

Data Collection of Japanese Dialects and Its Influence into Speech Recognition

Ikuo KUDO, Takao NAKAMA, Tomoko WATANABE and Reiko KAMEYAMA

Texas Instruments Tsukuba Research and Development Center Ltd.

17 Miyukigaoka, Tsukuba, Ibaraki 305, Japan.

kudo@trdc.ti.com

ABSTRACT

This paper reports the successful completion of Japanese POLYPHONE project, Voice Across Japan (VAJ) data collection project. The database has the following characteristic, 1) large speakers database (8,866 spk.) through telephone line, 2) to gather participant's personal information such as gender, age, growing place, and so on, and 3) to put data segmented by phone or word boundary. This paper describes several aspects of Japanese dialects and also, reports the results of experiments. How much percents do dialects make influence on speech recognition. In our result, dialects makes 2-4% influence on speech recognition rate. The results are useful information for building practical speech recognition system as well as data collection.

1. INTRODUCTION

This paper reports the successful completion of Japanese POLYPHONE project [1, 2], Voice Across Japan (VAJ) data collection project [3, 4], and also, describes some analysis of Japanese dialects and some experiments for speech recognition on the database. VAJ project collected 8,866 speakers data through telephone line with speakers' information such as gender, age and growing area. The personal information became very useful in sampling, because we could gather speakers in proportion to population in Japan for gender, age and growing area. Thus VAJ database covers most of areas in Japan. Such balanced speakers set does not exit in Japan.

The participant's personal information is useful for analysis of Japanese dialects, too. According to the previous studies, dialects are defined as sublanguage that includes the different usage for pronunciation, lexicon and syntax. In our database, different usage's for pronunciation and lexicon exist. For example, "0 (zero)" is pronounced as "dero", "jero", "zeru" and there are several types of accent patterns. We describes several aspects related with dialectal phenomena in Chapter 3.

Accent problems in speech recognition [5, 6, 7, 8] have been studied. Such studies are very important in practical system. We, also, experimented how much percents do dialects give birth to speech recognition rate. In our result, in the scale of 3,000 spks., dialects made influence on 2-4% speech recognition rate. In Japan, there are two typical accents, "Kansai (Osaka)" accent and "Kantou (Tokyo)" accent. Although "Kansai" accent is different

from "Kantou's", "Kansai's" training data can cover "Kantou" test data and vice versa. But the error rate increase in 2-4% degree. The preferable combination between training set and test set was selected from the same dialects. Thus dialects give birth to the influence for speech recognition in some degree. But the degree does not seem to be so big as the one of American English, because Japanese accent is not so dynamic as English's one [9]. We cleared the degree through experiments of speech recognition.

2. VOICE ACROSS JAPAN DATABASE

2.1 Collected data

VAJ project collected 8,866 spks. in total. 8,809 adults and 57 children (under 15 years old). Here, we describes the 8,809 spks. below.

- (1) Gender: man's data is 3,583 spks. (40.4%), female's data is 5,226 spks.(59.0%).
- (2) Age: the ratio is shown in Table 1.
- (3) Growing place: the number of speakers is shown in Figure 1. The left side on Figure 1 means the ratio of each area for Japanese population, and the right side is the number of our collected data. In some degree, our data is different from the ratio of Japanese population, but VAJ data covers all the age (over 15 years old) and all the areas except for "Sakishima" which is very small population area.

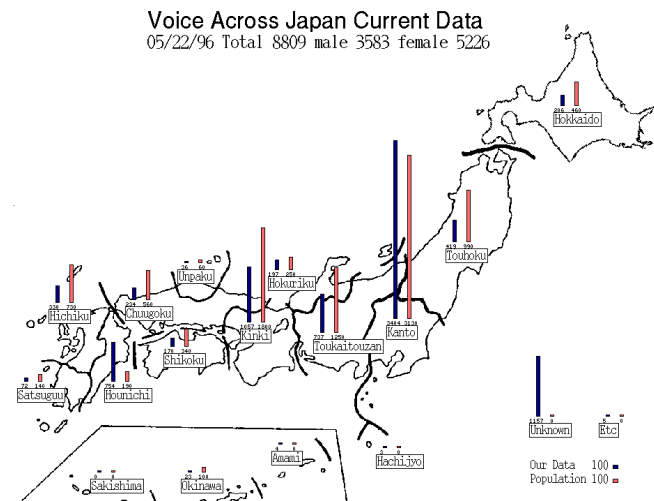


Figure 1. The numbers of speakers of each area.

