The Influence of Model Sounds on the Speech Production of Japanese Learners of English

Midori Iba

Abstract

It is too complicated to teach supra-segmental phonology, namely rhythm, intonation and the stress of a foreign language analytically. One tool for learning these may be the frequent imitation of model sounds. In creating authentic software or a textbook of English pronunciation training, voices of some native speakers of the language are usually recorded as models. In cooperation with staff members at Konan University, I am developing computer-based listening and pronunciation software which is designed to improve learners' ability to sense and monitor some prosodic features of English. Before practicing with the software, 20 students had taken a pretest. We have already created another set of software for the pretest which was designed to record students' voices. Ten sentences out of forty were recorded without model sounds. The same sentences were also recorded after they listened to model sounds. The aim of this study is to compare students' own creative reading sounds with their imitating reading sounds. To investigate which element of speech sounds changed or did not change is especially important when creating aural/oral training software. The fundamental frequencies of each sentence and the duration of some target consonants were measured by speech analyzers. Auditory impressions of two native teachers of English and two Japanese teachers of English were also examined. Fundamental frequencies apparently changed after listening to model sounds as predicted while some consonant confusion in voicing and manner was not reduced. Both groups of teachers observed that students' pronunciation had improved in most cases.

1. Introduction

1.1. Stress-timed VS mora-timed?

One of the difficulties in learning a foreign language is in mastering the prosodic aspects of the target language. For example, a native Japanese speaker generally has difficulty with stress and timing when speaking English. On the other hand, it is widely known that for a native English speaker it is hard to learn the mora-timing of the Japanese language. These two are very different languages, using different sounds in different ways. The stress-timed rhythm of the English utterance with the related obscuration of weak syllables is the prime distinguishing feature of the language's pronunciation. As Gimson (1981) noted, "For all learners, accentuation must provide the foundation of which any pronunciation course is built." The two languages differ greatly in prosodic realization, including temporal organization. The typological categorization of timing systems is rooted in the idea that temporal organization is based on some unit of timing, and Japanese is said to be mora-timed, whereas English is stress-timed (Dauer, 1983). In English sentences, the duration is alternately stretched and reduced, interacting with the other two correlates, namely, fundamental frequency and amplitude. The unit of English timing is the stress foot, that is, a string containing a stress accent followed by zero or more unstressed syllables. On the other hand, the unit of Japanese timing is the mora, a syllabification unit. As Ueyama (1996) noted, the duration of each mora is equal, abstracting away from the phrase-final lengthening, and the prosodic distinctions of Japanese are mainly conveyed by fundamental frequency.

However, some researchers are against the concept of stress-timing. Ladefoged (1982) describes it as only a 'tendency.' Roach (1991) considers that if stress-timing operates at all, it only occurs in very regular, formal speech. Regarding teaching pitch movement, Jenkins (2000) is doubtful. "Even if it were possible to teach pitch in the classroom, I do not believe that the use of 'native speaker' pitch movements matters very much for intelligibility in interactions among NBESs[1]." Nevertheless, I take a different view of teaching prosodic features. Even if we don't 'teach' them in a strict sense, we should demonstrate them to learners otherwise they may have no other chance to recognize them.

The audio-lingual method is less popular today because of its reliance on drills and habit-formation. Structural patterns in dialogues about everyday situations are imitated and drilled. These monotonous tasks are less interesting to learners who feel the need for more creative work in speech productions. More humanistic approaches are of course welcome but teaching pronunciation has a different story. To acquire the sound system of a different language, imitation and repetition of the target language are necessary.

1.2. Purpose of the Study

In this research, I have investigated how learners' utterances would differ after they listened to the model sounds. In 2003 I created and uploaded new software which I am planning to revise. To create appropriate software, careful research is needed to assign tasks to it. The pre-listening sounds and post-listening sounds were compared with model sounds. The duration of sentences, some selected consonants, and phrase-final vowels, and fundamental frequency were measured by a computer speech analyzer.

The auditory impressions of four professional English teachers were also examined by way of subjective evaluation. The results will be compared with computer assessment and discussed later.

My research questions in this study were as follows:

1. Is there any difference in the speech production of subjects after they listened to model sounds and reproduce the same sentences? The following items will be examined:
   - Fundamental frequencies
   - Duration of sentences
   - Duration of some consonants（fricatives）
   - Duration of phrase-final vowels

2. Do the results of subjective evaluation relate to those of the figures obtained through experiment?

2. Experiment

2.1. The materials

Test materials were comprised of eight sentences and two minimal pairs of words. In the pairs, the consonants /f, h, z, dz/ were embedded. The consonant /r/ and /l/ were inserted in the same context to ascertain whether or not the subjects already knew the difference between these two consonants before the test, and after the test whether or not their pronunciation changed（See Table 1, No.6 and No.7）. The voiceless dental fricative / θ / or the voiced form of the same consonant /ð/ were embedded because these two are unique to English sounds（See Table 1, No.1 to No.4, No.6, and No.8）The main purpose of the study is not to compare the isolated consonants but to focus on the change of prosodic features. Yet, if any changes were found in the above-mentioned consonants, it would be of more interest because no instruction in pronunciation was given to the subjects.

1	He went over the path.
2	They thought about it.
3	We've fired them.
4	Repeat the word.
5	I knew it was wrong.
6	He didn't collect the papers.
7	He didn't correct the papers.
8	Put all these things in the bag.
9	food / hood
10	cars / cards

Table 1. The test materials

2.2. Speaker

A male speaker of standard British English recorded the test items.

2.3. Subjects

The 20 subjects（12 female, 8 male）who participated in the experiment were native speakers of Japanese spoken in the Kansai area and students at Konan University in Kobe. They belonged to different faculties of the university. No one had spent more than two months in an English speaking country. Their ages ranged from 19 to 22. They reported normal hearing and vision.

2.4. Procedure

At the moment, our computer-based listening and pronunciation software is still under development, nevertheless a computer-based pretest[2] for this software was made and given to the subjects. In this pretest, forty sentences of each subject were recorded. The first twenty sentences were recorded without model sounds and the latter twenty were recorded after listening to model sounds.

The pretest was performed in a CALL room at Konan University. At the beginning of the pretest, I taught the students how to use the software.

From the first page to the 20th page, the subject was asked to read the text in the top box and record his/her voice by clicking the icon as seen in Figure 1. This pattern lasted until No.20. Then from No.21 to No.40, the subject listened to the model sound first and then recorded his/her voice trying to repeat the sounds in the same way as seen in Figure 2.

