

Development of a Program for Self Assessment of Japanese Pronunciation by English Learners

Chiharu Tsurutani*, Yutaka Yamauchi**, Nobuaki Minematsu***,
Dean Luo, Kazutaka Maruyama and Keikichi Hirose****

*School of Languages and Linguistics, Griffith University, **School of Business and Commerce, Tokyo International University

***Graduate School of Frontier Sciences, The University of Tokyo

****Graduate School of Information Science and Technology, The University of Tokyo

C.Tsurutani@griffith.edu.au

Abstract

A program for self assessment of Japanese pronunciation by English-speaking learners was developed using a language model built with input from a language teacher in collaboration with speech engineers. This collaboration enhanced the program's capacity for accurate assessment and provides practical support to users by linking evaluation with feedback, and an editorial function of error patterns. The program drew positive responses from participants in a trial run. This paper discusses the development of our language model, the function and evaluation of this self assessment program.

Index Terms: self assessment, CALL, Japanese, speech recognition, network grammar

1. Introduction

The number of learners of Japanese language has increased dramatically since 1980's and reached 2.35 million in the world [1]. The project seeks to support these students for the improvement of their pronunciation. Recent language education has had focus on communication skills; being able to communicate is most important, and has neglected teaching pronunciation under the trend of communicative approach [2][3]. However, comprehensible level of pronunciation is a foundation of communication. If a learner proceeds his learning path without being corrected for wrong pronunciation, habits of mispronunciation will embed easily and are difficult to remedy later. To prevent this problem we cannot resort to individual teachers because it is impossible for teachers to check and correct each student's pronunciation in every occasion. Even when teachers point out learners' errors, both learners and teachers might not know how to correct the error unless they have knowledge of phonetics and phonology. Besides, for some learners, it could be an embarrassing experience to have their pronunciation corrected in front of others and might discourage them to continue their study. If pronunciation is corrected by computer, learners do not have to worry about awkward repetition in front of a teacher and can practice pronunciation as many times as they like. In this respect, Computer Assisted Language Learning (CALL) is an ideal solution to resolve the dilemma around pronunciation practice. In recent years, interest in CALL in language teaching has flourished and several programs that can detect pronunciation errors at phoneme level have been developed [4]. However, a system which is equipped with both evaluation and feedback has not been produced. We brought together language teachers with experts in speech engineering to collaboratively develop a program for self

assessment of Japanese pronunciation by English learners. The process of development, overview and evaluation of the program will be discussed in the following sections.

2. Development of language model

In the past, the development of automatic speech recognition process for CALL systems was solely left in the hands of speech engineers and did not draw on the classroom experience of language teachers. We attempted to reflect language teachers' knowledge to improve the accuracy of an assessment program. First of all, based on the knowledge of an experienced language teacher, a list of possible errors English learners make were prepared proceeding from the most frequent error to the least frequent one in Table 1 below:

Table 1: Predicted error patterns

1	Insertion of a consonant / hito → hitto
2	Lengthening of a vowel / a → aa, u → uu
3	Omission of a consonant / itte → ite
4	Shortening of a vowel / aa → a
5	Simplification of a contracted sound/ CyV → CV
6	Vowel insertion in a contracted sound /CyV → CiyV
7	Substitution of a consonant / R (flap) → l (lateral approximant)
8	Substitution of a consonant / ts (alveolar affricate) → s (alveolar fricative)
9	Substitution of a consonant / i → ai
10	Insertion or omission of a nasal
11	Substitution of a consonant / F (bilabial fricative) → f (dental fricative)
12	Devoicing of a voiced consonant at the word initial position, Affrication of a fricative/ b → p, s → ts,
13	Diphthongization of vowels / o.i → oi
14	Devoicing of vowel /shimas → shimasu
15	Omission of a vowel / shiyuu → shoo,

To raise the accuracy of the program, actual speech data from learners were collected. 27 beginners and 19 intermediate Australian learners of Japanese participated in the data collection. They recorded their articulation of 30 Japanese sentences that contained the sounds and positions where the errors listed in Table 1 above were likely to occur.

The recordings were transcribed by Japanese university students majoring in linguistics who passed a short test to check the accuracy of their perception and transcription. Each sentence was transcribed by three people and when more than two people



agreed, the transcription was judged correct. After calculating the number of errors and error patterns in the final transcription, errors that occurred more than a few times were judged as systematic and added to the error pattern in Table 2.

