

## University of Tsukuba Multilingual Speech Corpus (UT-ML)

### Producer and Project

Machine Intelligence and Biomedical Engineering Laboratory, University of Tsukuba,

Special Research Project for the Typological Investigation of Languages and Cultures of the East and West, University of Tsukuba during 1997-2001.

### Contents

Isolated words (50 items)

1. Digits (14 items)
2. Month names of the calendar month (12 items)
3. Seven days of the week (7 items)
4. Words on weather (4 items)
5. Phrases of greetings (6 items)
6. Words of reply (3 items)
7. Words on time (4 items)

Continuous speech

8. Aesop's fable "North wind"

### Speakers

98 speakers from 11 countries

English: 4 males and 4 females

German: 4 males and a female

French: 5 males and 3 females

Spanish: 3 males and 5 females

Russian: 3 males and 2 females

Arabic: 3 males and 2 females

Chinese: 7 males and 7 females

Korean: 6 males and 4 females

Japanese: 7 males and 6 females

Indonesian: 6 males and 2 females

Thai: 7 males and 7 females

### Recording environment

Recorded in a soundproof room with a headset microphone

### Recording media

1 CD-ROM (16 kHz sampling, 16 bit quantization)

**Price**

No fee to be used for research purposes only.

**Comments**

**Collected languages:**

We chose 5 languages from Europe such as English, French, German, Russian, and Spanish, and 6 languages from Asia including Arabic, Chinese, Indonesian, Japanese, Korean, and Thai. These languages were chosen considering the number of students from overseas at University of Tsukuba and we tried to keep the number of languages of Asia and Europe almost even.

**Speakers:**

It is desirable to have two or more speakers for each language to see the differences in language than the differences in individual speaker. We aimed at having 5 or more speakers for each language, but we could not fulfill this requirement for some languages. Speakers other than Thai languages are the students and professors at the University of Tsukuba. Thai speakers are mostly the workers at NECTEC research center in Bangkok.

**Utterance Content:**

We have recorded utterances of the same semantic content for each language. We chose such content which was supposed to be understood widely in the world. As the first category, we chose 50 words consisting of 14 digits, 12 month names of the calendar month, seven days of the week, 4 words on weather, 6 phrases of greetings, 3 words of reply, 4 words on time. We also chose a text of Aesop’s Fables “The North Wind and the Sun”. The text in English is shown in Table 1 and Fig. 1.

Table 1 Utterance list of 50 items.

1	One	11	Zero	21	July	31	Friday	41	Good night
2	Two	12	one hundred	22	August	32	Saturday	42	Thank you
3	Three	13	one thousand	23	September	33	Sunday	43	I'm sorry
4	Four	14	ten thousand	24	October	34	fine weather	44	Yes
5	Five	15	January	25	November	35	a cloudy sky	45	No
6	Six	16	February	26	December	36	rain	46	Hello
7	Seven	17	March	27	Monday	37	snow	47	two o'clock
8	Eight	18	April	28	Tuesday	38	Good morning	48	quarter past ten
9	Nine	19	May	29	Wednesday	39	Good afternoon	49	four thirty
10	Ten	20	June	30	Thursday	40	Good evening	50	quarter to eight

The text of each language was prepared as follows. We asked at least two speakers of each language to translate Japanese text into the language of the speakers. We asked them to use expressions which are supposed to be used more widely in standard expressions, if there are one or more expressions corresponding to a Japanese word or a phrase. In some languages, they use the same words of greetings in the morning and in the afternoon, and so the content of the utterance text is not necessarily the same for all languages. For instance, French word ‘Bonjour’ corresponds to ‘Good Morning’ and ‘Good Afternoon’ and Korean people use the same expressions for the greetings in the morning, in the afternoon and in the evening.

Table 2 Number of speakers of each language

Language	Male	Female	Total
English	4	4	8
German	4	1	5
French	5	3	8
Spanish	4	5	9
Russian	3	2	5
Arabic	3	2	5
Chinese	7	7	14
Korean	6	4	10
Japanese	7	6	13
Indonesian	5	2	7
Thai	7	7	14
Total	55	43	98

A long time ago the North Wind and the Sun had a dispute as to who was stronger. A traveler passing by became the object of their conflict. They decided that whoever could remove his cloak would be the stronger. The North Wind blew and the cloak nearly came off. The man held his cloak tightly and it remained on him. Next the Sun shone fiercely. The heat from the Sun warmed the man quickly and he removed his coat. In this manner the Sun proved to be the more powerful of the two.

Fig. 1 Aesop’s Fable “North Wind”

## 1.1 Recording of Speech Data

Speech data were recorded onto DAT in a soundproof room. The speaker wore a headset microphone and read aloud the text. We asked the speaker to read the text in standard pronunciation as far as possible. The speaker first read the 50 word list as a test and then he/she read it again, and finally he/she read the Aesop's Fable. When an operator noticed the hesitation or erroneous utterances, he asked the speaker to repeat the utterance until the speaker makes correct utterances. The time for reading was about 5 minutes for each speaker. The recording environment is illustrated in Fig. 2.

We collected speakers' information such as age, gender, birthplace, mother tongue, history of domicile, birthplace of the parents. We also asked each speaker to accept that the recorded data

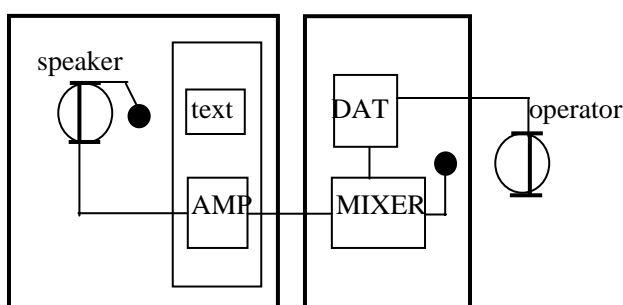


Fig. 2 Recording Environment

would be edited and converted onto CD-ROMs which would be used for research purposes by researchers at various research institutes.

The recorded speech data were taken into a workstation through a DAT interface with 16 kHz sampling and 16 bit quantization. Each of the 50 items and the fable were stored in a file, respectively.

## 1.2 Outline of the Recorded data

Table 2 shows the number of speakers of each language. There are many speakers of the ages around twenties and thirties. We intended to collect 10 speakers for each language, but the actual number of the speakers was 98. There are at least 5 speakers for each language.

## 2 Check of the Recorded Data

It is necessary to check the recorded data from the following points: difference between the utterance and the utterance list, degree of dialectal accent, speech rate, clarity of pronunciation, recording level, noise, etc. Especially, we found some speakers speak with very low speech level.

Fortunately, however, the recording was carried out in a soundproof room, and it was possible to magnify the amplitude of such speech data to a suitable level.