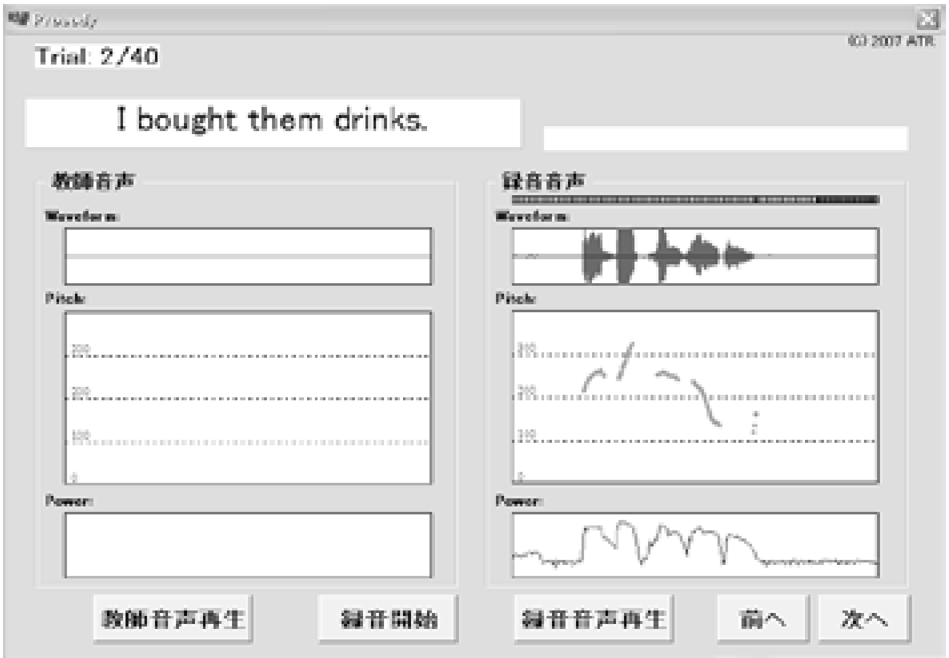


Figure 1. The second page of the pretest. In trial 1 to 20, if the icon（教師音声再生）were clicked, no voice would be heard. The subject was asked to read the sentence freely.

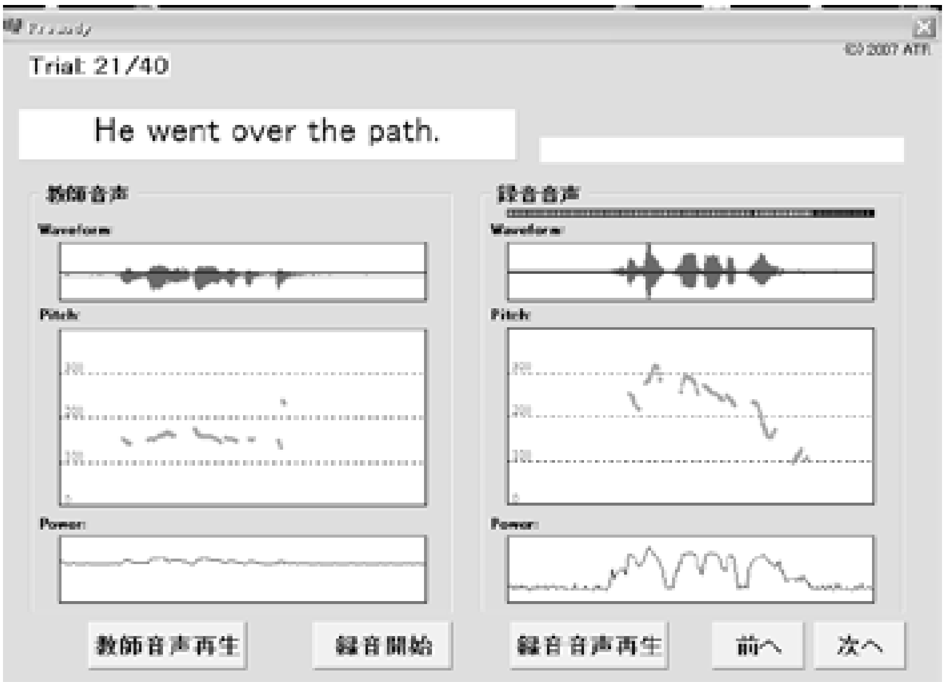


Figure 2. The interface of No.21.

In the test sentences, No.11 to 20 and No.21 to 30 were identical.（See Appendix）These twenty sentences were analyzed in this study. The voices of the subjects were recorded and stored in the computer as the WAV format. The WAV files were analyzed by a speech analyzer, WaveSurfer[3].

3. Results

3.1. Fundamental frequencies — Pitch analysis

As Ladefoged（2003）says, when discussing the pitch of the voice, it can usually be said to be the rate at which vocal fold pulses recur, and thus the fundamental frequency of the sound wave. Tone and intonation are manifested by pitch. You cannot literally measure the pitch of a recorded sound but you can measure the fundamental frequency of the sound wave.

Figure 3 shows the fundamental frequencies of the sentence “He went over the path” pronounced by the model speaker. There are clearly four features in his utterance as shown in the second panel from the top of Figure 3:（1）the sentence is divided into three parts with two pauses（①hɪwɛnt②/oʊvɚðə③pæθ/), with the main stress occurring on ‘went’ ;（2）the intonation pattern is ‘middle-fall’ ;（3）in ②/oʊvɚðə/, the two vowels /ɚ/, and/ə/ are relatively short;（4）the last vowel /æ/ is relatively long.

