T. Lee, P. C. Ching, L. W. Chan, Y. H. Cheng, and B. Mark. 1995. Tone recognition of isolated cantonese syllables. *IEEE Transactions on Speech Audio Processing*, 3(3):204–209.

B. C. J. Moore and B. R. Glasberg. 1983. Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America*, 74(3):750–753.

S. Potisuk, J. Gandour, and M. P. Harper. 1996a. Acoustic correlates of stress in thai. *Phonetica*, 53:200–220.

S. Potisuk, M. P. Harper, and J. Gandour. 1996b. Using stress to disambiguate spoken thai sentences containing syntactic ambiguity. In *Proc. Int. Conf. Spoken Language Processing*, pages 805–808.

S. Potisuk, M. P. Harper, and J. Gandour. 1999. Classification of thai tone sequences in syllable-segmented speech using the analysis-by-synthesis method. *IEEE Transactions on Speech Audio Processing*, 7(1):95–102.

A. C. M. Rietveld and C. Gussenhoven. 1985. On the relation between pitch excursion size and prominence. *Journal of Phonetics*, 13:299–308.

M. J. Ross, H. L. Shaffer, A. Cohen, R. Freudberg, and H. J. Manley. 1974. Average magnitude difference function pitch extractor. *IEEE Transactions on Acoustics, Speech, Signal Processing*, ASSP22:353–362.

N. Ström. 1997. Phoneme probability estimation with dynamic sparsely connected artificial neural networks. *The Free Speech Journal*, 1(5).

M. Swerts, E. Stranger, and M. Heldner. 1996. f_0 declination in read-aloud and spontaneous speech. In *Proc. Int. Conf. Spoken Language Processing*, pages 1501–1504.

N. Thorsen. 1980. A study of the perception of sentence intonation-evidence from danish. *Journal of the Acoustical Society of America*, 67:1014–1030.

N. Thubthong and B. Kijsirikul. 1999. A syllable-based connected thai digit speech recognition using neural network and duration modeling. In *Proc. Int. Symposium on Intelligent Signal Processing and Communication Systems*, pages 785–788.

N. Thubthong and B. Kijsirikul. 2000. Improving connected thai digit speech recognition using prosodic information. In *Proc. the 4th National Computer Science and Engineering Conference*, pages 63–68.

N. Thubthong and B. Kijsirikul. 2001. Stress and tone recognition of polysyllabic words in thai speech. In *Proc. Int. Conf. Intelligent Technologies*, pages 356–364.

N. Thubthong, A. Pusittrakul, and B. Kijsirikul. 2000a. An efficient method for isolated thai tone recognition using combination of neural networks. In *Proc. the 4th Symposium on Natural Language Processing*, pages 224–242.

N. Thubthong, A. Pusittrakul, T. Sookawatand, and B. Kijsirikul. 2000b. Tone recognition of continuous thai speech using half-tone model. In *Proc. the 4th National Computer Science and Engineering Conference*, pages 69–74.

A. Tungthangthum. 1998. Tone recognition for thai. In *Proc. IEEE Asia-Pacific Conf. Circuits and System*, pages 157–160.

C. Wang and S. Seneff. 1998. A study of tones and tempo in continuous mandarin digit strings and their application in telephone quality speech recognition. In *Proc. Int. Conf. Spoken Language Processing*, pages 635–638.

C. Wang and S. Seneff. 2000. Improved tone recognition by normalizing for coarticulation and intonation effects. In *Proc. Int. Conf. Spoken Language Processing*.

D. H. Whalen and Y. Xu. 1992. Information for mandarin tones in amplitude contour and in brief segments. *Phonetica*, 49:25–47.

Y. Xu and Q. E. Wang. 1997. What can tone studies tell us about intonation? In *ESCA Workshop on Intonation: Theory, Models and Applications*, pages 337–340.

