The Apparent Priority of Prosodic Features over Individual Sounds in Second Language Speech Learning

Midori IBA

1. Introduction

1.1. Background and outline of the study

Phonetics, phonology and other aspects of language have long been viewed as being best learnt through a bottom-up approach. Traditional works on phonetic pedagogy, such as Gimson (1981), began with the description of vowels, consonants, words and connected speech. It was natural for more practical textbooks on pronunciation to deal with individual speech sounds first, then morphemes, vocabulary, phrases, and discourse units. According to this approach, learners would reach a particular level of proficiency by accumulating the mastered entities of the target language. Over the years, however, there has been a shift towards a more holistic top-down approach in the field of English language teaching. The current emphasis in pronunciation teaching seems to reside in the prosodic features, or the suprasegmentals of language such as stress, rhythm, intonation, and pitch as opposed to the segmentals such as individual vowel and consonant sounds. Brown (1991) expressed the view that 'the suprasegmentals are more basic and contribute more to intelligibility and accent. They should therefore appear first in textbooks and be mastered first by learners.' However, investigations to support this claim were not carried out. Moreover, in Japan, there seems to be relatively little interest in pronunciation teaching and the bottom-up approach is still common in the English language classroom. Indeed, concerning both approaches, theory has not been sufficiently supported by empirical research to establish which approach is more effective in language acquisition.

This study aims to examine which approach is more effective in pronouncing English naturally. There were four groups of participants; Group A: learning consonants and vowels first, then prosody in phrases or sentences, Group B: learning prosody in phrases or sentences first then consonants and vowels, Group C: learning prosody and individual sounds together, and Group D: a control group which didn't practice but took the pretest, the midtest, and the posttest. The participants in this study were all Japanese and joined the experiment voluntarily. Most of them were enrolled in English listening courses at a university in Japan.

Sound data of the pre/mid/post-tests were collected using original software. Sentence or phrase duration and F0 ranges were measured using Praat software. Data were also judged by four raters as to whether they sounded natural as English. The results were analyzed in Analysis of Variance (ANOVA). Findings show that Group B achieved the highest results in both objective evaluation (the measurement of duration and pitch ranges) and subjective evaluation (raters' judgments). Therefore, the findings seem to indicate that repeating sentences or phrases at the beginning of a series of sessions played an important role in acquiring the prosody of a target language. The research in this area will shed much light on our understanding of the process of speech perception in general.

1.2. Purpose of the study

This study investigates the prosodic aspects of second language acquisition. My principal concern is order effects: Which group will acquire the prosodic features of English most effectively?

- Beginning training sessions with individual sounds followed by prosody
- Beginning training sessions with prosody followed by individual sounds
- Training individual sounds and prosody together

Sentence duration and F0 ranges of 17 stimuli were measured in pre/mid/posttest. The durational ratios and F0 range ratios of those stimuli were also figured and analyzed in ANOVA. Subjective evaluations were also conducted. The raters of the experiment were the same professional English teachers in Study 1 and 2. The only criterion for rating was "How natural does the utterance sound as English?" Participants' recorded productions were evaluated on a seven-point scale. ANOVA was used for analysis of the results and compared with the consequences of the objective evaluation stated above.

2. Experiment

2.1. Materials

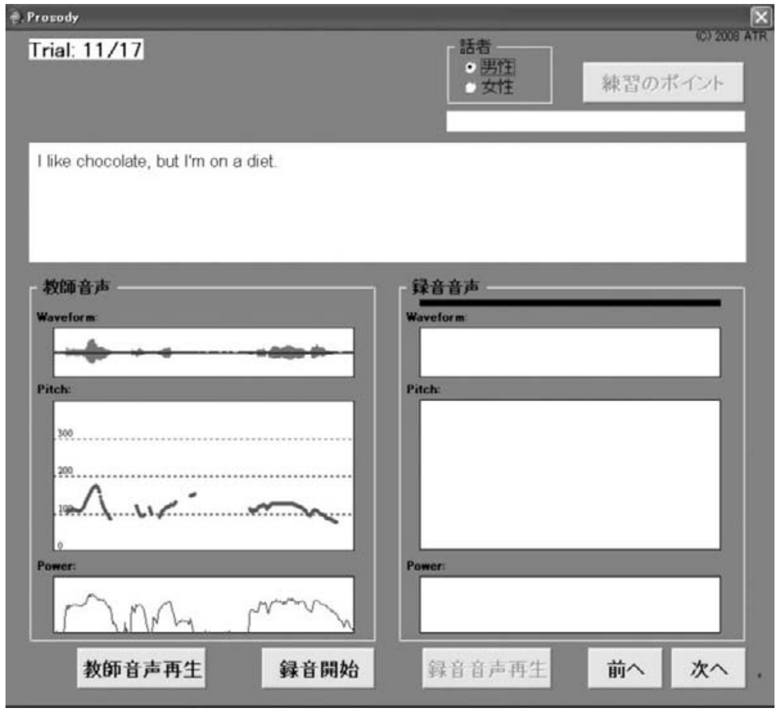
The same stimuli were used in the pretest, the midtest, and the posttest. The total number of stimuli in each test was 17. The stimuli were largely collected and selected by Professor Shinobu Mizuguchi at Kobe University. They were recorded by two native speakers of British English at a studio in Osaka. The recorded productions were inserted in the original software as model sounds. The test software was created by ATR. Table 1 shows the contents of the stimuli.

Table 1: The stimuli in the pre/mid/posttests

1	The BBC.
2	You'll have to take the tube.
3	Pardon?
4	Roads are rough in rural areas.
5	What a good idea!
6	What did Mary bring?
7	The wine.
8	Isn't she pretty!
9	Pork or beef?
10	Sorry, I don't eat meat.
11	I like chocolate, but I'm on a diet.
12	Milk, I believe, comes from cows.
13	Would you pass me the water?
14	Hey, are you going to return those books of mine you borrowed?
15	Which books? I can't remember borrowing any.
16	The ones about biology and language.
17	Oh, those books. Er --- could I keep them a few more days?