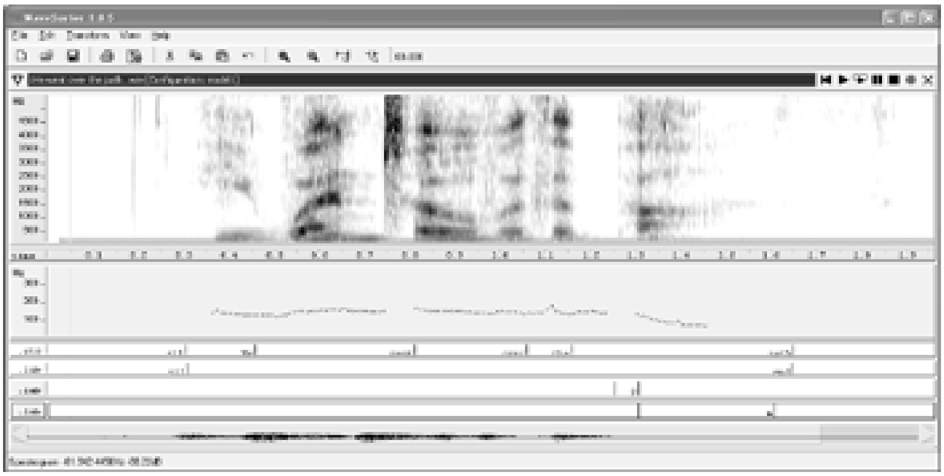


Figure 3. The sentence “He went over the path.” Pronounced by a model speaker.

The third feature indicates the stress-timed shortening, a typical effect in English speech. The fourth feature also shows the well-known phenomenon in English of phrase-final lengthening,[4] The stress-timed shortening and phrase-final lengthening are indications of the tendency toward isochronous spacing of prosodically strong syllables. Consequently, in ②, the vowels /ɚ/ and /ə/ are compressed in order to make the overall duration of the phrase “over the” closer to that of the contrasting monosyllable word “path” in ③.

Figure 4 shows a typical example of Japanese students.

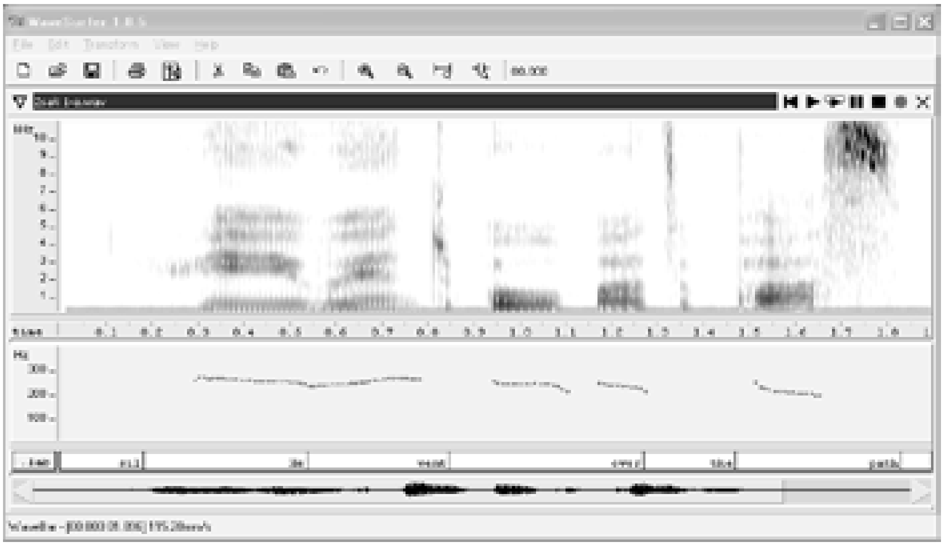


Figure 4. The same sentence as shown in Figure 3. It is pronounced by a Japanese student.

In her speech, the fundamental frequencies show that②/oʊvɚðə/ is not divided as one segment but as two segments.

As for the comparison of fundamental frequencies between the model sound and 20 subjects, I counted the number of segments to look for any difference in segmentation between the before-listening model sounds and the after-listening model sounds. If the model speaker’s reading of “He went over the path” is divided into three segments (as mentioned before), how are the results of the subjects? The following is the results of the segmentation. If a subject pronounced the sentence just as the model speaker, that would be counted as Segment1=1, Segment2=1, Segment3=1. If he pronounced the sentence differently such as //He/went// //over/the// //path//, the segmentation would be counted as Segment1=2, Segment2=2, Segment3=1.

	Segment1	Segment2	Segment3			
MODEL	1	1	1			
	Before			After		
	Segment1	Segment2	Segment3	Segment1	Segment2	Segment3
S1	1	2	1	1	1	1
S2	1	3	1	1	2	1
S3	1	2	1	1	3	1
S4	1	3	1	1	3	1
S5	2	3	1	1	3	1
S6	2	3	1	2	3	1
S7	1	3	1	1	3	1
S8	2	3	1	1	2	1
S9	1	2	1	1	1	1
S10	1	3	1	1	1	1
S11	1	2	1	1	1	1
S12	1	3	1	1	1	1
S13	1	3	1	1	1	1
S14	1	3	1	1	2	1
S15	1	3	1	1	2	1
S16	1	1	1	1	1	1
S17	1	3	1	1	1	1
S18	1	3	1	1	3	1
S19	1	2	1	1	2	1
S20	1	1	1	1	1	1
Mean	1.15	2.55	1	1.05	1.85	1

Table 2. The results of the segmentation of 20 subjects.

Table 2 shows the results of counting the number of segments by 20 subjects. As for segmentation, there was no difference between “Before” and “After” in Segment 3. There was not much difference in Segment 1. Yet, in Segment 2, there was some difference. To make the difference more obvious, mean averages were calculated as in Table 3. Figure 5 is a graph based on Table 3.

	Segment1	Segment2	Segment3
before	1.15	2.55	1
after	1.05	1.85	1

Table 3. The mean averages of segmentation of 20 subjects.

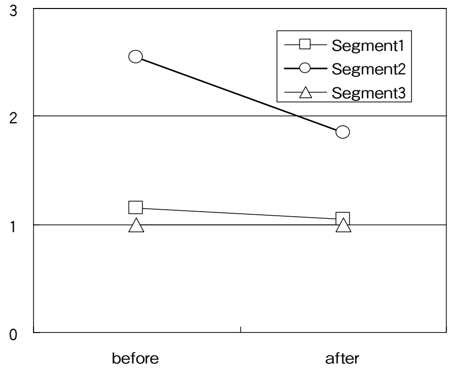


Figure 5. Graphic form of Table 3. The comparison of mean averages of the number of segments between “before-listening” and “after-listening.”

