data to computer systems has proved to be technically and practically feasible. At the same time the range of potential military and commercial applications of this medium appears extensive. All of these encouraged the author to initiate this study and, hopefully, to provide some useful information for further research and future possible applications of Chinese speech recognition/input.

B. THE LANGUAGE AND THE RECOGNITION

The language used in most studies mentioned above was English. There was, in fact, only one experiment that examined a second language- German. As described in [Ref. 2], the recognition system functioned equally well when training and testing used German as an input language. The same study also examined the capability of the

recognition system (Threshold Technology T600 voice recognition system) to function in a bilingual mode. However, significant degradation was observed when training and testing was bilingual in nature.

During one of his Man-Machine Interface laboratory projects, the author, under a programmed scenario, has successfully operated the DDN with Chinese speech. The DDN stands for Defense Data Network, a large distributed network of computers which are geographically located around the United States and other countries. From that preliminary experiment, the author has shown that Chinese speech can also be an effective input medium for command/control opertions.

C. THE PURPOSE AND THE SCOPE

Because of the imperfect phonetic system, Chinese speech has suffered a certain degree of difficulty. Due to the same reason, some confusion about the phonetic system has been raised during the past years. Although the difficulty itself will not influence the recognition of Chinese speech, the reasons that caused the difficulty will. In addition, all that confusion, if not clarified, will be the trouble area for Chinese speech recognition in the future.

The main effort of this study is, then, to do a thorough study on Chinese speech and the corresponding phonetic system. A brief discussion is provided in Chapter III. The detailed discussion, provided in Chapter II, on the English part is mainly for establishing a reference basis for the later discussions of Chinese speech. A further experiment on examining Chinese speech recognition was conducted. The description of the experiment itself and the results obtained are provided in Chapter IV and V respectively. Some suggestions on further studies are also discussed in Chapter V.

D. GENERAL INFORMATION ON THE STUDY

The studies on the two languages within this thesis focused on the sounds of the languages. Hence, it is necessary to point out *English* used here means American English while the *Chinese* means Mandarin Chinese. The presentation of the speech sounds during the discussion will be some selected letters quoted by special characters. To differentiate them, the author uses /..../ to present English pronunciations and <....> to present Chinese pronunciations.

The English phonetic system the author used is known as the *KK Phonetic System* established by two famous American linguists- Dr. John S. Kenyon and Dr. Thomas A. Knott. Their *A Pronunciation Dictionary Of American English* has been an international reference book for studying American English. The Chinese phonetic system the author used is the only system compiled by the Chinese Department of Education in 1918. The system is also known as <Droo In Foo Hao> in Chinese. Consult [Ref. 3] and [Ref. 4] for detailed description.

The KK System was so well established that it fully complied with the rule of thumb for constructing a phonetic alphabet system: *One symbol represents only one unique sound, and one sound only has one unique symbol on its behalf.* However, this is not the case in the Chinese phonetic system. There are symbols representing two or even three sounds, or two symbols actually representing the same sound. This is an important feature deserving special attention for those who want to apply the current Chinese phonetic system in Chinese speech recognition/input research. Further discussion will be provided in Chapter III.

II. AN EXAMINATION OF ENGLISH SPEECH

A. THE SOUNDS OF ENGLISH

According to the KK System, there are forty-one sounds used in English, which are called phonemes of English. Among them, seventeen are vowels and twenty-four are consonants. These forty-one sounds, depending on the way they are produced, have been sorted into ten groups. Each sound is associated with a unique phonetic alphabet formulated by the International Phonetic Association. (Consult Appendix A for more information on the original symbols used.) However, these phonetic alphabets are usually used only by linguists and therefore just several of them can be found on the NPS IBM 3800-3 printer system. For easing our discussion, the author constructed a symbol system to represent these forty-one sounds. Please see Table 1 for the general idea.

The phoneme is the smallest unit of significant distinctive sound. However, not all phonemes can form a syllable- the smallest unit of English words. To form a syllable, one and only one vowel sound is required as the base and may or may not be proceeded or followed by any consonant combinations. So /ei/, /bee/, /it/, /head/, and /spleen/ are all considered single syllable words.

The most reliable way to discriminate phonemes is to *first examine the manner and then the speech organs used to produce the speech.* [Ref. 5] has provided intensive discussions on the production of each phoneme and can be a very helpful reference. Human hearing is a good enough tool to tell the differences among sounds, but it is not always reliable in trying to differentiate certain similar sound pairs such as /ee/ and /i/, /oo/ and /o/, or /n/ and /ng/. We can use [Ref. 6] as a valuable source to obtain detailed information on those sound pairs.

Certain sounds may be recognized on one speech recognition system but not on another system. This is due to the algorithm design adopted by the recognizer manufacturers. Although it is beyond the scope of this study, it is proper to note that the algorithm of the recognizer has a dominant influence on the recognition performance.

TABLE 1

PHONEMES OF AMERICAN SPEECH

Front Vowels (FV):

1. ee.....bee
2. i.....bit
3. ei.....eight
4. ea.....head
5. au.....laugh *

Central Vowels (CV):

1. er.....letter *
2. ur.....hurt *
3. e.....the
4. u.....cut

Fricatives (FR):

1. f.....five
2. v.....very *
3. th.....think *
4. the.....bathe *
5. s.....six
6. z.....zoo *
7. sh.....she *
8. ge.....garage *
9. h.....him

Stops (ST):

1. p.....pool
2. b.....but
3. t.....tea
4. d.....do
5. k.....kiss
6. g.....give

Back Vowels (BV):

1. oo.....room
2. o.....woman
3. oa.....coat
4. aw.....law *
5. a.....car

Diphthongs (DI):

1. ai.....aisle
2. ow.....now *
3. oy.....boy *

Nasals (NA):

1. m.....make
2. n.....nice
3. ng.....king

Glides (GL):

1. y.....year
2. w.....wait
3. r.....right *

Affricates (AF):

1. ch.....chip
2. j.....joy

Lateral (LA):

1. l.....lay

* sounds not used in Chinese.

B. THE PRODUCTION OF SPEECH SOUNDS

The production of vowels is primarily done by adjusting the shape and size of the oral cavity, the main resonance chamber. Such adjustments are made by altering the position of the tongue, jaw and lips. The vocal tract[1], during speech production, remains relatively open and unobstructed. The production of consonants is done by adopting certain articulatory motions to produce different types of sounds. Therefore,

[1]Vocal tract is the area through which the breath stream passes during the production of the sounds.

we may discuss consonants by examining the place[2] of articulation and the manner[3] of articulation used to produce the sounds. During consonant production, some kind of obstruction of the vocal tract is observed.

In Table 1, some phonetic terminologies are being used. From these terminologies, one can easily obtain some information about the production of each category of English speech. Here is a brief introduction to these terminologies. More detailed information can be found in [Ref. 5.]

Front Vowel is a vowel which is pronounced with the front part of the tongue higher than the rest of the tongue. Front Vowel is also called Spread Vowel because it is also pronounced with the lips spread. Back Vowel is a vowel which is pronounced with the back part of the tongue higher than rest of the tongue. Back Vowel is also called Rounded Vowel because, of course, it is pronounced with the lips rounded. Central Vowel, then, is a vowel which is pronounced with the middle part of the tongue higher than the front or back of the tongue. The shape of the lips for Central Vowels is, as you can imagine, somewhat between spread and rounded.

All three categories of vowels mentioned above are considered single vowels. Diphthongs are sounds that appear to be formed from the blend of two single vowels spoken together in the same syllable. What actually happens here is that the articulator begins the syllable in the position for one vowel and then shifts with a smooth and continuous transition movement toward the position for some other vowel. One can easily learn to detect the first and second vowels of the diphthongs.

Fricative is a consonant consisting acoustically of friction noises. They are made by directing the breath stream with adequate pressure against one or more points of articulation and lead to the hissing noises of distinctive Fricatives. Stop is a speech sound which involves a complete blocking of the breath stream at some point and is subsequently released with a somewhat audible explosive puff. That is why Stop is sometimes also called Explosive. Nasal is chosen for the class because of the distinctive nasal resonance that those sounds uniquely contain. Glide is a consonant that consists primarily of the movement of an articulator which causes a rapid change of resonance. Glide is also called Semivowel, because the starting position of pronouncing each of them is a vowel. They are /ee/ for /y/, /oo/ for /w/ and /ur/ for /r/.

---

[2]Place of articulation includes bilabial, labiodental, linguadental, lingua-alveolar, linguapalatal, linguavelar and glottal.

[3]Manner of articulation includes nasal, stop, fricative, affricate, lateral and glide.

Usually the tongue moves from the position of each vowel to that for the following vowel in the same syllable. The sounds produced by the articulator movement between the two vowels are represented by each Glide respectively. Affricate is a consonant that is made up of two consonants- a Stop followed by a Fricative. Lateral is produced in a manner that the voiced breath stream escapes laterally over the sides of the tongue.

C. THE PITCH AND INTONATION OF ENGLISH

When you read an English word or a sentence composed of several words, your sound flow actually contains different pitches. Although each word has its unique pitch pattern in English, it has some variations when the same word is read with other words in a sentence. We use intonation as a term for the latter concept.

English has been described as using four pitch levels. They are extra-high, high, mid, and low. To simplify, numbers have been used to designate them. George L. Trager and Henry L. Smith, Jr., in their *An Outline of English Structure*, chose 1 to represent low. As the pitch level rises, the representation also increases in number. In normal speech, however, extra-high designated by 4 does not occur often. Extra-high usually indicates excitement.

Since pitch is determined by the frequency of the sound, the pitch level is, from the viewpoint of linguists, really a relative matter. There is no need to tell the difference between the pitches of the same syllable produced by two persons. Similarly, the attempt to tell the difference between the pitches of the same syllable produced by the same person at different moments is also meaningless. However, there are indeed certain rules regarding pitch which must be observed in order to generate understandable English words. These rules are as follow:

1. The principal stressed syllable of a word will be pronounced with high pitch (designated by 3).

2. All the syllables produced before the principal stressed syllable will be pronounced with mid pitch (designated by 2).

3. All the syllables produced after the principal stressed syllable will be pronounced with low pitch (designated by 1).

4. When the principal stressed syllable is the last syllable of a

word, the vowel sound of the syllable will present a 3-to-1 falling inflection of pitch.

5. An auxilliary stressed syllable will act similar to a principal one and the only difference is that its pitch level will be located between high and mid.

Some examples are provided in Figure 2.1, which apply those rules mentioned above. Again, one should keep in mind that the pitch relationship among syllables of a word is relative. As you can see, the first three examples are presented with an order that the principal stressed syllable appears at first, second, and then the last syllable of each word respectively. The last one is an example of a single-syllable word that will be pronounced like the last syllable of the third example. When a word with an auxilliary stressed syllable is encountered, you just insert that syllable into a level between 3 and 2, and pronounce it with a pitch higher than the mid pitch syllable but lower than the high pitch syllable of the word.

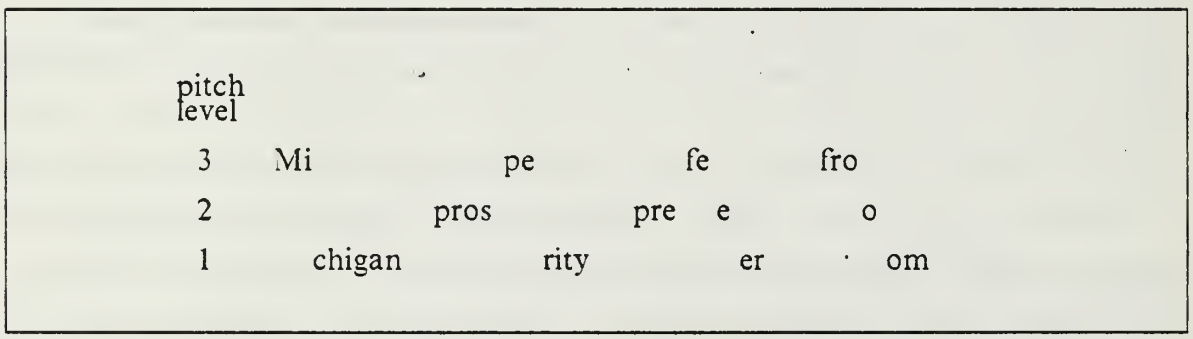


Figure 2.1 Examples of English Pitch Patterns.

The 3-level pitch system can also be applied in discussing intonation, where the whole sentence is put into a pitch frame having a wider frequency range for each level. To obtain the idea, see examples in Figure 2.2.