Selection of the stimuli for testing followed these guidelines: 1) familiar vocabulary, 2) structural variety, 3) sustained phonation which may provide a visually obvious display of pitch contour, 4) a variety of relatively short sentences or phrases and longer sentences.

Figure 1. The interface of pre/mid/posttest The same software was used in this study as with Study 1 and Study 2. The contents were customized by ATR.



2.2. Speakers

In the pre/mid/posttests, a male speaker and a female speaker of standard British English recorded the test items as models. Between the pretest and midtest, participants attended 10 sessions of pronunciation training, and following the midtest, they participated in the 10 additional sessions. Two pieces of software were used in the 20 sessions. In that used for training prosodic features, the same speakers' voices in the pre/mid/post tests were inserted. In that used for individual sounds, a male speaker of standard American English and a male speaker of standard British English recorded the models.

2.3. Participants

The total number of participants is 80. At first, approximately 120 native speakers of Japanese volunteered to participate in this study. All of them were undergraduate students at Konan University in Kobe. They belonged to different faculties of the university. None had spent more than two months in an English speaking country. Their ages ranged from 19 to 23. They reported normal hearing and vision. All of them except the students in the control group were taking

English listening courses. As for the control group, they were taking an English translation course.

The participants took a short version of the TOEIC to assess their English proficiency. The participants for the experiment were divided into three nearly homogeneous groups. Regarding the participants in the control group, they took the TOEIC to begin with, and then the pretest.

As the pre/mid/posttests and the 20 sessions were individually performed using 10 computers in a self-study room at the university, some of the data of the three tests were missing or some of them were not recorded clearly enough to analyze. Furthermore, the period of the training sessions was about two month long and the participants were supposed to attend 20 sessions. Consequently, the number of complete data was reduced. Barely 20 participants' data of each group can be analyzed as complete sets of three tests. Regarding the number of male/female participants, see Table 2. The number of female students who volunteered to join this experiment was originally higher than that of the male students. Accordingly, the number of female students who completed the three tests was higher.

Table 2. The number of the participants

Group	Male	Female	Total
A	6	14	20
B	6	14	20
C	7	13	20
D	8	12	20
Total	27	53	80

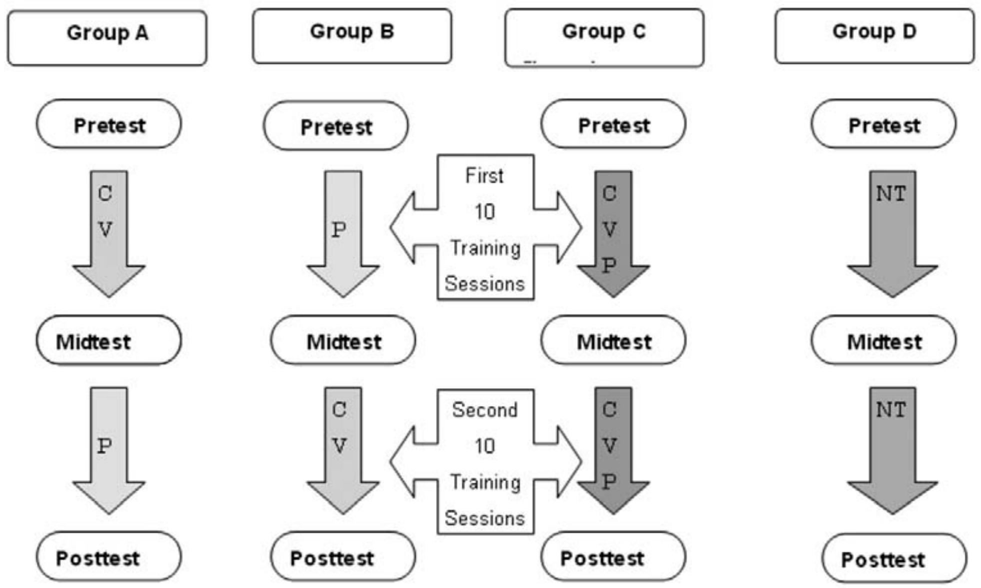
2.4. Procedure

A pretest-midtest-posttest design was used to measure the effects of two months training (20 sessions of about 40 to 60 minutes each) using computerized visual displays of pitch contours, wave forms and power as feedback (See Figure 1). As for the pre/mid/posttests, the same software that was provided by ATR (Advanced Telecommunications Research Institute International) was used. Users can customize it by inputting the stimuli. 17 stimuli for each test were set in the software. See Table 1. In the pre/mid/posttest, the participants recorded their voices which were saved in the computer server as WAV files. They were analyzed by computer software called Praat.

Participants were asked to come to a self-study room at any time during the training period, where the training sessions were performed on 10 computers. For the first time, they were asked to read instructions about using the software and took the pretest by computer. Their voices were automatically recorded and stored in the server. During the training period, they took the training sessions at any time they liked. Some of the participants came daily and finished the training sessions relatively early. Some of them came to the room as regularly as twice a week and others came quite irregularly.

Figure 2. The three tests and training sessions

In the arrows, 'CV' means Consonants and Vowels, i.e., individual sounds, 'P' means 'Prosody,' 'CVP' means 'Consonants, Vowels, and Prosody,' and 'NT' means 'No training.'



2.4.1. The training session groups

Figure 2 shows the three tests and the two-part training sessions. The following is the description of each group.

Group A: Individual-sound-first group: After taking the pretest, the participants of Group A trained individually with software to practice English pronunciation that I created in 2004 and made available on the web. Details of the software are mentioned in the next section. The participants took ten training sessions focusing on practicing particular individual sounds such as /r/ and /l/ in one session. After that, they took the midtest then participated in the second ten training sessions. This time they used different software the details of which are described below. The second ten training sessions focused on acquiring prosodic features. After they finished the second ten sessions, they took the posttest.

Group B: Prosody-first group: The participants of Group B trained in the opposite way to Group A. In the first 10 sessions, they practiced prosodic features and in the second 10 sessions, they practiced individual sounds.

Group C: Mixed training group: In the first and second training sessions, the participants of Group C trained with both pieces of software together. They practiced prosodic features and individual sounds in one session.

Group D: Control group: The participants of Group D didn't join the training sessions at all. They took the pretest first, and after three weeks they took the midtest, and finished the posttest three weeks later.