Table2: Additional error patterns

16	<i>kudasai</i> → <i>ku:dsai</i>
17	Substitution error / d→ r
18	Substitution error / geminate consonants→ a long vowel
19	Substitution error / a long vowel→ geminate consonants

In the entire list, the top four error patterns (1~4) are duration errors. It is well known that HMM-based acoustic modeling used in speech recognition system is not capable of handling the length of phoneme adequately. To secure the accuracy of assessment, some adjustment would be required as post processing when we work on error detection later.

3. Construction of network grammar

The analysis of error patterns is important for a technological reason as well as for the educational purpose to find out the tendency of learners' errors. When the program detects pronunciation errors using the automatic speech recognition system, it is necessary to narrow down the possibility of error patterns in advance. Without this consideration, the accuracy of error detection is extremely low. We need to create a network grammar that predicts the occurrence of errors. If we use a network grammar which considers every possible error, the detection rate would suffer. An example using the sentence "Okyakusan ni chotto matte moratte kudasai" (Please ask the guest to wait for a second) pronounced by one learner is identified as follows:

G.1) /o/k/i_ins_6/y/a/q_ins_1/k/u/s/a/N/n/i/ch/o:_sub_2/q/t/o
/m/a/q/t/e/m/o/_sub_9/R_sub_7/a/q/t/e/k/u/r_sub_17/a/q_ins_1/ts_sub_C/a/I

*sub=substitution, ins =insertion, q= geminate consonants

The actual utterance produced by the learner was:

"OkyaQ(1)kusan ni chotto matte moR(7)atte kudasai(13)"

*The number indicates the error number in Table 1 and 2.

The result does not sound like the original utterance at all. Therefore we need a network grammar which reasonably covers the possible error patterns. Next we will see the version after applying the 19 error patterns the teacher predicted, without any constraints.

G.2) /o/ky/a/q_ins_1/k/u/s/a/N/n/i /ch/o:_sub_18/t/o
/m/a/q/t/e/m/o/R_sub_7/a/q/t/e/k/u:_sub_2/r_sub_17/a/s/ai_s
ub_13

Some adjustment is still needed to make it close to the actual utterance of the learner. In order to eliminate some unrealistic errors that the system mechanically produced, the Japanese teacher restricted application of the network grammar to certain sounds, taking into consideration the tendency of English speakers' error patterns. In practice the Japanese teacher indicated where in the sentence a particular error pattern occurs and avoided listing of all possible incorrect sentences.

Okyaku(1)san ni(2,5,6) chotto(3,18,5,6) matte(3,18)
moratte(3,7,9,18) kudasai(13,16,17)

This information improves the error detection as follows:

G. 3) /o/ky/a/q_ins_1/k/u/s/a/N/n/i /ch/o/q/t/o /m/a/q/t/e
/m/o/R_sub_7/a/q/t/e /k/u/d/a/s/ai_sub_13

These examples demonstrate the importance of language teachers' input to the development of the language model.

We conducted two more operations to raise the accuracy of assessment, use of triphone and post processing. Australian learners' incorrect pronunciation of a Japanese word would contain an Australian English phoneme inserted between Japanese phonemes. We created these weird triphone HMMs, which are never found in native speech samples of Japanese and Australian English, by state clustering among Japanese HMMs and Australian English HMMs. Figure 1 shows that the difference between the detections, based on monophone and triphone (mono / tri). Three stages (G1), 2), 3)) mentioned above are also marked in the graph and the improvement after the teacher's input is observed. However, as mentioned before, the program inevitably leaves duration errors out of its error detection. Post processing that adjusts the incorrect judgement of duration further improved the detection rate as shown at G 4) in Figure 1 below.

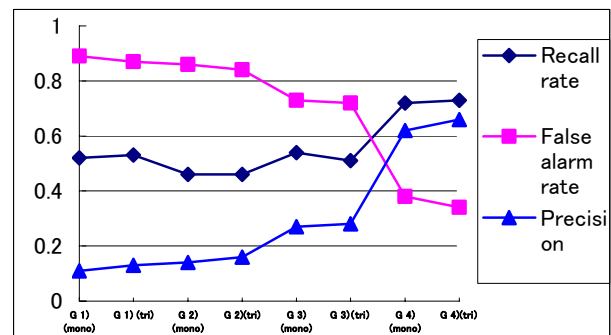


Figure 1: Detection rate

In this graph, recall rate, false alarm rate and precision were calculated using 110 sentences that were selected from the data collection in the previous section. The following formulas were used to obtain the three measurements.

Recall rate = the number of errors correctly detected / actual number of errors

False alarm rate = the number of errors incorrectly detected / the number of all errors detected by the program

Precision = the number of errors correctly detected / the number of all errors detected by the program

The results indicate that the program can detect errors at 70 % of accuracy. When we compare it with the accuracy of assessment by humans, this figure seems to be acceptable. However, further adjustment of assessment is under way by examining the consistency of assessment by humans.