As indicated in Table 3 and Figure 5, the segmentation in Segment 2 reduced appreciably. A possible reason for this phenomenon is that the Japanese subjects tended to separate into two or three parts when they pronounced “over the” because they are affected by the timing system of their first language, but they could change their pronunciation after they listened to the English model sounds. If they can imitate model sounds, they will read “over the” in a weak and fast way like the model. I predicted the results of Segment 2 but as for Segment 1, the results were interesting and I had not predicted them. I had thought most subjects would pronounce Segment 1 in two parts as “He/ went” before they listened to the model reading, yet only three subjects out of 20 did. In reality, most students pronounce it as one segment. In this case, we could say that most of the subjects are not so strongly influenced by their first language. Comparing Segment 1 with Segment 2, both segments consist of two words but there was a variety of segmentation in Segment 2. Perhaps because the word “over” in Segment 2 is a two-syllable word, the total number of syllables in Segment 2 is three whereas the total number of syllables in Segment 1 is two. There might be a considerable difference between two syllables and three syllables in one rhythm group in perception and production.

3.2. Duration of sentences

	Sentence1		Sentence2		Sentence3		Sentence4	
MODEL	1685		1158		1068		1101	
	Before	After	Before	After	Before	After	Before	After
S1	1297	1763	1206	1117	1112	1217	1100	1133
S2	1587	1562	1320	1108	1983	1412	886	1007
S3	1519	1668	1239	1369	937	1026	885	1198
S4	1778	1565	1580	1185	1292	1203	998	1099
S5	1424	1479	1252	1059	949	950	949	1090
S6	1605	1570	1326	1162	1407	1110	964	1152
S7	1475	1379	1170	1200	1209	1057	865	916
S8	1600	1869	1467	1100	1270	1225	1026	1039
S9	1920	1361	1261	1171	1214	1229	938	972
S10	1492	1716	1280	1195	1218	1236	902	1077
S11	1541	1667	1132	1131	1126	1145	976	1135
S12	1798	1535	1414	1066	1045	1101	984	942
S13	**2584**	**1604**	1613	1311	2117	1415	2565	1030
S14	1468	1323	1681	1160	1399	1275	1006	1095
S15	1404	1333	1217	1307	1141	1057	1170	966
S16	1617	1491	1616	1459	1506	1442	1104	1228
S17	1805	1546	1637	1375	1497	1201	999	978
S18	1404	1461	1292	1263	1123	1142	928	944
S19	1561	1464	1103	1062	1174	937	957	900
S20	1437	1467	1663	1238	1046	1159	898	1065
Mean	1615.8	1541.15	1373.45	1201.9	1288.25	1176.95	1055	1048.3

Table 4. The duration of Sentence 1, 2, 3 and 4.

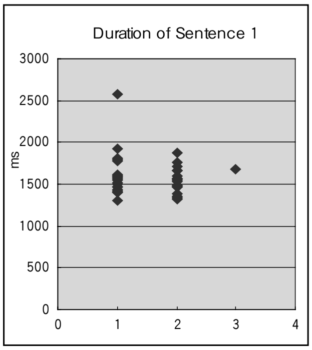
Table 4 and Table 5 show the results of measuring the duration of eight sentences pronounced by 20 subjects. After listening to model readings, how did the subjects change the duration of the sentences? The eight graphs in Figure 6 were made according to Table 4 and Table 5. On the x-axis, figure 1 means "before-listening-to-model-sound (BL)," figure 2, "after-listening-to-model-sound (AL)," and figure 3, "model-sound (M)." The graphs show the distribution of the sentence duration before/after listening.

These graphs visually indicate that the duration of the sentences pronounced by the subjects were influenced by model sounds. There is a tendency in BL sentences for the pronunciation to be rather longer than the model's. AL sentences tend to be shorter than BL sentences and to approximate M sentences. There were

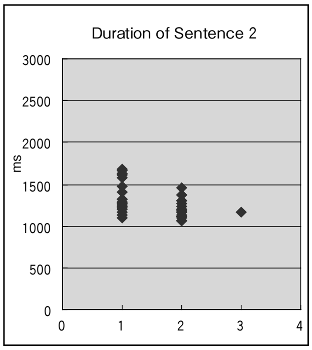
	Sentence5		Sentence6		Sentence7		Sentence8	
MODEL	1118		1608		1761		1683	
	Before	After	Before	After	Before	After	Before	After
S1	1236	1189	2117	1940	1791	1912	1697	1694
S2	1419	1385	1861	1941	1993	1948	2065	2104
S3	1256	1181	2325	2014	2226	1968	1976	1830
S4	1822	1464	2265	2193	2630	1970	1977	1785
S5	1135	1313	1752	1599	1528	1889	2082	1971
S6	1379	1130	2187	2138	2476	2232	2700	2142
S7	1477	1419	2375	1831	2782	1889	3126	1930
S8	1179	1228	1783	1666	1990	1842	2029	1790
S9	1200	1176	1796	1909	1908	1914	2126	1883
S10	1288	1215	1784	1901	1877	1910	2364	1923
S11	1440	1120	1912	1999	2136	1927	2383	2088
S12	1824	1445	1587	1625	1531	1965	3524	1849
S13	1732	1225	2667	1864	2367	2003	3554	2583
S14	1308	1440	1948	2271	2162	1806	1918	1627
S15	1302	1450	2113	1810	2216	1753	2223	1939
S16	1373	1402	2089	1994	2399	2113	2262	2151
S17	1514	1391	1993	2133	1888	2094	2159	2104
S18	1177	1182	1899	1865	2030	1724	2128	1691
S19	1388	1285	1739	1635	1720	1569	1872	1640
S20	1250	1360	2054	2180	2101	2260	2744	1951
Mean	1384.95	1300	2012.3	1925.4	2087.55	1934.4	2345.45	1933.75

Table 5. The duration of Sentence 5, 6, 7 and 8.

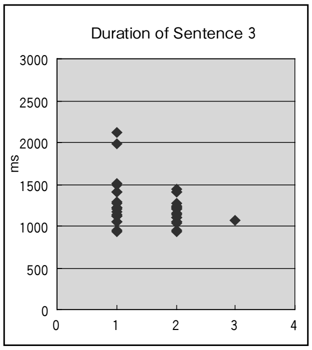
some sentences in BL which were far longer than the models, such as the highest dot（duration = 2584 ms）in Duration of Sentence 1 in Figure 6. This phenomenon shows the subject's uncertainty in pronunciation. After he listened to the model sound, the duration of the same sentence produced by him became 1604 ms which was similar to the duration of the model sound（1685）.