The first example represents the most common and colorless intonation pattern in English, which is designated with number 231. Simple statements and questions starting with question words always use this pattern. The second intonation pattern is used by what we called 'yes/no questions', and is designated with number 233. The last one is an example to show a simple statement colored by extra meaning, and is designated with number 223. Interested readers may consult [Ref. 7] for a complete

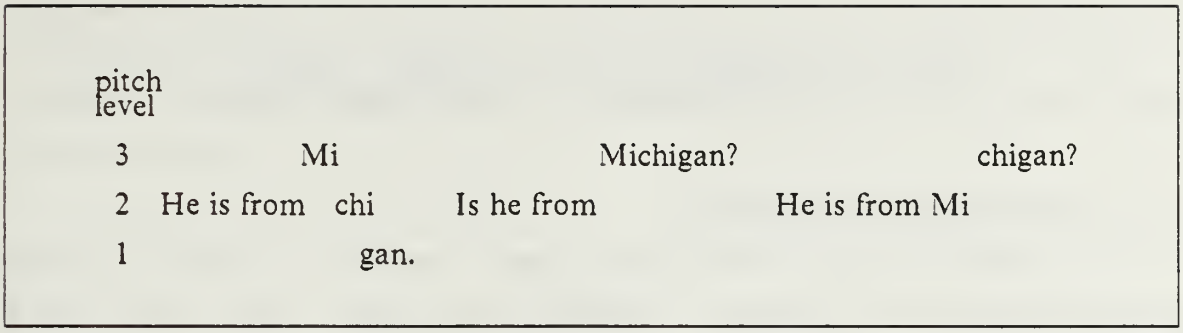


Figure 2.2 Examples of English sentences intonation.

discussion on this subject. The main point the author wants to address here is that the pitch pattern of an English word may change depending on how/where it appears within a sentence.

III. AN EXAMINATION OF CHINESE SPEECH

A. THE SOUNDS OF CHINESE

The original Chinese phonetic system had 41 symbols. However, the current system used only has 37 symbols. Four symbols were deleted. Two of them, exist in English as well, represent the sounds /ng/ and /v/. The reason for the deletions, however, was different. The symbol representing the sound /ng/ was deleted because the system had another symbol also representing the sound. The latter was simply because the Chinese does not have the speech sound /v/. The third one was a Nasal sound produced with tongue-front pushed against the hard palate, which does not exist in either English or Chinese. The fourth symbol, representing two very similar Front Vowels of Chinese , was deleted for, probably, the following two reasons. First, they are not able, as other finals, to form a syllable by themselves. They must follow a particular Fricative. Second, the articulation places of the two sounds are the same as that of the Fricatives which proceed them. This deletion causes Chinese characters to sound sometimes as being represented by a single consonant. As a remedy, the author uses <ih> to represent the two sounds and which will be shown as the 38th symbol of Table 2 .

Given the historical information mentioned above, the author constructed a 38-symbol table for the Chinese phonetic system, which actually can be seen as a romanization system. Appendix B has provided a table that simultaneously presented several current existing romanization systems, namely, Yale(YL), Wade-Giles(WG), Chinese Phonetic System Second Form(SF), PinYin(PY), and the system suggested by author (SG), for purposes of cross reference. The order of the symbols in Table 2 is exatly the same as that of the existing phonetic system. The first 21 symbols are consonants, also called *initials*, and the succeeding 17 symbols are vowels or combinations of a vowel and a Nasal, also called *finals*. The reason for the alias is due to the features of Chinese pronunciation. Chinese characters are always single-syllable sounds. They usually are an initial followed by a final, an initial and a Glide then followed by a final or just a final itself. In most situations, the characters end with a vowel sound. The only two consonants allowed to be produced at the end of a

character are sounds /n/ and /ng/. Although /n/ is also one of the 21 initials, the Chinese phonetic system has another symbol to represent the /n/ sound that appears at the end of a character. Hence, those 21 consonants will always be the initial part of a character sound.

TABLE 2

CHINESE PHONETIC SYSTEM

Initials:

Bilabials

1. b.....ST
2. p.....ST
3. m.....NA

Labiodental

4. f.....FR

Lingua-alveolars

5. d.....ST
6. t.....ST
7. n.....NA
8. l.....LA

Lingua-velars

9. g.....ST
10. k.....ST

Glottal

11. h.....FR

Lingua-palatals

12. j.....AF
13. ch....AF
14. hs....FR *
15. dr....AF *
16. tsh...AF *
17. sh....FR *
18. r.....FR *

Lingua-alveolars

19. dz....AF +
20. ts....AF +
21. s.....FR

Finals:

Single Vowels

22. a.....BV +
23. o.....BV
24. e.....CV
25. ea....FV

Diphthongs

26. ai....DI
27. ei....FV
28. ao....DI *
29. oa....BV

Combinations

30. an *
31. en +
32. ang *
33. eng +

Single Vowels

34. er....FV *
35. i.....FV +
36. oo....BV +
37. iu....FV *
38. ih....FV *

* sounds not used in English/further disscussion provided

\+ further discussion provided

A quick look Table 2 shows that the consonants of the Chinese phonetic system are grouped by the articulation organs used to make each sound. The reason for this was mentioned by Prof. Francis Dow in his work [Ref. 8: p.24], and quoted below:

1. The consonants of each category have their homorganic nature in articulation.
2. In the constitution of syllables, certain sets of initials occur before certain sets of finals (consult Table 5).
3. It is more convenient to compare the consonants of each category with those in other Chinese dialects.

In Table 2, twenty symbols followed by neither '+' nor '*' are sounds also used in English. The author selected exactly the same symbols shown in Table 1 to represent them respectively. Seven symbols followed by a '+' are sounds also used in English, but some details need to be clarified. Eleven Symbols followed by '*' are sounds not used in English; therefore, a brief introduction is provided for each of them. The following two sections provide detailed discussions on this.

B. CLARIFICATION OF CONFUSIONS IN THE CHINESE PHONETIC SYSTEM

The 20th and 19th symbols represent a pair[4] of affricates. Sound <ts> is voiceless and <dz> is the voiced counterpart of <ts>. They appear, in English, at the end of the plural form of nouns with ending sound /t/ or /d/ respectively such as hats and hands.

The 22nd symbol, <a>, represents three different sounds. All of them are used in English, but only two are considered phonemes. The first sound is /a/ of car and the second sound is /u/ of cut. The third one is the first half of diphthongs /ai/ and /ow/; however, it is, in Chinese, the most frequently used sound among the three. The author suggests using <aa> to represent this, since the lips, when producing the sound, are spread wider than when producing sound /a/. And the symbols for the remaining two sounds are, as their English counterparts, <a> and <u>.

[4]Two sounds are considered a pair when they adopt the same method and use the same articulator and point of articulation for pronunciation. The only difference is that one is voiceless and the other is voiced sound.

The symbol <a> has already caused an unrecoverable damage in Chinese. No one, at present time, is able to tell, when encountering a character with symbol <a>, which one of the three sounds should be used. Words in Chinese such as mother, <ma ma>, lama, <la ma>, or to punch a card, <da ka>, should actually be pronounced, from the author's limited-scale investigation, as <mu mu>, <laa mu> and <daa ka>. Since the situation is messed up already, no one ever has the authority to say which one of the three sounds should be the right sound for certain characters. A further wide-range investigation is needed if one is really anxious to use the right sound for characters with symbol <a>. And, probably, the end product of the investigation would only be the majority-used sounds of the general population in this age. However, since it is beyond the scope of this thesis, the author leaves the problem to future researchers. For the purpose of simplifying the following discussion, the author will, from now on, use only <a> to represent the three sounds.

Both the 31st and the 33rd symbols represent two different sounds. They represent sounds /n/ and /ng/ respectively in some cases and /e/ followed by /n/ or by /ng/ in some other cases. Although many people are confused by these two symbols, a careful study certainly helps to differentiate the usages of them. Symbol <en>, in most situations, represents sound /e+n/, except when appearing after the symbols <i> and <iu>. In the latter case, the <en> represents sound /n/. Symbol <eng>, the same as <en>, represents the sound /e+ng/ most of the time, but when appearing after symbol <i>, <oo> or <iu>, it represents sound /ng/ as well. See Table 6 for some examples.

Again, the 35th symbol, <i>, represents three sounds which are also used in English. They are /i/ and /ee/ of Front Vowels and /y/ of Glides. To tell when <i> representing sound /y/ is easy, because once one notes an <i> appearing before a final, he is almost sure that the symbol <i> represents sound /y/. However, the finals <en> and <eng> are two exceptions. In this situation, the symbol <i> represents sound /i/ or /ee/; with the two finals becoming consonants /n/ and /ng/.

In the case of telling whether /i/ or /ee/ is represented by symbol <i> for a certain character, one faces the same problem discussed earlier. It is again an unrecoverable damage which was caused many years ago. Secret, as an example, in Chinese symboled by <mi mi> should in fact be pronounced as <mee mi>. The author, for the same reason, leaves the problem to researchers for further study and, uses the symbol <i> to represent the two sounds through the following discussions.

Although the 36th symbol represents two sounds also used in English, we can easily tell them apart by examing the usages of the symbol. The two sounds represented by the symbol are /oo/ of Back Vowels and /w/ of Glides. Once an <oo> is found before a final, for most situations, we know that it is sound /w/. However, the final <eng> is the only exception. In this case and in the case that the <oo> itself is the final part of a character, we know that the sound /oo/ is represented.

TABLE 3
SUPPLEMENTARY TO THE CHINESE PHONETIC SYSTEM

NO	Original	Suggested	Articulation
22	<a>	<a>	BV
		<u>	CV
		<aa>	CV
31	<en>	<en>	CV+NA
		<n>	NA
33	<eng>	<eng>	CV+NA
		<ng>	NA
35	<i>	<i>	FV
		<ee>	FV
		<y>	GL
36	<oo>	<oo>	BV
		<w>	GL
37	<iu>	<iu>	FV
		<yw>	GL

Table 3 provides a summary of this section, which lists all the symbols that are easily confused. The first column of the table is the number of each symbol, which corresponds to the number appearing in Table 2 . The second column lists all the symbols, except 19 and 20, discussed in this section. Symbol 19 and 20 are not included because they are not confused at all. Symbol 37 is listed here too, but the discussion is provided in the next section, because it is a sound existing uniquely in Chinese. The third column is the author's suggestions that each symbol should actually be according to the discussions provided in this chapter. The last column provides articulation information on each symbol. Consult Table 1 and the discussions provided in Chapter II for a better understanding of the abbreviations used here.

C. INTRODUCTION TO UNIQUELY EXISTING SOUNDS IN CHINESE

The 14th sound of the Chinese phonetic system, <hs>, is a Fricative. To produce the sound, one needs to raise his or her tongue-front toward, but does not touch the hard palate, and let the tongue-tip stretch down against the lower teeth ridge. With the tongue held in this position, an unvoiced breath stream is directed against the hard palate, lower teeth ridge and teeth to produce the sound <hs>.

The 17th and 18th symbols, <sh> and <r>, represent a pair of Fricatives also. These two sounds do not appear in English, but they have some similarities to the sound pair /sh/ and /ge/ in English. The author directly 'borrowed' the symbols from English for reasons mentioned below.

The only difference between /sh/ and <sh> is the articulators used by the two sounds. The /sh/ sound requires raising the tongue-mid toward the hard palate; while the <sh> sound uses tongue-front to stretch toward the hard palate. Everything else is the same.

Just as /ge/ to the /sh/, the sound <r> is the voiced counterpart of the <sh>. The reasons we do not use <ge> is that /r/ also has some similarities to the sound <r> and /r/ appears more as an initial which is exactly the characteristic that <r> has. The way to produce the sound <r> is the same as producing <sh> except adding the vibration of the vocal cords, which is the main feature of voiced sounds.

The 16th and 15th symbols, <tsh> and <dr>, are a pair of Affricates. Their relationship with the sound pair of <sh> and <r> is just like that of /ch/ and /j/ to the sound pair /sh/ and /ge/ in English. That is why the author selected <t + sh> and <d + r> to represent the two sounds respectively. And, the way to produce the sound <tsh> is just as the symbol itself suggests: do a preparation action as if you were going to produce a <t> sound. When ready, actually produce an <sh> sound instead. It is similar to producing the English sound /ch/ except using a different articulator. To produce <dr> is the same as producing <tsh> except adding a vibration of the vocal cords since <dr> is its voiced counterpart.

Although the 28th sound, <ao>, does not appear in English, there is indeed a very similar sound in English. That is /ow/. The only difference between the two sounds is the first half starting position of the sound. The sound /ow/ is a diphthong formed by blending <aa> and <o> together; however, the sound <ao> is produced by blending <a> and <o> together. A careful examination of the lips' shape can certainly help to distinguish the two sounds, <a> and <aa>, without any difficulty.

The 30th symbol, <an>, is actually a combination of a Central Vowel and a Nasal. It does not appear in English because of the vowel part of the sound. It is <aa> which does not appear in single vowel form in English. However, one can find the sound in the first half of the diphthongs such as /ai/ and /ow/. In a similar manner, the 32nd symbol, <ang>, is classified as sound not existing in English for the same reason.

