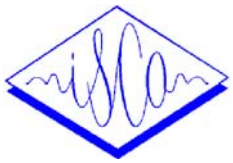

6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

A Study of Speech Recognition for the Elderly

Fumihiro YATO[+], *Naomi INOUE*[++] *and Kazuo HASHIMOTO*[++]

+ Teikyo University of Science & Technology
2525 Yatsusawa, Uenohara-machi, Yamanashi 409-0193 JAPAN

++ KDD Research & Development Laboratories
2-1-15 Ohara Kamifukuoka-shi, Saitama 356-8502 JAPAN

ABSTRACT

The elderly have needs for voice operated interfaces to manipulate various systems. However, only few results were reported about the performance of the current speech recognition technology with the elderly. In this paper we report our speech recognition experiments for 469 Japanese elderly persons belong to the age range 59 to 85. The speech recognition is executed on PC by using IBM Via Voice. The utterances of 600 place names are used for our speech recognition experiments. The average error rate becomes about 7 %, which seems to be a little bit lower than that of young adults. The paper presents the detailed analysis of the results of our experiments.

10.21437/Eurospeech.1999-29

1. INTRODUCTION

Recently we have many chances to manipulate a system controlled by a computer, e.g. cash dispenser and ticket vending. However such a system is not designed the user interface from the viewpoint of the elderly persons, therefore the elderly feels many troubles to use the system.

We believe that speech is effective and comfortable scheme to communicate each other for human being. So if we can apply speech recognition method for manipulating a system, we get a very comfortable man machine interface for the elderly. However, only few results were reported about the performance of the current speech recognition technology with the elderly [1][2]. In the proceedings of ICASSP 96, J. G. Wilpon and C. N. Jacobsen reported that the error rates of recognition increase dramatically for the speakers over 70 years old [1].

We are investigating the user interface design, especially the use of speech recognition technology, for the elderly sponsored by Telecommunication Advancement Organization of Japan, and we also recognized the difficulty to apply speech recognition for the elderly, therefore we executed the supplementary experiments for Japanese elderly.

In this paper we report our speech recognition experiments for 496 Japanese elderly persons belong to the age range 59 to 85.

2. SPEECH DATABASE

As the first step of the study, we have developed the speech database of the elderly. In this database, utterances of 469 Japanese elderly persons are stored with the label data.

The utterances include the 600 Japanese and foreign place names, phonetic balanced sentences, numbers and usual sentences for an international call. Each person pronounced only a part of the whole words and sentences. The utterances were recorded in the noise proof room.

Table 1 shows the distribution of ages of the elderly. The number of the persons aged over 75 years old becomes very few because of the difficulty to find out the suitable person. The elderly persons are registered to the employment agency for the elderly of Kawagoe or Kamifukuoka city. All persons are healthy and have strong will to work.

Table 1. Distribution of ages.

Age of speaker	Number of persons (Number of females)
Less than 65	114 (29)
From 65 to 69	177 (35)
From 70 to 74	127 (21)
From 75 to 79	40 (7)
Over 80	11 (1)

3. RECOGNITION METHOD

In our speech recognition experiment, we used the software IBM Via Voice. The reason to select this software as a platform is that Via Voice has the good phoneme models and supports the development kit under Microsoft Windows environment.

The recorded utterances of 600 place names were played on the PC through the Sound braster soundboad, and input to another soundcard directly of another PC as the input for Via Voice. The experiment system configuration is described in figure 1. Table 2 shows the part of place names used in our experiments.

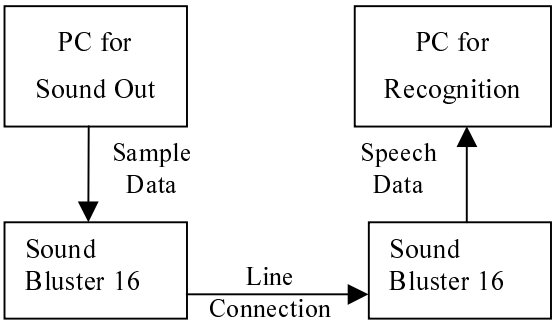

Figure 1. System configuration of experiments.

Table 2. Part of words used for experiments.

Sapporo	Sendai	Nagoya	Urawa	Yokohama
Kofu	Mito	Ohtsu	Matsue	Maebashi
Bejin	Manila	Paris	London	New York
Canada	France	Mexico	Peru	America

4. EXPERIMENT RESULTS

We also executed the speech recognition experiments for 32 adult persons aged from 20 to 56 as the reference. The average error rate of this experiment became 6.0 %. We think this value is the average performance of Via Voice for our task. In this experiment we used the headset microphone as an input device.

