

JAPANESE PRONUNCIATION-TRAINING USING A PERSONAL COMPUTER

Yukie Masuko*, Hiroshi Imagawa and Shigeru Kiritani

1 Word Accent Training System

1.1 Introduction

It has been said in the field of teaching Japanese that the pronunciation drills are difficult to give in the class. First, the degree of difficulty in acquiring pronunciation is different according to the learner's mother tongue and his auditory sensitivity. The speech sound disappears as soon as it is pronounced. It is often very hard for a learner with less auditory sensitivity to capture the difference between the teacher's speech sound and his own. In addition, the goal to be achieved is vague.

For the first problem, it might be useful to show learners the auditory information in a way more understandable using some supplementary means. Thus, teachers have tried to show the quantity of the acoustic features visually to learners. Formerly, it was difficult to develop such devices mainly because of the technical restrictions. Recently, however, as the efficiency of the personal computer has improved, various types of sound signal processing can be carried out in real-time with a handy device. We made a device that analyzes and shows the fundamental frequency in real-time using a personal computer and a DSP(digital signal processor). With this device, a Word Accent Training System for learners of Japanese was developed.

1.2 Construction

The device is composed of a personal computer, DSP board, AD converter board, and DA converter board. In advance, the standard speech sounds of the words in the teaching materials are analyzed and stored by the computer as the "teacher's" patterns for learners to aim at. When the device is used, the list of the words is shown on the screen of the computer. The learner chooses one of the words, and then the pitch and the sound wave of the teacher's pronunciation of the word are shown on the screen. The screen is divided into two parts, left and right, in order to put the two paired words side by side so the learner can compare them while practicing. The learner enters his voice through a microphone, and the speech sound is converted A/D through a low-pass-filter. The pitch of the speech sound is extracted and is overlapped with the teacher's pattern. When the learner listens to the speech sound of the teacher's pattern or his own, the speech sound is converted D/A, passes through a low-pass-filter and is output through a loud speaker.

*Tokyo University of Foreign Studies

1.3 Functions

The above device has the following functions. These functions are assigned to the function keys on the keyboard.

1 The speech sound of the teacher's pattern can be heard. (Assigned to keys f1,f6)

2 The speech sound of the learner's pattern can be heard. (Assigned to keys f2,f7)

3 The speech sound of the learner's pattern is recorded. (Assigned to keys f5,f10)

4 The pattern of the learner's pitch and the sound wave is normalized and displayed. (Assigned to keys f4,f9) Normalization is the result of two processing steps:

- (1) The total duration of the word pronounced by the learner is shortened/lengthened to fit the teacher's.
- (2) The axis of the fundamental frequency of the learner's pronunciation is shifted upward/downward so that the pitch contour of the learner's overlaps with that of the teacher's.

5 The list of words is shown. (Assigned to keys f3,f8) The learner types in the number of the word he wants to learn and presses the CR key. Then, the teacher's pattern is displayed in red. ROLL UP/ROLL DOWN key can scroll up/down the list of the words.

6 The right arrow key overlaps the teacher's pattern on the left part of the screen on the right part. The left arrow key overlaps the teacher's pattern on the right part of the screen on the left part. The downward arrow key erases an overlapped teacher's pattern.

1.4 Teaching Materials

Japanese word accent is a pitch accent and is different from stress accent, such as occurs in English. In Tokyo dialect, the pattern where a first high mora is followed by low morae is distinguished from that where a first low mora is followed by high morae. In the latter, patterns are further distinguished by how many morae are high continuously. In Tokyo dialect, the accent patterns are distinguished by the position of the mora with a falling tone, called the accent nucleus. Therefore, a word with n morae has $n+1$ patterns of accent. Words composed of the same segmental features can have different meanings due to different accent patterns. (For example ame:rain, but ame:candy) In order to master these accent patterns, it is necessary to control the pitch contour independent of the segmental features and the stress.