2.4.2. The two pieces of training software

A) "English Pronunciation Practice for Japanese Learners" for practicing individual sounds

As mentioned in the above section, two different pieces of software were used in this experiment. As for training individual sounds, software called "English Pronunciation Practice for Japanese Learners" was used. I created the software with financial assistance from Konan University and it is now available on the web. See Figure 3.

Figure 3. The interface of software, "English Pronunciation Practice for Japanese Learners."
(http://kccn.konn-u.ac.jp/ilc/english/)

Prosodic features can also be practiced with this software, but on this occasion, the software was used to train individual sounds. In the 10 training sessions with this software, the subject practiced the following items.

Table 3. The contents of 10 training sessions for individual sounds

Session 1	/r/ and /l/
Session 2	/f/ and /h/
Session 3	/b/ and /v/
Session 4	/θ/ and /s/
Session 5	/ð/ and /z/
Session 6	/s/ and /ʃ/
Session 7	/n/ and /ŋ/
Session 8	The difference between /i/, /ɪ/ and「い」
Session 9	The difference between /e/ and「え」・ The difference between /æ/, / ɘ/and「あ」
Session 10	The difference between /ɒ/, /ɑ/ and「お」・ The difference between /u/, /ʊ/ and「う」

In a single training session, participants read the explanation on the computer display about how to produce a given consonant or vowel, and then performed five to ten exercises. The participant was asked to repeat each exercise at least five

times. As the software is not equipped with a recording function, the production of the participants at any session was not recorded.

B) "Prosody" for practicing prosodic features

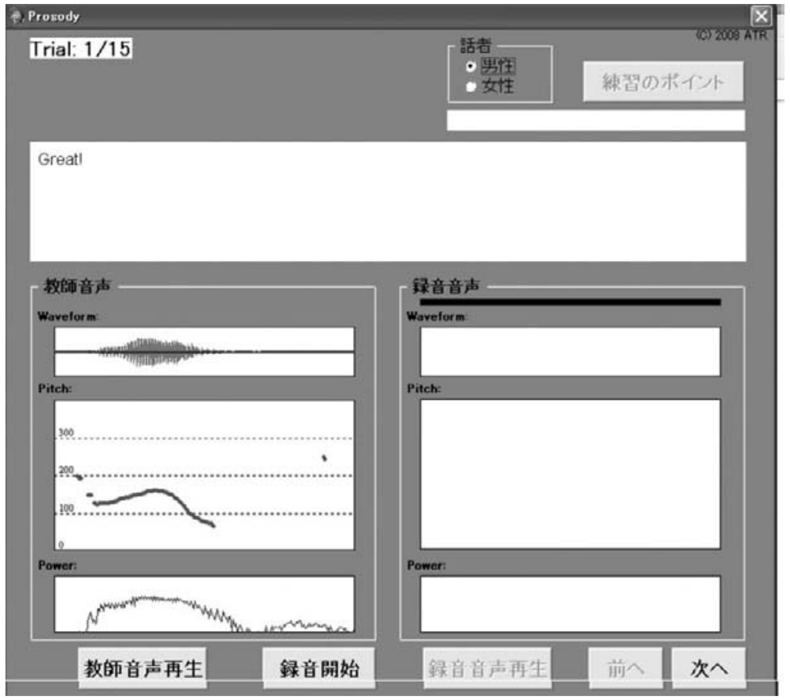
As for the training sessions in prosodic features, different software named "Prosody" supported by Grants-in-Aid for Scientific Research was used. The format is the same as pre/mid/posttest but the inputted contents were different. The software can be customized according to the needs of users. Table 4 shows an example of stimuli used in the first and second sessions of prosody training. See Appendix for the entire stimuli of all sessions using this software.

Table 4. Example of stimuli in two sessions of prosody training

Session 1 Tones	1	Great!	
	2	Thanks!	
	3	Pardon?	
	4	Yes.	
	5	Bass	
	6	A: I'll be there by five.	B: Great!
	7	A: Care for a /drink?	B: Thanks!
	8	A: You'll have to take the tube.	B: /Pardon?
	9	A: You were there, |weren't you?	B: Yes.
	10	A: He sings tenor.	B: Bass.
Session 2 Statements	1	This is a pen.	
	2	I think it's great.	
	3	A: When'll they finish?	B: Next Wednesday.
	4	I won't eat anything.	
	5	I won't eat anything.	
	6	Will you eat /anything?	
	7	Roads are rough in rural areas.	
	8	It's not hot,| it's cold.	
	9	A: Who's that?	B: I know her face.
	10	A: She's working in Oxford.	B: Cambridge.

As mentioned before, the interface of this software is the same as the pre/mid/posttest. Participants were asked to repeat the model utterance for which the text was shown on the screen at least 10 times. Along with the text, the participant saw the waveform, pitch and power of the model sound. Then he/she pushed the recording button and read the text aloud. The waveform, pitch and power of the subject were also shown on the same screen. See Figure 4. The production during the sessions with this software was saved on the computer automatically but not used for the analysis of this experiment. Only the production of the pre/mid/posttest was analyzed later.

Figure 4. The interface of software "Prosody."



2.4.3. Acoustic analysis

Sentence duration and F0 ranges were measured using waveform displays and wideband spectrograms of Praat. See Figure 5. There are 8,160 data (4,080 for sentence duration, 4,080 for F0 ranges) in total.

Figure 5-a. Praat interface. A model pronounced Stimulus 2 in the pre/mid/posttest: "You'll have to take the tube."