The 34th symbol, <er>, is directly 'borrowed' from the English phoneme /er/. Although the two sounds have some similarities, they are not the same. The sound /er/ is a short, lax,[5] mid-central, r-colored[6] vowel which can be produced by tongue retroflexion[7]. The sound <er> is short, r-colored too, but it is a high-front and tense[8] vowel. The tense is caused by keeping the tongue retroflexed and stretching the tongue forward to the hard palate simultaneously.

The 37th symbol, <iu>, actually represents two sounds. One is a vowel and the other is a Glide whose start position of production is the vowel. The author suggests using <yw> to represent the said Glide. The sound <iu> is a Front Vowel but not a Spread Vowel. When pronouncing the sound, one must hold the tongue in the position of producing sound <i> and, at the same time, round the lips as if producing sound <oo>. The sound, hence, can be described as a lower high-front, rounded, tense vowel. The relationship between <iu> and <yw> is just as <ee> to <y> or <oo> to <w>.

Although the 38th symbol, <ih> represents two different vowels, the author does not intend to differentiate them with two symbols. Because, first, they are very similar; second, the speech organs used by each of them are identical to those of a Fricative respectively; third, each of them can only follow a particular group of sounds that are formed by that same Fricative; fourth, they don't independently exist as other finals.

---

[5]Lax vowel is a vowel which is pronounced with the muscles of the throat, tongue and corresponding mouth lax.

[6]The r-color is an acoustic effect of a simultaneously articulated 'r' imparted to a vowel by retroflexion or bunching of the tongue.

[7]Retroflexion is the articulation with or involving the participation of the tongue tip raised and retracted toward the hard palate.

[8]Tense vowel is a vowel which is pronounced with the muscles of the throat, tongue and corresponding mouth tense.

To produce the first sound, one needs to stretch the tongue-front toward the hard palate, same articulation position of producing the sound <sh>, and then vibrate the vocal cords and let the sound resonate in the oral cavity. This sound can only follow the sound <sh>, <r>, <tsh> and <dr>. Similarly, to produce the second sound, one needs to stretch the tongue-front toward the alveolar ridge with the same articulation position of producing the sound <s>, and then vibrate vocal cords and let the sound resonate in the oral cavity. This sound can only follow the sound <s>, <ts> and <dz>.

Since the ending position of sounds <sh>, <r>, <tsh> and <dr> is the position for producing <sh> and the articulation position of the <ih> that follow these sounds is also <sh>, when we produce the syllable <sh + ih>, for example, we actually produce the consonant first and then maintain the same articulation position and produce the vowel. Because the ending position of sounds <s>, <ts> and <dz> is, similarly, the position for producing <s> and the articulation position of the <ih> that follow these sounds is also <s>, when we produce the syllable <s + ih>, for example, we actually produce the consonant first and then maintain the same articulation position and produce the vowel. That is why the author intends to use the same symbol to represent the two similar sounds. It is probably, as mentioned earlier, also the main reason why they deleted this symbol in the first place.

Table 4 concludes the discussions provided in last two sections. The table provides the complete information about the Chinese speech phonemes. There are 25 consonants in the table and 21 of them are the initials of the original phonetic system. The three of the remaining ones are the Glides which use the same symbol with three finals, namely <i>, <iu> and <oo>. And the last one is the <ng> separated from final sound <eng>. There are 16 vowels in the table too. Only 12 of them are from the original system. The four symbols, <an>, <en>, <ang> and <eng>, are dismissed because they are simply combinations of two phonemes. The four new vowels in this table are <ee> separated from <i>, <u> and <aa> separated from <a> and sound <ih>. As their English counterpart, these phonemes are grouped into ten categories. The three groups of single vowels are put in an order that the sound produced with the highest tongue posture of each group is the first one and the lower the latter. The four groups of consonant, namely Nasal, Stop, Fricative and Affricate, are put in an order that the sound produced at the most outside of the vocal tract is the first one and the inner the latter. The remaining groups, however, have no special order at all.

TABLE 4

PHONEMES OF CHINESE SPEECH

Front Vowels (FV):

1. ih.....(38) *
2. ee.....(35)
3. er.....(34) *
4. i.....(35)
5. iu.....(37) *
6. ei.....(27)
7. ea.....(25)

Back Vowels (BV):

1. oo.....(36)
2. o.....(23)
3. oa.....(29)
4. a.....(22)

Central Vowels (CV):

1. e.....(24)
2. u.....(22)
3. aa.....(22) *

Diphthongs (DI):

1. ai.....(26)
2. ao.....(28) *

Stops (ST):

1. p.....(2)
2. b.....(1)
3. t.....(6)
4. d.....(5)
5. k.....(10)
6. g.....(9)

Nasals (NA):

1. m.....(3)
2. n.....(7,31)
3. ng.....(33)

Fricatives (FR):

1. f.....(4)
2. s.....(21)
3. hs.....(14) *
4. sh.....(17) *
5. r.....(18) *
6. h.....(11)

Affricates (AF):

1. ts.....(20)
2. dz.....(19)
3. tsh.....(16) *
4. dr.....(15) *
5. ch.....(13)
6. j.....(12)

Glides (GL)

1. y.....(35)
2. w.....(36)
3. yw.....(37) *

Lateral (LA):

1. l.....(8)

* sounds not used in English.

D. THE SOUND COMBINATIONS OF CHINESE

We have devoted a lot of effort to studying the phonemes of Chinese speech and we are now ready to make a further step to examine the sounds of Chinese characters. As we know already, Chinese characters are always single syllable and formed by an

initial followed by a final, an initial plus a Glide and then followed by a final or just a final itself. This statement now needs a minor amendment. Since Glides can also function as intials, we know that a Glide followed by a final can also form a Chinese character sound.

Table 5 provides a matrix of Chinese character sounds formed by initials followed by finals. The total possible sound combinations are 374. This number is obtained by multiplying 21(initials) by 17(finals, including <ih>) and then adding 17. The extra 17 represents the character sounds formed by only finals themselves. However, according to information provided in [Ref. 3: p. 30], only 220 are actually existing in Chinese speech. In Table 5, letter x represents those sounds that are actually existing. Letter c(hange) and d(elete) represent the sounds that the author has modified. In the author's opinion, the sounds <bo>, <po>, <mo>, <fo> and <lo> should actually be <bwo>, <pwo>, <mwo>, <fwo> and <lwo> respectively. The letter n(ew) represents the sounds not appearing in the source table. See also Appendix A for the original table used, which, however, has been reformatted by the author for easy observation.

Table 6 provides a matrix of character sounds formed by an initial plus a Glide and then followed by a final. The total possible sound combinations are 484. There are, among them, 22 that are actually sounds formed by a Glide followed by a final. From the same information source, however, there are only 190 actually existing. There are, hence, 858 total possible sound combinations and only 410 of them actually exist in Chinese speech.

E. THE TONES AND INTONATIONS OF CHINESE

Mandarin Chinese is a tone language, because it uses pitches to distinguish lexical meaning[9]. There are four lexical tones in Chinese. Usually they are refered to as tone-1 through tone-4. They are also called, in Chinese, <in1 ping2> for tone-1, <yang2 ping2> for tone-2, <shang4 sheng1> for tone-3 and <chiu4 sheng1> for tone-4. These tones may be associated with any sound combination to form at least four different Chinese Characters if all of them exist. Chinese character <ma>, for example, associated with tone-1 means 'mother'; with tone-2 means 'numb'; with tone-3 means 'horse'; with tone-4 means 'to scold'.

---

[9]Lexical meaning is the meaning of the base (as the word play) in a paradigm (as plays, playing and played).

TABLE 5

INITIAL+FINAL SOUND COMBINATIONS

finals	initials																					
		b	p	m	f	d	t	n	l	g	k	h	j	ch	hs	dr	tsh	sh	r	dz	ts	s
a	x	x	x	x	x	x	x	x	x	x	x	x				x	x	x		x	x	n
o	x	c	c	c	c				d													
e	x			x		x	x	x	x	x	x	x				x	x	x	x	x	x	x
ea	x																					
ai	x	x	x	x		x	x	x	x	x	x	x				x	x	x		x	x	x
ei	x	x	x	x	x	x		x	x	x		x				x		x		x		x
ao	x	x	x	x		x	x	x	x	x	x	x				x	x	x	x	x	x	x
oa	x		x	x	x	x	x	x	x	x	x	x				x	x	x	x	x	x	x
an	x	x	x	x	x	x	x	x	x	x	x	x				x	x	x	x	x	x	x
en	x	x	x	x	x			x		x	x	x				x	x	x	x	x	x	x
ang	x	x	x	x	x	x	x	x	x	x	x	x				x	x	x	x	x	x	x
eng	x	x	x	x	x	x	x	x	x	x	x	x				x	x	x	x	x	x	x
er	x																					
i	x	x	x	x		x	x	x	x			•	x	x	x							
oo	x	x	x	x	x	x	x	x	x	x	x	x				x	x	x	x	x	x	x
iu	x							x	x				x	x	x							
ih																x	x	x	x	x	x	x

c for change, d for delete and n for newly add.

There are two features deserving special attention. First, not all sound combinations are associated with all four tones. According to an early investigation described in [Ref. 9], there are only 1272 out of a total of 1640 sound-tone combinations actually existing in the Chinese language. Secondly, there are more than forty-eight thousand Chinese characters. Among these characters, 4808 are frequently used. In either case, a severe homonymic problem occurs. Take the sound <i> as an example. There are 173 Chinese homonyms[10] existing in the Chinese language. It will be impossible for a recognizer to distinguish these characters. Therefore, a vocabulary formed by at least more than one character is recommended.

According to a 5-point tone system established by Dr. Drao Ywan Ren many years ago, the Chinese tone can be expressed with a 5-level pitch matrix. Imagine that the matrix is in the first quadrant of a rectangular coordinate system. On the vertical axis, five points, from one to five, represent the pitch of a Chinese character from low

[10]Homonyms are words/characters that are spelled and pronounced alike but are different in meaning.

TABLE 6

INITIAL+GLIDE+FINAL SOUND COMBINATIONS

G + F	initials																						
		b	p	m	f	d	t	n	l	g	k	h	j	ch	hs	dr	tsh	sh	r	dz	ts	s
ya	x								x				x	x	x							
yo	x																					
yea	x	x	x	x		x	x	x	x				x	x	x							
yai	x																					
yao	x	x	x	x		x	x	x	x				x	x	x							
yoa	x			x		x		x	x				x	x	x							
yan	x	x	x	x		x	x	x	x				x	x	x							
in*	x	x	x	x				x	x				x	x	x							
yang	x							x	x				x	x	x							
ing*	x	x	x	x		x	x	x	x				x	x	x							
wa	x									x	x	x				x	x	x				
wo	x	n	n	n	n	x	x	x	x	x	x	x				x	x	x	x	x	x	x
wai	x									x	x	x				x	x	x				
wei	x					x	x			x	x	x				x	x	x	x	x	x	x
wan	x					x	x	x	x	x	x	x				x	x	x	x	x	x	x
wen	x					x	x		x	x	x	x				x	x	x	x	x	x	x
wang	x									x	x	x				x	x	x				
oong*	x					x	x	x	x	x	x	x				x	x		x	x	x	x
ywea	x							n	x				x	x	x							
ywan	x								x				x	x	x							
iun*	x								x				x	x	x							
iung*	x												x	x	x							

* are sounds actually formed by (Vowel + Nasal).
n for newly add.

to high. On the horizontal axis, five points, again from one to five, represent the elapsed time unit for pronouncing the particular character. Tone-1, then, can be graphed as the line connecting the points (1,5), (2,5), (3,5) and (4,4). Tone-2 is the line connecting the points (1,3) and (4,5). Tone-3, a little strange, is the line connecting the points (1,2), (2,1), (3,1), (4,1) and (5,4). Tone-4 is the line connecting the points (1,4), (2,3), (3,2) and (4,1). Consult [Ref. 3: p. 34] for detailed information. The appendix of [Ref. 7] has provided an intensive discussion on the Chinese tones from the viewpoint of spectrographic evidence.

The 5-point Chinese tone system is, in this author's opinion, achieved by adding an extra level between level 3 and level 2 and between level 2 and level 1 of the English pitch system. A simplified version can be used to sufficiently express these tones. In this new version, similar to English, tone-1 is given a symbol of 55; tone-2, 35; tone-3,

214; tone-4, 51. Two utterances, numbered 01 and 50, selected from Appendix C are displayed as examples in Figure 3.1 and Figure 3.2. Again, the pitch changes occur only at the vowel sound of each character.

