Japanese Spoken Language Learning System Using Java Information Technology

Jouji Miwa, Hiroshi Sasaki and Kazunori Tanno

Department of Computer and Information Science, Iwate University
4-3-5 Ueda Morioka, Iwate, 020-8851, Japan

ABSTRACT

This paper describes a system for Japanese spoken language learning using Java information technology (IT). The system is automatically evaluated Japanese speech uttered by trainee as computer assisted language learning (CALL) on the Internet. For the Internet, the system can be used at any time, in anywhere and for anyone so that it is very suitable for busy learners in the world.

Phonetically, Japanese is a mora-timed language but English is a syllable-timed language. Also Japanese is a pitch-accented language but English is a stress-accented language. So many learners of Japanese as a second language can hardy hear and speak for special mora and word accent such as super-segmental feature.

Japanese speech uttered by a learner for special mora is automatically scored as 0 to 100 degree and a type of Japanese accent for speech uttered by a learner is automatically decided by the system. If learner's score is less than 50 degree and/or learner's accent type is wrong, learner must train until the score goes up with the system.

1. INTRODUCTION

For the globalization, computer assisted language learning (CALL) is important. A system of computer language learning is consists of four skills such as reading, writing, listening and speaking.

For listening, we developed a Japanese language education system for speech on an on-demand network (LESSON/J)[1]. A system for speaking had also been developed[2, 3]. The system is programmed by Java language so that the system can be used for multi-platform such as a personal computer, a work station and an information toy, and different OS such as Windows, Macintosh and UNIX. But the system for speaking was stand-alone only and couldn't process for speech on the Internet.

For the Internet, the system can be used at any time, in anywhere and for anyone so that it is very suitable for busy learners in the world. So a system of CALL with the information technology (IT) is just as important such as E-commerce. Table 1 shows a comparison between E-commerce and CALL. Java applet can not send learners speech to a CALL server because of the limitations on the security. So we have developed a new CALL system with Java servlet.

Phonetically, Japanese is a mora-timed language but English is a syllable-timed language. Also Japanese is a pitch-accented language but English is a stress-accented language.

Table 1 Comparison between E-commerce and CALL.

	E-commerce	Language Learning
Server	Web server	CALL server
Users	Customers	Learners
Products	Goods	Tests
Database	Price	Score

So many learners of Japanese as a second language can hardy hear and speak for special mora and word accent such as super-segmental feature.

In Japanese special mora, there are long vowel, double consonant and mora nasal. Japanese speech uttered by a learner for long vowel and double consonant such as [ozi:san] and [otto] which are a grand father and a husband in English, respectively.

Japanese speech uttered by a learner for special mora is automatically scored as 0 to 100 degree with the system. And a type of Japanese accent for speech uttered by a learner is automatically decided by the system. For an example, "ame" in Japanese has different meaning just to be consistent "rain" and "candy" in English. The former type of accent is 1, the later is 0. If learner's score is less than 50 degree and/or learner's accent type is wrong, learner must train until the score goes up with the system.

2. SYSTEM OVERVIEW

2.1 SYSTEM CONFIGURATION

Fig. 1 shows a schematic diagram of Japanese spoken language learning on the Internet. Evaluation of pronunciation for Japanese speech is carried out with a servlet.

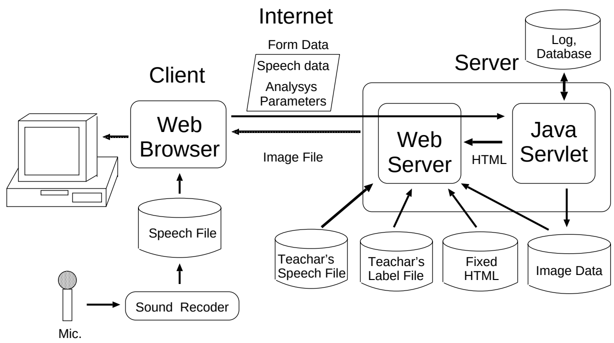

Figure 1: Block diagram of the learning system.

In the first stage, a learner saves a file of own speech wave on the local disk with an application program such as the sound recorder for Windows OS. In this stage, the format for a wave file is shown in Table 2. In the second stage, speech file is sent to Web server from a form tag in a top page. In the server, Java servlet program is processed. In the last stage, evaluated image file made with GifEncoder.class is displayed on own screen. So learner can save own evaluated image file to own computer.

In Java applet, learner's speech file could not be sent server because of the limitations on the security. But the servlet can get learner's speech file with post method. In the system, user's history and score are stored in a log file and a database.

Table 2: Wave format for the system.