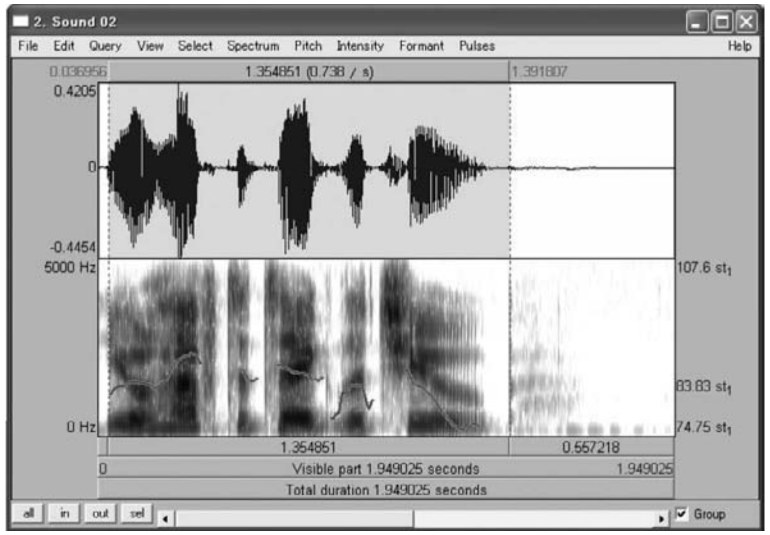
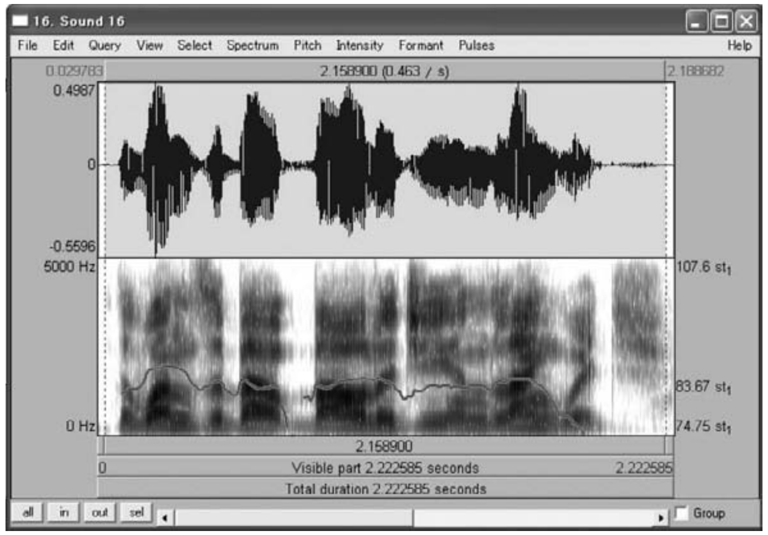


Figure 5-b. Praat interface. A model pronounced Stimulus 16 in the pre/mid/posttest: " The ones about biology and language."



Sentence duration

There were four groups in all. Each group had 20 participants and took the three tests. Each test had 17 stimuli. Table 1 (Table 1-a to Table 1-d) in Appendix 2 shows the sentence duration of each group.

In order to examine the contrasts in duration among the participants and the models, duration was analyzed proportionally as well. See Table 2 (Table 2-a to Table 2-d) in Appendix. Then the average of the proportion of duration was calculated.

F0 ranges

The highest and lowest F0 values in the whole utterance in the pre/mid/posttest were measured in semitones using Praat. The semitones are logarithmic scales of Hertz. Usually F0 values and F0 ranges of male and female speakers are quite different. However, there is no more difference between the ranges of the two genders when they are converted to semitones. Strictly speaking, pitch and F0 should be categorized differently though they are widely taken to be the same. In this study, F0 is used for considering pitch.

The total number of the data of F0 ranges was 4,080. It was measured using Praat displays in the same way as sentence duration. See Figure 5. There were four groups in all. Each group had 20 participants and took the three tests. Each test had 17 stimuli. Table 10 in Appendix shows the pitch ranges of each group.

In order to examine the contrasts of pitch ranges among the participants and the models, F0 was analyzed proportionally as well. See Table 11 in Appendix. Then the average of the proportion of pitch ranges was calculated.

Subjective evaluations

The procedure of subjective evaluations is the same as in Study 1 and Study 2. The raters are also the same group of people in Study 1 and Study 2. The participants' recorded productions were evaluated on a seven-point scale by a total of four teachers of English at a university in Kobe, Japan. Two of them are native speakers of English and the other two are Japanese. Rating sessions were done individually. Raters were presented with the files of each subject from Stimulus 1 to Stimulus 17, and the order of presenting the files was random. They were required to judge how natural the utterance sounded as English. If a rater felt an utterance was as natural as English spoken by a native speaker or near-native, seven points would be added to the utterance. As there were 80 participants and each subject's recorded productions were 51, each rater listened to 4,080 sound files on computer. The total number of the rating results of the four raters is 16,320.

1 -------- 2 -------- 3 -------- 4 -------- 5 --------- 6 -------- 7

↓ ↓

(most unnatural as English) (most natural as English)

3. Results

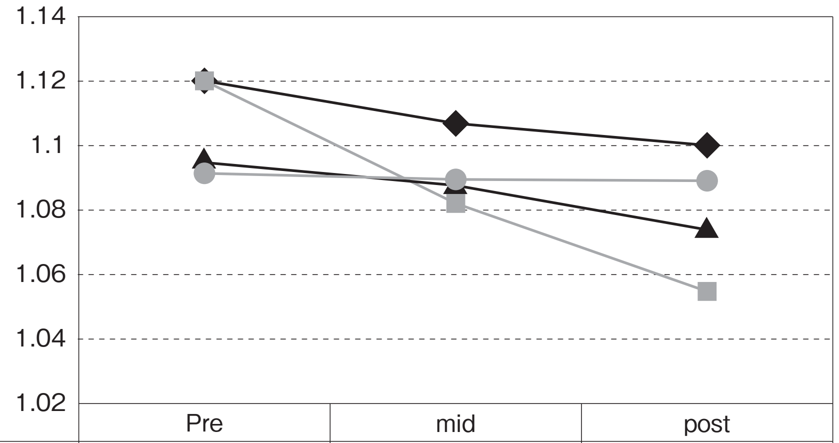
3.1. Durational ratios of the 17 stimuli of the three tests

ANOVA was conducted on the results of the average of durational ratios to test differences between means for significance. See Table 5.

