A PHONETICALLY ORIENTED SPEECH DATABASE FOR MANDARIN CHINESE

Chiu-yu Tseng
Institute of History and Philology & Institute of Information Science, Academia Sinica, Taipei, Taiwan, R.O.C.

ABSTRACT

A phonetically oriented database for Mandarin Chinese speech is designed. Using two electronic corpora of 77,324, lexical entries and 5,353 sentences, 1,455 lexical items and 599 phrases/ sentences in discourse/short stories that cover all possible segmental, syllabic plus tonal combinations in Mandarin Chinese were generated. Tailored software is designed to perform phonetic and acoustic analyses for collected speech samples .

INTRODUCTION

The need to establish a large scale database of Mandarin Chinese speech has been existing ever since research in speech synthesis and speech recognition began in Taiwan over a decade ago. Synthetic speech and automatic speech recognition by computers offer the most optimal and efficient method of communication between humans and computers. [2, 3] While researchers in Taiwan have been actively conducting research in both speech synthesis and speech recognition in Mandarin Chinese without a large scale database, a consensus has been reached that a database that would provide orthographic, phonetic as well as acoustic information would be essential. The paper reports part of an ongoing project toward that goal. The project consists of a knowledge database, a corpus database, a parsing database, a speech database and finally an application end. Resources and specialties from various sectors in Academia Sinica, Taipei, Taiwan has been delegated. This report is the first attempt to describe the speech database only.

Although officially beginning in the fall of 1994, researchers at Academia Sinica have initiated and participated in several previous efforts to collect of speech database in Mandarin Chinese. We realize such a database would be crucial because the acoustic realizations of segments and tones and their interactions depend on complex interactions among many factors. At the present stage, we are prepared to deal with factors that are phonetic and acoustic, emphasizing the tonal aspect of Mandarin Chinese in particular and therefore using syllable as the basic unit . The long term goal is to establish a large scale database that would incorporate intra-speaker and inter-speaker factors. However, the present focus is a phonetically-oriented database that aims to include all possible intra- and inter-syllabic and tonal combinations in most frequently used words so that speech collected under such guidelines would enable us to investigate phonetic properties that would be of use for developing a speech synthesis and recognition system.

THE SPEECH DATABASE

The database consists two types: (1) a word database and (2) a continuos speech database. Both types are now being developed by collecting speech data from different speakers.

DESIGN OF THE DATABASE

Both the word database and the continuous sentence database are designed to be phonetically balanced.

For the word database, an electronic dictionary corpus called Modern Chinese Corpus [1] that include more than 80,000 lexical items were used. Software was designed to first select lexical entries of at most four syllables in structure. A total of 77,324 items were derived. Software was then designed to select items that cover all possible intra- and intra-syllabic plus tonal combinations from three sets of sub-corpus, i.e., the most frequently used 20,000, 40,000, and 77,324 lexical items from the text corpus. Table 1 summarizes the results.

Table 1. Statistical analysis of phonetically specified lexical items from 3 text corpora.

Total # of lexical items		20,000	40,000	77,324
mono-syllabic words	#	3,177	4,369	6,974
	# of possible tones	5	5	5
di-syllabic words	#	14,276	27,431	48,349
	# of possible tonal combinations	20	20	20
tri-syllabic words	#	1,674	4,380	11,562
	# of possible tonal combinations	78	92	97
quadri-syllabic words	#	873	3,820	10,349
	# of possible tonal combinations	235	300	354
# of possible inter-syllable combinations		1,351	1,536	1,649

Results demonstrate that choosing lexical items with the above-mentioned phonotactic and phonetic specifications from a corpus of only 20,000 most frequently used words would suffice. Therefore, the chosen word database consists of 1,455 frequently used words that the include 393 monosyllabic, 676 disyllabic, 145 trisyllabic and 241 quadrisyllabic words of altogether 3,144 syllables. Table 2 summarizes the results.

Table 2. Statistical analysis that shows the distribution of the chosen words that formed the described database.

Total # of lexical items		1,455
monosyllabic words	#	393
	# of possible tones	5
disyllabic words	#	676
	# of possible tonal combinations	20
trisyllabic words	#	145
	# of possible tonal combinations	78
quadrisyllabic words	#	241
	# of possible tonal combinations	235
# of possible inter-syllable combinations		1,351

The continuous speech database, on the other hand, consists of 599 sentences that are constructed from 5,353 sentences that included ten stylistic variations of narratives and/or speech. Duration of sentences/discourse varies from 2 to 180 syllables.

DATA COLLECTION

The initial goal of the database is intended to collect homogeneous speech to set up standard references phonetically and acoustically because large-scale speech data to be collected in later stages will include a variety of inter- and intra-speaker differences due to dialectical pronunciations. We recruited professional Mandarin language teachers whose production of Chinese is of the standard of professional narrators. Sound proof chambers equipped with PC486 and beyerdynamic M69N(C) microphone were used during recording sessions. The words and sentences were read at a normal speaking rate. Each complete set of speech data by each speaker came to 7 hours of recording time. Table 3 summarizes the speakers of our standard references.