Table 1 VAJ participants.

AGE	-10	20	30	40	50	60	70	80+	UN
Spks	770	2569	2481	1104	760	300	68	14	744
%	8.7	29.2	28.2	12.5	8.6	3.4	0.8	0.2	
Male	272	948	1115	446	253	134	28	7	
%	3.1	10.8	12.7	5.1	2.9	1.5	0.3	0.1	
Female	498	1621	1366	658	507	166	40	7	
%	5.7	18.4	15.5	7.5	5.8	1.9	0.5	0.1	

2.2 Validated data

In VAJ database, each session sheets includes 8 sentences for digits such as telephone number, 4 tri-phone balanced sentences and 2 yes-no answers. In total, 122,570 files were validated and 91,268 files (74.5 %) are available for training. The other files includes 1) background noise (11.4%) such as TV, other persons voice, music, fun, and so on, 2) unnecessary words, 3) hesitation to say, and so on.

2.3 Segmented data

Some files have word or phonetic segmentation. 16,294 files have word segmentation and they includes 157,268 digits sampled at random. About 6,000 tri-phone balance sentences, also, were segmented by a phonetic unit. Each sentence includes about 80 phonetics.

2.4 Signal to Noise Ratio (SNR)

Signal to Noise Ratio was shown in Figure 2. The median is 35 dB. The shapes of graphs are not dependent on gender, age, area code.

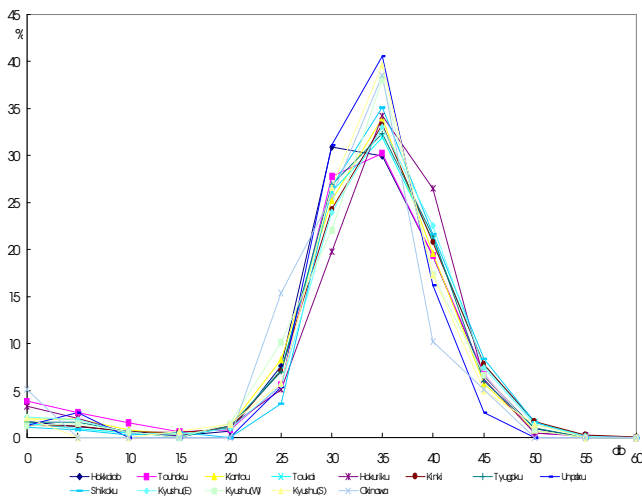


Figure 2 Signal to Noise Ratio (SNR).

3. Dialects in Japan

3.1 Different pronunciations

The same words are sometimes pronounced in different way. For example, as shown in Table 2, "0 (zero)" is changed into "dero", "jero", or "zeru". "1 (ichi)" is pronounced as "iji". "6 (roku)" is called "rogu". Linguists studied the classification of Japanese dialects, and then they said, "to make dialects map is not so easy". Our data, also, proves the words. That is, there is tendency of the area where a dialect is used, but there is not clear boundary of the area. The reason is why although "dero" is used in the "Kinki area (= Kansai)" mainly, it appears in the other areas such as "Kantou", "Toukai", "Kyusyu (East) area".

Table 2 Different pronunciations.

GENDER	dero	jero	zeru	iji	rogu	AREA	dero	jero	zeru	iji	rogu
male	13	4	0	0	0	Hokkaido	0	0	0	0	0
female	52	17	1	2	2	Touhoku	0	2	0	0	0
AGE	dero	jero	zeru	iji	rogu	Kantou	1	0	1	2	2
10	0	0	0	0	0	Toukai	1	0	0	0	0
20	6	1	0	0	0	Hachijo	0	0	0	0	0
30	10	0	0	0	0	Hokuriku	0	0	0	0	0
40	15	1	0	0	0	Kinki	55	2	0	0	0
50	17	4	1	0	0	Tvugoku	1	2	0	0	0
60	10	4	0	0	0	Unpaku	0	0	0	0	0
70	2	11	0	0	0	Shikoku	0	2	0	0	0
80	2	0	0	0	0	Kvusyu(E)	7	10	0	0	0
UNKNOWN	3	0	0	2	2	Kvusyu(W)	0	0	0	0	0
						Kvusyu(S)	0	0	0	0	0
						Amami	0	0	0	0	0
						Okinawa	0	0	0	0	0
						Sakishima	0	0	0	0	0
						UNKNOWN	0	1	0	0	0