Our teaching materials are composed of nonsense words and sense words. The nonsense words are used to develop sensitivity

to pitch contour related to the perception of mora. Using two-mora words, the learner can practice the rise or fall in the pitch contour. Using three-mora words, the learner can practice maintaining pitch contour high/low for two morae, or rising in the second mora and falling in the third mora. Words are prepared with two speeds: slow and fast. Real pairs of words, composed of the same segmental features, but with the different accent patterns, are also included. They have different meanings because of their different accent patterns. Two-mora words and three-mora words are included as materials to expand the pitch contours mastered by using nonsense words.

The list of all the words is shown in Table 1. The accent patterns are shown using lines over/under the letters.

Table 1 Two-mora and Three-mora Word Pairs

two-mora pairs		three-mora pairs	
<u>a</u> ka	āka	a <u>o</u> i	a <u>o</u> i
<u>a</u> me	āme	i <u>t</u> a <i>i</i>	i <u>t</u> a <i>i</i>
<u>u</u> mi	ūmi	ka <u>sh</u> ly <i>a</i>	ka <u>sh</u> ly <i>a</i>
<u>sh</u> i <u>r</u> o	shīro	ha <u>r</u> er <u>u</u>	ha <u>r</u> er <u>u</u>
<u>ts</u> u <u>y</u> u	tsūyu	sh <u>i</u> me <u>r</u> u	sh <u>i</u> me <u>r</u> u
<u>t</u> o <u>s</u> hi	tōshi	ka <u>k</u> er <u>u</u>	ka <u>k</u> er <u>u</u>
<u>k</u> a <u>u</u>	kāu	i <u>sh</u> oku	i <u>sh</u> oku
<u>k</u> i <u>r</u> u	kīru	sh <u>i</u> ga <i>i</i>	sh <u>i</u> ga <i>i</i>
<u>s</u> a <u>k</u> u	sāku	ka <u>d</u> an	ka <u>d</u> an
<u>f</u> u <u>r</u> u	fūru	i <u>i</u> dou	i <u>i</u> dou
<u>n</u> i <u>l</u> wa	nīwa	ka <u>sh</u> ima	ka <u>sh</u> ima
<u>t</u> e <u>n</u>	tēn	ka <u>t</u> ou	ka <u>t</u> ou
		de <u>n</u> ki	de <u>n</u> ki
		ke <u>i</u> ki	ke <u>i</u> ki
		ko <u>sh</u> ou	ko <u>sh</u> ou
		sa <u>t</u> ou	sa <u>t</u> ou
		ni <u>h</u> on	ni <u>h</u> on
		ka <u>e</u> ru	ka <u>e</u> ru

1.5 Results

Using the system, the instructor sits next to the learner and appropriately explains the pitch contour displayed on the screen and shows the points to be improved. The instructor first says that the learner's pattern should be close to that of the teacher's. After the learner has the meaning of the pitch contour, it is easy for the learner to imitate the teacher's pattern. The learner can reach a stable similarity after about ten trials. When the learner gets accustomed to the display, he can find out the points to be improved by himself.