Table 3. Summary of speakers whose speech serves as standard reference for the database.

age	gender	# of speakers
65 years and above	male	1
	female	1
35 - 65 years	male	1
	female	1
under 35 years	male	1
	female	1

SEGMENTATION AND LABELING

Segmentation and transcription of the database were done by hand in order to keep the quality of the reference speech data as high as possible so that it could serve as our basis for designing the software that would perform the initial segmentation and labeling when large-scale speech data is collected. Four windows displayed (1) waveform of the utterance; (2) spectrogram; (3) peak of auto correlation function, root mean square, probability of voicing, and fundamental frequency patterns; and (4) phonetic labeling respectively on one screen. are also displayed. A trained personnel inspects the display of the top three windows while segmenting the speech signal phoneme-by-phoneme. Note that at the current stage, only phonetic transcription is provided. Since it is a difficult task to define boundaries between phonemes, especially between two adjacent vowels, boundaries were defined as the center of the formant transitions between the two phonemes [4] while listening through headset at the same time. Figure 1 shows an example of segmentation and labeling.

When establishing the electronic files, tagging system was designed following specifications from the Linguistic Data Consortium (LDC) with additional tags for tonal information. Phonetic information is yielded to provide possible in-depth investigation of spoken Mandarin Chinese in general. Statistical analyses are were also performed to further yield results of phonetic phenomena that not available otherwise. Table 4a and 4b illustrates the kind of statistical results from analyzed speech data of one speaker.

Table 4a. Statistical analysis of Mandarin Chinese consonants from the described database produced by one speaker.

phone	mean (ms)	std (ms)	phone	mean (ms)	std (ms)
b	15	11	j	75	22
p	79	23	q	164	33
m	90	25	x	172	41
f	106	28	zh	73	98
d	15	13	ch	128	47
t	85	23	sh	202	107
n	71	24	r	90	65
l	70	25	z	129	90
g	24	9	c	188	113
k	101	20	s	204	97
h	111	33			

Table 4b. Statistical analysis of Mandarin Chinese vowels from the described database produced by one speaker.

phone	mean (ms)	std (ms)	phone	mean (ms)	std (ms)
i	236	97	iao	298	107
u	211	108	iou	336	104
yu	308	122	ian	304	116
a	249	87	in	291	92
o	247	80	iang	317	117
e	237	118	ing	283	88
ai	279	94	ua	311	115
ei	256	106	uo	275	127
ao	294	111	uai	296	84
ou	263	121	uei	265	103
an	268	103	uan	337	112
en	253	92	uen	289	98
ang	302	110	uang	333	117
eng	276	98	ong	283	100
er	278	99	yue	316	83
ia	304	104	yuan	341	97
ie	291	98	yun	296	92
iai	577	1	yung	313	111

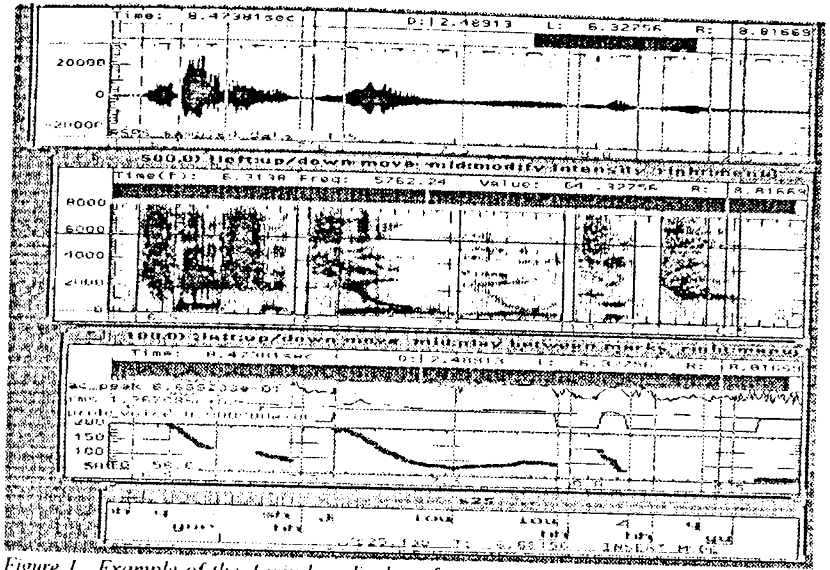

Figure 1. Example of the 4-window display of an utterance from the screen. From top to bottom, window 1 displays the waveform; window 2 the spectrogram; window 3 auto correlation functions, probability of voicing, and the fundamental frequency patterns; window 4 the phonetic labeling.

CONCLUDING REMARKS

An outline of a Mandarin Chinese speech database is described. At the current stage, it consists of two types of databases. Speech data are transcribed with fine acoustic-phonetic labels to meet a variety of needs for speech research. So far, data from six speakers, three males and three females have been completely digitized. The project is at its first year of a 5-year endeavor. Next year efforts will be devoted to collected speech data using statistical methods so that a large number of speakers, each providing a fraction of the above designed set, will participate.

REFERENCES

[1] Chen, K-J, and Huang, C-R, Modern Chinese Corpus (ongoing project at Institution of Information Science, Academia Sinica, Taipei, Taiwan, R.O.C.)

[2] Zue, V., Seneff, S, and Glass, J., "Speech database development: Time and beyond" Paper presented at the Workshop on Speech input/output Assessment and Speech Databases, Amsterdam, The Netherlands, 20-23 September, 1989.

[3] Lamel, L., Kassel, R.H. and Seneff, S. "Speech database development: Design and analysis of the acoustic-phonetic corpus" Proceedings of the Speech Recognition Workshop, Palo Alto, Ca., February 19-20, 1986.

[4] Kuwabara, H., Takeda, K, Sagisaka, Y, Katagiri, S., Morikawa, S, and Watanabe, T. "Construction of a large-scale Japanese speech database and its management system" S10b.12,1989 IEEE.