3.2 Different accents

There is two representative accents in Japan, Tokyo ("Kantou") accent and Osaka ("Kansai") accent. "Kantou" accent is like standard accent. "Kansai" area has the second population (18.0%) in Japan and has business centers such as "Oosaka", "Kyoto" and "Kobe". "Kansai accent" often appear in TV programs through all Japan. Thus it is an outstanding phenomena in Japanese dialects. A sample spectrograph is shown in Figure 3. "2511-14-7085" is pronounced as "ni, go, ichi, ichi, no, ichi, yon, no, nana, zero, hachi, go". You can listen to the example in CD-ROM version of this proceedings. Please compare it to "Kantou" accents. "Kantou" accent is more flat. According to the previous study for Japanese accent [9], Japanese accent is not so dynamic as English's one (see English examples in [10]). According to our experiments of speech recognition (Training data = "Kansai", Test data = "Kinki", and vice versa.) as shown in Chapter 4, "growing place" is not so big influence as "age".

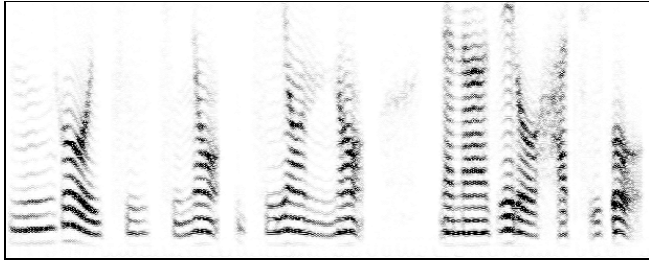


Figure 3 "Kansai" accent spectrograph.

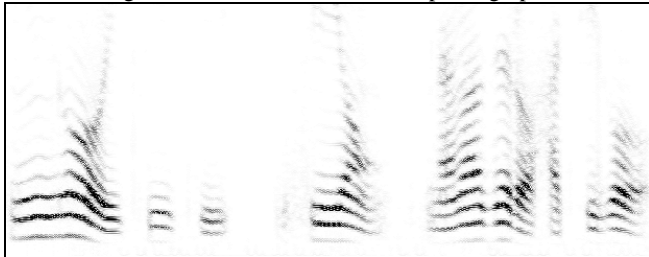


Figure 4 "Kantou" accent spectrograph.

3.3 Speaking speed

(1) The relationship between growing area and the speaking speed: "The Southern drawl" in U.S.A. is known as a dialect of which speaking speed is slower. Whether such phenomena exists in Japanese dialects or not? In our experiment, there is no dialect in Japan which has different utterance speed like "the Southern drawl". We used 157,268 segmented digits sampled from telephone number at random and counted the duration of each words by the area codes. The results are shown in Figure 5. It shows there is no difference of speaking speed between Japanese dialects in our experiments.

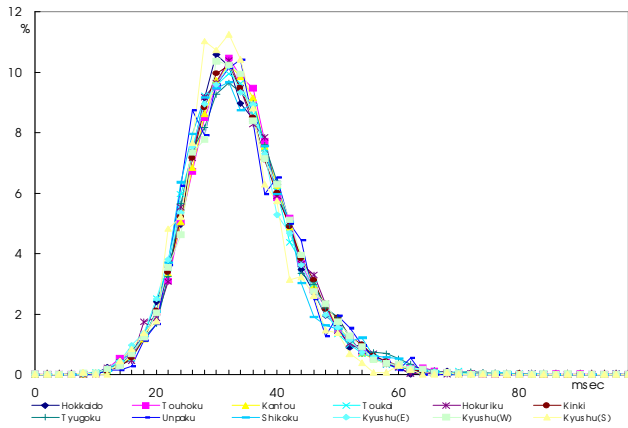


Figure 5 Speaking speed by each area.