We show the relation between the average error rate and age of speaker in table 3.

Table 3. Relation between average error rate and age of speaker.

Age	Average error rate
Less than 65	6.29 %
From 65 to 69	6.59 %
From 70 to 74	7.27 %
From 75 to 79	5.79 %
Over 80	14.20 %

Table 4. Relation between number of persons and average error rate in each age group

(a) From 59 to 64 Years Old Group

Error Rate	Number of persons	Rate
$0 \leq E \leq 5$	58	0.51
$5 < E \leq 10$	34	0.30
$10 < E \leq 15$	14	0.12
$15 < E \leq 20$	4	0.04
$20 < E$	4	0.04

(b) From 65 to 69 Years Old Group

Error Rate	Number of persons	Rate
$0 \leq E \leq 5$	70	0.40
$5 < E \leq 10$	77	0.44
$10 < E \leq 15$	18	0.10
$15 < E \leq 20$	8	0.05
$20 < E$	4	0.02

(c) From 70 to 74 Years Old Group

Error Rate	Number of persons	Rate
$0 \leq E \leq 5$	50	0.39
$5 < E \leq 10$	50	0.39
$10 < E \leq 15$	13	0.10
$15 < E \leq 20$	9	0.07
$20 < E$	5	0.04

(d) From 75 to 79 Years Old Group

Error Rate	Number of persons	Rate
$0 \leq E \leq 5$	18	0.45
$5 < E \leq 10$	15	0.38
$10 < E \leq 15$	6	0.15
$15 < E \leq 20$	1	0.03
$20 < E$	0	0.00

(e) Over 80 Years Old Group

Error Rate	Number of persons	Rate
$0 \leq E \leq 5$	2	0.18
$5 < E \leq 10$	2	0.18
$10 < E \leq 15$	2	0.18
$15 < E \leq 20$	2	0.18
$20 < E$	3	0.18

(f) From 20 to 58 Years Old Group

Error Rate	Number of persons	Rate
$0 \leq E \leq 5$	15	0.47
$5 < E \leq 10$	14	0.44
$10 < E \leq 15$	3	0.09
$15 < E \leq 20$	0	0.00
$20 < E$	0	0.00

From table 3 we can find that the average error rates increase gradually according to the age of person except the group from 75 to 79 years old. Wilpon reported that the error rates increase dramatically over 70 years old. However in our result the error rates increase gradually compared with the reference value 6.0 %. Moreover the average error rate of the persons from 75 to 79 years old decreases and is small value than the reference value.

We analyzed the relationship between number of persons and average error rate in each age group for resolving the reason why the average error rate of persons from 75 to 79 years old decreased in our experiment. The average error rates are classified in every 5 %. Table 4 shows the results in each age group. From this table the persons from 75 to 79 years old group do not include the person with average error rate more than 20 %. The average rate of an appearance of such a person is estimated more than 2 % from other age group. Therefore if the number of persons is 40, then such a phenomenon will happen with probability more than 45 %. So in the speech recognition experiment, we need more than 114 subjects to avoid such a phenomenon with 10 % risk probability.

We can conclude that we need more subjects to obtain reliable results. However, actually it seems very difficult for us to collect so many speech data of elder persons more than 75 years old. So we need some collaboration with the organizations related to the study of elderly person's speech.

The reason why the average error rate of adult persons is less than that of the elderly from 75 to 79 years old can be explained as followings. In the reference experiment we

used microphone as an input device, so the reliability of the speech detection affects to the error rate. However, in the case of the elderly we used a line input as the input speech, so the speech detection does not affect so much to the error rate compared with the case of the reference experiment.

5. CONCLUSION

We have executed the speech recognition experiment for 496 Japanese elderly persons belong to the age range 59 to 85. From this experiment, we find out the average error rates of the elderly increase gradually according to the increase of age. However, we need more subjects to obtain more reliable results.

In next step, we will examine the recognition experiment by using the phoneme model created from the speech database of the elderly. Moreover we will collect more speech data of the elderly.

6. ACKNOWLEGEMENT

This research was supported by Telecommunication Advancement Organization of Japan. The authors thank the employment agencies for the elderly of Kawagoe city and Kamifukuoka city.

References

[1] J. G. Wilpon and C. N. Jacobsen, “A Study of Speech Recognition for Children and the Elderly”, Proc. of ICASSP96, pp.349-352 (1996).

[2] T. Konuma, H. Kuwano, T. Kimura and T. Watanabe, “A study of the elder speech recognition”, Proc. Autumn Meet. Acoust. Soc. Jpn., 2-Q-1 (1997-9) (In Japanese).