Y. Xu. 1997. Contextual tonal variations in mandarin. *Journal of Phonetics*, 25:61–83.

Y. Xu. 1998. Consistency of tone-syllable alignment across different syllable structure and speaking rate. *Phonetica*, 55:179–203.

W. J. Yang, J. C. Lee, Y. C. Chang, and H. C. Wang. 1988. Hidden markov model for mandarin lexical tone recognition. *IEEE Transactions on Speech Audio Processing*, 36(7):988–992.

J-S. Zhang and K. Hirose. 1998. A ıoboust tone recognitoin method of chinese based on sub-syllabic f_0 contour. In *Proc. Int. Conf. Spoken Language Processing*, number 3, pages 703–706.

Speech Database Construction for Japanese As Second Language Learning

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[4]

[1]International Student Center, Tokyo Institute of Technology,
2-12-1 Oookayama, Meguro-ku, Tokyo-to, Japan 152
Tel.: +8103-5734-3373 email: knishina@ryu.titech.ac.jp
[2]Toyohashi University of Technology [3]Tohoku University [4]Kyoto University
[5]National Institute for Japanese Language [6]University of Tokyo

Abstract

This paper describes the construction of a speech database for Japanese language learning and teaching by a research project of Grant-in-Aid for Scientific Research on Priority Areas (A). The aim is to propose a new technological and pedagogical method in the speech area of Japanese language teaching. This database recorded 140 non-native speakers, who were all overseas students at 8 universities in Japan, in order to recognize their distinguished features in pronunciation and prosody. The corpus of the database is distributed on 5CD-ROMs, and includes the reading text sentences, words and dialogues. The corpus is distributed into 141 speakers' files, which include the following 4 files each:

1. Reading data of approximately 100 sentences in ATR text in order to compare with native Japanese speakers.
2. Reading data of 115 words, which include difficult pronunciations for learners chosen by experienced Japanese language teachers.
3. Reading data of 108 sentences, which include the same difficult pronunciation words as those listed in number 2 above.
4. Reading data of dialogue, including 11 types of prosody.

1. Introduction

With increased globalization, people have more opportunities to speak with foreign peoples than ever before, meaning that we have more oral communication than ever before. Although the Communicative approach method in Japanese language teaching has gained some attention recently, phonetic, particularly prosodic items in the syllabus, have not been seriously considered.

2. Purposes

This paper aims to construct a speech database of non-native Japanese speakers in order to contribute to research and education for Japanese language learning.

In the field of phonetics, there are significant problematic areas in the present teaching method, as identified in the following statements:

1) It is difficult for learners to adopt an appropriate course for learning accurate pronunciation and prosody because language courses are usually divided by the learner's ability, such as knowledge of grammar and Kanji characters, rather than other factors. However, these kinds of knowledge and their speech ability are not identical.

2) Individual vowels and consonants have been emphasized more than total prosody in the syllabus. Too much time is required for learners to master correct and natural prosody. At the same time, teachers have to teach a lot of other items in a limited class hour. We can say these factors are the reason why complicated phonetic items such as prosody are not taught in the classroom activities.

Under these conditions, we seek to contribute an automatic system of appropriate instruction and accurate evaluation to the phonetic field of Japanese language teaching.

Moreover, network technology, as represented by the Internet, has shown remarkable progress in speech processing, particularly in such areas as speech recognition, and speech synthesis.

As our ultimate objective, we seek to integrate our pedagogical experience, linguistic knowledge and computational technology to develop an intelligent and flexible speech exercise system.

Our priority is to construct a speech training system that includes prosody, and one that will enable students to learn using the Internet. In order to construct such a system, we must first collect Japanese language speech data from as many international students as possible.

Then we must analyze the data in order to isolate the phonetic distributions that are easily and often mistaken by non-native speakers.

Lastly, we must construct standard phonetic and prosodic patterns based on these data. The following section describes how the database of non-native speakers is constructed for this purpose.

3. Informants

We asked 141 overseas students at eight universities* in Japan to read four different kinds of tasks. The tasks and number of recordings are shown below.