(2) The relationship between age and the speaking speed: speaking speed is dependent on speakers' age. We counted duration length of 156,268 segmented digits data by each age. Then the speaking speed is slower, as the age is higher. The results are shown in Figure 6. The peak by teenagers is 28ms, while one by sixties is 36 ms. The speaking speed by sixties is

about 28% slower. This phenomena makes influence on speech recognition rates in our other experiments as shown in Chapter 4.

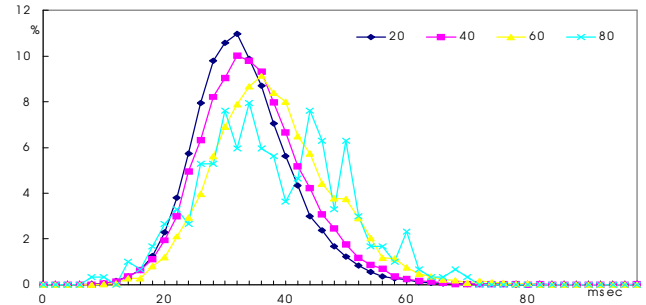


Figure 6 Speaking speed by each age.

(3) The relationship between gender and the speaking speed: Male's speed is faster than female's one. We compared male and female speaking speed by each generation. An example (twenties) is shown in Figure 7. Females' speed is slower about 20 ms. Such tendency is found in all the ages.

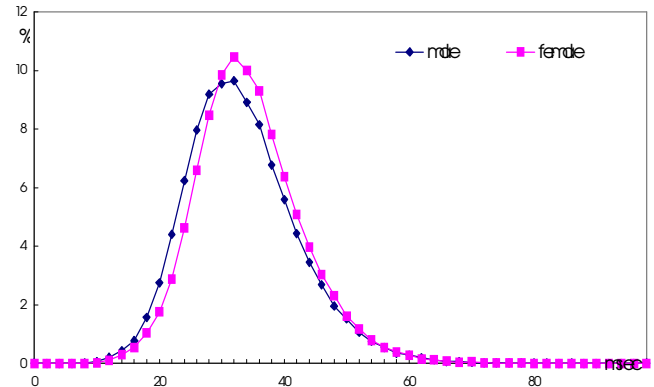


Figure 7 Speaking speed by gender.

4. Experiments of speech recognition

Here, we report experiments, how much percents do dialects make influence on speech recognition.

4.1 Experiments for speakers' age

Accoding to our experiments, age is very important factors. Figure 8 shows the error-ratio of each age. The shapes of the graphs are tend to be lower in twenties, thirties, and forties, and to be higher in teens, fifties, sixties, seventies and eighties. The same shapes appear, although the training set is different. The same trend appears in female's data. Thus "age" gives influence on recognition rate.

The training sets are three sets, "All", "Kantou" and "Nishi". "All" is a set selected from all the areas in proportion to population in Japan. "Kantou" is a data set extracted from only "Kantou" area. "Nishi" includes all the west areas in Japan. Each set includes 500 speakers' data (female 500 spks.) which are all

the ages in proportion to population in Japan. All the training data are digits and HMM models were created from segmented data of each sets with 8 kHz sampling rate, 30 ms Hamming windows, 20 ms frame, and LPC 10th order. The test data was open data, 3037 females.

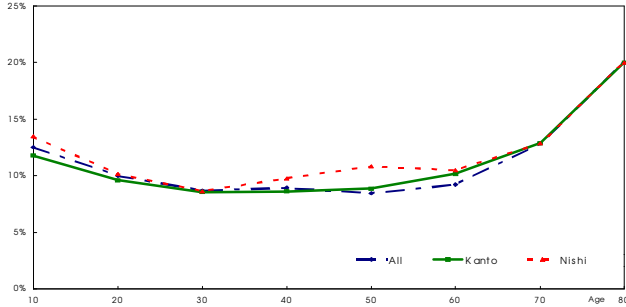


Figure 8 Error-ratio of speech recognition by each age.