Wave format	RIFF, .WAV
Sampling	11.025 kHz
No. of channel	1 channel (Mono)
No. of bits	16 bits
PCM format	Linear
Speech length	3 seconds

2.2 TOP PAGE FOR SPECIAL MORA

Fig. 2 shows an example of a top page for special mora learning. In the example, the upper side is for long vowel and the other is for germinate. In the page, learner select own speech file by a button "Browse" or a text field and send the file to a servlet by a button "Evaluate". In the example, a learning word is [oba:san] (grand mother in English) and the file is c:\tmp\obaasan.wav.

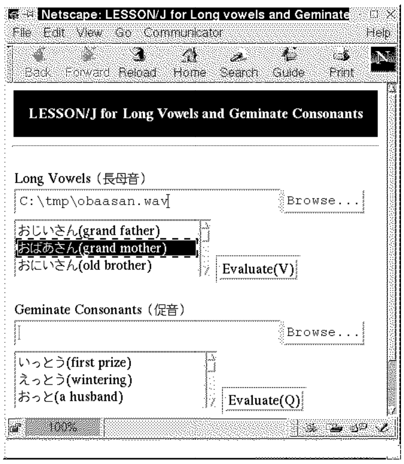

Figure 2: Top page for special mora learning.

2.3 TOP PAGE FOR WORD ACCENT

Fig.3 shows an example of a top page for word accent learning. In the page, learner select own speech file by a button "Browse" and send the file to servlet by a button "Submit". In the example, a learning word is [ame] (rain in English) and a file is c:\tmp\rain.wav. If a post-positional particle does not exist, check a radio button "None" in "Particle", otherwise check "Exist". If a learner is a male, check a radio button "Male" in "Select F0 range", otherwise check "Female". The range of fundamental frequency for "Male" is 100 to 200 Hz. The range for "Female" is 200 to 400 Hz.

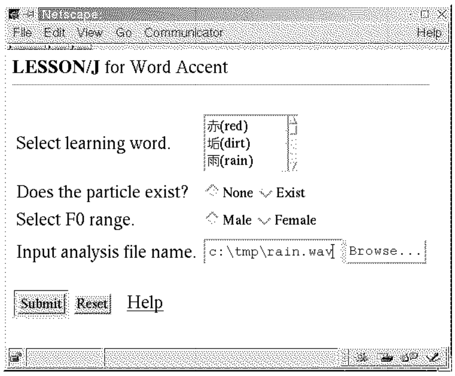

Figure 3: Top page for word accent learning.

3. CALL FOR JAPANESE SPECIAL MORA

3.1 JAPANESE SPECIAL MORA

Japanese special morae have three types such as long vowel, germinate and mora nasal. These have phonetically super-segmental features. It is difficult for foreigner to hear and speak these sounds. In the system, both long vowel and germinate are processed for learning.

An example of minimal pair in long vowels is [ozi:san] (grand father in English) and [ozisan](uncle in English). And an example of minimal pairs in germinate is [otto](husband in English) and [oto](sound in English).

3.2. PROCESSING FOR SPECIAL MORA

A processing ot the evaluation for special more is carried out the following stage.

1. Speech input
2. Estimation of word length (l_w)
3. Spectral Analysis(16 channel BPF)
4. Dynamic programming
5. Back tracking for optimal path
6. Estimation of duration of special mora using label of a teacher's sound (x)
7. Calculation of score (s_1)

Features l_w, x and score s_1 are used as follows.

3.3. SCORING FOR SPECIAL MORA

When ω_1 is special mora and ω_2 is non special mora, score s_1 for special mora is calculated based on a *posteriori* probability as follows.

$$s_1 = \frac{100 \times p(x|\omega_1)}{p(x|\omega_1) + p(x|\omega_2)} \tag{1}$$

In the equation (1), conditional probabilities for ω_1 and ω_1 are calculated as follows.

$$p(x|\omega_1) = \frac{1}{\sqrt{2\pi}k\sigma_1} \exp\{-\frac{(x-\mu_1(l_w))^2}{2(k\sigma_1)^2}\} \tag{2}$$

$$p(x|\omega_2) = \frac{1}{\sqrt{2\pi}k\sigma_2}\exp\{-\frac{(x-\mu_2)^2}{2(k\sigma_2)^2}\} \quad (3)$$

where $\mu_1(l_w)$, μ_2, σ_1, σ_2, x and k are normalization function for special mora, mean for non special mora, standard deviation for special mora, standard deviation for non special mora, estimated duration for special mora and scale factor, respectively.

For scoring, normalized mean functions by word duration l_w are as follows.

$$\mu_Q(l_w) = 0.093(l_w - 722) + 247 \quad (4)$$

$$\mu_V(l_w) = 0.154(l_w - 779) + 251 \quad (5)$$

Table 3 Means and standard deviations for special mora.