* The eight universities are as follows:
1.Iwate University 2.Kyoto University, 3.Osaka University, 4.Tokyo Institute of Technology 5.Tohoku University 6.Toyohashi University of Technology 7.University of Tsukuba 8.University of Tokyo.

A1～A6: 6 sets of ATR set tasks; each task has 50 sentences except one set, which has 53 sentences.

B1, B2: represent two sets of our original sentences. We composed 108 sentences, which include 115 words that are difficult to pronounce. We divided the total number of sentences by half in order to reduce the speaker's requirement. Each informant is assigned to read 54 sentences.

C: These are 42 sets of dialogues, which include 11 items for prosodic evaluation.

D: 115 minimal pair words, which include difficult pronunciations.

The recording data includes 72 male informants and 69 female, totaling 141 students in all.

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
B (total)	72	69	141
C	72	69	141
D	72	69	141

Their native tongues include more than ten different languages, such as Chinese, Korean, Thai, Vietnamese, Malaysian, Indonesian, Arabic, Spanish, French, and English. Language abilities ranged from the intermediate to the advanced levels at their universities, whose learning terms are from six months to three years.

4. Composing Reading task materials

In this section we will describe in more detail the four different types of texts mentioned above.

4.1. ATR database

Advanced Telecommunications Research Institute International (ATR) developed Phonetic Balanced Japanese Sentences for reading tasks. It contained ten sets and each set has been composed with 50 or 53 sentences, which becomes 503 sentences in total. All utterances and sentences are in the Japanese language.

As they have recorded native Japanese speakers using these tasks, we can compare both native speakers and non-native speakers. We decided to use these tasks for this reason.

ATR extracted these sentences at random from newspapers, magazines, novels, letters, textbooks, etc., and these were arranged by statistical logic without any particular context. These sentences are quite difficult to pronounce even by native adult speakers.

We used 303 sentences of the above 503 sentences, for which we chose rather simple sentences among the six sets of ATR sentences that were easy to read for non-native speakers. Therefore we estimated that we had to prepare supplements for the informants before making a speech recording. We made Japanese–English vocabulary lists for these tasks.

The following A101 ~ A103 are examples in set A1, which corresponds to J tasks in the original ATR sets.

Example 1:
A101 *Chiisana Unagi ya ni wa nekki no yoona mono ga minagiru.*
(The small unagi (eel) shop was filled with hot stuff).

Example 2:
A102 Doroboo demo haittaka to isshun boku wa omotta.
(For a moment, I thought that a burglar or someone had come in.)

Example 3:
A103 Gakusee wa repooto o oku to chotto atama o sagete dete itta.
(As soon as a student put down the report, he bowed his head a bit and left.)

4.2. Minimal pair tasks for difficult pronunciations

It has been observed that non-native speakers find 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 widely differentiated into fourteen different groups as shown below.

1) Vowel versus long vowel
The Japanese phonetic system is a pitch accent language. All syllables are pronounced high or low pitch while English is pronounced with a stress accent. Japanese syllables are pronounced approximately in equal length and stress.
Example4: "*biru*" (building) versus "*biiru*" (beer)

2) Voiceless vowel
Example5:"*tsu-ki*" (moon) " *kishi*" (bank)
Some learners of Japanese cannot distinguish voiceless "*u*", "*i*" in these words depending upon their mother tongues.