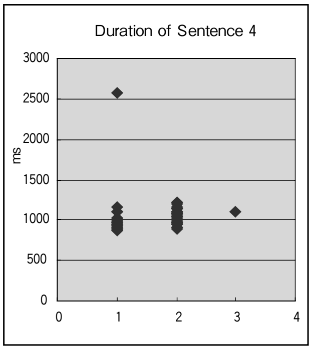
1. Duration of Sentence 1



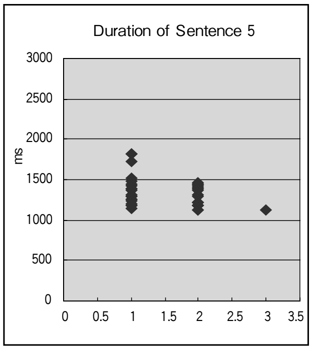
2. Duration of Sentence 2



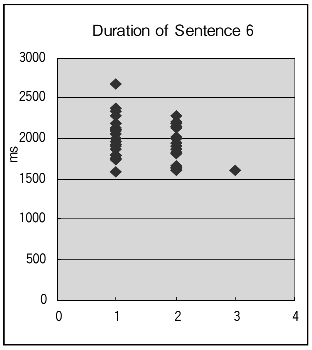
3. Duration of Sentence 3



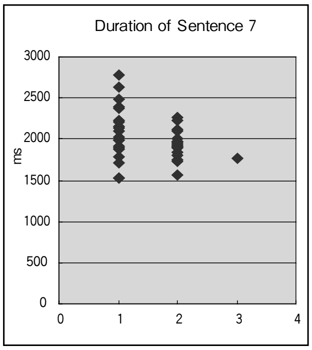
4. Duration of Sentence 4



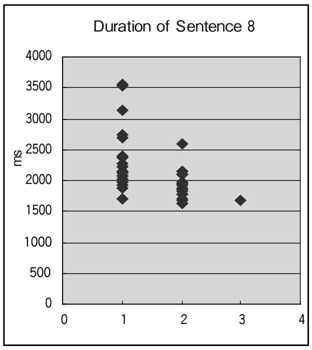
5. Duration of Sentence 5



6. Duration of Sentence 6



7. Duration of Sentence 7



8. Duration of Sentence 8

Figure 6. These eight graphs show the comparison among the duration of BL sentence, AL sentence and M sentence. On an x-axis, figure 1 means “before,” figure 2 “after,” and figure 3 “model.” The unit of the y-axis is millisecond (ms).

3.3. Duration of consonants (fricatives)

The following are the results of measuring the duration of some consonants that appeared in the test sentences. Fricatives were chosen for observation because they are conspicuous in the sound spectrogram.

	Sen.2 θ in thought		Sen.3 f in fired		Sen.8 ð in these	
MODEL	141		176		72	
	Before	After	Before	After	Before	After
S1	122	126	199	177	44	96
S2	95	120	217	187	97	70
S3	118	164	74	163	35	59
S4	186	163	125	188	57	64
S5	121	124	132	186	98	99
S6	140	170	162	142	41	70
S7	105	116	140	147	101	76
S8	211	147	226	197	62	40
S9	120	179	150	134	87	70
S10	117	191	148	166	50	94
S11	105	131	159	174	42	80
S12	152	117	135	125	157	158
S13	170	127	165	170	59	87
S14	141	109	108	120	26	93
S15	92	137	131	152	55	45
S16	42	133	154	198	64	63
S17	158	148	230	164	32	88
S18	104	115	132	125	20	54
S19	63	83	93	105	79	71
S20	73	141	111	189	36	53
Mean	121.75	137.05	149.55	160.45	62.1	76.5

Table 6. Duration of consonants prior to stressed vowels

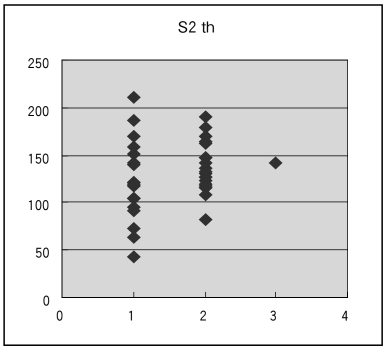
Table 6 lists the results of measuring the duration of consonants prior to stressed vowels. Table 7 shows the results of measuring the duration of consonants prior to unstressed vowels. In the English language, consonants that are placed prior to stressed vowels tend to be longer than when they are prior to unstressed vowels. Table 8 is made according to the results of Table 6 and Table 7 to prove this phenomenon.

	Sen.1ð in the		Sen.4 ð in the		Sen.8 ð in the	
MODEL	64		38		20	
	Before	After	Before	After	Before	After
S1	62	42	57	40	28	21
S2	27	63	88	32	57	31
S3	47	63	27	35	58	16
S4	21	29	68	55	33	16
S5	58	83	57	46	30	30
S6	68	78	98	39	49	38
S7	95	65	33	59	73	21
S8	52	62	63	65	54	21
S9	49	49	47	63	50	37
S10	39	83	41	45	87	50
S11	54	59	51	38	46	28
S12	39	30	173	59	50	25
S13	131	96	72	59	79	72
S14	94	83	55	137	25	38
S15	49	41	39	30	48	20
S16	80	53	51	38	25	21
S17	46	72	51	40	25	26
S18	63	57	110	37	26	22
S19	52	72	55	68	35	29
S20	46	43	68	43	53	30
Mean	58.6	61.15	65.2	51.4	46.55	29.6

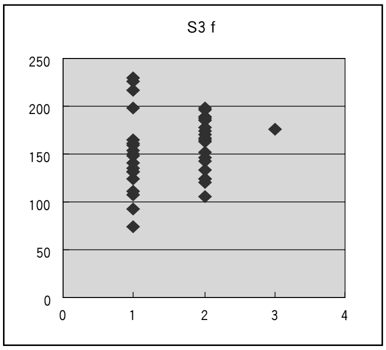
Table 7. Duration of consonants prior to unstressed vowels

	Stress (±)	Before (ms)	After (ms)
θ	+	121.76	137.05
f	+	149.55	160.45
ð	+	62.1	76.5
Mean		111.137	124.6667
θ	-	58.6	61.15
θ	-	65.2	51.4
θ	-	46.55	29.6
Mean		56.7833	47.38333