	Germinate ω_1 / ω_2	Long Vowel ω_1 / ω_2
Mean $\mu_{1,2}$	247 ms / 91 ms	251 ms / 118 ms
Standard deviation $\sigma_{1,2}$	24 ms / 21 ms	37 ms / 20 ms

3.4. RESULTS FOR SPECIAL MORA

Fig. 4 shows a scoring result for special mora [oba:san] uttered by Japanese speaker. In this case, score is 99 degree, so the utterance is good.

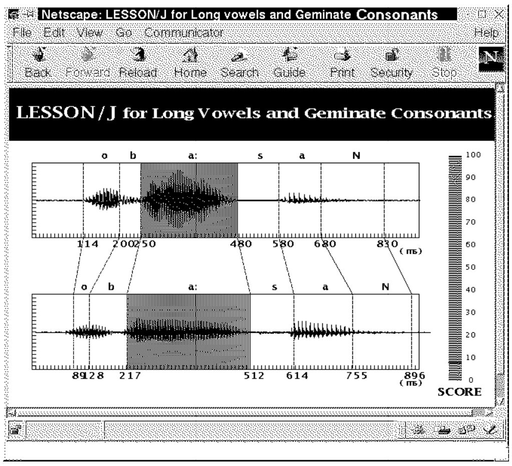

Figure 4: An example of evaluation [oba:saN].

Fig. 5 shows an evaluation for special mora [ozi:san] uttered by a foreign learner. In this case, score is 62 degree. So the learner must repeatedly train until score is up.

4. CALL FOR JAPANESE WORD ACCENT

4.1. JAPANESE WORD ACCENT

Japanese is a pitch-accented language but English is a stress-accented language. In Japanese word accent, no. of types for N morae is $N+1$. For example, no. of the patterns of fundamental frequency for two morae $N=2$ and

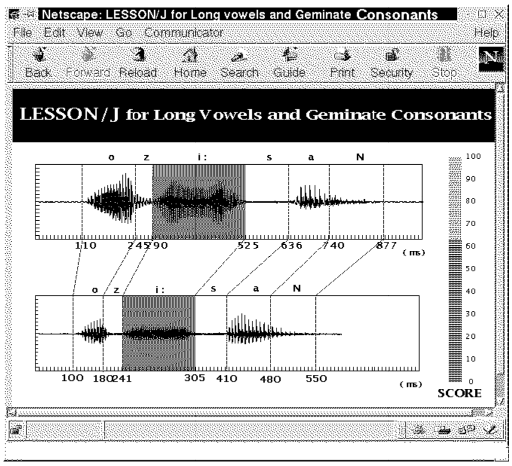

Figure 5: An example for evaluation [ozi:saN].

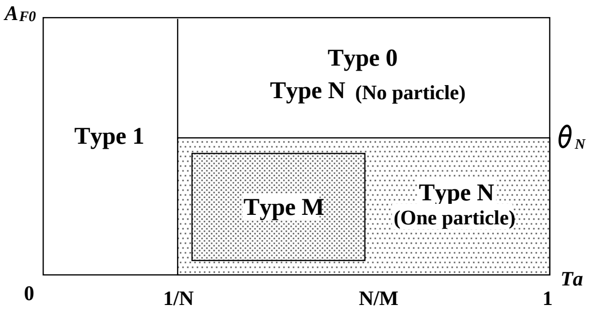

Figure 6: The discriminant function for accent types.

one particle such as "ha-si-ga" is 3 such as HL(L), LH(L) and LH(H), the types for the accent are 1, 2 and 0, respectively. In a system for accent, the type is automatically decided from learner's speech.

Fig. 6 shows discriminant functions of accent type for N mora word. T_a is relative time for accent position and the value in the system is 0 to 1. A_{F0} (octave/second) is decay for locus of logarithmic fundamental frequency. The threshold θ_N for A_{F0} in the system is -3.9.

4.2. PROCESSING FOR ACCENT

The processing of decision for accent type is carried out the following stage.

1. Speech input
2. Estimation of word end points
3. Extraction of fundamental frequency by MABC (Moving Average and Band-limitation based on Cepstrum) method [4]
4. Interpolation and smoothing for F0 locus [4]
5. Estimation T_a
6. Calculation A_{F0}
7. Decision of accent type by discriminant function
8. Display of accent type, F0, T_a and A_{F0}.

