Development of Japanese Speech Database Read by Non-native Speakers for Constructing CALL System

NISHINA Kikuko[*1] *YOSHIMURA Yumiko*[*2] *SAITA Izumi*[*3] *TAKAI Yoko*[*4]
MAEKAWA Kikuo[*5] *MINEMATSU Nobuaki*[*6] *NAKAGAWA Seiichi*[*2]
MAKINO Shozo[*3] *DANTSUJI Masatake*[*7]
*1: International Student Center Tokyo Institute of Technology knisina@ryu.titech.ac.jp
*2：Toyohashi University of Technology *3：Tohoku University
*4：Tianjin Foreign Studies University *5： National Institute for Japanese Language
*6：University of Tokyo *7：Kyoto University

Abstract

This paper describes the construction and evaluation of Japanese speech database read by non-native speakers in order to develop CALL systems by a research project. The project has been organized with interdisciplinary members such as specialists in second language acquisition, phonetics and speech processing. The database recorded 140 non-native speakers, who were all overseas students at eight universities in Japan, to recognize distinctive features of pronunciation and prosody.

The project also constructed a native Japanese speech database, which was used as a control data based on the same text read by the non-native speakers. The corpus of the database is distributed on 5CD-ROMs, and includes the reading text sentences, words and dialogues. Professional language teachers and linguists evaluated this database.

1. Introduction

This paper describes the construction and evaluation of a speech database in order to develop CALL systems for Japanese language learning and teaching by a research project of Grant-in-Aid for Scientific Research on Priority Areas.

The aim is to propose a new technological and pedagogical method in the speech area of Japanese language teaching. From the pedagogical point of view, there are some problems for pronunciation teaching such as

(1) The necessity to adapt the courses differently according to the level and the native languages of the students,
(2) Total prosody teaching in the syllabus,
(3) Compromising between teaching natural pronunciation and teaching other syllabus items.

Teachers have to teach a lot of other items in a limited class hour. Under these conditions, an automatic system of appropriate instruction and accurate evaluation will contribute effectively to the phonetic field in Japanese language teaching. In order to solve these problems, we hope to develop an intelligent and flexible speech exercise system.

This database includes recordings of 141 non-native speakers and of 41 native speakers. The non-native speakers (overseas students at eight universities in Japan), were recorded in order to recognize distinctive features of their pronunciation and prosody. As control data, the utterances of male and female native speakers, all university students of the same age, were also recorded.

2. Contents of database

We will introduce the contents and methods of the construction of the database; discuss the evaluation of the data by professional teachers and the construction of the native speakers' database as a control data. The corpus includes the following four files:

1) Reading texts of approximately 100 sentences from ATR corpus in order to compare with native Japanese speakers.
2) Reading 115 words, chosen by experienced Japanese language teachers, including words that are difficult to pronounce.
3) Reading 108 sentences, which include the same words as those listed in number step 2.
4) Reading dialogues, including 11 types of prosody.

The corpora of the non-native speakers are distributed on five CD-ROMs, and those of the native speakers are on two CD-ROMs.

3. Method of the construction

3.1. Composing Reading task materials

Although we realized that accent and intonation in a word is an important part of speech, we did not include those features in the database. The reason for this was that the volume of tasks had to be reduced so as to lessen the burden of the informant.

Hence, we chose the items that were the most important for the non-native speakers to be understood by native speakers.

3.1.1. Tasks from ATR database

The Advanced Telecommunications Research Institute International (ATR) developed Phonetic Balanced Japanese Sentences for reading tasks. It contained 10 sets, each of which comprised 50 or 53 sentences, for a total of 503 sentences.

We have decided to use these tasks to compare native and non-native Japanese speech. We used 303 sentences of the above 503 sentences. We chose the six sets of ATR sentences that were deemed the easiest for non-native speakers to read.

3.1.2. Tasks of minimal pair with difficult pronunciations

It has been observed that non-native speakers have difficulties in pronouncing certain words in the Japanese language ([1], [2]). As such, 115 minimal pair words were prepared and tested during the speech recording. These minimal pairs can be roughly differentiated into 14 different groups as shown below.