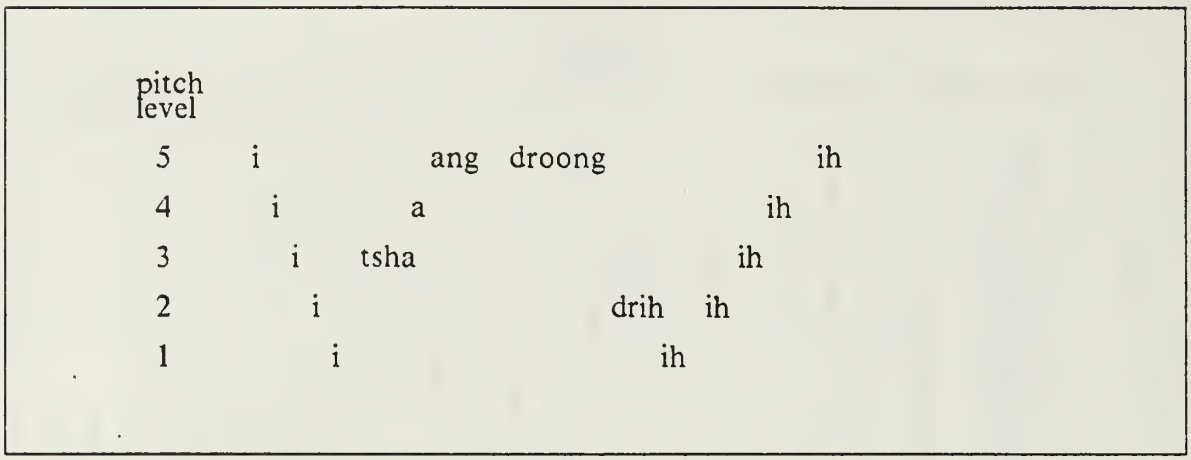


Figure 3.1 Examples of Chinese Tone Patterns.

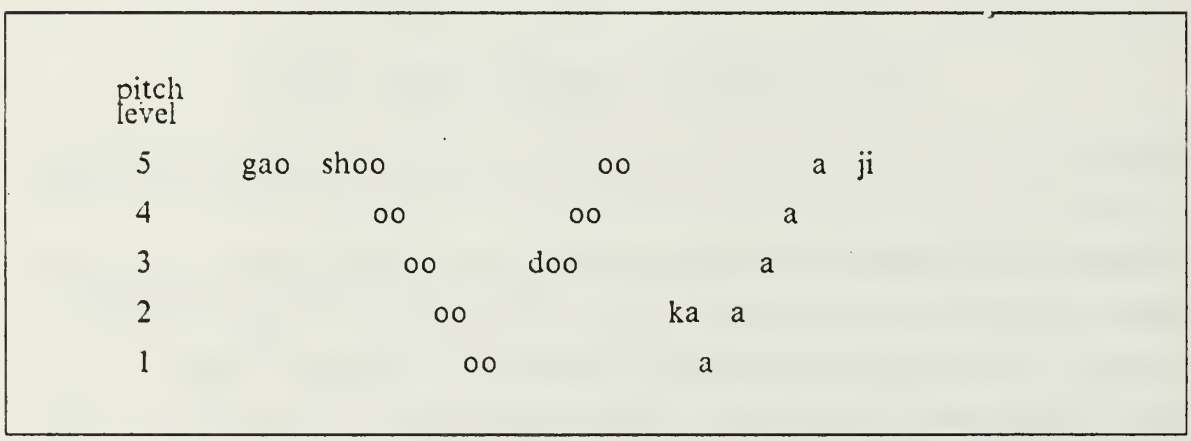


Figure 3.2 Examples of Chinese Tone Patterns.

The rule for the intonation of Chinese, in a sense, is relatively simple. Basically, each character in a sentence remains the same tone pattern as they independently appear. Therefore, when combining the two examples shown in Figure 3.1 and Figure 3.2, a complete imperative sentence is obtained, which means in English, 'Abort the high speed reader.' However, the spectrographic evidence showed that although the

tone pattern was generally maintained, both the elapsed time units and pitch levels of the character sound passing through were both slightly shortened when appearing within a sentence.

IV. A DESCRIPTION OF THE EXPERIMENTS

A. OBJECTIVES

The experiment was actually a package of three related subexperiments. The first one was to examine the recognition of Chinese phonemes. A similar experiment examining the recognition of English phonemes was also conducted to obtain a comparison reference. The second part was to examine the recognition of a set of ninety Chinese utterances[11] in a simulated speaker independent[12] mode. More information can be found in [Ref. 10]. The third part of the experiment was to examine the recognition of the same set of Chinese utterances in speaker dependent[13] mode.

The objective of the experiment was to determine if Chinese speech could be an effective communication medium between human beings and computer systems. Since no similar study had been conducted, especially using Chinese phonemes, the information obtained would, hopefully, serve as a basis for the further Chinese speech recognition/input studies.

B. SUBJECTS

The first part of the experiment was conducted by the author himself, because it required a thorough understanding of the articulatory phonetics. Ten subjects participated in the remaining parts of the experiment on a volunteer basis. All of the subjects were male students from the Republic of China and studying at the Naval Postgraduate School. Two civilian students were working on their doctor's degree. The remaining eight subjects were naval officers and were working on their master's

---

[11]An utterance can be spoken words, phrases or any form of voice that is meaningful to the speaker.

[12]A speaker independent system contains algorithms which supposely can handle many different voices and dialects. The system should be able to recognize the voice of anyone who tries to use it. Since it requires no previous samples of a given user's voice, then, theoretically, we would not expect the speaker independent system to work as perfectly as a speaker dependent system.

[13]A speaker dependent system requires samples of the potential user's voice to be in memory in order to work properly. Because it is tuned to the user's voice, the speaker dependent system should work better than an independent system.

degree. The ranks for those officers ranged from Lieutenant Junior Grade to Lieutenant inclusive. All subjects were between the ages of 24 and 32 inclusive. Only two of the subjects had ever heard about voice recognition before. However, none of the subjects had any previous experience on the voice recognition system used in the experiment.

C. EQUIPMENT

A T600 voice recognition system of Threshold Technology Inc. (TTI) was used as the recognizer for the experiment. The model T600 is a speaker-dependent, isolated utterance recognizer. The recognition unit contained memory which allowed a maximum of 256 spoken utterances to be stored. The length of each utterance, required by the T600, is between one tenth second and two seconds. A pause of at least one tenth second between utterances is also required to signal that the first utterance has ended and the second utterance may be coming. Each utterance can be associated with a maximum 16-character ASCII string as a recognizer output to a host computer system. In this experiment, however, the output string was only displayed on the screen of a local terminal for purposes of verifying a correct recognition.

The system comprises a TTI 8036-3 model main processor unit, a TTI 7020A tape cartridge unit, a TTI 8013 speech level control unit, an Ann Arbor 400 model large-character keyboard/display terminal and a Shure SM-10 noise-cancelling microphone with headset. Please consult [Refs. 11,12] for more information.

D. VOCABULARY

The vocabulary used in this experiment was a group of ninety computer related terms selected from [Ref. 13]. The first priority for the selection was to cover as many sounds as possible. Tables 7 and 8 allow readers to have an idea about which sounds were used in the experiment. The numbers in both tables represent the times that particular sound was used. The second priority was to equally distribute a certain number of utterances into different length categories. However, this attempt was not successfully achieved, because Chinese have an intention to form their terms with two or four characters to obtain a sense of symmetry. Hence, the vocabulary came up with 28 two-character utterances, 16 three-character utterances, 26 four-character utterances, 15 five-character utterances and 5 six-character utterances. Appendix C

lists all the Chinese utterances in their romanization forms. The numbers following each character represent the tone of the character and the last number, enclosed in parentheses, is the number of syllables in the utterance. Of course, the last number also represents the number of characters in the utterance. Appendix D provides the English meaning of those Chinese utterances used and lists them in an alphabetical order.

TABLE 7

I+F SOUND COMBINATIONS USED IN THE EXPERIMENT

finals	initials																					
		b	p	m	f	d	t	n	l	g	k	h	j	ch	hs	dr	tsh	sh	r	dz	ts	s
a	x	x	x	3	1	1	x	x	x	x	2	x				x	3	x		1	x	n
o	x	c	c	c	c				d													
e	x			x		x	x	x	x	1	1	2				x	x	2	x	1	1	x
ea	x																					
ai	x	1	1	2		2	1	x	x	x	x	x				x	x	x		1	x	x
ei	x	1	1	x	x	x		x	1	1		x				x		x		x		x
ao	x	2	1	x		1	1	1	x	2	2	1				1	x	1	x	x	1	x
oa	x		x	x	x	x	x	x	1	1	x	x				1	x	x	x	x	x	x
an	1	1	x	x	2	1	1	x	x	1	1	x				1	1	1	x	x	1	x
en	x	1	x	x	2			x		x	x	x				2	x	x	1	x	x	x
ang	x	x	x	x	1	2	x	x	x	x	x	x				x	1	x	x	x	x	x
eng	x	x	x	x	1	1	x	2	x	x	x	x				3	3	1	x	1	x	x
er	1																					
i	3	1	1	x		1	2	x	5				9	5	4							
oo	3	1	x	x	4	2	2	x	2	1	1	1				3	6	6	1	1	x	x
iu	4							x	1				1	2	1							
ih																9	1	10	x	5	2	1

E. PROCEDURE

The entire experiment was conducted in the evening or on weekends to avoid any possible noise interruption. All subjects were gathered and provided a brief orientation on the experiment itself and the procedure of the experiment in advance. Discussions were also provided to ensure the subjects sensed the flavor of the experiment. Subjects were asked to come, one at a time, to the Man-Machine Interface Lab to conduct the experiment. First, the recognition system was input with voice samples of the author,

TABLE 8

I+G+F SOUND COMBINATIONS USED IN THE EXPERIMENT

	initials																						
G + F		b	p	m	f	d	t	n	l	g	k	h	j	ch	hs	dr	tsh	sh	r	dz	ts	s
ya	x								x				x	x	x							
yo	x																					
yea	1	x	x	x		1	x	x	1				1	x	1							
yai	x																					
yao	x	4	x	2		x	1	x	3				2	1	1							
yoa	1			x		x		x	1				1	x	x							
yan	3	3	1	1		4	x	x	1				4	x	2							
in*	3	x	x	x				x	1				1	x	3							
yang	1							x	2				x	x	1							
ing*	1	1	1	x		1	1	x	2				1	x	3							
wa	x									x	x	1				x	x	x				
wo	x	n	n	n	1	x	x	x	x	x	x	x				x	x	x	x	1	1	1
wai	x									x	1	1				x	x	1				
wei	6					1	x			1	1	1				x	1	x	x	1	x	x
wan	1					1	x	x	1	1	1	1				1	1	1	x	x	x	x
wen	1					x	x		x	x	x	x				1	3	x	x	x	x	x
wang	x									1	1	x				1	x	1				
oong*	x					2	4	x	x	2	2	x				3	2		x	1	x	x
ywea	x							n	x				x	1	1							
ywan	3								x				x	1	1							
iun*	1								x				x	1	2							
iung*	2												x	x	x							

which was prerecorded in a training session. The subjects, then, read in each utterance three times through a microphone using the author's reference templates[14]. The author recorded the outputs shown on the terminal screen. Two out of three or more wrong outputs (including no output displayed, in this case the system provided a beep sound) for each utterance was considered an incorrect recognition; otherwise, a correct recognition. After this simulated speaker independent mode was completed, the subjects were instructed to retrain all ninety utterances by introducing individual voice samples into the recognition system. When the training was done, the subjects started to read in, again, each utterance five times. Since, at this time, a speaker dependent mode recognition was conducted, the criterion was escalated. Unless five correct outputs in series were recorded at the first trial, the recognition was considered incorrect. When some utterances couldn't be correctly recognized at all, a retraining

[14]A template is the digital representation or matrix of the utterance which is stored by the recognizer and used later as a reference to perform recognition.

was allowed until a correct recognition was finally obtained. After all results were recorded, the experiment was concluded.

V. THE RESULTS, DISCUSSIONS AND THE SUGGESTIONS

A. THE RESULTS OF PHONEME RECOGNITION

Tables 9 and 10 have respectively provided a copy of record of the recognition of English and Chinese phoneme performed by T600 VRS. It is a 21-session experiment conducted at most once a day within a period of two months. The author read every phoneme five times during each session and recorded the number of times, out of every five trials, that the phoneme was correctly recognized. The author also retrained the recognizer for those phonemes that cannot be properly recognized at the end of the 1st, 5th, 9th, 13th and 17th session. A further discussion is provided in the following paragraphs.

The information provided in Table 11 is directly derived from Tables 9 and 10. The second and fifth columns of the table are the total number out of 105 trials that each phoneme was correctly recognized during the entire experiment. The third and sixth columns are the averages of each total number over 21 sessions. The percentage of each total number is also listed in columns four and eight. From the information provided in this table, we may obtain some idea about the recognition of phonemes.