Error patterns could differ depending on learners' proficiency level and individual learner's pronunciation habits. It is ideal if language teachers can modify the network grammar according to the level of learners. To this end, we organized Graphical User Interface as well. Teachers can use this editorial system to add a new sentence they want to use, and choose necessary errors to be detected in the sentence depending on students' level. It does not only raise the accuracy of detection of errors, but also provides language teachers with the knowledge of how the system is organized. In the next section, we will explain the

process flow of the program which is equipped with these functions discussed in this section.

4. Overview of the assessment program

Our assessment program consists of seven stages: (1) preparation (2) recording (3) evaluation (4) diagnosis (5) exercise (6) comparison and (7) network-based practice.

4.1 Preparation

As a first step, the language teacher (evaluator) accesses the website of the assessment program and modifies the parameters of the network grammar which predicts error types so that the assessment is appropriate for the pronunciation level of the participants. The teacher then creates ID numbers of participants on the website.

4.2 Recording

The participant starts up the recording program and types in his/her ID. S/he chooses the suitable level for his/her Japanese proficiency, whether beginner or intermediate. Then s/he reads aloud the 10 sentences that appear one by one on the monitor screen. When s/he is unsure about the meaning of the sentence, s/he can click the button “English translation” and see the meaning of the sentence. While the participant is recording a sentence, sound waveforms are displayed on the screen. After recording s/he can listen to his/her pronunciation by clicking the “play” button. If s/he is not satisfied with the recording, s/he can record the same sentence as many times as s/he can. After s/he presses the “submit” button, the recorded sentence is automatically compressed and sent to the Griffith University server through the Internet. The server stores the recorded data with the participant’s ID, date and time. Through this procedure learners of Japanese language anywhere in the world can record and send his/her pronunciation through this network and obtain immediate feedback.

4.3 Evaluation Procedure

After completing reading of the ten sentences the participant accesses the website of the program and types in his/her ID and password. S/he can see the ten sentences with the overall scores. By clicking each sentence, s/he can view the results in detail.



Figure 2: Computer screen displaying pronunciation assessment

We attempted to provide scores for each phrase as well as the total score for a sentence. In order to provide the score for phrases, we used the assessment of a Japanese teacher, and built the framework to determine the score by setting acoustic parameters obtained from acoustic analysis as explanatory variables, based on the CART (Classification and Regression

Tree) algorithm here. The work to raise the accuracy of phrase scores is currently in progress.

The participant can compare his/her own pronunciation with the model pronunciation of a native speaker by clicking “play” and “model” buttons. So s/he can check the difference between the model speaker’s and his/her own. “Feedback” button appears under the phrase where pronunciation errors are found.

4.4 Diagnosis

“Feedback” reveals to the participant the diagnostic information on which part and what kind of error s/he has made as shown below. Palatograms and video clips are also available to explain how to use his/her vocal organs and change manners of articulation and so forth.



Figure 3: Computer screen for feedback and exercises

4.5 Exercise

A few phrases that include the same type of error the participant has made appear on the screen for the participant to practice his/her pronunciation.

4.6 Comparison

Since the server stores access information such as recording date, time, performance score and error types, the participant can view the previous performance and compare it with his/her current trial. This facilitates the learner the awareness of his/her own progress and motivates continued use of this program.

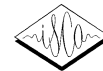
4.7 Network-based practice

After the participant finishes practicing his/her pronunciation using this program, s/he can access the website for audio chat and videoconference through the link button. In audio chat or videoconference the participant can practice what s/he has learned through this program in a real conversation. S/he can check how intelligible his/her pronunciation is and how well s/he can communicate with his/her interlocutor. After participating in the audio chat or videoconference the participant is encouraged to take notes on his/her pronunciation and points to be aware of for improvement. This recording note helps the participant realize his/her progress and conduct self-evaluation.

5. Evaluation of the program

14 Australian learners of Japanese (five beginners, five intermediate and four advanced learners) participated in the trial and filled in an evaluation sheet. The evaluation sheet asked the following questions:

- 1) Was the program easy to use? 1)-1. Recording / 1)-2. Evaluation



- 2) Was the program fun to use?
- 3) Did the program clearly tell weak points in your pronunciation?
- 4) Did the program explain well what needs to be done to improve your pronunciation?
- 5) What other functions do you think the program should have?
- 6) Do you think the program is useful for your study?
- 7) Which do you prefer for practicing pronunciation, with (a teacher or a computer)?
- 8) Do you want to try the program again? (Y / N)
- 9) Any other comments?

