Title	Preliminary testing of "SpeechTeach"/l/ and /r/ learning software on university students : Preparation for final study
Author(s)	Smith, F. Antonio
Citation	大阪外国語大学英米研究. 2005, 29, p. 63-78
Version Type	VoR
URL	https://hdl.handle.net/11094/99295
rights	
Note	

Preliminary testing of "SpeechTeach"/l/ and /r/ learning software on university students: Preparation for final study

Antonio F. Smith

1. Introduction

Everybody knows that Japanese have difficulty distinguishing between English /l/ and /r/ and that this poses problems for Japanese students of English. The problems are of four types:

1) The dilemma of deciphering ambiguous utterances

Although one might think of this first, it is surprisingly rare. This is because there are only a few hundred minimal pairs involving /l/ and /r/ in English, and context usually indicates the appropriate menber of a pair. For example, a Japanese listener is unlikely to wonder if someone has said "lentil car" or "rental car". In other words, Japanese students of English rarely hear utterances that are truly ambiguous due to /l/ and /r/. An example of such an ambiguous utterance might be, as follows: Devoutly religious child A and religious child B, who is Japanese, belong to a church with a playground; A phones B and says, "Let's go to the church and play."

2) Poor pronunciation

Japanese students of English often do not pronounce /l/ and /r/ distinctly. This could be in part because of ignorance of the means of production, but it is also

in part because of not hearing the distinction. Although I have not verified it experimentally, my opinion is that students who can hear /l/ and /r / well can pronounce them well.

3) Poor spelling

Japanese students of English often misspell words containing "l" and "r" even when they are not members of minimal pairs. In fact, it is such a common phenomenon, that there is even an Internet site dedicated to the phenomenon www.engrish.com.

4) Jokes/stigmatization as poor English speakers

Unfortunately, native speakers of English, such as some people at the above site no doubt, make fun of people who cannot produce /l/ and /r/ well. They make jokes like, "I went to a Japanese restaurant and the waitress asked me if I wanted steamed lice or flied lice." This kind of insensitivity is most likely due to the fact that most native speakers of English have not studied a foreign language seriously, so they do not sympathize with the difficulty of mastering a foreign sound system.

Because of these problems, it would be a very nice benefit to Japanese students of English if there were a convenient and efficient way for them to master /l/ and /r/. "SpeechTeach" aims to be just that.

2. Literature review

It has been shown that the forced-choice-and-feedback listening task involving /l/ and /r/ does produce small but measurable and lasting results among monolingual laboratory subjects (Logan et al. 1991, Lively et al. 1994). But it can be assumed that such monolingual subjects lack practical motivation to learn the

distinction between /l/ and /r/, as they are probably not students of English.

In his pilot study, Smith (2003) tested his autodidactic /l/ and /r/ software, "SpeachTeach", on three adults and found that with consistent use the two subjects in their late thirties could achieve a steadily increasing learning curve, while the seventy-year-old subject's learning curve improved a few percentage points quickly and then flattened out completely. However, given the time constraint of the study, it could not be verified whether or not native-like ability to distinguish between /l/ and /r/ was a real possibility.

3. Present study

3.1 Goal

The aim of the present study is to find out whether SpeechTeach can be used by motivated young adult subjects to consistently achieve native-like ability to distinguish between /l/ and /r/, or if/when it cannot, how the program and/or subjects' method of study might be modified in order to reach that result. Once these facts are known, a final study of university students can be prepared and conducted.

3.1 Subjects

49 Osaka University of Foreign Studies 1st-year English majors, aged approximately eighteen or nineteen, about 75% female, submitted data.

3.2 Methodology

Three 1st-year classes of students, with about 25 students per class, were given both a copy of the SpeechTeach program to use by themselves and also a log-sheet on which to record for each practice session the date, amount of time spent, number of words listened to and percentage correct.

1) Subjects see on a screen a minimal pair involving /l/ and /r/, such as "play"

and "pray."

2) They hear one of the words pronounced by one of four speakers, selected at random, and they can have that word repeated as many times as they wish.