Table 5. The average of durational ratios.

	model	1	1	1
group	subject	pre	mid	post
	1	1.104618	0.980418	0.97161
	2	1.05195	1.021759	1.026426
	3	1.098796	1.201647	1.132223
	4	1.044493	1.070632	1.008742
	5	1.10582	1.104653	1.093069
	6	1.248186	1.231877	1.219038
	7	1.00596	0.956913	0.994716
	8	1.182816	1.155855	1.264735
	9	1.151067	1.233346	1.17655
A	10	0.978713	0.997244	0.967382
	11	1.180559	1.130626	1.113421
	12	1.153247	1.136693	1.056812
	13	1.345105	1.206466	1.187826
	14	1.037759	1.009649	1.006457
	15	1.033983	1.018244	1.048096
	16	1.101871	1.133224	1.039997
	17	1.154069	1.18524	1.157466
	18	1.198946	1.199708	1.221174
	19	1.141069	0.977201	1.12763
	20	1.082772	1.168554	1.187352
	21	1.108177	1.107892	1.065409
	22	1.116231	1.047944	1.046668
	23	1.090027	0.988608	0.990034
	24	1.187751	1.169669	1.14454
	25	1.060517	1.022047	1.005743
	26	1.137591	1.02269	1.071228
	27	1.061695	1.018868	0.973861
	28	1.13783	1.111857	1.063067
	29	1.225165	1.282991	1.18245
B	30	1.084957	1.030036	0.991026
	31	1.087377	1.021804	0.997932
	32	1.185809	1.132502	1.051573
	33	1.184484	1.130311	1.076543
	34	1.100927	1.073073	1.072277
	35	1.097381	1.086615	1.04106
	36	1.051373	1.011626	1.001122
	37	1.074581	1.044094	1.032477
	38	1.246726	1.290204	1.255894
	39	1.023946	0.978181	0.999122
	40	1.132558	1.056145	1.007992
	41	1.029	1.0271	0.9928
	42	1.059	1.031	1.0066
	43	1.0881	1.055	1.1133
	44	1.1155	1.1136	1.1258
	45	1.103	1.0547	1.0269
	46	1.1867	1.1748	1.1514
	47	1.1132	1.0958	1.1135
	48	1.2092	1.1107	1.1355
	49	1.0551	1.1341	1.1159
C	50	1.053	1.0587	1.0665
	51	1.0512	1.1687	1.0958
	52	1.0618	1.1708	1.0368
	53	1.2761	1.1316	1.104
	54	1.1365	1.1337	1.183
	55	1.0011	1.0944	1.0394
	56	1.1298	1.1342	1.08
	86	1.204	1.1678	1.1665
	58	1.0751	1.0497	1.0382
	59	0.9359	0.8711	0.906
	60	1.0173	0.9815	0.9824
	61	1.1095	1.0993	1.0864
	62	1.0779	1.051	1.0429
	63	1.0566	1.0369	1.0423
	64	1.0946	1.0784	1.0674
	65	1.0914	1.0892	1.042
	66	1.0623	1.0527	1.0677
	67	1.0885	1.0998	1.0805
	68	1.1239	1.0853	1.0768
	69	1.0816	1.0771	1.0765
D	70	1.0902	1.0829	1.0873
	71	1.0892	1.0777	1.0792
	72	1.1192	1.0902	1.0879
	73	1.1248	1.1242	1.1067
	74	1.0969	1.0767	1.0372
	75	1.0886	1.1072	1.0751
	76	1.0937	1.0712	1.0494
	77	1.0795	1.1016	1.075
	78	1.0707	1.0236	1.0224
	79	1.0778	1.0429	1.0353
	80	1.11	1.0662	1.0369

Figure 6. The numbers in the table under the line graph are means of durational ratios of each group.



	Pre	mid	post
Group A	1.12	1.106	1.1
Group B	1.12	1.081	1.054
Group C	1.095	1.088	1.074
Group D	1.091	1.089	1.088

A two-factor ANOVA with group (A, B, C, and D) and phase (pre, mid, and posttest) as factors showed a significant main effect of phase, [F (2, 76)= 20.351, p<.001], a significant group×phase interaction [F(6, 152)= 2.234, p<.05].

The interaction between group and phase was further explored. The significant simple main effect was observed only for the factor phase for Group B, [F (2, 152) = 19.720, p<.001]. See Figure 6. Observing the above line graph, the lines other than Group D appear to lean enough to suggest that the difference among pre/mid/post is significant. However, regarding main effects, only the result of Group B was significant. The model sounds were treated as 1 in ratio. Post hoc pairwise comparisons using Ryan's method, where .05 as a significance level, showed that there were significant differences between pretest vs. midtest, midtest vs. posttest, and pretest vs. posttest for Group B.

The results show that only durational ratios of Group B significantly approximate the model ratio from the pretest to the posttest.

3.2. F0 range ratio ranges of the 17 stimuli

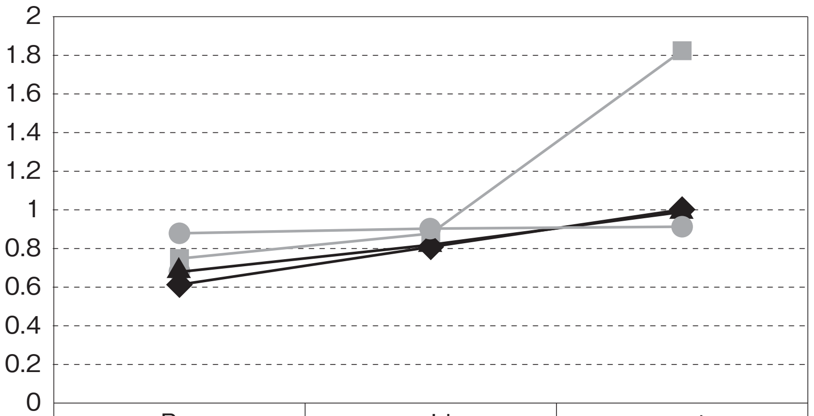
ANOVA was conducted on the results of the average of F0 range ratios to test differences between means for significance.