1) Vowel versus long vowel
2) Voiceless vowel
3) Voiceless consonant versus voiced consonant
4) "*shi*" versus "*hi*"
5) "su" versus "tsu"
6)"chi" versus "tsu"
7)"da, de, do" , "ra, ri, ru, ru, re, ro", and "na, ni, nu, ne, no"
8) The consonants in "ka, ki, ku, ke, ko" versus "ga, gi, gu, ge, go"
9) Contracted sounds (*yoo-on*) and plain sounds (*choku-on*)
10) The syllabic nasal "n" sound
11) Double consonants
12) The combination of nasal and double consonant sounds
13)"za" versus "ja"
14)"*tu*"versus "*chu*"

3.1.3. Tasks of sentences including minimal pairs with difficult pronunciations

Although we extracted 115 minimal pair words for recording, we considered that utterances should be evaluated in a sentence context.

Hence, we composed 108 sentences including 115 minimal pair words. Because 108 sentences are too lengthy for informants to read, these sentences were divided into two sets, A and B respectively. Set A contains the 54 odd-numbered sentences, while Set B contains the 54 even-numbered sentences. Complicated sentences such as idiomatic expressions were avoided and the lengths of the sentences were kept to a minimum. In addition, the accents of the minimal pairs were kept as similar as possible. Below is one of the sentences of Set B:

Example 1(B-18)
Tenki ga warui node , denki o tsuketa.
(Because the weather was bad, I switched on the lights.)
'Tenki' (weather) and 'denki' (electricity) are respectively examples of a voiceless word and a voiced word.

3.1.4. Tasks for prosody

In order to evaluate the prosody of non-native speakers, dialogue tasks were included in the test. For the purpose of effective evaluation, these tasks can be having been divided into 11 different items ([3]) as indicated below.

1) Simple Yes/No questions
2) Wh- Interrogative sentences
3) Interrogative sentence with an interrogative word in the middle of a sentence, the answers and the return questions
4)*"nanika"* versus *"nanimo"*
5) Right base structure 1
6)Left base structure 1
7) Right base structure 2
8) Left base structure 2
9) Contrastive emphasis
10) Final particle(s)
11) Filler expressions

We will show a contrastive pair of examples from the above items.

Example (Text No.C-25) (an example of (5) right base structure):
A:Jiro wa donna ie ni sunde imasu ka?
(What type of house does Jiro live in?)
B:(Aoi yane) no ie desu. (A house with a blue roof.)

Example (Text N0.C-26) (an example of (6) left base structure)
A: Yumiko wa donna ie ni sunned imasu ka?
(What type of house does Yumiko live in?)
B: Aoi (ookina ie) desu. (A big blue house,
Literally; A big house which is blue.)

A speaker is not expected to put a prosodic boundary between 'Aoi'and 'yane'in the sentence of No.C-25. On the other hand, he/she is expected to put a prosodic boundary between 'Asian 'ookina'in sentence of No.C-26. Thus, these two examples indicate that one has to know how to put prosodic boundaries between words depending on the relationship between a noun modifier and the modified noun in the sentence.

3.2. Informants

We asked 141 overseas students at eight universities in Japan to complete four kinds of tasks. The universities are listed below.

1.Iwate University,
2. Kyoto University,
3.Oska University,
4. Tokyo Institute of Technology,
5. Tohoku University,
6. Toyohashi University of Technology,
7. University of Tsukuba,
8. The University of Tokyo

The tasks and number of recordings are shown on Table.1.

Table 1: Reading task list

Task	Male	Female	Total
A1	14	12	26
A2	12	12	24
A3	13	13	26
A4	13	12	25
A5	10	11	21
A6	10	9	19
A total	72	69	141
B1	41	37	78
B2	31	32	63
C	72	69	141
D	72	69	141

A1〜A6: Six sets of ATR tasks;
B1, B2: The two sets of sentences we composed;
C: 42 sets of dialogues including the items for prosodic evaluation as mentioned in 3.1.4.
D:115 minimal pair words;

The recording data includes 72 male informants and 69 female informants, totaling 141 students in all. More than ten different native languages were spoken among the pool of students that collaborated with us. These included, in balanced proportions, Chinese, Korean, Thai, Vietnamese, Malaysian Indonesian, Arabic, Spanish, French, and English. The language abilities of the students ranged from the intermediate to the advanced levels of their respective universities, in which learning terms range from six months to three years.

4. Recording

Informants were given the task lists a week before the day of recording. They were instructed to practice beforehand so as to familiarize themselves with the tasks. On the day of the recording, the following procedure was followed:

1) Reading the ATR sentence list
2) Reading sentences containing words that are difficult to pronounce, (sets B1 or B2).
3) Reading prosody

The recording of the dialogue was carried out with only the informant as the sole speaker. The informant was expected to understand the meaning of the dialogues and read and correctly convey the mood according to the context. (Corresponding to C intable.1 in 3.2)

4) Reading words with difficult pronunciations

5. Evaluation

We asked six professional Japanese language teachers to evaluate the recordings of the non-native speaker's. A common evaluation criterion was not established. Instead each expert evaluated the data using a scale of one to five point ranking system based on their own personal judgment.