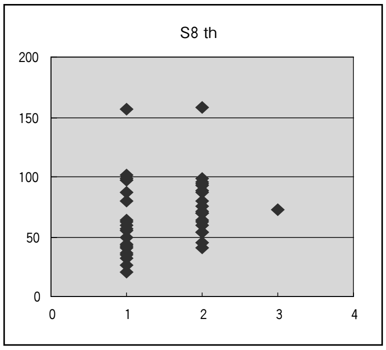
Table 8. Comparison of the duration of consonants between stressed and unstressed



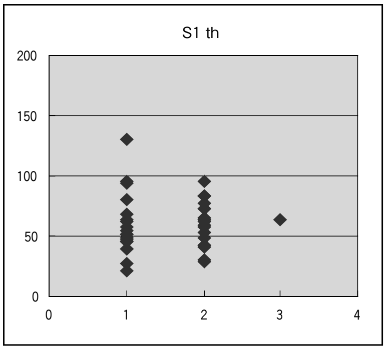
1.Sentence2 / θ / in "thought"



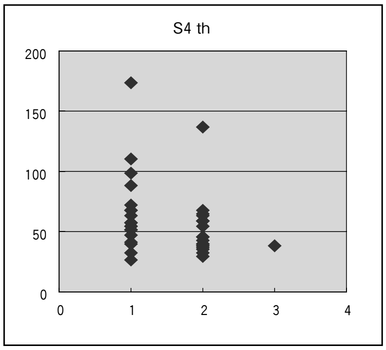
2. Sentence3 /f/ in "fired"



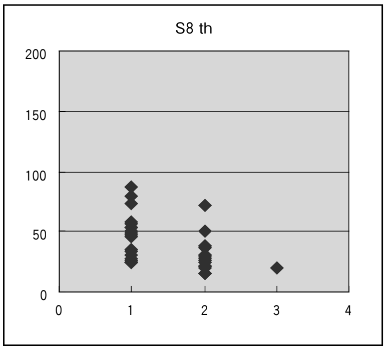
3. Sentence8 /ð/ in "these"



4. Sentence1 /ð/ in "the"



5. Sentence 4 /ð/ in "the"



6. Sentence8 /ð/ in "the

Figure 7. These six graphs show the distribution of the duration of consonants in BL, AL and M." On an x-axis, figure 1 means "before," figure 2 "after," and figure 3 "model." The unit of the y-axis is millisecond(ms).

Table 8 shows that even without listening to model sounds, the duration of consonants prior to stressed vowels produced by Japanese subjects (mean average=111.137ms) is already long compared with that prior to unstressed vowels (mean average=47.333ms). After listening to the model reading, long consonants became longer (111.137ms to 124.666ms) and short consonants shorter（56.7833ms to 47.38333ms）.

To investigate more precisely, the six graphs in Figure 7 were made based on Table 6 and Table 7. These graphs indicate that after listening to the model reading, the duration of consonants produced by Japanese subjects tended to approximate the model sounds.

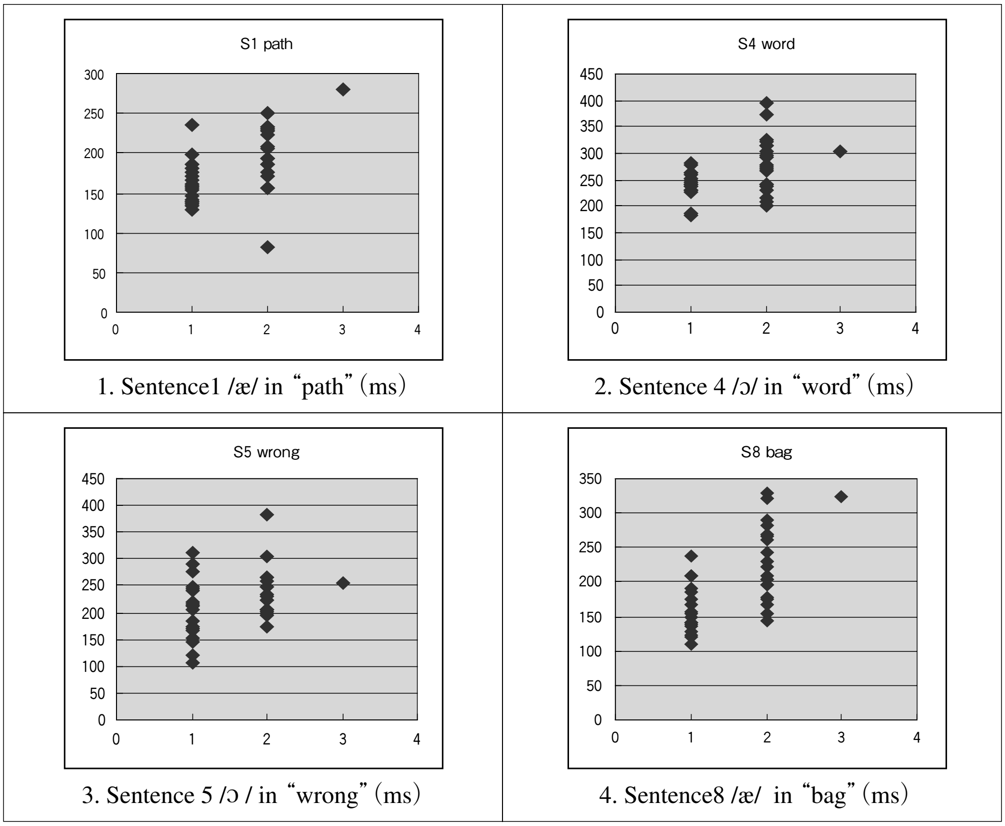
3.4. Duration of phrase-final vowels

Table 9 shows the results of measuring the duration of phrase-final vowels. As noted earlier, English vowels tend to increase in the duration when they occur at the end of a phrase.