The answers for 1)~5) indicated in scores 1(good)-5(poor), were averaged and summarized in Table 3.

Table3: Participants' responses to program evaluation

Proficiency	1)-1.	1)-2.	2)	3)	4)	5)
Advanced	1.20	1.80	1.40	1.40	1.80	1.25
Beginner	1.50	2.00	1.50	1.25	1.75	1.25
Intermediate	1.40	1.80	1.00	2.00	2.00	1.80
Total	1.36	1.86	1.29	1.57	1.86	1.46

Overall, participants gave high scores on both the usefulness and usability of the program. However, the relatively low score in 1-2 and 4) indicates that the display of the web screen needs more improvement. When the results were examined according to the learners' proficiency levels, it was found that advanced learners were more supportive of the CALL program than beginners. In general, advanced learners who have acquired autonomy in their own study positively use CALL programs than beginners who still need guidance and encouragement from teachers [5]. Results from the evaluation by the small number of users in our trial are consistent with the tendency. The computer screen needs to be designed to substitute encouragement and support from teachers, which is vital for beginners' study.

We considered the participants' reaction to the computer's assessment and noted the following points.

- Learners accept the score after comparing their own pronunciation with the model pronunciation a few times
- The scoring system motivates learners to score as highly as they can
- Learners prefer to compare their pronunciation with the model pronunciation rather than reading the description in the feedback.
- Some learners commented that the program seems to be too sensitive.

The generation that has grown up with modern technology find it appealing to obtain the assessment of their pronunciation immediately and accurately from the computer. However, when it comes to practice, they prefer to practice with humans. When participants experience difficulties in their speech or repeat some phrases, the computer is not tolerant or supportive. After all a machine cannot compete with the flexibility of human conversation partners. At the same time people cannot assess their partner's pronunciation precisely on the spot while having a conversation. Machine and human can compensate for each other's weakness in the practice of pronunciation. Some students voiced a view along the lines that, "When I made an error in our pronunciation, the teacher lets me know by repeating the word I mispronounced, but does not tell me exactly which sound was wrong and how I should correct it. This assessment program can do it. I still like to practice

conversation with a teacher and my friends in class, but it is very convenient if I can practice my pronunciation using this program at home."

Our assessment program is clearly suitable for self study, when learners need to know weak points in their pronunciation. The ultimate goal of pronunciation practice in a second language is, however, to be able to converse with native speakers. We leave that role to the chat with native speakers on the program's web page. Lastly, some learners' comments alert that the false alarm rate is still high. We need to continue to work on this problem.

6. Conclusions

In this paper, we have discussed the development process of a CALL program which assesses Japanese pronunciation. Collaboration between language teachers and engineers has made the product more readily adaptable to the classroom situation. The improvement of network grammar, and the editorial system which allows teachers to select error patterns to apply are certainly advantageous for the development of a CALL system in this area. The evaluation of the program by the 14 users was generally very favorable, although it also revealed the following shortcomings that require improvement:

- 1) The web display needs to be easier to read and use
- 2) Feedback on the web screen needs to be more stimulating
- 3) Speech recognition technology should be applied to parts of the pronunciation exercise
- 4) Learners' progress could be monitored through chat sessions or teleconferences
- 5) The accuracy of assessment needs to be further improved

In addition to attending these shortcomings, we plan to further develop the program so that it can detect prosodic errors as well as segmental errors. This program targeted only English learners but we would like to expand the scope of the program to Chinese or Korean learners of Japanese in the future. Overall we believe that the collaboration between language teachers and engineers has been very fruitful, allowing us to produce a useful program and as such, a valuable tool for learners of Japanese language.

7. Acknowledgement

This project was funded by Matsushita Education Foundation.

8. References

- [1] Japan Foundation,
http://www.jpff.go.jp/j/japan_j/oversea/img/2003-01.pdf
- [2] Celce-Murcia, M., Brinton, D.M. & Goodwin J.M.
Teaching pronunciation Cambridge University Press, 1996
- [3] Tsurutani, C. "Speech rate normalization strategies of moraic timing –The study of English learners of Japanese-
Journal of Japanese Language Teaching No. 119, 51-64, 2003
- [4] Kei Ohta and Seiichi Nakagawa
A statistical method of evaluating pronunciation proficiency for Japanese words, *Proc. Eurospeech' 2005*, 2233-2236, 2005
- [5] Yamauchi, Y. *Multimedia and EFL Classroom in the IT era*, Kenkyusha, Tokyo, 2001