3) They choose the word they believe they heard with a click of the mouse.

4) If they choose correctly, they get positive feedback; if they choose incorrectly, the correct word is indicated and pronounced again, once.

5) The program keeps track of the total number of words and percentage correct.

3.3 Data[1]

There are five data categories (number of each session, date, time, numbers of words and % correct), 49 subjects returned log sheets, and on average, one subject completed about 80 practice sessions. Altogether, that is about 20,000 bits of data.

3.3.1 Excluded data

There are two types of subjects that are excluded in the analysis.

1) Those whose log sheets were not completed properly

   6 of the 49 students who returned log sheets, did not fill in their log sheets properly. Some data looked obviously "bad", because it diverged significantly from the appearance of other data, and in addition, the same calendar date was used multiple times.

2) Those who started with native-like ability

   A native speaker using the program, such as the author, will almost always get a score of 100% on the program, but if I go through the program very rapidly I occasionally click on the wrong word, or if I listen without headphones, on my portable computer, the volume is sometimes too low to hear well, and I make a mistake, so I occasionally score below 100%.

Therefore, at the risk of being a little too generous, for the purpose of the present study, an average score of 90% or better is considered "native-like" ability *using the program*.

Three of the forty-three subjects with good data, numbers 41, 42 and 43, started with scores above 90% (93%, 93% and 100% respectively). They showed no significant improvement as a result of the training sessions. This might be because they used the program very few times compared with other students (One of those who started with 93% used the program 33 times and the other 34 times) and/or because the program does not focus on the few phonetic environments that are still problematic for such students. The subject who started at 100% spent several years in an English-speaking country before coming to Osaka University of Foreign Studies. The following discussion will exclude the three above-mentioned subjects, leaving a total of forty.

3.3.2 Unsubmitted data

Although it can be assumed that virtually all English majors at Osaka University of Foreign Studies are highly motivated to improve their English, about one third of the total 1st-year class did not submit log-sheets.

Probably, the non-submitters did not complete their log sheets well enough to feel comfortable submitting them. Probably, they did not use the program very many times and therefore did not achieve a very good result. It can be assumed that such students felt their time better spent in studying something else, at a part-time job or in recreation.

It must be recognized that in any voluntary study such as this, that not all subjects will submit high quality data. For my part, I feel lucky that I received as much data as I did, and I thank all of those diligent 1st-year English majors who submitted. Hopefully, in the future, their excellent results will encourage others to follow suit.

3.4 Results

3.4.1 The "A" students: those who achieved native-like ability

There are fifteen subjects who started below 90% accuracy and rose above it, plus one other subject, Subject 40, who started at 90% and moved up to 95%, becoming significantly "more" native-like. Below, I list these successes by subject number and juxtapose their average starting and finishing scores. In cases where a starting or finishing score is not representative of a subject's actual performance, I include an average of the last few sessions, unless the last sessions follow a long break, in which case I use the last few session before the break. I have also included the numbers of training sessions, long breaks, days in training period, approximate average words per session, approximate average time per session and approximate total hours in training.

	Sub #	Start score	Av. start score	Final score	Av. final score	# ses-sions	# bre-aks	# days in training period	~Av.# words/ session	~Av. minutes. / session	Approx. total hours
i	1	**30%**	41%	96%	95%	101	1	176	300	30	**50.5**
ii	4	**47%**		90%		92	0	135	80	10	**15**
iii	7	**58%**		95%		91	1	141	90	13	**20**
iv	8	**59%**		92%		100	2	165	200	15	**25**
v	10	**60%**		94%		106	2	185	150	15	**26.5**
vi	19	**66%**		91%		119	1	185	300	20	40
vii	24	74%		91%	*96%	84	0	106	160	10	14
viii	25	75%		97%		120	0	169	250	11	22
ix	29	76%		93%		64	0	75	150	10	11
x	30	76%		97%		60	0	70	120	10	10
xi	32	76%	80%	88%	90%	60	3	165	*50	10	10
xii	33	78%		92%	93%	120	0	175	75	11	22
xiii	34	**80.5%**		93%		60	0	95	150	15	**15**
xiv	36	**82%**		98%		55	0	72	200	10	**9**
xv	38	**84%**		94%		74	2	115	200	10	**12**
xvi	40	**90%**	excep-tion	95%		56	1	185	70	5	**5**

Table 1

The data from Subject 1 shows that about 50 hours of training is enough to produce native-like ability in even the subject with the lowest starting score. Congratulations Subject 1, and thanks for pushing the limits of the Speechteach program!