Table 6. The average of F0 range ratios.

	model	1	1	1
group	subject	pre	mid	post
	1	0.708	0.841	1.065
	2	0.683	0.943	1.205
	3	0.578	0.786	0.956
	4	0.626	0.805	0.994
	5	0.533	0.802	0.992
	6	0.694	0.828	1.001
	7	0.554	0.72	0.896
	8	0.664	0.73	1.02
	9	0.501	0.734	0.911
A	10	0.623	0.735	1.01
	11	0.725	0.817	1.099
	12	0.558	0.836	0.991
	13	0.645	0.782	0.925
	14	0.58	0.704	1.024
	15	0.606	0.754	0.91
	16	0.617	0.733	0.96
	17	0.563	0.871	0.94
	18	0.565	0.839	1.05
	19	0.565	0.822	0.988
	20	0.637	0.823	1.062
	21	0.534	0.773	0.972
	22	0.64	0.748	0.977
	23	0.596	0.73	0.897
	24	0.539	0.724	0.965
	25	0.639	0.769	0.933
	26	0.624	0.772	0.98
	27	0.557	0.85	0.94
	28	0.641	0.785	1.061
	29	0.762	0.792	0.965
B	30	0.575	0.744	0.978
	31	0.826	0.942	1.004
	32	0.856	0.989	1.091
	33	0.881	0.931	1.005
	34	0.836	0.955	1.054
	35	0.664	0.798	0.967
	36	0.751	0.891	1.005
	37	0.663	0.879	1.04
	38	0.657	0.877	0.943
	39	0.584	0.796	0.891
	40	0.652	0.849	0.962

	model	1	1	1
group	subject	pre	mid	post
	41	0.728	0.873	1.068
	42	0.729	0.877	1.054
	43	0.823	0.888	1.172
	44	0.906	1.035	1.169
	45	0.73	0.843	1.181
	46	0.638	0.818	0.975
	47	0.716	0.854	1.048
	48	0.726	0.87	1.002
	49	0.801	0.826	1.14
C	50	0.821	0.903	1.092
	51	0.681	0.906	1.022
	52	0.629	0.834	7.072
	53	0.796	0.908	1.053
	54	0.688	0.8	1.121
	55	0.758	0.974	1.063
	56	0.742	1.003	1.047
	86	0.735	0.874	0.968
	58	0.823	0.894	9.986
	59	0.702	0.903	1.161
	60	0.738	0.86	1.042
	61	0.815	0.991	0.973
	62	0.814	0.845	0.933
	63	1.022	0.952	0.89
	64	0.894	0.952	0.921
	65	1.257	0.862	0.891
	66	0.879	0.875	0.866
	67	0.844	0.894	0.901
	68	0.857	0.764	0.92
	69	0.87	0.913	0.84
D	70	0.752	0.98	0.838
	71	0.899	0.877	1.008
	72	0.84	0.893	0.862
	73	0.885	0.885	0.858
	74	0.878	0.867	0.862
	75	0.823	0.881	0.937
	76	0.807	0.81	0.976
	77	0.854	0.873	0.832
	78	0.829	0.797	0.899
	79	0.908	1.019	1.008
	80	0.821	0.871	1.012

Figure 7. Means of F0 ratios of each group. The numbers in the table under the line graph are means of F0 range ratios of each group.



	Pre	mid	post
Group A	0.611	0.795	1
Group B	0.746	0.887	1.822
Group C	0.674	0.83	0.981
Group D	0.877	0.89	0.911

A two-factor ANOVA with group (A, B, C, and D) and phase (pre, mid, and posttest) as factors showed a significant main effect of group [F (3, 76)= 3.302, p<.05] and phase [F (2, 76)= 9.442, p<.001] . A significant group×phase interaction [F(6, 152)= 2.234, p<.05] was also found.

The interaction between group and phase was further explored. The significant simple main effect was observed for the factor phase for Group B, [F (2, 152) = 14.819, p<.001]. There was also a significant difference between groups in posttest, [F (3, 228) = 8.015, p<.001] See Figure 7. Observing the above line graph, the lines other than Group B do not appear to rise enough to suggest that the difference among pre/mid/post is significant. The model sounds were treated as 1 in ratio. Post hoc pairwise comparisons using Ryan's method, where .05 as a significance level, showed that there were significant differences between pretest vs. posttest and midtest vs. posttest for Group B.

As seen in Figure 7, the results show that F0 range ratios of Group B significantly changed throughout the tests especially from the midtest to the posttest. Group D, the control group, did not show any difference throughout the three tests. As for the results of the posttest of the other groups, Group A and Group C approximate to the model ratio.

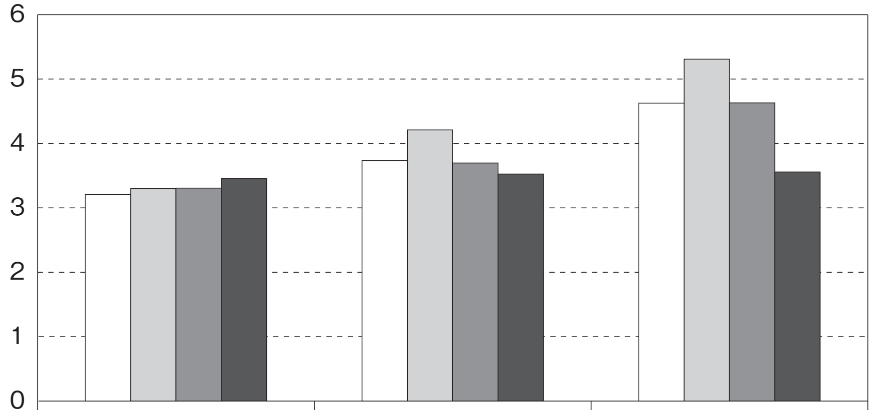
3.3. Subjective evaluations

Regarding the results of all the ratings, see Appendix. Table 9 shows mean subject evaluations of the four raters.

Table 9. Mean subject evaluations of four raters.