First, the recognition of vowels is better than the recognition of consonants. The table is designed in a format that presents single vowels first, diphthongs second and consonants the last. A comparison between the upper half and the lower half of the table helps to illustrate this findings. Second, among vowels, the recognition of diphthongs is better than the recognition of single vowels. The diphthongs include </ai/>, /ow/, /oy/ and <ao>. Third, among single vowels, the recognition of Tense Vowels is better than the recognition of Lax Vowels. The Tense Vowels include </ee/>, <er>, <iu>, </ei/>, /au/, </oo/>, </oa/> and /ur/. Fourth, among the consonants, the recognition of voiced sounds is better than the recognition of voiceless sounds. The voiceless sounds include </p/>, </t/>, </k/>, </f/>. /th/, </s/>, <hs>, <sh>, /sh/, <ts>, <tsh> and </ch/>. A comparison between those sound pairs can help one to appreciate this finding.

An overall comparison between the phoneme recognition of the two languages is shown at the end of the Table 11. There were 89.59% of the total trials correctly

TABLE 9

RECOGNITION OF ENGLISH PHONEMES

	session																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
ee	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
i	5	5	5	4	5	5	3	5	5	5	5	5	5	4	5	4	5	5	5	5	5
ei	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
ea	5	5	3	4	5	4	5	3	4	5	5	5	5	4	5	5	5	5	5	5	5
au	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
oo	5	5	5	4	5	4	3	4	5	5	4	4	4	4	5	5	5	5	5	4	5
o	5	5	4	5	5	5	5	5	5	5	4	5	5	5	4	5	5	5	5	5	5
oa	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
aw	5	4	5	4	5	3	5	5	5	5	5	5	4	5	4	3	3	5	5	5	5
a	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
u	5	5	5	5	5	5	5	5	5	5	4	5	5	5	5	5	5	5	5	5	5
e	5	5	4	5	5	4	5	5	5	5	3	5	5	4	4	4	5	5	4	5	5
er	5	5	5	5	4	4	5	2	4	4	4	5	3	3	5	4	5	5	4	3	5
ur	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
ai	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
ow	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
oy	5	5	5	5	5	5	5	5	5	5	5	5	4	5	5	5	5	5	5	5	5
m	5	3	3	3	5	3	5	3	3	4	3	4	3	5	4	5	4	3	4	4	5
n	5	5	3	5	5	4	3	5	5	5	4	5	5	5	5	5	5	5	5	5	5
ng	4	5	3	4	5	4	5	4	3	2	0	4	3	4	5	4	4	3	3	4	5
p	5	3	3	3	4	5	2	4	5	5	2	5	5	5	5	4	4	5	5	5	5
b	5	4	5	3	3	5	3	4	5	5	3	3	0	4	3	4	5	3	3	5	4
t	5	3	5	4	4	3	3	5	4	4	5	3	1	4	3	4	4	3	4	2	4
d	5	5	5	5	5	5	5	4	5	5	4	5	5	5	3	5	5	5	5	5	5
k	5	3	5	5	3	4	3	4	3	3	5	3	1	5	3	4	5	3	4	5	5
g	5	2	2	3	3	3	5	4	3	2	5	5	0	5	5	5	4	5	4	5	5
f	5	3	2	3	3	2	3	4	4	3	0	3	2	4	5	3	5	5	4	5	4
v	5	5	4	4	4	4	5	4	3	4	5	3	5	5	5	5	5	5	4	3	4
th	5	5	5	5	3	1	3	5	4	5	3	3	3	5	3	3	5	5	5	5	5
the	4	3	4	5	5	5	4	5	4	5	5	5	5	5	5	5	5	5	5	5	4
s	4	3	2	3	5	3	4	4	4	3	3	2	4	3	5	5	3	5	4	3	4
z	5	4	4	5	4	4	5	5	5	4	4	5	5	4	5	5	5	5	5	5	5
sh	5	5	4	5	5	5	5	5	4	5	5	5	2	5	5	5	5	4	4	4	5
ge	5	3	3	3	2	3	3	3	3	3	4	5	2	4	3	4	2	3	4	4	4
h	5	5	3	5	3	4	5	4	4	3	3	4	4	5	5	5	4	5	5	5	5
ch	4	5	5	5	5	5	5	5	5	5	4	4	4	5	5	5	5	5	3	5	4
j	4	5	5	4	3	5	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5
l	5	4	5	5	5	3	5	4	5	4	5	5	3	5	5	5	5	4	3	4	4
y	5	5	5	4	5	3	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
r	5	5	5	5	5	5	5	5	5	5	5	4	5	5	5	4	4	5	5	4	5
w	5	4	5	5	4	4	4	4	3	5	5	5	5	5	5	5	5	4	5	5	5

Figures here represent the number of times correctly recognized out of five trials.

TABLE 10

RECOGNITION OF CHINESE PHONEMES

	session																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
ih	4	4	5	5	4	4	4	4	4	5	4	5	5	5	4	5	5	5	5	5	4
ee	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
er	5	5	4	5	4	5	5	5	5	5	5	5	5	5	5	5	4	5	5	5	5
i	5	5	5	5	5	5	5	5	4	4	3	4	5	5	5	5	5	5	5	5	4
iu	5	5	5	5	5	5	5	5	5	5	5	4	5	5	5	5	5	5	5	5	5
ei	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
ea	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
oo	5	5	5	5	5	5	5	4	5	5	5	5	5	4	5	5	5	5	5	5	5
o	5	3	3	3	3	5	3	5	3	2	3	3	3	5	3	5	3	5	5	5	5
oa	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
a	4	5	5	5	5	5	4	5	5	4	5	3	5	5	5	5	5	5	4	4	5
aa	5	5	4	5	5	4	3	4	5	5	4	5	3	4	4	5	5	5	5	5	5
u	5	5	5	5	5	4	5	4	5	5	5	5	5	5	5	5	5	4	5	5	5
e	4	0	2	4	5	5	5	5	4	5	3	5	5	5	5	3	5	5	5	4	4
ai	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
ao	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
m	4	2	5	3	4	3	4	3	3	3	4	4	4	4	4	4	4	4	3	4	5
n	4	5	5	4	5	3	5	5	5	5	5	5	5	5	4	5	5	4	5	5	5
ng	4	5	3	3	5	3	4	3	4	4	4	4	3	5	3	5	2	5	4	5	5
p	5	3	5	5	4	3	3	4	3	5	3	4	4	3	5	4	4	5	3	3	5
b	5	5	2	5	4	5	5	5	4	5	5	5	5	5	5	5	5	5	4	5	5
t	4	4	3	3	3	5	5	5	5	4	4	5	5	4	5	4	3	5	4	4	5
d	5	5	5	5	5	3	5	5	5	5	4	3	5	5	5	4	5	5	3	5	5
k	5	4	5	4	5	4	3	4	3	4	5	4	4	5	5	5	3	5	3	4	5
g	5	4	2	3	1	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
f	5	4	4	5	4	4	4	3	5	5	3	5	5	5	4	3	4	3	3	4	5
s	4	4	4	4	2	3	3	4	3	3	3	3	4	5	5	4	3	4	4	3	4
hs	4	5	3	4	3	5	3	3	4	3	4	5	3	5	3	5	4	4	3	4	5
sh	4	4	4	5	4	4	3	2	4	3	5	3	4	4	3	4	4	5	4	5	4
r	4	3	4	4	3	5	4	3	4	4	4	5	4	4	4	5	4	4	3	5	4
h	5	5	5	4	3	3	5	5	4	4	5	5	5	5	5	3	5	3	3	4	4
ts	4	2	4	3	3	3	3	4	3	2	3	4	3	5	4	4	5	4	5	3	4
dz	5	5	5	5	5	5	5	5	4	5	5	5	5	5	5	5	5	5	5	5	5
tsh	5	3	3	2	0	3	3	4	4	4	3	4	3	5	4	4	3	4	4	5	4
dr	5	4	4	3	4	3	3	4	3	4	5	3	4	5	4	4	5	5	5	4	4
ch	5	5	4	4	3	5	4	4	4	5	5	5	4	5	4	4	3	4	3	5	3
j	5	3	4	4	3	5	4	3	4	2	2	3	4	4	5	5	5	5	5	5	5
l	4	3	3	4	5	5	5	5	5	5	4	5	5	5	5	5	5	3	5	5	3
y	5	5	3	5	4	3	4	4	4	3	4	3	5	4	5	3	5	5	5	5	5
w	5	5	5	5	5	5	5	5	5	5	4	5	5	5	5	4	4	5	5	5	5
yw	4	5	5	5	5	4	3	4	4	5	5	5	4	5	4	5	4	5	5	4	5

Figures here represent the number of times correctly recognized out of five trials.

TABLE 11

RECOGNITION PERFORMANCE OF PHONEMES

English	TOT	AVE	%	Chinese	TOT	AVE	%
--				ih	95	4.52	90.48
ee	105	5.00	100	ee	105	5.00	100
--				er	102	4.86	97.14
i	100	4.76	95.24	i	99	4.71	94.29
--				iu	104	4.95	99.05
ei	105	5.00	100	ei	105	5.00	100
ea	97	4.62	92.38	ea	105	5.00	100
au	105	5.00	100	--			
oo	95	4.52	90.48	oo	103	4.90	98.10
o	102	4.86	97.14	o	80	3.81	76.19
oa	105	5.00	100	oa	104	4.95	99.05
aw	95	4.52	90.48	--			
a	105	5.00	100	a	98	4.67	93.33
u	104	4.95	99.05	u	102	4.86	97.14
e	97	4.62	92.38	e	88	4.19	83.81
er	89	4.24	84.76	--			
ur	105	5.00	100	--			
--				aa	95	4.52	90.48
ai	105	5.00	100	ai	105	5.00	100
ow	105	5.00	100	--			
oy	104	4.95	99.05	--			
--				ao	105	5.00	100
m	81	3.86	77.14	m	78	3.71	74.29
n	99	4.71	94.29	n	99	4.71	94.29
ng	78	3.71	74.29	ng	83	3.95	79.05
p	89	4.24	84.76	p	83	3.95	79.05
b	79	3.76	75.24	b	99	4.71	94.29
t	77	3.67	73.33	t	89	4.24	84.76
d	101	4.81	96.19	d	97	4.62	92.38
k	81	3.86	77.14	k	89	4.24	84.76
g	80	3.81	76.19	g	95	4.52	90.48
f	72	3.43	68.57	f	87	4.14	82.86
v	91	4.33	86.67	--			
th	86	4.10	81.90	--			
the	98	4.67	93.33	--			
s	76	3.62	72.38	s	76	3.62	72.38
z	98	4.67	93.33	--			
--				hs	82	3.90	78.10
--				sh	82	3.90	78.10
--				r	84	4.00	80.00
sh	97	4.62	92.38	--			
ge	70	3.33	66.67	--			
h	91	4.33	86.67	h	90	4.29	85.71
--				ts	75	3.57	71.43
--				dz	104	4.95	99.05
--				tsh	74	3.52	70.48
--				dr	85	4.05	80.95
ch	98	4.67	93.33	ch	88	4.19	83.81
j	99	4.71	94.29	j	85	4.05	80.95

TABLE 11 (CONT'D)
RECOGNITION PERFORMANCE OF PHONEMES

English	TOT	AVE	%	Chinese	TOT	AVE	%
l	93	4.43	88.57	l	94	4.48	89.52
y	102	4.86	97.14	y	89	4.24	84.76
--				yw	95	4.52	90.48
r	101	4.81	96.19	--			
w	97	4.62	92.38	w	102	4.86	97.14
TOTAL	3857	4.48	89.59		3799	4.41	88.25

recognized in English phoneme recognition and 88.25% of the total trials in Chinese phoneme recognition. There is only a 1.34% of difference between the two languages. This finding highly suggests that Chinese speech should also capable of being a bridge between human beings and computer systems.

Table 12 is derived from Table 11 by deleting those phonemes that are uniquely existing in one language only. Therefore, the 28 phonemes, which is 68.29 % of the total number, shown in Table 12 are mutually used by the two languages. Since the author used the same recognizer to examine the phonemes of the two languages, the results of the recognition for each pair of phonemes should be very similar or even the same. However, shown in the table, this is not true. Some significant degradations between the pair are observed. The reason for this is, probably, the existance of certain sounds having similar characteristics. That is, the more similar the phonemes are in the group, the worse the recognition of the phonemes will be.

During the experiment, the author noticed and recorded some consistent substitution errors that occured with certain sounds. A substitution error is when an input utterance was calculated as a closer match to a different template in storage and caused an incorrect recognition. The information about these substitution errors has been put in columns titled 'SUB' of the table. Although not all observed degradations have recorded a consistent error, the existing information provides a possible explaination for the degradations.