Regarding the mean averages of the duration of vowels in Table 9, the duration of every vowel after listening to the model sounds increases. The reason for this phenomenon can be related to the fact that in most cases, the duration of the

	Sen.1 æ in path		Sen.4 ɔ in word		Sen.5 ɔ in wrong		Sen.8 æ in bag	
MODEL	280		305		254		324	
	Before	After	Before	After	Before	After	Before	After
S1	138	251	249	291	204	196	141	195
S2	140	209	240	327	221	**384**	155	203
S3	236	251	277	372	242	248	210	328
S4	198	230	228	322	249	203	191	268
S5	171	176	236	275	105	206	139	267
S6	176	207	243	238	277	197	208	231
S7	130	155	248	203	212	222	124	223
S8	162	231	282	243	214	258	237	260
S9	159	194	232	303	211	175	128	322
S10	141	156	229	270	215	204	174	290
S11	158	227	239	277	**292**	264	158	283
S12	140	224	182	208	151	233	137	205
S13	134	186	242	230	313	200	140	178
S14	147	83	265	315	186	205	148	196
S15	180	170	246	243	169	205	210	243
S16	185	232	263	267	174	305	167	155
S17	136	227	252	396	245	264	142	208
S18	154	194	186	298	122	248	120	144
S19	165	234	260	216	166	232	186	167
S20	156	187	230	314	146	225	110	174
Mean	147	219	241.45	280.4	205.7	233.7	161.25	227

Table 9. Duration of phrase-final vowels



1. Sentence1 /æ/ in "path" (ms)
2. Sentence 4 /ɔ/ in "word" (ms)
3. Sentence 5 /ɔ / in "wrong" (ms)
4. Sentence8 /æ/ in "bag" (ms)

Figure 8. These four graphs show the distribution of the duration of phrase-final vowels in BL, AL and M." On an x-axis, figure 1 means "before," figure 2 "after," and figure 3 "model." The unit of the y-axis is millisecond(ms).

model is longer than that of the Japanese subjects. There is one exception found in Table 9 (See the bold figure 292ms). Judging by the subject's utterance in this case, she was not very confident about pronouncing the word. After listening to the model reading, the duration became quite similar to the model's. (Subject11-264ms, Model-254ms).

As for Figure 8, the four graphs indicate that if the duration of a model vowel is longer than that of subjects seen in Graph 1, 2 and 4, AL consonants tend to be longer than BL. If the duration of the model vowel is not so long as in Graph 3, the duration of AL consonant would not increase. In Graph 3, we can observe one exceptionally high dot (identical figure 384ms in Table 9). Judging from listening to the recording sound, the subject seemed to pronounce the word "wrong" exaggeratedly after she listened to the model reading.

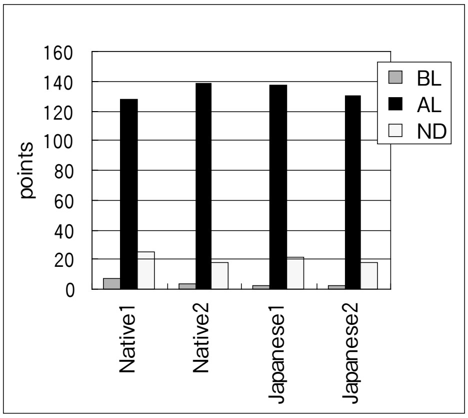
In this section, the duration of test sentences, some consonants and vowels have been measured and the comparison was made between BL, AL and M. In consequence, the results demonstrate the significant influence of model sounds. AL sounds tend to approximate model sounds. In this experiment, subjects were asked to listen to the model reading at least once and most subjects listened once for each sentence before recording their voice. Judging from this, we may conclude that only one or two exposures to model sounds has an influence on speech production of subjects. In the next section, the auditory impressions of the four raters will be stated. I will compare the results of the speech analyzer and human ratings later.

3.5. Auditory impressions

Subjects' recorded productions were evaluated on a three-point scale by a total of four teachers of English. Two of them are native speakers of English and the other two are Japanese. Rating sessions were done individually. Raters listened to the subjects' BL sound and AL sound, and then they were asked to judge which

	Sentence1 - 8			Minimal pair 9,10		
	BL	AL	Same	BL	AL	Same
Native 1	7	128	25	3	14	23
Native 2	3	139	18	0	9	31
Japanese	12	137	21	0	10	30
Japanese	22	130	18	0	6	36
Mean	3.5	134	20.5	0.75	9.75	30
Sum	14	534	82	3	39	120

Table 10. Auditory impressions by raters



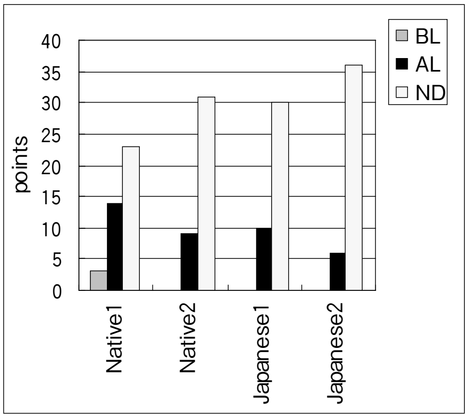


Figure 9. Graphic form of Table 10. The left graph covers Sentence1 to 8. The right graph covers minimal pairs.

sounded more naturally as English. The three rating scales were BL, AL, no difference（ND）. If a rater felt AL sounds to be more natural, one point would be added to AL. As there were 20 subjects and each subject's recorded productions were 10 pairs of AL and BL, each rater listened to 400 sound files on computer.

Results of auditory impressions are shown in Table 10 and Figure 9. Regarding sentences（Sentence 1-8 in Table 10 and Figure 9, left graph）, AL sounds were rated quite highly compared to BL. On the other hand, the minimal pairs such as

Student	BL	AL	ND	Student	BL	AL	ND
S1	0	31	1	S11	0	29	3
S2	0	27	5	S12	2	29	0
S3	0	24	8	S13	0	26	6
S4	0	28	4	S14	0	32	0
S5	0	20	12	S15	1	29	2
S6	0	18	14	S16	3	23	6
S7	0	30	2	S17	1	31	0
S8	5	23	4	S18	0	27	5
S9	1	24	7	S19	0	31	1
S10	0	30	2	S20	1	31	0

Student	BL	AL	ND	Student	BL	AL	ND
S1	0	0	8	S11	0	2	6
S2	0	0	8	S12	0	2	6
S3	0	0	8	S13	0	4	4
S4	0	2	6	S14	0	4	4
S5	0	2	6	S15	1	4	3
S6	0	1	7	S16	1	4	3
S7	0	4	4	S17	0	2	6
S8	0	5	3	S18	0	2	6
S9	1	2	5	S19	0	3	5
S10	0	1	7	S20	0	1	7

Table 11. Individual difference of ratings. The left table is based on the data of Sentence 1-8. The right table is based on the data of two minimal pairs, food/hood and cars/cards (Test material 9 and 10).