		PRE	MID	POST			PRE	MID	POST
	1	3.4411765	3.9852941	4.7941176		41	3.2941176	3.8676471	4.5882353
	2	3.3676471	3.8382353	4.7058824		42	3.5441176	3.6617647	4.3529412
	3	3.5294118	3.8823529	4.5735294		43	3.3529412	3.6470588	4.5147059
	4	3.3529412	4.1764706	5.1911765		44	3.2941176	3.6764706	4.6911765
	5	3.2794118	3.9117647	4.6176471		45	3.6029412	3.6176471	4.4705882
	6	3.1617647	3.6911765	4.6617647		46	3.1176471	3.7205882	4.9558824
	7	3.3088235	3.8676471	4.5		47	3.2941176	3.7352941	4.5735294
	8	3.1470588	3.5882353	4.3970588		48	2.9852941	3.5735294	4.3529412
	9	3.3970588	3.6176471	5.0588235		49	3.4117647	3.6176471	4.5441176
A	10	3.3088235	3.8970588	4.9558824	C	50	3.25	3.5588235	4.2647059
	11	3.1911765	4	4.7058824		51	3.4117647	3.5588235	4.8823529
	12	3.3970588	3.9264706	4.6911765		52	3.3823529	3.9411765	4.7941176
	13	3.1617647	3.7794118	4.5294118		53	3.2205882	3.6323529	4.75
	14	3.4117647	3.7205882	4.5588235		54	3.2352941	3.9117647	4.75
	15	3.5588235	3.4852941	4.3823529		55	3.3235294	3.8235294	4.5441176
	16	3.3676471	3.6911765	4.3529412		56	3.3970588	3.7941176	4.7794118
	17	3.2058824	3.3676471	4.1176471		57	3.1470588	3.3823529	4.7352941
	18	3.25	3.4705882	4.1323529		58	3.4411765	3.6176471	4.4411765
	19	2.8676471	3.9411765	5.1176471		59	3.1029412	3.6911765	4.9264706
	20	3.1029412	3.4264706	4.4852941		60	3.3235294	3.8823529	4.6617647
	21	3.2205882	4.25	5.3676471		61	3.5	3.5	3.6911765
	22	3.2941176	4.1323529	5.0882353		62	4.0588235	4.3529412	4.4117647
	23	3.5294118	4.3382353	5.6323529		63	3.5	3.5588235	3.6764706
	24	3.7941176	4.2058824	5.5147059		64	3.4705882	3.5588235	3.4705882
	25	3.1323529	4.1176471	5.4705882		65	3.5294118	3.6764706	3.7205882
	26	2.9411765	4.1323529	5.3235294		66	3.4558824	3.5294118	3.3970588
	27	3.4852941	4.1911765	5.4852941		67	3.4264706	3.4705882	3.5441176
	28	3.75	4.4117647	5.4852941		68	3.4852941	3.5588235	3.6323529
	29	3.4411765	4.1323529	5.3088235		69	3.5441176	3.6323529	3.7205882
B	30	3.2941176	4.6323529	5.3235294	D	70	3.6617647	3.5147059	3.1323529
	31	3.2941176	4.1323529	5.2352941		71	3.8235294	3.7647059	4.0588235
	32	3.3088235	4.1176471	5.0882353		72	3.5294118	3.6764706	3.3970588
	33	3.25	4.0735294	5.2941176		73	3.3382353	3.8382353	3.6470588
	34	3.2058824	3.8823529	5.4705882		74	3.3676471	3.6911765	3.8235294
	35	3.0882353	4.1323529	5.2647059		75	3.4117647	3.2941176	3.3529412
	36	3.4852941	4.0735294	5.0588235		76	3.4411765	3.1176471	3.25
	37	3.1323529	4.7205882	5.1176471		77	3.1617647	3.1764706	3.1764706
	38	3.4117647	4.25	5.3382353		78	3.1617647	3.1323529	3.3235294
	39	2.9264706	4.2352941	5.1323529		79	3.2352941	3.3235294	3.4558824
	40	3	4.0294118	5.1764706		80	3	3.1323529	3.25

Figure 7-a. Mean subject evaluations by four raters.



	Pre	mid	post
Group A	3.209	3.736	4.626
Group B	3.299	4.21	5.309
Group C	3.307	3.696	4.629
Group D	3.455	3.525	3.557

A two-factor ANOVA with group (A, B, C, and D) and phase (pre, mid, and posttest) as factors showed significant main effects of phase [$F(2, 76)=1008.674$, $p<.001$] and group [$F(3, 76)=62.731$, $p<.001$]. The interaction between phase and group was also significant [F (6, 152)=113.872, $p<.001$].

The interaction between group and phase was further explored. The significant simple main effects were observed in most cases except for the effect of group for pretest (tendency, $p<.10$), and the effect of phase for Group D (NS). Post hoc pairwise comparisons using Ryan's method, with .05 as a significance level, were conducted. There were significant differences between groups in the midtest and posttest phases except for the Group A vs. Group C in the midtest. In contrast, there were no significant differences between groups in the pretest phase. There were differences between pretest vs. midtest, midtest vs. posttest and pretest vs. posttest in all groups except Group D, which was a control untrained group. Group B achieved the highest evaluation in the posttest.

Figure 7-b. Mean subject evaluations by four raters.

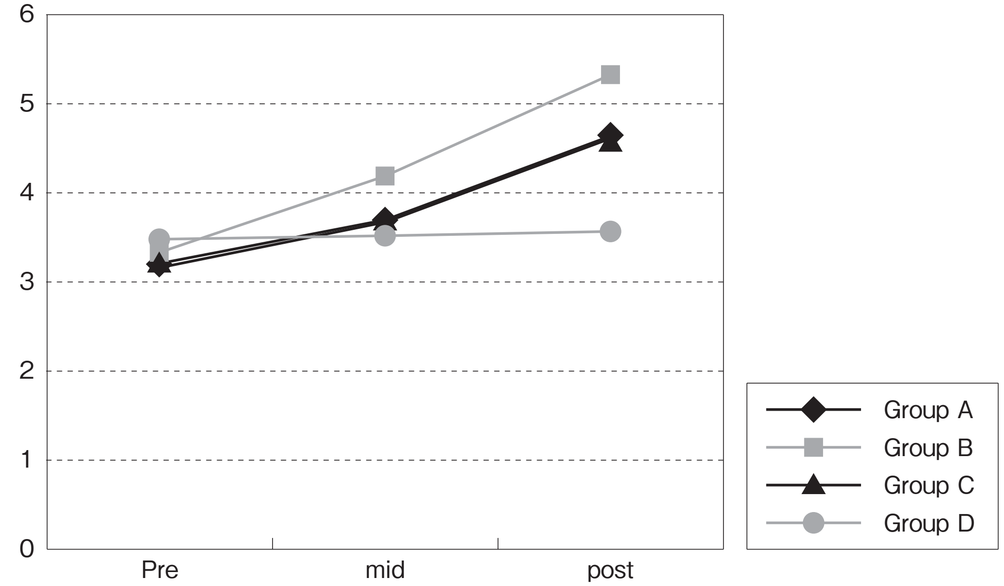


Figure 7-b is the line graph version of Figure 7-a. Observing the inclination of the lines of each group, we note that Group B obtained the highest evaluation among the raters in posttest. Group A and Group C also showed steady growth in evaluation. As their lines are almost overlapped, it can be said that they had the same growth tendency.