TABLE 12

RECOGNITION PERFORMANCE OF SHARED PHONEMES

	English			Chinese		
	TOT	AVE	SUB	TOT	AVE	SUB
ee	105	5.00		105	5.00	
i	100	4.76		99	4.71	
ei	105	5.00		105	5.00	
ea	97	4.62	au	105	5.00	
oo	95	4.52		103	4.90	
o	102	4.86		80	3.81	eng
oa	105	5.00		104	4.95	
a	105	5.00		98	4.67	
u	104	4.95		102	4.86	
e	97	4.62		88	4.19	
ai	105	5.00		105	5.00	
m	81	3.86		78	3.71	
n	99	4.71		99	4.71	
ng	78	3.71	aw	83	3.95	
p	89	4.24		83	3.95	
b	79	3.76		99	4.71	
t	77	3.67		89	4.24	
d	101	4.81		97	4.62	
k	81	3.86		89	4.24	
g	80	3.81		95	4.52	
f	72	3.43	th	87	4.14	
s	76	3.62	th	76	3.62	ts
h	91	4.33		90	4.29	
ch	98	4.67		88	4.19	tsh
j	99	4.71		85	4.05	dr
l	93	4.43		94	4.48	
y	102	4.86		89	4.24	
w	97	4.62		102	4.86	

B. THE RESULTS OF CHINESE UTTERANCE RECOGNITION

Table 13 provides general information about the performance of the subjects during the Chinese utterance recognition experiment. The first column lists the number of the subject. The second column provides the number of correct recognitions by the recognizer for each subject in the speaker independent mode of the experiment. The third column is the percentage of that number against the total

vocabulary, which is ninety, stored in the recognizer. The fourth column provides the number of incorrect recognitions by the recognizer for each subject in the speaker dependent mode. The fifth column is the percentage again. The last column lists the number of vocabulary items that each subject had to retrain to finally obtain a correct recognition in the speaker dependent mode experiment. An overall information is provided at the end of Table 13 , which showed that 74.67% of 900 trials of simulated speaker independent mode recognition were correctly recognized by the recognizer and 12.44% of 900 trials of speaker dependent mode recognition, on the first attempt, were incorrectly recognized by the recognizer. Only 16 utterances required a retraining to eventually obtain a correct recognition.

The speaker independent mode, as mentioned earlier, is supposed to have a worse performance because the contemporary technique cannot fully support the function. A correct recognition in this mode is more meaningful, hence, deserves more attention. On the other hand, speaker dependent mode is supposed to have a better performance because it has been fully supported by the subject's own voice. Therefore, an incorrect recognition in this mode certainly conveys more information and deserves more attention. Table 14 lists all ninety utterances used in the experiment. The first and fifth columns are the number of each utterance. The second and sixth columns are the number of syllables that each utterance has. The third and seventh columns are the times of the correct recognition for each utterance. A percentage of the correct recognition times against total trials, which is 10, of each utterance is provided in columns four and eight. Table 15 is similar to Table 14. The only difference is that the third and seventh columns provide the times of the incorrect recognition for each utterance.

Table 16 provides information on the relationship between the syllable numbers of each utterance and the recognition performance. The first column lists the numbers of the syllables for each Chinese utterance. The second column lists the numbers of utterances formed by that certain number of syllables. The third column, I(ndependent) and C(orrect), was derived from Table 14 by selecting those with more than nine correct recognitions inclusive. The fourth column is the percentage of the correct number against the total number of utterances for certain syllable lengths. The fifth column, D(ependent) and I(ncorrect), was derived from Table 15 by selecting those with more than two incorrect recognitions inclusive. The sixth column, again, is the percentage.

The author has no intention to make any further statistical analysis because of the limited scale of the experiment itself and the data collected. However, the author has reached the goal that he set for the experiment. That is, by the support of existing information collected, the Chinese speech, accompanied with speaker dependent recognition, could be a good communication medium with computer systems. Applications of Chinese speech recognition/input are expected in the future. Some possible applications in the near future are production line routing, quality control, inventory control, package sorting and some military applications such as weapon systems control and Combat Information Center operations, etc.

TABLE 13

GENERAL PERFORMANCE OF THE SUBJECTS

Subject	Indep Correct	%	Depen Incorre	%	Retraining Needed
1	56	62.22	19	21.11	0
2	64	71.11	6	6.67	2
3	82	91.11	6	6.67	1
4	68	76.56	7	7.76	2
5	47	52.22	3	3.33	0
6	80	88.89	4	4.44	3
7	67	74.44	31	34.44	2
8	67	74.44	21	23.33	2
9	64	71.11	4	4.44	1
10	77	85.56	11	12.22	3
TOTAL	672	74.67	112	12.44	16

C. SUGGESTIONS FOR THE FUTURE

This thesis has contributed to clarifying some existing confusion in the Chinese phonetic system. The main purpose was to establish a solid, error-free basis for researchers in their future studies of Chinese voice recognition and voice input. By doing so, those researchers will no longer base their studies on a questionable phonetic system.

This study on the phonetic system is also quite future oriented. In speaker dependent recognition, the voice samples the system stores are directly obtained from

TABLE 14

CORRECT RECOGNITIONS IN INDEPENDENT MODE

NO	SYL	NUM	%	NO	SYL	NUM	%
01	4	9	90	46	5	10	100
02	4	7	70	47	5	8	80
03	3	7	70	48	5	10	100
04	2	7	70	49	4	9	90
05	2	5	50	50	5	8	80
06	4	7	70	51	5	6	60
07	6	10	100	52	4	7	70
08	5	6	60	53	5	8	80
09	3	6	60	54	5	8	80
10	2	9	90	55	2	8	80
11	4	7	70	56	3	6	60
12	3	10	100	57	3	5	50
13	3	2	20	58	2	10	100
14	3	6	60	59	2	7	70
15	4	6	60	60	2	4	40
16	5	7	70	61	2	5	50
17	2	3	30	62	2	10	100
18	5	5	50	63	3	9	90
19	2	10	100	64	4	6	60
20	3	5	50	65	2	6	60
21	6	10	100	66	3	10	100
22	4	8	80	67	2	4	40
23	2	6	60	68	3	7	70
24	5	9	90	69	2	8	80
25	2	4	40	70	4	10	100
26	2	5	50	71	4	10	100
27	2	9	90	72	2	8	80
28	4	9	90	73	4	10	100
29	6	10	100	74	4	10	100
30	3	4	40	75	5	8	80
31	4	10	100	76	4	6	60
32	4	2	20	77	5	10	100
33	4	4	40	78	4	7	70
34	2	6	60	79	2	8	80
35	4	4	40	80	2	9	90
36	2	4	40	81	4	10	100
37	6	9	90	82	4	10	100
38	2	10	100	83	5	9	90
39	6	10	100	84	3	10	100
40	4	8	80	85	4	6	60
41	5	8	80	86	4	7	70
42	2	7	70	87	2	10	100
43	4	4	40	88	3	10	100
44	2	6	60	89	3	8	80
45	3	7	70	90	2	10	100

the user himself. Therefore, no matter how the user pronounces his input, so long as it matches the way he pronounced it during the training session, the system will correctly recognize it. So, a phonetically wrong pronunciation will cause no trouble in a speaker

TABLE 15

INCORRECT RECOGNITIONS IN DEPENDENT MODE

NO	SYL	NUM	%	NO	SYL	NUM	%
01	4	1	10	46	5	1	10
02	4	3	30	47	5	0	00
03	3	3	30	48	5	1	10
04	2	2	20	49	4	0	00
05	2	1	10	50	5	0	00
06	4	3	30	51	5	3	30
07	6	3	30	52	4	0	00
08	5	3	30	53	5	1	10
09	3	2	20	54	5	1	10
10	2	0	00	55	2	0	00
11	4	3	30	56	3	0	00
12	3	2	20	57	3	2	20
13	3	1	10	58	2	0	00
14	3	1	10	59	2	1	10
15	4	2	20	60	2	2	20
16	5	0	00	61	2	1	10
17	2	0	00	62	2	2	20
18	5	2	20	63	3	1	10
19	2	4	40	64	4	0	00
20	3	1	10	65	2	3	30
21	6	1	10	66	3	0	00
22	4	1	10	67	2	1	10
23	2	2	20	68	3	0	00
24	5	1	10	69	2	2	20
25	2	2	20	70	4	1	10
26	2	4	40	71	4	1	10
27	2	2	20	72	2	4	40
28	4	0	00	73	4	0	00
29	6	0	00	74	4	0	00
30	3	2	20	75	5	1	10
31	4	1	10	76	4	0	00
32	4	0	00	77	5	5	50
33	4	0	00	78	4	0	00
34	2	1	10	79	2	0	00
35	4	0	00	80	2	0	00
36	2	2	20	81	4	0	00
37	6	1	10	82	4	0	00
38	2	1	10	83	5	0	00
39	6	1	10	84	3	0	00
40	4	4	40	85	4	1	10
41	5	0	00	86	4	0	00
42	2	3	30	87	2	1	10
43	4	1	10	88	3	1	10
44	2	7	70	89	3	1	10
45	3	0	00	90	2	0	00

dependent recognition system. On the other hand, once the voice recognition technique enters the phase of speaker independent mode, this articulation problem will be a factor requiring thorough considerations. The further studies relating to this

TABLE 16

RELATIONSHIP BETWEEN NUMBER OF SYLLABLES AND PERFORMANCE IN CHINESE UTTERANCE RECOGNITION

SYL	NUM	IaC	%	DaI	%
2	28	9	32.14	14	50.00
3	16	5	31.25	5	31.25
4	26	10	38.46	5	19.23
5	15	5	33.30	4	26.66
6	5	5	100	1	20.00

future-oriented problem on both English and Chinese voice recognition are highly encouraged to conduct as soon as possible.

Due to the same reason, two Chinese phonemes, <i> and <a>, will also be a potential trouble area when speaker independent recognition is applied in the future. Let us take <i> as an example. The word 'secret' in Chinese is <mi mi>. According to author's argument, there are, in fact, four possible ways to pronounce; namely, <mee mee>, <mi mi>, <mee mi> and <mi mee>. In dependent mode, so long as the user remembers which one is the voice sample he input into the system, he will have no trouble at all. However, when facing a speaker independent recognition system, the situation requires us to answer the questions such as: Can the three others than the one in memory be properly recognized? Is vocabulary design a possible alternative to solve the problem? Studies to answer these questions are certainly needed for the development of a speaker independent voice recognition system in the near future.

Vocabulary design is also an important factor in the performance of voice recognition systems. As described in [Ref. 10: p. 4], Prof. Gary Poock suggested using longer vocabulary phases for better recognition performance. The results shown in Table 16, however, only partially support his statement. This is probably because the author, when selecting the utterances, concentrated his efforts on covering more sound combinations. Hence, a further study, with more careful vocabulary design strategy, to research the relationship between the number of syllables of an utterance and the performance of the recognition system will be helpful.

A study, also stimulated by Table 16, on the relationship between phonemes and the recognition performance will also be appropriate. The experiment results suggested that the syllable numbers might not be the only determinant to recognition performance. Therefore, this new direction of study might provide an alternative way to obtain important information to seek better recognition performance.

Last but not least, this study was heavily based on knowledge absorbed from articulatory phonetics. The author, in fact, has used the knowledge to help people to produce more acceptable American English pronunciation in the past several years. By using exact speech organs and articulations, his students did establish a much better articulation custom, and, eventually, produce more acceptable pronunciation. Human beings can improve their pronouncing skill by the help of articulatory phonetics. Can we, then, apply this knowledge to help a voice recognition system obtain a better performance? The study to answer the question, to the author himself, will certainly be a very interesting one and deserve his constant devotion in the future.