3) Voiceless consonant versus voiced consonant
Some language speakers, such as Chinese and Korean, have difficulty distinguishing between voiced and voiceless pronunciations in the Japanese language.
Example6: "*pin*" (pin) versus "*bin*" (bottle)
"*matsu*" (pine tree) versus "*mazu*" (first of all)

4)"*shi*" versus "*hi*"
Example 7: "*shikaku*"(qualification) versus "*hikaku*" (comparison)

5) "*su*" versus "*tsu*"
Example 8: "*tsuki*" (moon) versus "*suki*" (like)

6)"*chi*" versus "*tsu*"
Example 9: "*chiru*"(to fall) versus "*tsuru*"(to hang)

7) "*da, de, do,*" "*ra, ri, ru, ru, re, ro*" and "*na, ni, nu, ne, no*"
Example10:"*hade*"(flashy) versus "*hane*" (shuttlecock) and "*hare*" (sunny)

8) nasal consonant in "g*a, gi, gu, ge, go*" versus "*ka, ki, ku, ke,ko*"
Example 11: "*kaikai*" (opening a ceremony) versus "*kaigai*" (oversea)

9) contracted sound (*yoo-on*) and plain sound (*choku-on*) sound
Example 12: "*kyaku*"(guest) versus "*kaku*"(angle)

10) The syllabic nasal "n" sound
Syllabic nasal "*N*" has one syllabic
Example 13: *Hon o yomu* . (I read a book.)
Some nonnative speakers pronounce "*Ho-no- yo-mu,*" that is to say, "*ho*" and "*no*" are almost the same pitches without "*N*". However, "*N*" must be pronounced one syllabic word, it is difficult for some nonnative speakers to take a certain syllable length.

11) Double consonant
Example 14: "*kite*" (to come)" versus "*kitte*" (postage stamp)

12) Nasal and double consonant sound
Example 15: "*akeru*" (to open),"*ageru*" (to raise) and "*agatta*" (to have risen)
"za" versus "ja"
Example 16:"*kanzashi*"(hair pin) versus "*kanja*" (patient)
"*tu* "and "*chu*"
Example 17: "*tsuushin*" (communication) versus "*chuushin*" (center)

4.3. Reading tasks with words difficult for non-native speakers to pronounce
Although we extracted 115 minimal pair words for recording, we considered that utterances should be evaluated in a sentence context. Hence we made 108 sentences including 115 minimal pair words. Because reading 108 sentences may be too burdensome for the informants, these sentences were divided into two sets, A and B respectively. Set A contains 54 odd-numbered sentences while Set B contains 54 even-numbered sentences. Words that have difficult meanings were avoided and the length of the sentences was kept to a minimum. In addition, the minimal pairing of accents was made as similar as possible. Special attention was also given to ensure that natural sentences and onomatopoeias with easy-to-understand meanings were chosen. Below are some of the example sentences:
Example 18: *Tenki ga warui node, denki o tsuketa.*
(Because the weather was bad, I switched on the lights.)
'*Tenki*' (weather) and '*denki*' (electricity) are examples of a voiceless word and a voiced word.
Example 19: *Obasan to obaasan ni atta.*
(I met the aunt and the grandmother.)
'*Obasan*' (aunt) and '*obaasan*' (grandmother) are also examples of minimal pairs, which show a vowel versus a long vowel. (Ref to 4.2.1)
Example 20: *Suibun o, zuibun takusan totta.*
(I took a considerable amount of water.)
'*Suibun*' (water) and '*zuibun*' (considerably) are examples of minimal pairs that show a voiceless

word versus a voiced word.

4.4 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 divided into 11 different items ([3]) as indicated below.

1) Simple Yes/No questions

We intend to evaluate if an informant puts prominence in the predicate, and if he or she rises in pitch at the beginning of the word.

Example 21:

A: *Jiroo wa odoru*? (Does Jiro dance?)

B: *Iie, odorimasen. Odoru no wa Yumiko desu.*

(No, he does not dance. The one who dances is Yumiko)

Underlines are markers of prosody.

In this example, a pitch range of "*odoru*" must be wider than "*Jiroo.*"

A: Jiroo wa oyogu?

B: Iie. oyogimasen. Oyogu nowa Yumiko desu.

(Does Jiro swim? No, he doesn't. The one who swims is Yumiko.)