4. Discussion and Concluding remarks

The main goal of this study was to investigate prosodic aspects in second language speech perception and production, especially concerning the order effects of training prosodic features and individual sounds.

This study provides some encouraging new data regarding the acquisition of non-native prosodic features and individual sounds. The findings are as follows.

Effects of training

Regarding the effects of the training, particularly in the subjective evaluations, all the experimental groups showed significant differences from pretest to posttest whereas the control group did not show any changes throughout the tests. Raters detected an improvement from the pretest to the posttest productions of young adult learners. The results support the report by Hakuta, Bialystok, and Wiley

(2003) that adult learners who have passed the putative end of a critical period exhibit overall L2 success.

In the training sessions of this study, the participants were exposed to a great many model sounds. That means acoustic templates must have played an important role in monitoring the articulatory output.

As for the results of acoustic analysis of Group A, the individual-sound-first group, and Group C, the mixed training group, the durational ratios did not change significantly but F0 range ratios approximated the model sound's ratios throughout the tests. These two groups obtained better subjective evaluations in the posttest by raters.

The importance of the order of training: the apparent priority of training prosodic features over individual sounds

Group B, the prosody-first group, showed remarkable results in the three analyses of durational ratios, F0 range ratios and subjective evaluations. In the pretest, all groups were homogeneous in English proficiency as confirmed by acoustic analysis and subjective evaluations. However, as the training sessions went by, the participants of this group began to exhibit differences from the other groups. What does this indicate? The present data may not provide conclusive evidence to support the hypothesis that prosody training should be prior to individual sounds training because the tasks in this study were quite different. Two different pieces of software were used. In the prosody training, participants can check their voice recordings acoustically and visually against pitch contours, wave forms, and intensity on the computer display. Thus they had some feedback. On the other hand, in the individual sounds training, participants read instructions on the screen and practice without recording. In the latter task, the quality of feedback that the participants received was different. Further experiment is needed to confirm the priority of training prosody over segmentals by using the same task.

Although further experiments are expected, it may be said that priority of suprasegmentals in the early stages of pronunciation training did not hinder the learning process of the participants, and I believe that this training order will contribute to modify the structure of the learner's phonetic system. Previous studies have investigated the relationship between perception and production of non-native contrasts such as /r/ and /l/. For example, in perceptual training for non-native contrasts, Yamada et al (1994) used the high-variability training technique and found significant improvements in the Japanese trainees' production

of /r/ and /l/ as a result. Thus, the effectiveness of training individual sounds is clearly apparent. If the order-effect hypothesis that prosody training should be prior to individual sounds (i.e. non-native contrasts) training is proven correct, then this study will have contributed to develop a more effective training program on computer or one that can be used in the classroom.

REFERENCES

Best, C. T., McRoberts, G. W., and Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 345-360.

Bohn, O. S., and Munro, M.J. (2007). *Language Experience in Second Language Speech Learning: In honor of James Emil Flege*. Philadelphia: John Benjamins Publishing Company.

Bradlow, A. R., Pisoni, D. B., Yamada, R. A., and Tokuhara, Y. (1997). Training Japanese listeniers to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America, 101*, 2299-2310.

Brown, A. (1991). Pronunciation Models. Singapore: Singapore University Press.

Celce-Murcia, M., Briton, D.M., and Goodwin, J.M. (1996). *Teaching Pronunciation*. New York: Cambridge University Press.

Cristal, D. (1997) *English as a Global Language*. Cambridge: Cambridge University Press.

Flege, James (1988). The production and perception of foreign language speech sounds. In H.Wintz (Ed.), *Human communication and its disorders*. Norwood, N.J.: Ablex.

Flege, James (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). Timonium, MD: York Press.

Flege, James (2002). Factors affecting the pronunciation of a second language, In *PMLA-2002*, 136-136.

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

Hakuta, K., Bialystok, E., & Wiley E., (2003). Critical evidence: A test of critical period hypothesis for second-language acquisition. *Psychological Science*, 14, 31-38.

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

Lightbown, P., and Spada, N. (2006). *How Languages Are Learned* 3rd edition. New York: Oxford University Press.

McKay, S.L. (2002) *Teaching English as an International Language*. New York: Oxford University Press.

Munro, Murray J. & Tracey M. Derwing (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45*: 73–97.

Rochet, B. L. (1995). Perception and production of second-language speech sounds by adults. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp379-410). Timonium, MD: York Press.

Rvachew, S. (1994). Speech perception training can facilitate sound production learning. *Journal of Speech, Language, and Hearing Research,* 37, 347-357.

Spolsky, B. (1999). *Concise encyclopedia of educational linguistics*. Oxford: Elservier Science Ltd.

Strange, W., and Dittmann, S. (1984). Effects of discrimination training on the perception of /r/ and /l/ by Japanese adults learning English. *Perception and Psychophysics* 36, 131-145.

Yamada, R. A. (1993). Effects of extended training on /r/ and /l/ identification by native speakers of Japanese. *Journal of the Acoustical Society of America, 93*, 2391(A).

Yamada, R. A., Strange, W., Magnuson, J.S., Pruitt, J.S., and Clarke III, W. D. (1994) The intelligibility of Japanese speakers' productions of American English /r/, /l/, and /w/, as evaluated by native speakers of American English. *Proceedings of the International Conference on Spoken Language Processing, Yokohama, Japan, 1994,* 2023-2026.