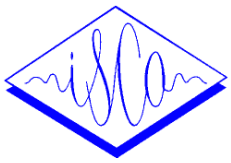

First International Conference on
Spoken Language Processing (ICSLP 90)
Kobe, Japan
November 18-22, 1990

CONSTRUCTION OF A LARGE KOREAN SPEECH DATABASE AND ITS MANAGEMENT SYSTEM IN ETRI

Joon-Hyuk Choi, Kyung-Tae Kim

Signal Processing Section
Elec. and Telecom. Research Inst., Korea

Abstract

A large size Korean speech database under construction at ETRI is introduced. We have three kinds of speech databases. They are 35 connected-4-digits,144 CV monosyllables which are segmented into phonemes(C+V)combination, and 445 phoneme-balanced words. The first two are collected from 10 male speakers and the last, from 4 male and 4 female speakers. For the easy application of the database in speech research, we proposed a 2-level acoustic-phonetic transcription to express almost all possible phonetic environments in Korean. And the transcriptions were carried out manually. For the effective management of these databases, a relational database management system, which helps fast access and easy manipulation, was used.

1. Introduction

In the field of speech study, a speech database is necessary and important. First, speech researchers need to evaluate many kinds of speech analysis and recognition algorithms without loss of generality in order to develope better ones. Second, their activities are more dependent on the size and quality of the speech database. The necessity of a common speech database has been continually emphasized in the US, Japan, and Europe. In the US, many speech databases collected within DARPA family became available to the public through NIST. And they have produced a prototype CD-ROM version of the DARPA TIMIT Acoustic-Phonetic Speech Database. In CMU, SPHINX system is based on more than 8500 words of database[1]. In France, BDLEX and BDSONS are being carried out as database building projects for written and spoken French[2]. The NCSR(National Center for Scientific Research) is doing these projects as parts of GRECO assignment[3]. Other countries(Sweden, Norway ...) have made their own speech databases. In Japan, a speech database of about 8500 words was constructed by ATR[4]. ATR, since its foundation in 1986, is playing a major role in speech research in Japan.

In Korea, still no generally acceptable speech database for public domain is available. Therefore it is very hard to expect any significant progress in the speech related research in Korea. A few existing databases are not so reliable to provide the generality for other researchers. Since 1987, ETRI has been trying to build a speech database that satisfies the needs of an acceptable criterion in various studies. This paper describes a speech database which is now being built at ETRI, in section 2, the overall scheme and contents of the speech database are introduced. These speech data are transcribed in 2-ways which are described in Section 3. In Section 4, the DBMS we used for the easy and efficient access and manipulation of the database is mentioned.

2. Speech database

The speech database consists of numbers and monosyllables, 4-connected digits, and phoneme-balanced words.

2.1 4-connected digit word

The 4-connected digit database was built in the hope of being utilized for the voice dependent/independent dialing system and for the training of HMM. This includes about 100 phonetic environment(10 numbers X 10 numbers) and each connected digit was pronounced 4 times.

2.2 Numbers and monosyllables

22 numbers and 144 CV monosyllables were selected. They were used in the study of basic characteristics of Korean phonemes, recognition of monopthongs and the training process for manual segmentation. Especially this speech database is very helpful for investigating the characteristics of initial consonants and fundamental variation of successive phonemes. The CV monosyllables are made by combining of 16 consonants with basic 8 monopthongs. This covers all spoken Korean except dipthongs.

2.3 Phoneme-balanced word

445 phoneme-balanced words were collected. They were selected from 4114 most frequently used words which appear in elementary school text books. The selection is made with following algorithms. First, the words of two phoneme sequence that appears only once are chosen(Unique Word Set : 131 words). Second, for the remaining, the following measure function is applied to evaluate the degree of phoneme balance. Based on this measure, other 314 words were selected.

$$S = -\sum_{n=1}^{N} Pn \log_2 Pn$$

Pn: the occurrence probability of the n-th phoneme sequence
N: the number of different sequences
S: the entropy which has a maximum value when Pns are all equal

As a consquence, a set of 445 phoneme-balanced words were formed. Though phoneme balanced words are selected from the elementary school text books, the distribution of phoneme sequence covers a large vocaburary dictionary.

2.4 Recording environments

In recording, we used PCM/VCR technology. All speech signals are converted into digital data and stored on a video tape. Namely, speech signal was passed through a analog band pass filter(70 Hz - 8 kHz), then A/D converting is executed at 20 kHz sampling rate with the 12 bit resolution. All procedures are performed on the MASSCOMP-5500 workstation, and spectrograms with other parameters are plotted for manual labelling.

3. Labelling of the database

3.1 Labelling

It is not easy to locate a precise boundary between two successive phonemes in the real speech sound. But, still labelling of various layers are essential to the study of the phonetic and acoustic phenomena[5].

For the effective analysis of speech sound, the spectrogram information is essential. Hence, our data sheet shows the spectrogram with speech waveform, spectral difference, and energy. A 10 ms Hamming window spaced every 2.5 ms and 512-FFT were used in spectrum analysis. The spectrogram which ranged between 0 kHz and 7 kHz was shown in 49 gray scale levels.