We intend to evaluate if he or she distinguishes a non-nucleus verb "*oyogu*" to a nucleus verb "*odoru*"

2) Wh- Interrogative sentences

We intend to evaluate if an informant puts prominence on Wh-interrogative words.

Example 22:

A: *Dare ga odoru?* (Who dances?)

B: *Jiroo ga odoru.* (Jiro dances.)

We intend to investigate interrogative sentences with an interrogative word in the middle of the sentence, as well as the answers and the return questions. Interrogative words can be put at the beginning or in the middle of a sentence in Japanese, but the prominence must be put on the word wherever it appears. We also intend to evaluate if (1) he or she puts a prominence on an interrogative word, (2) he or she raises a target word in a return question.

Example 23:

A: *Jiroo wa nani o tabemashita?*

(What did you (Jiro) eat?)

B: *Yakisoba o tabemashita.*

(I ate fried noodles.)

A: *Yakisoba?* (Fried noodles?)

B: *Soodesu.* (Yes.)

3) *"nanika"* versus *"nanimo"*

("Anything" versus "nothing")

We intend to evaluate if he or she put prominence on infinitives, *"nanika," "nanimo."*

Example 24:

A: *Yumiko-san wa nanika tabemashitaka?*

(Did you (Yumiko) eat anything?)

B: *Iie, nanimo tabete imasen.*

(No, I ate nothing.)

4) Right base structure 1

We intend to evaluate if he or she put a prosodic boundary between *"Aoi"* and *"yane."* It should be uttered without any pause.

Example25:

A: *Jiroo wa donna ie ni sunde imasuka?*

(What kind of house does Jiro live in?)

B: *Aoi yane no ie desu.*

(He lives in a house with a blue roof.)

5)Left base structure 1

We intend to evaluate if he or she puts a prosodic boundary between *"aoi "* and *"ookina "*. It should be pronounced with a pause between the two words.

Example26:

A: *Yumiko wa donna ie ni sunde imasuka?*

(What kind of house does Yumiko live in?)

B: *Aoi ookina ie desu.*

(She lives in a big, blue house.)

6)Right base structure 2

Example 27

A: *Ueno sensei wa nomu to odoridashimasu. Shittemasuka?*

(Ueno sensei starts to dance whenever he drinks. Do you know?)

B: *Shittemasuyo. Yuumei desukara.*

(Yes, I know. He is famous for it.)

In this dialogue, we evaluate:

(1) if he or she put a prosodic boundary just before the phrase "*nomu to*".

(2) if he or she maintain the final particle "*yo*" in low level in prosody.

(3) if he or she raises up final parts with LH% in case of taking a raising pitch.

7)Left base structure 2

Example 28:

A: *Yamada sensei ga shaberuto nemuku narimasu yone.*

(When Yamada sensei speaks, we get sleepy, right?)

B: *Souunan desu yo.* (Yes, I think so too)

In this dialogue, we evaluate:

if he or she raises the final particle *"ne,"*

if he or she keeps the final particle *"yo"* low level

in the prosody.

8)Contrastive emphasis
Example29:
A: *Etoo-san wa doko no shusshin desuka?*
(Where does Mr. Eto come from?)
B: *Itoo-san wa Nagoya no shusshin desu.*
(Mr. Ito comes from Nagoya.)
A: *Itoo-san janakute, Etoo-san desu.*
(Not Mr. Ito, but Mr. Eto.)
B: *Aa, Etoo-san desuka. Etoo-san nara Sendai no shusshin desu.*
(Oh, Mr. Eto? Mr. Eto is from Sendai.)
In this dialogue, we evaluate:
if he or she puts a prominence on *"Nagoya"*,
if he or she puts a prominence on *"Etoo san"*
if he or she puts a prominence on *"Sendai"*.
if he or she performs prosodic boundaries in certain places.