4.2 Experiments for dialects

Dialects gives birth to the differences on recognition rate 2-4%. In Japan, there are 16 dialects. But as Hachijyo, Amami, and Sakishima is few data in our database, the test was carried out for the other 13 areas. The same models in the previous experiment were used. The test is open data set in Table 3. The data was extracted from twenties, thirties, and forties, because the other ages give influence on the result. (See 4.1) The results are shown in Figure 9. The important point is that although "Nishi-nihon's" accent is very different from "Kantou's", "Nishi-nihon's training data" or "Kantou's training data" can cover all the area. But the differences between three graphs in Figure 9 are 2-4%. The error rates in "Saninn" in Figure 9 are big. The SNR of "Saninn" is the same as the other areas. Thus this is not a SNR problem. Therefore dialects give birth to the influence.

Table 3 Test data sets (twenties, thirties and forties).

	Hokkaido	Tohoku	Kanto	Tohoku	Hokuriku	Kinki	Tyugoku
male	388	388	5906	1055	288	1623	336
female	440	763	10188	1901	346	2545	474
	Saninn	Shikoku	Kyushu(E)	(W)	(S)	Okinawa	
male	110	167	1496	396	139	10	
female	72	341	2060	445	157	70	

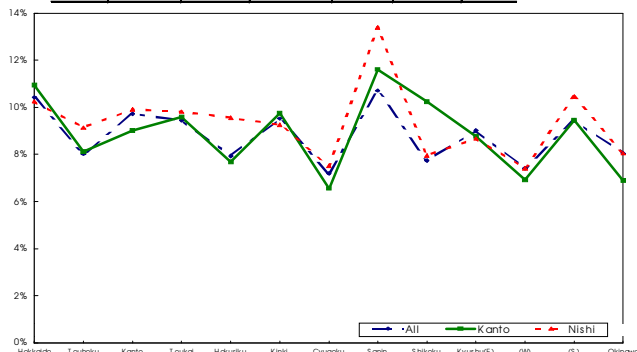


Figure 9 Error-ratio of speech recognition by each growing area.

5. CONCLUSION

This paper reported the successful completion of Japanese POLYPHONE project, Voice Across Japan (VAJ) data collection project and, also, reported some features for Japanese dialects. Furthermore, we experimented how much percents do dialects make influence on speech recognition and then make it clear. The results are useful for building of a practical speech recognition system as well as data collection.

6. REFERENCES

- [1] Report on the Cocosda Workshop, Berlin, Germany, 23-25, September, 1993.
- [2] Tom Staples, Joseph Picone and Nozomi Arai: "The voice Across Japan Database -- The Japanese Language Contribution to Polyphone", Proc. of ICASSP'94, Vol.1, pp.89-92, Australia, (1994).
- [3] Ikuo Kudo, Takao Nakama, Nozomi Arai and Nahoko Fujimura, : " The data collection of Voice Across Japan (VAJ) project", Proc. of ICSLP'94, Yokohama, pp.1799-1802, (1994).
- [4] Ikuo Kudo, Takao Nakama and Tomoko Watanabe : " An Estimation of Speaker Sampling in Voice Across Japan Database", Proc. of ICASSP'96, Atlanta, (1996).
- [5] J.Miwa and K. Hoshi : "On the characteristic between fundamental frequency and speech power in Japanese dialects speech," Proc. of fall meeting of Acoustic Society Japan, pp.287-288, (In Japanese), (1990).
- [6] H. Fujisaki, K. Hirose and N. Takahashi : "Statistical distribution of fundamental frequency in standard and dialectal Japanese", Proc. of fall meeting of Acoustic Society Japan, pp.251-252, (In Japanese), (1991).
- [7] Shuichi Itahashi, Tsutomu Yamashita : "A Discrimination Method between Japanese Dialects," Proc of ICSLP'92, Vol.2, pp.1015-1018, Banff, CANADA, (1992).
- [8] Shuichi Itahashi, Kimihito Tanaka : "A Method of classification among Japanese Dialects," Proc of EUROSPEECH'93, Berlin, Germany, (1993).
- [9] Miyoko Sugitou : "Nihongo Accent no Kenkyu (Study of Japanese accent)" Sanseidou, in Japanese, (1982).
- [10] Barbara Wheatley and Joseph Picone : "Voice Across America: Toward Robust speaker-Independent Speech Recognition for Telecommunications Applications", Digital Signal Processing, Academic Press, Inc. Vol.1, Num.2, April, (1991).

Sound File References:

[a308s01.wav]

[a308s02.wav]