Three examples are shown in Figs.1,2 and 3. The left portion

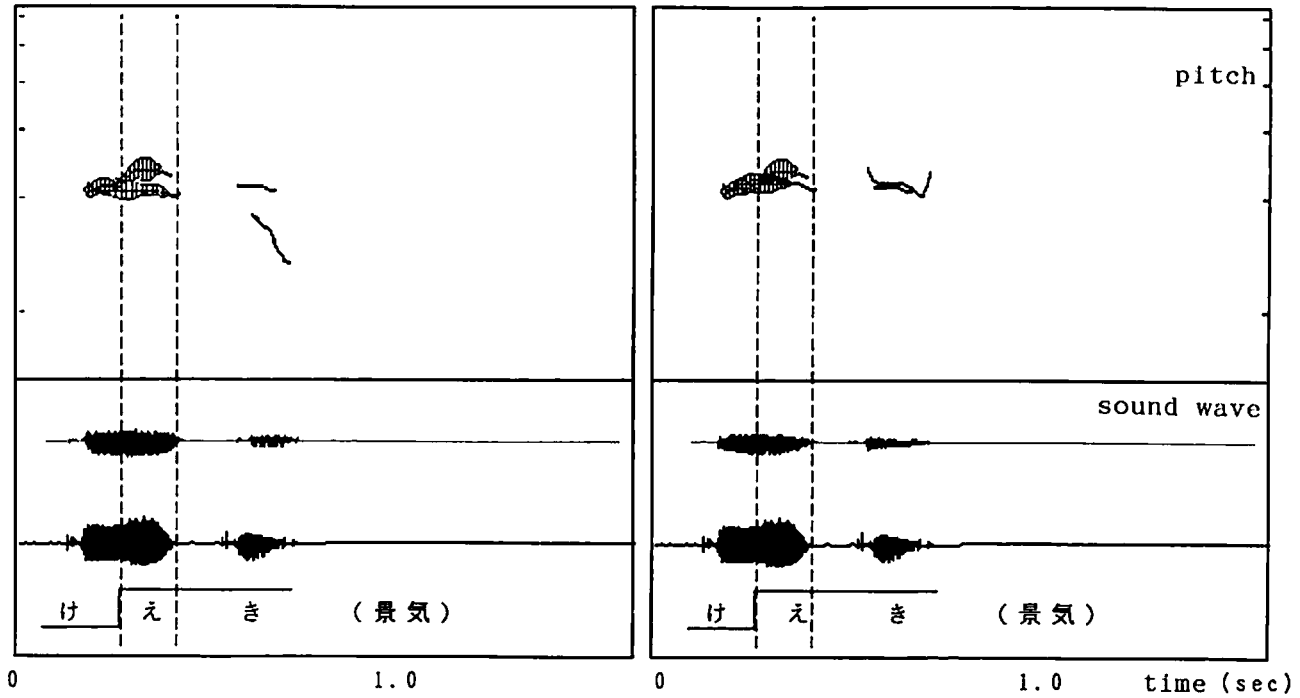


Fig.1 The two sound waves are in the lower half of the screen. The lower is the teacher's, and the upper the learner's. The two pitch patterns are in the upper half of the screen. If they overlap, the learner succeeded in imitating the teacher's pattern. The pitch contour should be kept high for the two last morae, but the learner tended to fall from the second mora to the third. At the right of the screen, the learner succeeded, and the pitch contour overlaps.