By inspecting the above informations, phonetic segmentation is done and other informations such as transition and voice offset are also marked. In many cases, it is hard to point out the exact boundaries in successive phonemes especially when vowel-to-vowel, semi-vowel-to-vowel, and vowel-to-final consonant transition occur. Figure 3 - 1 shows an example of manual segmentaion.

3.2 Labelling symbol

We define 2-layer transcriptions in order to satisfy the variety of requirements in speech research. The first layer is the Korean alphabetical symbols. And the second is phonetic environmental one reflecting acoustic characteristics of phonemes including

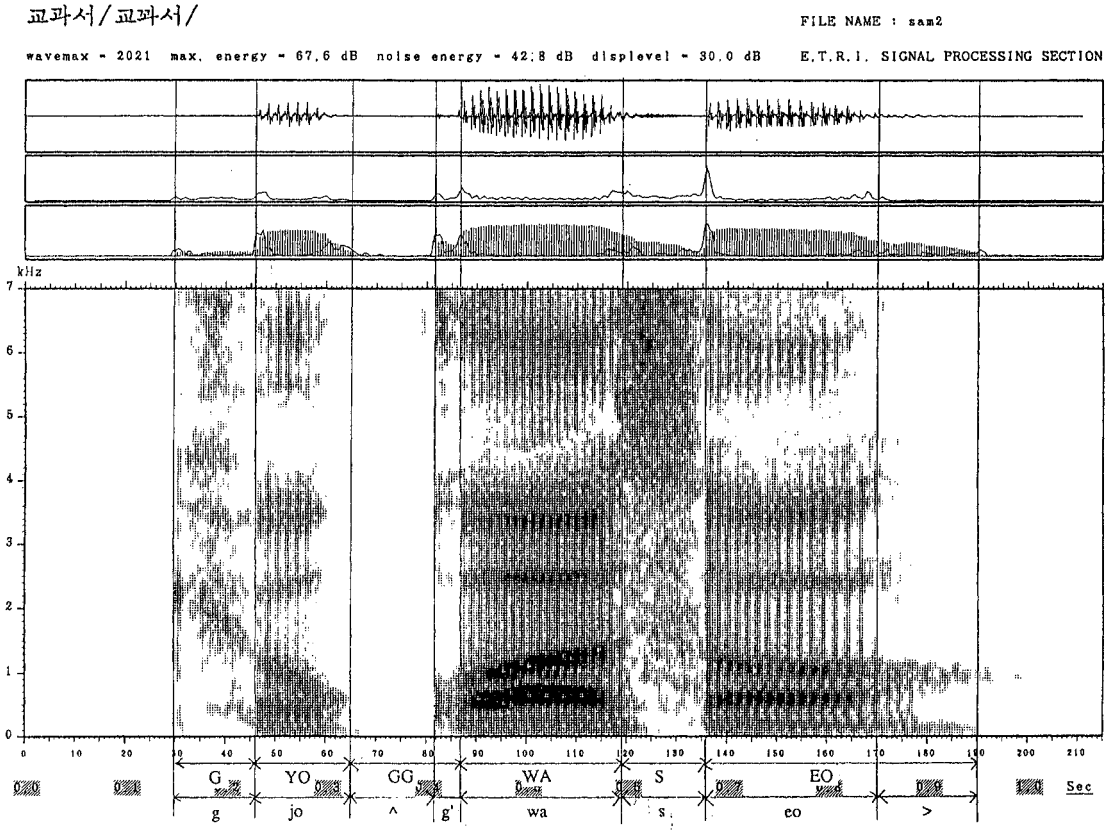

Figure 3 - 1 An Example of Manual Segmentation of /GYO GGWA SEO/

allophonic variations. Though most of the symbols are the same as those used in phonemic description, their corresponding parts are slightly different in some cases. In this layer, for instance, unvoiced plosive is usually divide into two parts: the unvoiced sound and the unvoiced closure. Table 3 - 1 shows symbols for the description for two layers.

First Layer

Phoneme	Symbol	Phoneme	Symbol
g (ㄱ)	G	a (아)	A
k (ㅋ)	K	ʌ (어)	EO
g' (ㄲ)	GG	o (오)	O
d (ㄷ)	D	u (우)	U
t (ㅌ)	T	ɯ (으)	EU
d' (ㄸ)	DD	i (이)	I
b (ㅂ)	B	ɛ (애)	AE
p (ㅍ)	P	e (에)	E
b' (ㅃ)	BB	ø (외)	OE
j (ㅈ)	J	ja (야)	YA
c (ㅊ)	C	jʌ (여)	YEO
j' (ㅉ)	JJ	jo (요)	YO
s (ㅅ)	S	ju (유)	YU
s' (ㅆ)	SS	je (얘)	YAE
h (ㅎ)	H	je (예)	YE
ŋ (ㅇ)	NG	wa (와)	WA
n (ㄴ)	N	wʌ (워)	WEO
r/l (ㄹ)	R/L	wi (위)	WI
m (ㅁ)	M	we (왜)	WAE
		we (웨)	WE
		ui (의)	EUI