APPENDIX A
ORIGINAL TABLES USED IN TEXT

PHONEMES OF AMERICAN SPEECH (ORIGINAL)

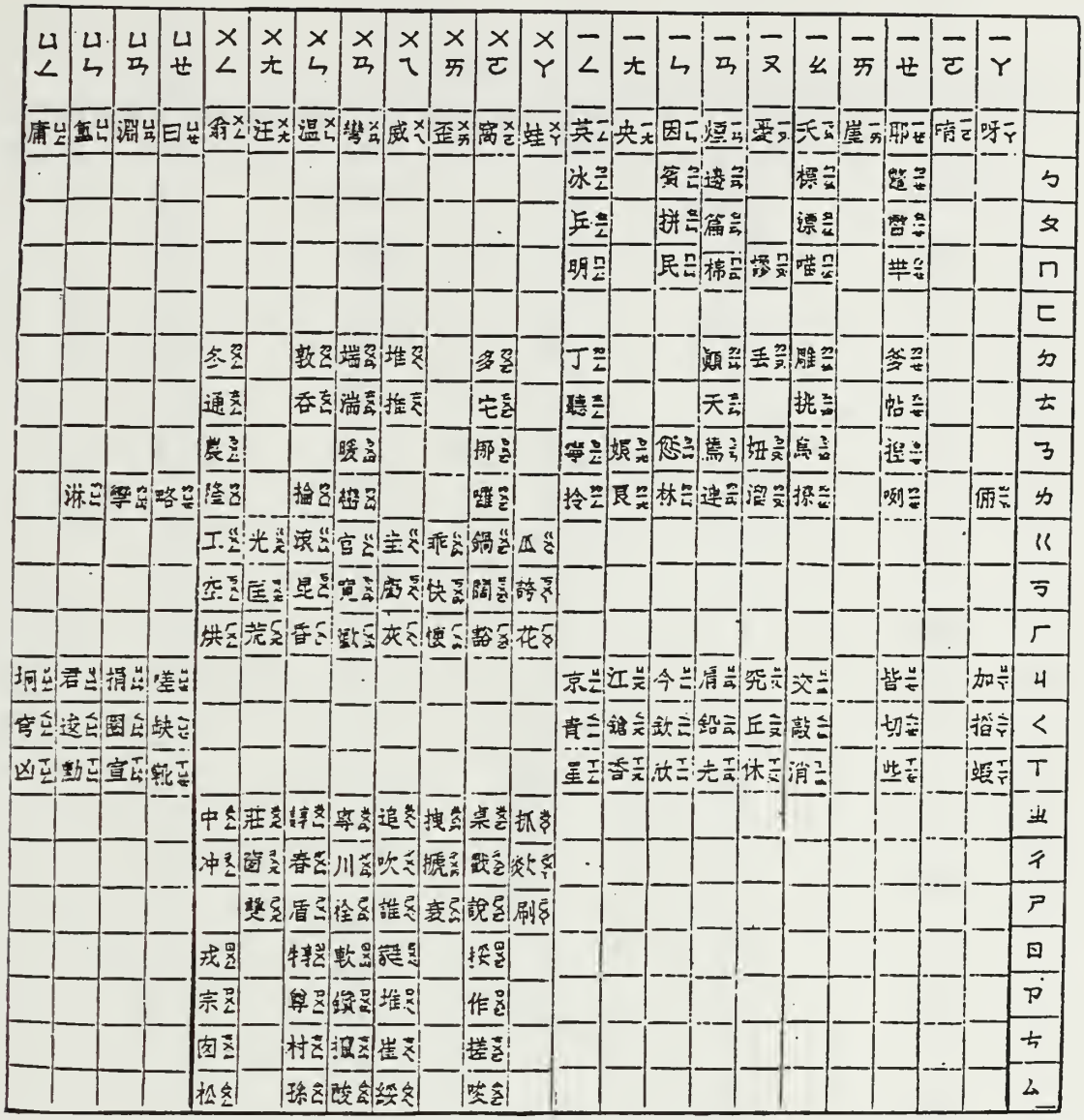
Vowels

Front vowels				Back vowels			
SYMBOL	KEY			SYMBOL	KEY		
[i]	heed	[hid]		[u]	who'd	[hud]	
[ɪ]	hid	[hɪd]		[ʊ]	hood	[hʊd]	
[e]	hayed	[hed]		[o]	hoed	[hod]	
[ɛ]	head	[hɛd]		[ɔ]	hawed	[hɔd]	
[æ]	had	[hæd]		[ɑ]	hod	[hɑd]	
Central vowels				**Diphthongs†**			
[ɝ-ɜ]*	hurt	[hɝt]		[aɪ]	file	[faɪl]	
[ʌ]	hut	[hʌt]		[aʊ]	fowl	[faʊl]	
[ɚ-ə]*	under	[ʌndɚ]		[ɔɪ]	foil	[fɔɪl]	
[ə]	about	[əbaʊt]		[ju]	fuel	[fjul]	
Consonants							
Stops				**Fricatives**			
[p]	pen	[pɛn]		[f]	few	[fju]	
[b]	Ben	[bɛn]		[v]	view	[vju]	
[t]	ten	[tɛn]		[θ]	thigh	[θaɪ]	
[d]	den	[dɛn]		[ð]	thy	[ðaɪ]	
[k]	Kay	[ke]		[h]	hay	[he]	
[g]	gay	[ge]		[s]	say	[se]	
[tʃ]	chew	[tʃu]		[ʃ]	shay	[ʃe]	
[dʒ]	Jew	[dʒu]		[z]	bays	[bez]	
				[ʒ]	beige	[beʒ]	
Nasals and lateral				**Glides**			
[m]	some	[sʌm]		[w]	way	[we]	
[n]	sun	[sʌn]		[hw]	whey	[hwe]	
[ŋ]	sung	[sʌŋ]		[j]	yea	[je]	
[l]	lay	[le]		[r]	ray	[re]	

* [ɝ] and [ɚ] are the "*r*-colored" vowels. [ɜ] and [ə] are the pronunciations typical of *r* vowels in Eastern, Southern, and English speech.

† Does not include the "nondistinctive" and centering diphthongs.

(帀)	ㄩ	ㄨ	ㄧ	ㄦ	ㄥ	ㄤ	ㄣ	ㄢ	ㄡ	ㄠ	ㄟ	ㄞ	ㄝ	ㄜ	ㄛ	ㄚ	
	迂ㄩ	烏ㄨ	衣ㄧ	兒ㄦ	鞥ㄥ	骯ㄤ	恩ㄣ	安ㄢ	歐ㄡ	熬ㄠ	欸ㄟ	哀ㄞ	誒ㄝ	疴ㄜ	喔ㄛ	啊ㄚ	
		晡ㄅㄨ	逼ㄅㄧ		崩ㄅㄥ	邦ㄅㄤ	奔ㄅㄣ	般ㄅㄢ		包ㄅㄠ	卑ㄅㄟ	掰ㄅㄞ			玻ㄅㄛ	巴ㄅㄚ	ㄅ
		鋪ㄆㄨ	批ㄆㄧ		烹ㄆㄥ	滂ㄆㄤ	噴ㄆㄣ	潘ㄆㄢ	剖ㄆㄡ	拋ㄆㄠ	胚ㄆㄟ	拍ㄆㄞ			坡ㄆㄛ	趴ㄆㄚ	ㄆ
		母ㄇㄨ	咪ㄇㄧ		蠓ㄇㄥ	忙ㄇㄤ	悶ㄇㄣ	顢ㄇㄢ	牟ㄇㄡ	貓ㄇㄠ	枚ㄇㄟ	埋ㄇㄞ		麼ㄇㄜ	摸ㄇㄛ	媽ㄇㄚ	ㄇ
		夫ㄈㄨ			鋒ㄈㄥ	方ㄈㄤ	分ㄈㄣ	番ㄈㄢ	否ㄈㄡ		非ㄈㄟ				佛ㄈㄛ	發ㄈㄚ	ㄈ
		都ㄉㄨ	低ㄉㄧ		登ㄉㄥ	當ㄉㄤ		丹ㄉㄢ	兜ㄉㄡ	刀ㄉㄠ	得ㄉㄟ	呆ㄉㄞ		得ㄉㄜ		搭ㄉㄚ	ㄉ
		禿ㄊㄨ	梯ㄊㄧ		謄ㄊㄥ	湯ㄊㄤ		攤ㄊㄢ	偷ㄊㄡ	滔ㄊㄠ		胎ㄊㄞ		特ㄊㄜ		他ㄊㄚ	ㄊ
	女ㄋㄩ	奴ㄋㄨ	尼ㄋㄧ		能ㄋㄥ	囊ㄋㄤ	嫩ㄋㄣ	囡ㄋㄢ	耨ㄋㄡ	撓ㄋㄠ	餒ㄋㄟ	乃ㄋㄞ		訥ㄋㄜ		那ㄋㄚ	ㄋ
	閭ㄌㄩ	嚕ㄌㄨ	哩ㄌㄧ		楞ㄌㄥ	狼ㄌㄤ		蘭ㄌㄢ	摟ㄌㄡ	撈ㄌㄠ	勒ㄌㄟ	來ㄌㄞ		勒ㄌㄜ	咯ㄌㄛ	拉ㄌㄚ	ㄌ
		姑ㄍㄨ			耕ㄍㄥ	缸ㄍㄤ	根ㄍㄣ	干ㄍㄢ	勾ㄍㄡ	糕ㄍㄠ	給ㄍㄟ	該ㄍㄞ		哥ㄍㄜ		嘎ㄍㄚ	ㄍ
		枯ㄎㄨ			坑ㄎㄥ	康ㄎㄤ	肯ㄎㄣ	刊ㄎㄢ	摳ㄎㄡ	尻ㄎㄠ		開ㄎㄞ		科ㄎㄜ		咖ㄎㄚ	ㄎ
		乎ㄏㄨ			亨ㄏㄥ	夯ㄏㄤ	痕ㄏㄣ	酣ㄏㄢ	齁ㄏㄡ	蒿ㄏㄠ	黑ㄏㄟ	咳ㄏㄞ		呵ㄏㄜ		哈ㄏㄚ	ㄏ
	居ㄐㄩ		基ㄐㄧ														ㄐ
	區ㄑㄩ		欺ㄑㄧ														ㄑ
	須ㄒㄩ		希ㄒㄧ														ㄒ
知ㄓ		朱ㄓㄨ			征ㄓㄥ	章ㄓㄤ	珍ㄓㄣ	氈ㄓㄢ	州ㄓㄡ	招ㄓㄠ	這ㄓㄟ	齋ㄓㄞ		遮ㄓㄜ		查ㄓㄚ	ㄓ
蚩ㄔ		初ㄔㄨ			稱ㄔㄥ	昌ㄔㄤ	嗔ㄔㄣ	攙ㄔㄢ	抽ㄔㄡ	抄ㄔㄠ		釵ㄔㄞ		車ㄔㄜ		叉ㄔㄚ	ㄔ
尸ㄕ		舒ㄕㄨ			升ㄕㄥ	傷ㄕㄤ	申ㄕㄣ	山ㄕㄢ	收ㄕㄡ	稍ㄕㄠ	誰ㄕㄟ	篩ㄕㄞ		奢ㄕㄜ		沙ㄕㄚ	ㄕ
日ㄖ		如ㄖㄨ			扔ㄖㄥ	攘ㄖㄤ	人ㄖㄣ	然ㄖㄢ	柔ㄖㄡ	繞ㄖㄠ				惹ㄖㄜ			ㄖ
咨ㄗ		租ㄗㄨ			曾ㄗㄥ	臧ㄗㄤ	怎ㄗㄣ	簪ㄗㄢ	鄒ㄗㄡ	糟ㄗㄠ	賊ㄗㄟ	災ㄗㄞ		則ㄗㄜ		臢ㄗㄚ	ㄗ
雌ㄘ		粗ㄘㄨ			噌ㄘㄥ	倉ㄘㄤ	參ㄘㄣ	參ㄘㄢ	湊ㄘㄡ	操ㄘㄠ		猜ㄘㄞ		側ㄘㄜ		擦ㄘㄚ	ㄘ
私ㄙ		蘇ㄙㄨ			僧ㄙㄥ	桑ㄙㄤ	森ㄙㄣ	三ㄙㄢ	嗖ㄙㄡ	搔ㄙㄠ	塞ㄙㄟ	腮ㄙㄞ		瑟ㄙㄜ			ㄙ



APPENDIX B

TABLE OF CURRENTLY EXISTING CHINESE ROMANIZATION SYSTEMS

NO	YL	WG	SF	PY	SG
01	b	p	b	b	b
02	p	p'	p	p	p
03	m	m	m	m	m
04	f	f	f	f	f
05	d	t	d	d	d
06	t	t'	t	t	t
07	n	n	n	n	n
08	l	l	l	l	l
09	g	k	g	g	g
10	k	k'	k	k	k
11	h	h	h	h	h
12	j	ch	j(i)	j	j
13	ch	ch'	ch(i)	q	ch
14	sy	hs	sh(i)	x	hs
15	j	ch	j	zh	dr
16	ch	ch'	ch	ch	tsh
17	sh	sh	sh	sh	sh
18	r	j	r	r	r
19	dz	ts,tz	tz	z	dz
20	ts	ts',tz'	ts	c	ts
21	s	s,ss,su	s	s	s
22	a	a	a	a	a
23	o	o	o	o	o
24	e	e,o	e	e	e
25	e	eh	e	e	ea
26	ai	ai	ai	ai	ai
27	ei	ei	ei	ei	ei
28	au	ao	au	ao	ao
29	ou	u,ou	ou	ou	oa

30	an	an,en	an	an	an
31	en	en	en	en	en
	n	n	n	n	n
32	ang	ang	ang	ang	ang
33	eng	eng	eng	eng	eng
	ng	ng	ng	ng	ng
34	er	erh	er	er	er
35	i,yi	i,yi	i,yi	i,yi	i
	y	y,i	i	y,i	y
36	u,wu	u	u,wu	u,wu	oo
	w	w,u	u	w,u	w
37	yu	yu,u	iu,yu	yu,u,io	iu
	yw	yu,u	iu	yu,u,	yw
38	r,z	u,ih	r,z	i	ih