In this group of 16, out of 40 total accepted log sheets, subjects starting at under 70% accuracy reach native-like proficiency in 15-50 hours, those starting between 70% and 80% do it in 10-22 hours and those starting above 80% do it in 5-15 hours. As would be expected, it appears that the higher the starting score, the less time it takes to reach native-like proficiency.

The average number of sessions is 85.

The average number of words per session is 160.

The average minutes per session is 13.

The average total time is 19 hours.

Future subjects who wish to achieve native-like proficiency can use these numbers to guide their training.

3.4.2 The "B" Students

This group of subjects achieved final scores in the range of 80% to 89% accuracy. Fifteen subjects are in this category. Most of the subjects below improved 20-29 percent.

	Sub #	Start score	Av. start score	Final score	Av. final score	# ses-sions	# bre-aks	# days train-ing period	~Av.# words/ session	~Av. min./ ses-sion	Ap-prox total hours
xvii	6	55%		80%	85%	60	0	65	60	13	13
xviii	12	60%	57%	81%	86%	49	*0	80	60	5	4
xix	14	61%		86%	85%	126	0	185	60	11	23
xx	16	64%		86%		143	1	180	140	12	29
xxi	17	65%		82%	85%	60	0	80	60	30	30
xxii	18?	65%		84%	89%	45	3	165	50	NA	NA
xxiii	20	70%		83%		45	0	185	60	9	7
xxiv	21?	70%		85%	88%	43	0	106	30	NA	NA
xxv	23	73%	71%	81%	85%	60	1	165	60	10	10
xxvi	26	75%	63%	80%	86%	57	0	95	30	5	5
xxvii	27	75%	70%	84%	85%	93	0	175	100	14	22
x	28	75.5%		90%	88%	118	0	185	100	10	20
xxiv	31	76%		84%		61		70	100	10	10
xxx	37	83%	70%	85%		118	0	155	100	10	20
xxxi	39	86%	70%	80%	82%	115	1	170	180	15	28

Table 2

The average number of sessions is 80. (vs. 85 in the A group)

The average number of words per session is 79. (vs. 160)

The average minutes per session is 12. (vs. 13)

The average total hours per person is 17. (vs. 19)

3.4.2.1. Fewer words per session=under 90% accuracy

On average, the A group had just five more sessions than the B group and worked one minute longer per session. However, what appears significant is that the B group listened to half as many words on average as the A group. Therefore, future subjects should be told that a "fast" strategy works best for most subjects.[2]

By and large, it appears that the learning curves of the B subjects level out significantly after 82-83%. Probably, this is due to subjects' first learning "easily" mastered phonetic environments and then proceeding more slowly with the remaining "hard-to-master" phonetic environments.

Since the program presents phonetic environments only randomly, most of the environments subjects hear are easy ones that they can already do, and few are hard ones that subjects have yet to master. Perhaps the B group simply failed to achieve native-like proficiency because they did not get enough practice with the tough environments.

Therefore, it would be advisable to create a special training program containing only tough environments for students to use after they achieve 82%. Subjects could also use the regular SpeechTeach program periodically to see how training with the advanced program affects their regular SpeechTeach learning curve.

3.4.3 The "C" Students

These are the students who achieved a final score of less than 80% or those who showed no improvement. Nine students fall into this category.