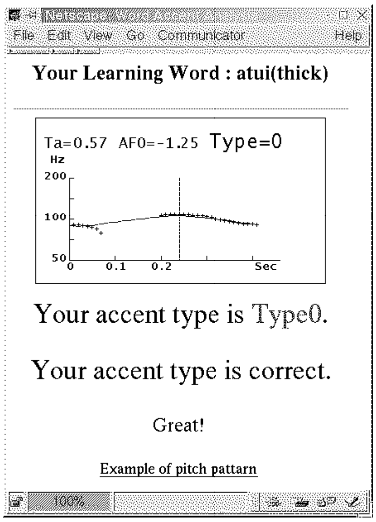

Figure 7: An example of result for Japanese learner.

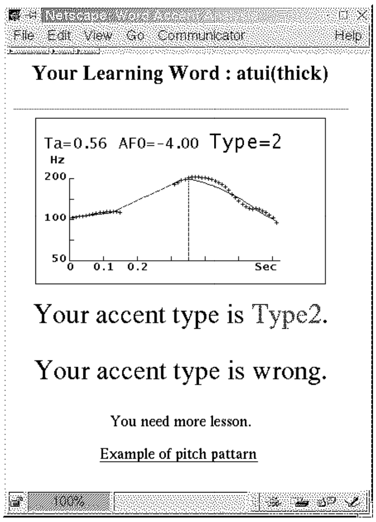

Figure 8: An example of result for Chinese learner.

4.3. RESULTS FOR ACCENT

Fig. 7 shows an example of discriminant result for speech [atui] (thick in English, type 0) uttered by Japanese student. In the case, the parameters are T_a = 0.57 and A_{F0} = -1.25, the result is correct.

Fig. 8 shows an example of discriminant result for same word uttered by Chinese student. In the case, the parameters are T_a = 0.56 and A_{F0} = -4.00, the result is wrong. The learner must train.

Fig. 9 shows the results [atui] (thick in English) uttered by a teacher and a student for comparison of acoustic parameter. In the figure, the left part is correct from teacher's speech and the right part is wrong from student's input.

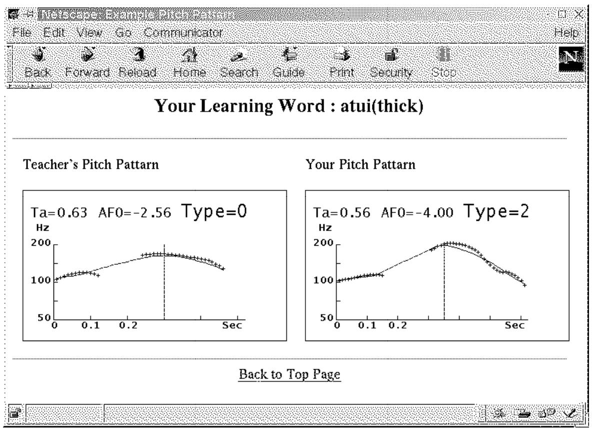

Figure 9: A comparison between teacher and learner.

5. CONCLUSION

We have developed a system for Japanese spoken language learning using Java information technology. The system is automatically evaluated Japanese speech uttered by trainee as computer assisted language learning (CALL) on the Internet. In the system, learner's pronunciation is evaluated for Japanese special mora and word accent.

For the Internet, the system can be used at any time, in anywhere and for anyone so that it is very suitable for busy learners in the world. But we must improve in order to make it user-friendliness. You will be able to access the system at the following URL.
http://sp.cis.iwate-u.ac.jp/sp/lesson/j/

ACKNOWLEDGEMENT

This project is partially supported by the grant-in-aid for scientific research(12558019) in Japan Society for the Promotion of Science.

REFERENCES

[1] J. Miwa, Y. Kumagai, J. Tian and M. Imaishi: Language Education System for Japanese Speech Using On-demand and Network, Report of Speech Technical Committee, IEICE, SP97-17, pp.55-62 (June 1997). (in Japanese)

[2] Y. Kumagai, K. Yoshida and J. Miwa: On a Decision Method of Accent Type for Japanese Learning, Report of Speech Technical Committee, IEICE, SP98-139, pp.1-8 (Feb. 1999). (in Japanese)

[3] Masato Yamamoto and Jouji Miwa: Computer Assisted Learning System for Japanese Special Mora and Its Evaluation, Trans. Tech. Comm. Psychol. Physiol. Acoust., The Acoustical Society of Japan, H2000-1, pp.1-8 (Jan. 2000). (in Japanese)

[4] S. Kato and J. Miwa: Fundamental Frequency Extraction Using Moving Average and Band-limitation in Cepstrum Method and Its Application. Report of Speech Technical Committee, IEICE, SP94-95, pp.1-8 (Feb. 1995). (in Japanese)