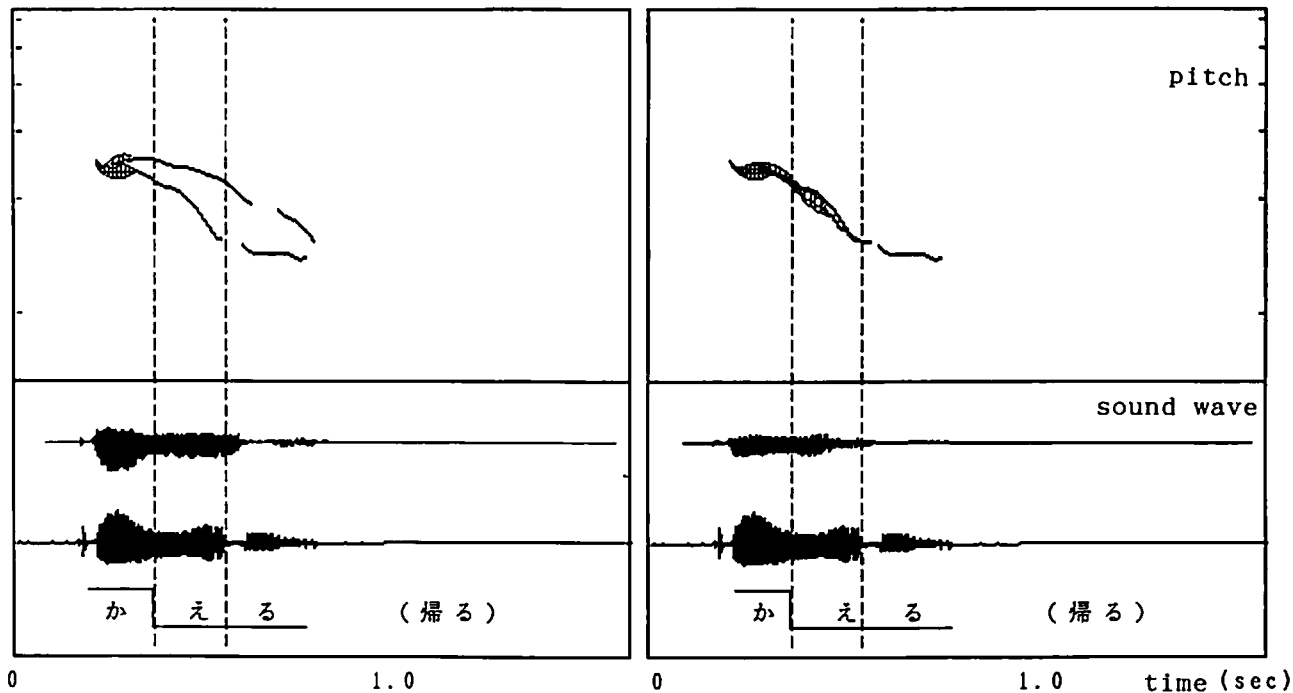


Fig.2 The two sound waves are in the lower half of the screen. The lower is the teacher's, and the upper the learner's. The two pitch patterns are in the upper half of the screen. If they overlap, the learner succeeded in imitating the teacher's pattern. The pitch contour should begin falling in the middle of the first mora and reach bottom in the third mora. However, the learner's pitch contour falls slowly. At the right of the screen, the learner succeeded, and the pitch contour overlaps.