Second Layer

p, t, k b, d, g b', d', g'	voiced/unvoiced plosive
cl cl*	closurefor unvoiced closure for voiced
a,eo,o,u,eu, i,ae,e,oe, eui	vowel
s,ch,j,s',j'	affricative, fricative
r,l	glide
y,w	semi-vowel
n,m,ng	nasal
h	aspiration
?	inseperable portion
~	transition
^	pause
>	voice offset

Table 3 - 1 Description Symbols

4. Database management system

4.1 Design of database structure

For the effective access and the easy manipulation of a large database, database management system is necessary and a commercial database management system UNIFY is used. This database management system is based on the relational data expression and supports fast access. Each data structure consists of 5 basic units, that is to say, speaker, word, syllable, phoneme, and phonetic environment. And all of these units are connected to the raw speech data file. Figure 4 - 1 shows speech database structure with UNIFY.

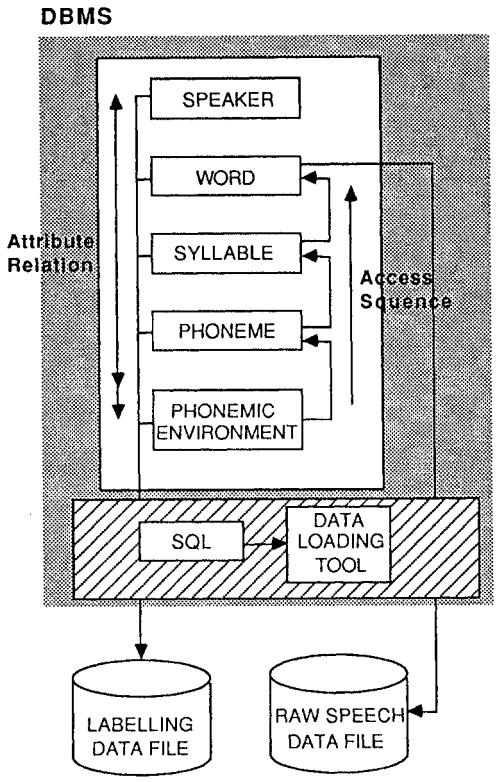

Figure 4 - 1 Speech Database Structure with UNIFY

4.2 Contents of data structure

As mentioned above, each data structure(table) has 7 - 12 fields and can also access other corresponding tables by inspecting the current reference field. Each table is summarized in Table 4 - 1.

SPEAKER

S_ID	Speaker no.
S_NM	Name
S_AG	Age
S_BR	Birthday
S_BP	Birth place
S_DL	Dialect
S_SX	Sex
S_RD	Rec. date
S_RC	Rec. condition
S_RV	Reserved

WORD

W_NO	Word no.
W_RF	Reference
W_NM	Name
W_ST	Start point
W_LG	Length
W_SC	Syllable count
W_GM	Part of speech
W_RV	Reserved

SYLLABLE

SY_ID	Syllable id.
SY_RF	Reference
SY_ST	Start point
SY_LG	Length
SY_PC	Phoneme count
SY_NM	Name
SY_PO	Position
SY_RF	Reserved

PHONEME

P_ID	Phoneme id.
P_RF	Reference
P_NO	Code no.
P_PO	Position
P_PR	Pre_ phoneme
P_FL	Fol_phoneme
P_ST	Start point
P_ED	End point
P_DR	Duration
P_GR	Group
P_NM	Name
P_RV	Reserved

PHONEME ENVIRONMENT

PE_ID	Environ. id.
PE_RF	Reference
PE_NM	Name
PE_ST	Start point
PE_ED	End point
PE_DR	Duration
PE_RV	Reserved

Table 4 - 1 Table Contents

5. Conclusion

Here we introduced ETRI speech databases which are being built. Currently, the collection of speech data from about 4 - 10 male speakers is completed. But we have only one database for female now. Hence we need to continue the collection of female data. The way of improving the labelling accuracy(including auto labelling) has to be intensively sought. Up to present we have developed speech databases for the research purpose only, but now, are planning to construct other databases for voice recognition and speaker verification in the public domain

Reference

[1]. Kai-Fu Lee : Automatic Speech Recognition, "The Development of the SPHINX SYSTEM" Kluwer Academic Publishers, 1989.

[2]. G. Perennou : "B.D.L.E.X. : A Data and Cognition Base of Spoken French" Proceedings of ICASSP, 7.5.1(1986)

[3]. R. Carre, et al : "The French Language Database : Defining, Planning, and Recording a Large Database" Proceedings of ICASSP, 42.10.1(1984)

[4]. H. Kuwabara, K.Takeda, Y.Sagisaka, et al : "Construction of a large-scale Japanese speech database and its management system" Proceedings of ICASSP, S10b.12(1989)

[5]. Shigeru Katagiri : "Speech Labelling Using a Spectrogram" , IEICE technical report, SP87-115, 1987(In Japanese)