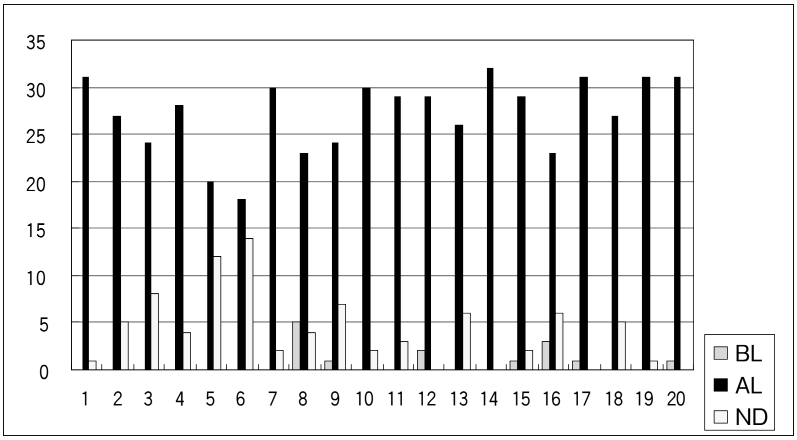


Figure 10. Graphic form of Table 11. The x-axis indicates 20 subjects and the y-axis shows the total amount of points that each subjects gained from four raters.

food/hood and cars/cards didn't show much difference in BL and AL. Since no instruction concerning pronunciation was given during the experiment, it must have been difficult for subjects to perceive the difference in the unfamiliar consonants of these pairs. As for prosodic features, however, such as rhythm or intonation, all raters felt AL sounds were more natural.

Table 11 and Figure 10 show the individual differences among the subjects. They tend to be masked by averaging, but the variation between individuals is important. There were a group of subjects who were fairly good at adjusting their pronunciation to the model reading such as Subject 1, 14, 17, 19, and 20 (whose scores were over 30). On the other hand, there were a few subjects whose pronunciation didn't change significantly as shown in Subject 5 and 6. This is because their pronunciation of English was already natural before practicing. In the present experiment, there happened to be no subject whose accent of English is strongly influenced by the first language.

4. Discussion and Concluding remarks

The results of the study revealed the significant effects of the model reading on the subjects' production. Especially in prosody, the duration of sentences, consonants, and vowels approximated that of the model reading. As for the unfamiliar consonants, there was no significant change between BL and AL. That means the subjects could recognize the prosodic features by listening to the model reading only once or twice and produce the sound in a similar way. Yet, regarding unfamiliar consonants, they could not recognize the difference and, as a consequence, they couldn't produce them. The pedagogical implication of this is that both the suprasegmentals (mainly rhythm, intonation) and segmental features (vowels, consonants) are indispensable in teaching pronunciation of a foreign language.

Regarding the approximation to the model reading that the subjects showed in the study, it is too early to conclude that they acquired or improved the prosody of the target language because they were exposed to the model reading only once or twice for each stimulus. How will the subjects change their production if they participate in further pronunciation training? More extensive experiments, including studies of retention, would contribute to our understanding.

In addition to the quantifiable results, I concluded from the experiment that learners' confidence was key to success in learning a foreign language. As I stated earlier, some researchers have doubts about "teaching" suprasegmentals. Even if their contentions are right in theory, we can still nevertheless introduce them to learners. Then learners may gain confidence in producing the target language.

I used computer software for this study and I am developing pronunciation training software at the moment. However, this is not to say that computer-assisted training is superior to other approaches. A question might arise as to whether non computer-based training approaches such as traditional teacher-led instruction would be as or more effective. One might attempt to compare these approaches experimentally. However, such a comparison is irrelevant because there are numerous elements that make up an approach. Simply using the same materials for the same period of time would not provide a basis of comparison. In my opinion, they should complement each other.

Acknowledgements

I wish to thank Prof. Reiko Yamada at Kobe University for her advice on the analysis of this data. I also acknowledge the assistance of Mr. Makoto Kishimoto and Mr. Shinya Matsuo, technical staff members of the Institute for Language and Culture at Konan University, in developing the software for the study, the model speaker, and those who served as raters.

Notes

1) NBES is an abbreviation that Jenkins created. It stands for 'non-bilingual English speaker.'
2) The basis of the software was provided by ATR (Advanced Telecommunications Research Institute International). We are permitted to use it on a research basis.
3) WaveSurfer is an open source tool for sound visualization and manipulation created in the School of Computer Science and Communication, The Royal Institute of Technology in Sweden.
4) In speech, slight decelerations and pauses frequently occur at the ends of phrases.

References

Dauer, R.（1983）. Stress-timed and syllable-timing reanalyzed. *Journal of Phonetics* 11, 51-62.

Gimson, A.C.（1981）. *An introduction to the pronunciation of English*. 3rd edn. London: Edward Arnold.

Jenkins, J.（2000） *The phonology of English as an International language*. Oxford: Oxford University Press.

Ladefoged, P.（1982）. *A Course in phonetics*. 2nd edn. New York: Harcourt Brace Jovanovich.

_____（2003） Phonetic data analysis. Oxford: Brackwell Publishing.

Roach, P.（1991）. *English phonetics and phonology*. 2nd edn. Cambridge: Cambridge University Press.

Ueyama, M.（1996）. Phrase-final lengthening and stress-timed shortening in the speech of native speakers and Japanese learners of English. *Proceeding of ICSLP*, Volume 2, 610-613.

Appendix

Test sentences

1 Thank you very much for everything.
2 I bought them drinks.
3 What's she saying?
4 They are all afraid
5 What would you like to do?
6 Give it to him.
7 That is better than that.
8 I know it's true.
9 think/sink
10 right light

11 He went over the path.
12 They thought about it.
13 We've fired them.
14 Repeat the word.
15 I knew it was wrong.
16 He didn't collect the papers.
17 He didn't correct the papers.
18 Put all these things in the bag.
19 food/ hood
20 cars/ cards

21 He went over the path.
22 They thought about it.
23 We've fired them.
24 Repeat the word.
25 I knew it was wrong.
26 He didn't collect the papers.
27 He didn't correct the papers.
28 Put all these things in the bag.
29 food/hood
30 cars/ cards

31 Look at the train.
32 Don't disturb them while they are praying.
33 Humpty Dumpty sat on a wall
34 Humpty Dumpty had a great fall
35 All the king's horses and all the king's men
36 Couldn't put Humpty together again
37 What a wonderful life he lived!
38 How beautiful you are!
39 ban/van
40 deaf/death

11－20 BL（Before Listening to the model reading）
21－30 AL（After Listening to the model reading）