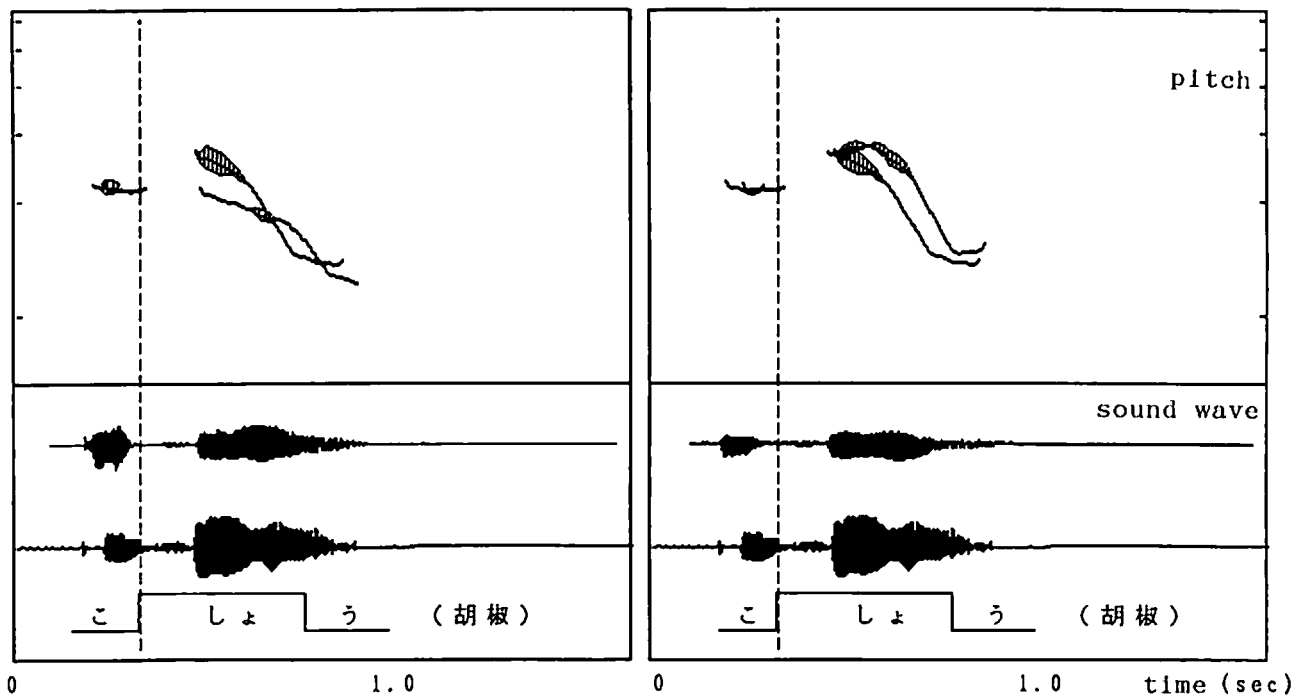


Fig.3 The two sound waves are in the lower half of the screen. The lower is the teacher's, and the upper the learner's. The two pitch patterns are in the upper half of the screen. If they overlap, the learner succeeded in imitating the teacher's pattern. The pitch contour should rise from the first mora to the second, and in the middle of the second mora it should begin to fall rapidly. However, the learner fails to raise the pitch and also cannot make it fall rapidly enough.

is the trial a learner has failed and the right one where he succeeded. The two sound waves are in the lower half of the screen; the lower is the teacher's, and the upper the learner's. The two pitch patterns are in the upper half of the screen. If they overlap, the learner has succeeded in imitating the teacher's pattern. In Fig.1, the pitch contour should be kept high for the two last morae. But the learner tended to fall from the second mora to the third. After about ten trials, she succeeded in keeping the pitch contour high to the third mora, and the contour overlapped with that of the teacher's. In Fig.2, the pitch contour should begin to fall in the middle of the first mora and reach bottom in the third mora. But the learner's pitch contour falls slowly. After about ten trials, she succeeded in falling rapidly. In Fig.3, the pitch contour should rise from the first mora to the second, and in the middle of the second mora it should begin to fall rapidly. But the learner fails to raise the pitch higher and also, cannot fall rapidly. After about ten trials, she succeeded in raising the pitch, but was sometimes late in beginning to fall.

After using this system, the learner became able to catch the pitch and imitate the accent patterns quite correctly. Thus, this system is thought to be efficient for learners who cannot catch pitch independently.

1.6 Problems

This learning system using has three problems. The first is that an effective procedure for learning must be investigated to make this system a "self-teaching system". Basic data have been gathered for this purpose. The second is how the effect of this system lasts without the visual information must be examined. The procedures for learning must be adjusted according to the result. The third is that the use of this device should also be extended to the accent of connected words and the intonation of sentences. More data will have to be gathered.

2 Listening Exercise System

2.1 Introduction

When learning a foreign language, difficulty in mastering pronunciation is caused mainly by the following two factors. One is that the speech sounds used in one language are a little different from those used in another language. The other is that the context for one speech sound to appear is often restricted in a language. A careful comparison must be made between the learner's mother tongue and the foreign language from the viewpoint of phonemics, the distribution of allophones and phonotactic rules. The learner must re-acquire speech sounds in strange contexts. The more the learner is exposed to the speech sounds of the language in the course of the exercise, the more effective the exercise for pronunciation and listening becomes. However, much time cannot be spent on drills of pronunciation and listening in

the class. Tapes made for learners of Japanese can provide the various speech sounds of Japanese outside the class.

Recently a system with these teaching materials was produced experimentally for listening exercises using the device mentioned above. In the listening test, the speech sounds are output at random and the learner can check which speech sounds are hard to distinguish for him. The learner can choose the speech sounds which are hard to distinguish and freely make a set to practice intensively. As the results of a test can be printed out easily in the form of a confusion matrix, data are easily be collected.

2.2 Functions

In the listening exercises, the speech sounds focused on must be indicated clearly. If features other than those focused on change simultaneously, the learner might be puzzled about which features he should pay attention to. One set of words consists of three to six words whose compositions are all the same, except one, which is easily confused with the others. The speech sounds of each word are in advance analyzed and stored by the computer. In each set, the learner uses the functions below.