APPENDIX C
THE CHINESE UTTERANCES USED IN THE EXPERIMENT

NO	Chinese Romanization
01	i4 tshang2 droong1 drih3 (4)
02	chiu3 tshwen2 shih2 jyan1 (4)
03	hsyan4 iung4 dang3 (3)
04	wei4 drih3 (2)
05	pei4 drih4 (2)
06	ing4 iung4 tsheng2 shih4 (4)
07	dzih4 doong4 dzih1 lyao4 tshoo3 li3 (6)
08	foo3 droo4 tshoo2 tshwen2 ti3 (5)
09	ping2 dai4 kwan1 (3)
10	tyao2 ma3 (2)
11	dreng3 pi1 tshoo3 li3 (4)
12	ji1 drweng3 dyan3 (3)
13	er4 jin4 ma3 (3)
14	dzih1 lyao4 dwan4 (3)
15	boo4 lin2 dai4 shoo4 (4)
16	chi4 pao4 shih4 pai2 hsiu4 (5)
17	neng2 lyang4 (2)
18	droong1 yang1 tshoo3 li3 ji1 (5)
19	dzih4 ywan2 (2)
20	ma3 drwan3 hwan4 (3)
21	dyan4 nao3 foo3 droo4 jyao1 hsywea2 (6)
22	koong4 drih4 dan1 ywan2 (4)
23	tsih2 droo4 (2)
24	dzih1 lyao4 koo4 gwan3 li3 ywan2 (5)
25	drih4 yan2 (2)
26	shan1 tshoo2 (2)
27	she4 ji4 (2)
28	shoo4 wei4 hsyan3 shih4 (4)
29	tsih2 dyea2 dzwo4 yea4 hsi4 toong3 (6)
30	ding4 i4 iu4 (3)

31 ting2 ji1 shih2 jyan1 (4)
32 doong4 tai4 fen1 ge1 (4)
33 byan1 ji2 tsheng2 shih4 (4)
34 fang3 dren1 (2)
35 dang3 an4 droong1 dyan3 (4)
36 deng1 loo4 (2)
37 oo4 tsha1 dren1 tshe4 hsi4 toong3 (6)
38 drih2 hsing2 (2)
39 wen2 jyan4 tshwan2 dren1 hsi4 toong3(6)
40 shih1 oo4 iu4 goo1 (4)
41 ke3 hsing2 hsing4 yan2 jyoa4 (5)
42 ren4 ti3 (2)
43 foo2 dyan3 iung4 shwan4 (4)
44 lyoa2 tsheng2 too2 (2)
45 goong1 neng2 byao3 (3)
46 chywan2 myan4 hsing4 byan4 shoo4 (5)
47 too2 hsing2 shoo4 wei4 chi4 (5)
48 ban4 shwang1 goong1 toong1 dao4 (5)
49 gao1 jyea1 iu3 yan2 (4)
50 gao1 shoo4 doo2 ka3 ji1 (5)
51 ing3 hsyang4 tshoo3 li3 ji1 (5)
52 mai4 tshoong1 dza2 in1 (4)
53 dzeng1 lyang4 byao3 shih4 fa3 (5)
54 swo3 ing3 dran4 tswen2 chi4 (5)
55 drih3 ling4 (2)
56 jyao1 tan2 shih4 (3)
57 tsao1 dzoong4 gan3 (3)
58 gwang1 bi3 (2)
59 lyan4 chiun2 (2)
60 dzai3 roo4 (2)
61 ben3 di4 (2)
62 hwei2 loo4 (2)
63 bai3 wan4 droa1 (3)
64 he2 bing4 fen1 lei4 (4)
65 mwo2 dzoo3 (2)

66 hao2 wei1 myao3 (3)
67 kan3 tao4 (2)
68 dreng4 gwei1 hwa4 (3)
69 hsywan3 dze2 (2)
70 hsiun4 hsi2 bao1 feng1 (4)
71 tsa2 hsiun2 ji4 chyao3 (4)
72 bao3 hoo4 (2)
73 mai4 tshoong1 shwai1 jyan3 (4)
74 da3 koong3 ka3 pyan4 (4)
75 lwan4 shoo4 tshan3 sheng1 chi4 (5)
76 fan4 wei2 he2 dwei4 (4)
77 byan4 shih4 jing1 chywea4 doo4 (5)
78 tsan1 kao3 lyea4 byao3 (4)
79 fan3 she4 shao3 myao2 (2)
80 jiu4 tshih4 (2)
81 sih4 foo2 ji1 goa4 (4)
82 kwai4 drao4 kao3 bei4 (4)
83 loa4 tsha2 tswo4 oo4 liu4 (5)
84 yoa3 hsyao4 hsing4 (3)
85 tshwei2 drih2 kwei4 gei3 (4)
86 dzwei4 hwai4 drwang4 kwang4 (4)
87 wei2 hsyea3 (2)
88 ling2 chi2 byao1 (3)
89 ling2 i4 drih4 (3)
90 chiu1 iu4 (2)

APPENDIX D
THE ENGLISH EQUIVALENT USED IN THE EXPERIMENT

NO	English Vocabulary
01	abort (2)
02	access time (3)
03	active file (3)
04	address (2)
05	allocation (4)
06	application program (6)
07	automatic data processing (9)
08	auxiliary storage (7)
09	bandwidth (2)
10	bar code (2)
11	batch processing (4)
12	benchmark (2)
13	binary code (4)
14	block (1)
15	boolean algebra (5)
16	bubble sort (3)
17	capacity (4)
18	central process unit (6)
19	character (3)
20	code conversion (4)
21	computer aided instruction (8)
22	control unit (4)
23	cylinder (3)
24	database administrator (8)
25	delay (2)
26	delete (2)
27	design (2)
28	digital display (5)
29	disk operating system (7)
30	domain (2)

31 downtime (2)
32 dynamic partitioning (7)
33 editor (3)
34 emulation (4)
35 end of file (3)
36 entry (2)
37 error detection system (7)
38 execution (4)
39 facsimile document system (9)
40 failure prediction (5)
41 feasibility study (7)
42 firmware (2)
43 floating-point operation (8)
44 flowchart (2)
45 function table (4)
46 global variable (6)
47 graphic digitizer (6)
48 half-duplex channel (5)
49 high level language (5)
50 high speed reader (4)
51 image processor (5)
52 impulse noise (3)
53 incremental representation (9)
54 index register (5)
55 instruction (3)
56 interactive (4)
57 joystick (2)
58 light pen (2)
59 link group (2)
60 load (1)
61 local (2)
62 loop (1)
63 megacycle (4)
64 merge-sort (2)
65 module (2)

66 nanosecond (4)
67 nest (1)
68 normalize (3)
69 option (2)
70 packet (2)
71 polling technique (4)
72 protection (3)
73 pulse decay (3)
74 punch card (2)
75 random number generator (8)
76 range check (2)
77 recognition accuracy (8)
78 reference listing (5)
79 reflective scan (4)
80 rejection (3)
81 servomechanism (6)
82 snapshot copy (4)
83 undetected error rate (7)
84 validity (4)
85 vertical feed (4)
86 worst-case (2)
87 write only (3)
88 zero flag (3)
89 zero suppression (5)
90 zone (1)

LIST OF REFERENCES

1. Poock, Gary K., *Experiments with Voice Input for Command and Control: Using Voice Input to Operate A Distributed Computer Network*, Naval Postgraduate School, Monterey, CA, April 1980.

2. Neil, D. E., and Andreason, T., *Examination of Voice Recognition System to Function in A Bilingual Mode*, Naval Postgraduate School, Monterey, CA, March 1981.

3. Na, Dzoong-hsyung, *Mandarin Chinese Phonetics*, Taiwan Kaiming Book Co., 1959.

4. Chi Sheng Book Co., *Guidance of Chinese Teaching, Book One*, Hongkong, 1981.

5. Carrell, James, and Tiffany, William R., *Phonetics: Theory and Application to Speech Improvement*, McGraw-Hill Book Company, Inc., 1960.

6. Nilsen, D. L. F., and Nilsen, A. P., *Pronunciation Contrasts in English*, Regents Publishing Company, 1973.

7. Shen, Yao, *English Phonetics*, University of Michigan, 1962.

8. Dow, Francis D. M., *An Outline of Mandarin Phonetics*, Australian National University Press, 1972.

9. Suzuki, Takuroo, *Chinese Speech*, Doong Ya Toong Wen Institute, 1938.

10. Poock, Gary K., *Speech Recognition Research, Applications and International Efforts*, Invited paper for the 1986 Human Factors Society, published in The Proceedings of the Human Factors Society, Dayton, OH, October 1986.

11. Threshold Technology Inc., *Threshold 600 User's Manual*, June 1978.

12. Threshold Technology Inc., *Threshold 600-CRT User's Manual (Interim) With The RS 232 Adapter*, February 1979.

13. Liu, P. J., Hsyao, G. D., and Drwang, H. Y., *English-Chinese Computer terminology Dictionary*, Chwan Hwa Technology Book Co. Ltd., 1984.

BIBLIOGRAPHY

Allen, R. L., and Shute, Margare, *English Sounds and Their Spelling*, Thomas Y. Crowell Company, 1966.

Armstrong J. W., *The Effects of Concurrent Motor Tasking On Performance of A Voice Recognition System*, Master Thesis, Naval Postgraduate School, Monterey, CA, September 1980.

Armstrong, J. W., and Poock, G. K., *Effect of Task Duration On Voice Recognition System Performance*, Naval Postgraduate School, Monterey, CA, September 1981.

Butterworth, B., *Language Production Vol. One: Speech and Talk*, Academic Press, 1980.

Elster, R., *The Effects of Certain Background Noise On The Performance of A Voice Recognition System*, Naval Postgraduate School, Monterey, CA, September 1980.

Hall, R. A., Jr., *Sound And Spelling in English*, Chilton Company- Book Division, 1961.

Institute for Information Industry, *Status Report On Chinese Computers Researching Institutes*, January 1986.

Institute for Information Industry, *Status Report On Chinese Computers Market*, January 1986.

Institute for Information Industry, *Status Report On Chinese Computers Manufacturers*, January 1986.

Jay, G. T., *An Experiment In Voice Data Entry for Imagery Intelligence reporting*, Master Thesis, Naval Postgraduate School, Monterey, CA, March 1981.

Ling I Technology Co. Ltd., *Tsang Jyea III Chinese Character Input Method*, July 1985.

Ma, J. B., Hsyao, H. N., and Tsheng, T. M., *Data Processing of Chinese Computers*, Oo Nan Book Company, June 1985.

Poock, G. K., *To Train Ramdonly Or All At Once...That Is The Question*, Proceeding of Voice Data Entry Systems Applications Conference, sponsored by Lockheed Missiles and Space Co., Santa Clara, CA, October 1981.

Poock, G. K., *A Longitudinal Study of Computer Voice Recognition Performance and Vocabulary Size*, Naval Postgraduate School, Monterey, CA, June 1981.

Poock, G. K., and Roland, E. F., *Voice Recognition Vocabulary Lists for The Army's TACFIRE System*, Naval Postgraduate School, Monterey, CA, January 1983.

Poock, G. K., and Roland, E. F., *Voice Recognition Accuracy: What is Acceptable?* Naval Postgraduate School, Monterey, CA, November 1982.

Poock, G. K., Schwalm, N. D., Martin, B. J., and Roland, E. F., *Trying for Speaker Independence in The Use of Speaker Dependent Voice Recognition Equipment*, Naval Postgraduate School, Monterey, CA, December 1982.

Prator, C. H., Jr., *Manual of American English Pronunciation*, Holt, Rinehart and Winston, 1957.

Tong, Chin Wan, *An Outline of Chinese Phonetics*, Sinology Press, 1963.

Vallins, G. H., *Spelling*, Revised by D. G. Scragg, Andre Deutsch Limited, 1965.

Yee, Dennis K., *Chinese Romanization Self-study Guide*, The University Press of Hawaii, 1975.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center Cameron Station Alexandria, VA 22304-6145	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, CA 93943-5002	2
3.	Professor Willis R. Greer, Jr., Code 54Gk Department of Administrative Sciences Naval Postgraduate School Monterey, CA 93943-5000	1
4.	Curricular Officer, Code 37 Computer Technology Naval Postgraduate School Monterey, CA 93943-5000	1
5.	Professor Gary K. Poock, Code 55Pk Department of Operations Research Naval Postgraduate School Monterey, CA 93943-5000	5
6.	Professor Richard A. McGonigal, Code 54Mb Department of Administrative Sciences Naval Postdraduate School Monterey, CA 93943-5000	1
7.	Commander Liu, I Kang 17, Lane 37, ChungHsiao St. ChungHo, Taipei County Taiwan, Republic Of China	2

Thesis
L7135 Liu
c.1 A feasibility study
using Chinese speech as
a command/control tool
for computer systems.