	Sub#	Start score	Av. start score	Final score	Av. final score	# of sessions	# of breaks	# days training period	Av. # words/ session	Av. min./ session	Approx total hours.	Lev
xxxii	2	42%	48%	64%	58%	49	0	100	50	11	9	yes
xxxiii	3	46%	53%	61%	64%	89	1	170	135	11	16	no
xxxiv	5	55%	57%	79%	78-9%	80	2	185	*70			no
xxxv	9	59.8%	72%	75%	73%	39	5	170	NA	7	4	NA
xxxvi	*11	60 %		71%	70%	22	3	145	NA	NA	NA	NA
xxxvii	13	61%	59%	75%	75%	60	0	75	100	12	12	no
xxxviii	15	61%		75%	68%	60	0	95	50	10	10	no
xxxiv	22	72%	65%	77%	75%	57	0	120	90	13	12	no
xxxx	35	81%	79%	71%	81%	60	3	155	90	13	13	yes

Table 3

The average number of sessions is 57 (vs. 80 in group B and 85 in group A)

The average number of words per session is 83 (vs. 79 in group B and 160 in group A)

The average minutes per session is 11 (vs. 12 in group B and 13 in group A)

The average total hours per person is 11 (vs. 17 in group B and 19 in group A)

However, as Subject ♯3 has a slight speech impediment, which might indicate a hearing impediment, and he worked significantly harder than the average members in this group, excluding his data should provide more accurate results about the average

students. They are as follows.

The average number of sessions is 53 (vs. 80 in group B and 85 in group A)

The average number of words per session is 75 (vs. 79. in group B and 160 in group A)

The average time per session is 11 minutes (vs. 12 in group B and 13 in group A)

The average total hours per person is 10 (vs. 17 in group B and 19 in group A)

3.4.3.1 Fewer sessions and total time training = under 80% accuracy

From these averages, it is clear that the most significant deficiencies of the C group compared to the B group is their low average number of sessions and their low total hours in training. The average number of sessions is just 66% that of the B group (and 62% that of the A group). The total time, 10 hours, is just 59% that of the B group (and 53% that of the A group). The average number of words persession is about the same as the B group.

4. Conclusion

4.1 Summary

A significant proportion of subjects did achieve native-like ability to distinguish between /l/ and /r/. At least they did it on the Speechteach program if, on average, they listened to 160 or more words per session and had 85 or more sessions within about 185 days. Their success was achieved regardless of starting score, so presumably, other highly motivated students of English around the same age should be able to do the same.

With an advanced training program focusing on tough environments, such as after a consonant cluster, the learning curve should flatten out less after 82-3%, and more students should be able to reach native-like proficiency more quickly.

4.2 Implications

1) This study has strongly implied that young Japanese "adults" can achieve native-"like" ability to recognize a foreign phonetic distinction. This, in turn, leads one to propose that many (most, all?) young-adult-students of a foreign language might have the ability to "acquire" all its phonemic distinctions. Future studies should test the limits of this proposition, for it brings into question the conventional wisdom about "critical period".

2) If Japanese students can learn to distinguish /l/ and /r/ using the SpeechTeach program, then they should be able to use a similar program to learn other phonemic distinctions in English, such as /si/ and /shi/ or the various vowels, as well as phonemic distinctions in other languages. For example, students of Chinese or Thai could use such a program to learn to recognize the different tones instantaneously and consistently. Also, for example, American students of Japanese could use such a program to learn to recognize Japanese phonemic pitch-accent distinctions such as in the three variations of "hashi" as well as distinctions due to double vowels and consonants, like "obasan" (aunt) vs. "obaasan" (grandmother) and "ite" (stay) vs. "itte" (go).

4.3 Future study

I should conduct the experiment again with similar subjects but add:

1) a pretest and post-test administered under controlled conditions using different speakers and different minimal pairs than those on the SpeechTeach program / CD, in order to verify the program's effectiveness in preparing listeners to distinguish between /l/ and /r/ in the world outside the program and to verify the results recorded on log sheets

2) the condition that students with pretest scores of 92% or higher be asked to

use essentially only the "advanced" CD (see below), but also the regular CD periodically in order to record scores

3) the condition that students reaching scores of around 82% be asked to use mainly the advanced CD, but also the regular CD periodically in order to record scores and/or practice easy environments

4) a group of "average" subjects who only use the "advanced training CD" and whose results can be used for comparison purposes

5) a group of "average" students who train a long time per session, such as one hour, to be used for comparison purposes

6) a second post-test some time after the first to test students' retention of their ability to distinguish between the categories /l/ and /r/.