5.1. Selection of evaluating data

We submitted the following from each recording for evaluation.

1) 5 phonetic balanced sentences from ATR data
2) 10 words among minimal pairs
3) 54 sentences from sets B1 and B2
4) 12 prosodic sentences

5.2. Evaluation Method

The questionnaires were prepared to evaluate the adequacy of the entries such as pronunciation, intonation and prosody of each recording. The six experts evaluated the data according to questionnaires written on the Web.

After receiving the evaluation results through the Website from the evaluators, we totaled the scores of as shown in fig. 1.

The results show that prosodic items attained higher points than phonetic balanced items.

From this, we interpreted that phonetic items had been evaluated more strictly than prosodic items. This indicates that people tend to give a better evaluation if they understand the whole utterance enough, rather than details of the utterances. Hence, we assume that non-native speakers are evaluated better if they speak in proper prosody although they do not pronounce accurate phonemes.

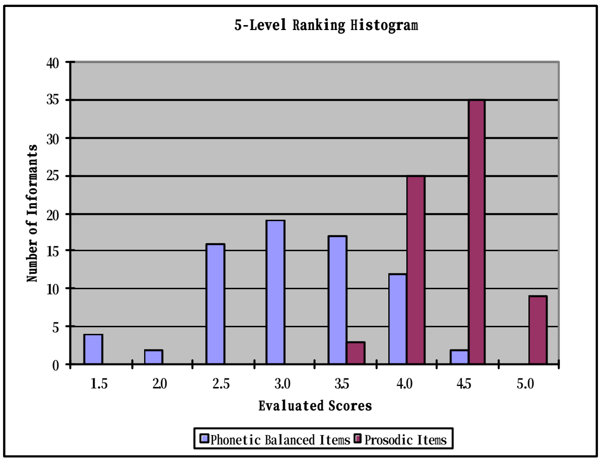

Figure 1. 5-Level Ranking Histogram

6. Japanese speech data read by native speakers

We recorded 41 Japanese native speakers reading the same texts as the non-native speakers. All of the Japanese native speakers were university students including 20 male students and 21female students, who speak Tokyo dialect. Using this as a control data, a comparison between Japanese native speakers and non-native speakers will be undertaken. We recognize the necessity of undertaking a detailed analysis in order to solve both common problems of non-native speakers, and specific problems related to each different language native language.

7. Conclusion

We have composed four kinds of reading tasks and recorded 141 informants to develop CALL systems. Also in order to determine the special features of non-native Japanese speech from a pedagogical, technical and linguistic point of view, we asked experts to evaluate the data.

From the evaluation results, we assumed that non-native speaker have tendency to be evaluated better if they speak in proper prosody even if they do not pronounce accurate phonemes.

We expect that the construction of the speech database will serve as a useful tool for the research of a technological and pedagogical method in the speech area of Japanese language teaching. Moreover we plan to accurately detail the evaluations to determine the various special features of non-native speech depending on the native language of the speakers. We expect to contribute to speech education CALL systems development through our future research.

8. Acknowledgement

The cooperation of the staff that recorded the data, the evaluation experts, the students and their universities is gratefully acknowledged. We would also like to acknowledge the financial support from Utilization of Multimedia to Promote High educational Reform, the Grant in-Aid of Scientific Research on Priority Areas (A) by Ministry of Education, Culture, Sports, Science and Technology.

9. References

[1] Bunka-cho National Institute for Japanese Language: "Kokugo Series, Bessatsu 3, Nihongo to Nihongo Kyooiku, Hatsu-on/Hyoogen-hen",5th Edition 1985

[2] Bunka-cho "Nihongo Kyooiku Shi-doo Sankoo-sho (1), Onsei to Onsei Kyooiku", 15th Edition 1988

[3] Spoken Language Working Group, Kikuo Maekawa "Speech and Grammar", Intonational characteristics of WH-questions in Japanese, pp.45-53 1977

[4] Itahashi,I.Yamamoto,M. Takezawa,T .Kobayashi,T., "Development of ASJ Continuous Speech Corpus", Jspanese Newspaper Article Sentences (JNAS), COCOSDA'97 1977

[5] Minematsu,N .Nishina,K. Nakagawa,K. Read Speech Database for Foreign Language Learning Association of Phonetics Vol11.No21 pp.45-53 2003