9)Final particle(s)
Example30:
A: *Nee, shitteru? Yamada sensei gane, daigaku o yameru rashiiyo.*
(Hey, do you know? Professor Yamada is retiring from the university.)
B: *Hontoo?* (Really?)
A: *Un. Daigaku o yamatene, kaisha o tsukuru rashiiyo.*
(Yes. He is going to retire and set up a company.)
B: *Kaisha o? Shinjirarenai naa. Hontoo desuka?*
(A company? That's hard to believe. Really?)
In this dialogue, we evaluate:
if he or she falls "*nee*" down,
if he or she starts "*hontoo*" in a low level tone and raises it up in the later,
if he or she utters "*hontoo desu ka*?" in doubtful tone.

10) Filler expressions
Japanese filler are pronounced "*eetto, anoo, maa* (well), *sodesu nee* (Let me see)" and so on. Additionally final particles "*ne, yo*" function as filler.
Example31:
A: *Chuuoo yuubinkyouku ni wa doo ikeba ii deshoo ka?*
(How do I get to the Central Post Office?)
B: *Yuubinkyoku nara, eeto, futatsu saki no shingoo o migi ni magatte kudasai.*
(To get to the post office, uh, turn right at the second traffic lights in front there.)
A: *Shingoo kara dore gurai arimasuka?*
(How far is it from the traffic lights?)
B: *Soo desu nee, futsuu ni arukeba, maa go fun gurai kana.*
(Well, let me see. By foot, it will take about 5 minutes, I think.)

In this dialogue, we evaluate:
if he or she keeps *"eeto", "maa" plain* tone,
if he or she keeps "*(soodesu)nee*" in a low tone or falls it down.

5. Recording

Informants were given the task list a week before the day of recording. They were instructed to practice beforehand so that they could familiarize themselves with the pronunciation of the words. Furthermore, they were expected to look up the meanings of the words that they did not understand.
On the day of the recording, the following procedures were followed:
ATR sentence list
Recordings were carried out under the supervision of the data base personnel. The informant read the sentences that were previously provided.
Words with difficult pronunciation
The recording of 115 words was carried out.
Sentences containing words with difficult pronunciation
The recording of one of the sets, A or B (as mentioned earlier) was carried out.
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 dialogue and read and express it according to the situation.

6. Conclusion

We constructed four tasks and recorded 141 types of data. We observed that accent and intonation in a word are important parts of speech, however, we did not include these in this database. We chose the items considered the most important for non-native speaker to be understood by other people.

Non-native speakers who have different mother tongues were observed to have distinctive features in their pronunciation of the Japanese language.

We expect that the construction of the speech database will serve as a useful tool for the research of technological and pedagogical methods in the speech area of Japanese language teaching.

Moreover, we intend to provide a more accurate evaluation and show various special features of non-native speakers depending upon their mother tongues.

References

[1] *Bunka-cho Kokuritsu Kokugo Kenkyuu-jyo* (1985) *"Kokugo Series, Bessatsu 3, Nihongo to Nihongo Kyouiku, Hatsu-on/Hyougen-hen"*, 5th Edition

[2] *Bunka-cho* (1988) *"Nihongo Kyouiku Shi-dou Sankou-sho (1), Onsei to Onsei Kyouiku"*, 15th Edition

[3] Spoken Language Working Group, Kikuo Maekawa (1997) "Speech and Grammar", Intonational characteristics of WH-questions in Japanese, pp.45-53

[4] S. Itahashi, M. Yamamoto, T. Takezawa, T. Kobayashi (Sep. 26-27th, 1997) "Development of ASJ Continuous Speech Corpus", Japanese Newspaper Article Sentences (JNAS), COCOSDA'97

Acknowledgements

The development of this database construction was carried out with financial support from Utilization of Multimedia to Promote High educational Reform, the Grant in-Aid for Scientific Research on Priority Areas (A) by Ministry of Education, Culture, Sports, Science and Technology. We particularly acknowledge, with great gratitude, the extremely fruitful cooperation with both staffs that recorded the data and the 141 overseas students who participated at the eight universities.

SHORT

PAPERS