4.4 Future advanced CD

Leveling out of the learning curve can often be observed as scores become high (after 82-83%), so I should create an advanced SpeechTeach CD that focuses only on the most difficult environments in which to distinguish /l/ and /r/, such as following a consonant. This advanced CD can be used by subjects with very high starting scores, subjects who eventually achieve high scores after training with the regular SpeechTeach program, and a set of subjects with a full range of pretest scores, since it is possible that students who just study the advanced CD's troublesome phonetic environments will also make advances in the easier environments "for free".

The exact content of the advanced CD should be determined by having several of the former subjects, especially the B subjects, go through the program and write down the words they miss. Those words will form the content of the advanced program.

4.5 Improvements to the basic SpeechTeach program

Some subjects have requested that I add a feature to the program that will allow them to hear the correct word repeated on command after the program tells them they have made a wrong choice. This feature would increase feedback and is therefore likely to improve the rate of learning.

4.6 Additional possibilities

I should conduct the experiment with high school students. Their youth will be an advantage in learning, but most will lack motivation. It would be necessary to find a highly motivated group of students in order to receive enough quality data from them to make the study worthwhile.

A "child friendly" version should be created for and tested on elementary, junior high school and even high school students. Because of their youth, such students should be able to learn to distinguish between /l/ and /r/ more quickly than the subjects of the present study, but they will probably only want to use the training program if it has the look and feel of a video game, which has motivation to play built in.

One would have to contract with a company specializing in such software to make a child's version of the game. Perhaps, Mombukagakusho would be interested in financing it for use in the public school system.

Mombukagakusho's decision to start English education in elementary school is a good one because the younger the child, the more easily and quickly s/he can acquire language, especially pronunciation and articles, but only if s/he has sufficient exposure to native, or at least native-like, models. Given the scarcity of elementary school English teachers with native-like pronunciation of /l/ and /r/, a program modeled after SpeechTeach could be made a valuable part of elementary school English education in Japan. If an effective program could be made for the Japanese elementary school system, it could solve the Japanese /l/ and /r/ problem once and for all.

References

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura Y., and Yamada, T. 1994. Training Japanese listeners identify English /r/ and /l/. III. Long term retention of new phonetic categories. *Journal of acoustic society of America* 96 (4) : 2076-87.

Logan, J. S., Lively, S. E., and Pisoni, D. B. 1991. Training Japanese listeners to identify English /r/ and /l/ : A first report. *Journal of acoustic society of America* 89 : 874-86.

Smith, Antonio. 2003. Improving Adult Japanese' Ability to Perceive English /l/ and /r/ via Self-Study with a Forced-Choice-and-Feedback Computer Game.*Eibeikenkyu* 27: 67-85.

1   I would like to thank Ms. Kyoko Bando for transferring all original data from the student logs to Microsoft Excel. Her help is greatly appreciated. I would also like to thank Ms. Naoko Yamashita for distributing Ms. Bando's data onto a "calendar" format in Excel so that I could make a graphical representation of subjects' training session scores over the period from 4/22/03 to 10/13/03. In this format, subjects' learning curves can be easily seen and understood. I would also like to thank Ms. Yamashita for arranging subjects sequentially from lowest to highest starting score. Thus, Subject 1 has the lowest starting score, 30% correct, and the last subject, 43, has the highest, 100% correct.

2   Unfortunately, it does not work well for every subject. Notice that Subject 39 averaged 180 words per session (and did the program 115 times). Subject 39 forces one to acknowledge that the program does not always work well for everyone who uses it properly. However, it *might* be the case that Subject 39 had poor quality sound coming from her computer or has a hearing deficiency. It might also be the case that it is especially hard to get an average over 82%.