1 All-Word Test

In this test, all the words are output at random. One word has twelve tokens pronounced by four speakers three times each. The number of repetitions can be chosen by the learner. The choice of #1 outputs twelve sounds (three times by four speakers) per word. The results can be printed out in the form of a confusion matrix if function #4 is chosen.

2 Selected-Word Test

In this test, all the words the learner chooses are output at random. The number of repetitions can be chosen by the learner. The results can be printed out in the form of a confusion matrix if function #4 is chosen.

3 Listening

This is only a listening exercise. The words the learner chooses are output in the order he chooses. The set of words can be heard again and again.

4 Print Out

The result of function #1 and function #2 are printed out in the form of a confusion matrix.

5 Other Exercises

The learner can choose another set of exercises.

2.3 Teaching Materials

One set consists of three to six words. The words in a set are composed of the same features except one. The features focused on are thought to be confused. Ten to fifteen sets are prepared for an exercise involving one feature. The sets are called

"groups" here. As the difficulty is different according to the learner's mother tongue, some features may not be confused by some learners, and thus may not be necessary. However, such groups could be used as screening tests to help the learner understand which sounds and which features are difficult for him. The features focused on and the sets of words are shown in Table 2. Groups #1, #2 and #4 are divided into two parts, one part has words whose first feature is different, the other has words whose focused feature is not in the first position but another. The words of each set are pronounced by four Japanese speakers of Tokyo dialect, two males and two females, each of whom pronounces each word three times. (That is, each word has twelve tokens.)

2.4 Present Situation

This system is now in trial use for learners of Japanese mainly from Southeast Asia. Basic data are being gathered on difficulties caused by their mother tongue and on how great these difficulties are.

Table 2 Sets of Words

Group #1 Dental Consonants

	[n l d dz t (z nd ndz)]						
Set #1	raikou	dalkou	naikou	zaikou	taikou		
Set #2	rouka	douka	nouka	zouka	touka		
Set #3	raimei	daimei	naimel	zaimel	taimei		
	:						
Set #8	karou	kadou	kanou	kazou	kandou	kanzou	
Set #9	irou	idou	inou	izou			
Set #10	furou	fudou	funou		fundou		
	:						

Group #2 Voiceless Fricatives and Voiceless Affricates

	[ç i ç i ç u s u t s u t ç i]						
Set #15	hiku	shiku	shuku	suku	tsuku	chiku	
Set #16	hiru	shiru		suru	tsuru	chiru	
Set #17	hiki	shiki	shuki	suki	tsuki	chiki	
	:						
Set #23	shihi	shishi	shishu	shisu		shichi	
Set #24	hihi	hishi	hishu	hisu			
Set #25	kahi	kashi	kashu	kasu	katsu	kachi	
	:						

Group #3 Mora Phonemes

	/N//H//Q/					
Set #30	kenshou	kesshou	keishou	keshou		
Set #31	konka	kokka	kouka	koka		
Set #32	kemmei	ken-ei	keimei	keiei		
	:					

Group #4 Voiced/Voiceless Plosives

	[p b t d k g]						
Set #47	pasu	basu		dasu	kasu	gasu	
Set #48	pakku	bakku	takku	dakku		gakku	
Set #49	pen	ben	ten	den	ken	gen	
	:						
Set #53	zenpou	zenbou	zentou	zendou	zenkou		
Set #54	senpai	senbai	sentai	sendai	senkai	sengai	
Set #55	kinpen	kinben	kinten		kinken	kingen	
	:						

Group #5 Nasals

	[m n ŋ k kk ŋŋ ŋk]						
Set #63	gogo	gono		goko	gokko	gongo	
Set #64	age	ane	ame	ake	akke		
Set #65	kagi	kani	kami	kaki	kakki	kanki	
	:						

Group #6 Devoiced Vowels

Set #67	ikimasu	kimasu	ukimasu	shikimasu	hikimasu	fukimasu	
Set #68	utsuru	tsuru	suru	shisuru	yusuru		
Set #69	kikuchi	hukuchi	hikichi	shikichi			
	:						