Ann.Bull.RILP(1989)No.23, 51-57

A COMPARISON OF EVALUATIONS BY AMERICAN AND JAPANESE LISTENERS OF ENGLISH SPOKEN BY JAPANESE SPEAKERS

Hiroshi Suzuki*, Ghen Ohyama** and Shigeru Kiritani

1. Intoduction

The first series of this study represented an attempt to find out which feature of the English speech uttered by Japanese should be modified in order to be judged more English-like, the duration of each sound, the pitch change or the intensity change. Each of these three prosodic features and various combinations of the three were replaced with the same prosodic feature or combinations of features of the same English sentence read by an American. The modified utterances were tape recorded, and a group of Americans listened to the recording and graded the "Englishness" of each utterance. The results obtained showed that the duration of each sound and pitch change contribute more to the utterance's being rated more English-like than intensity change does.

The present study attempts to discover if there is any difference between the judgment of American listeners and that of Japanese listeners in this respect.

2. Speech Samples

The sample sentence was so constructed that it would reveal weak spots in the speech of ordinary Japanese learning English: "None of us can leave as long as he stays with us." A native speaker of American English and 10 Japanese college students read aloud and recorded the sentence. Three utterances out of the ten by the Japanese students were judged as having a typical Japanese rhythm and were used as samples for the present study. While ordinary American speakers place stress on the four words, 'none,' 'leave,' 'long,' and 'stays' and not on any other, the three Japanese students put stress on these four words and also on the others, which made their utterances sound less English-like.

3. Method of Speech Analysis and Conversion

The recorded utterances were analyzed and converted by means of the PARCOR analysis and synthesis technique. The frequency range of the recorded utterances was limited with a 4.5kHz-low-pass filter. They were A/D converted with a sampling frequency of

* College of Arts and Sciences, University of Tokyo
** Hearing and Speech Perception Department, ATR

10 kHz and were stored in the computer. The following three parameters were estimated every 10 msec with analysis conditions of a 30-msec hamming window and 12 poles.

- o A parameter representing spectral components:
  K parameter (K)
  This parameter is closely related to articulation.

- o Parameters related to the source of the speech sound:
  Fundamental frequency or pitch (P)
  Intensity (Amplitude) (I)
  These parameters and the duration of each sound (D) are related to the prosody of speech.

The measurement of the duration of each sound was done by closely looking at the wave forms, the sound spectrograms, the formant frequencies, and the intensity (amplitude). The minimum time unit for each measurement was 10 msec, so that the measurements would synchronize with the parameters obtained by the PARCOR analysis. When a difference between the duration of each sound in the utterance by the American and that by the Japanese had to be made identical, each parameter was linearly interpolated. Based on the newly obtained parameters, speech was synthesized by means of the PARCOR synthesis technique. For example, if the duration of each sound in the speech of the Japanese was replaced with that of the American, the synthesized speech had sounds identical in duration to those of the American. However, the articulation, fundamental frequencies, and amplitudes of the sounds remained unchanged, that is, as in the original speech of the Japanese.

4. Synthesized Speech Samples and Listening Test

The assessment of the modified utterances was carried out in regards to the following five aspects.

1) Which contributes more to the utterance's being judged more English-like, the pronunciation of each sound or the combination of three prosodic features, i.e., duration, fundamental frequency change and intensity change?

2) Which contributes more, duration or fundamental frequency change?

3) How much does fundamental frequency change contribute?

4) How much does the duration component contribute?

5) How much does the intensity component contribute?

In order to get clues to these questions, the modified utterances were arranged in pairs, and the listeners were asked to judge which utterance in each pair sounded more English-like. The listeners were forced to choose one member in all the pairs.

Two groups of adult Americans and two groups of adult Japanese listened to the tape recording once. The seven American listeners in Group A (American Group A) were staying in Japan, and the 10 American listeners in Group B (American Group B) were staying in the U.S.A. The seven Japanese listeners (Japanese Group A) were students at a college in Tokyo, whereas the other Japanese Group (Japanese Group B) were in the Kansai district.

The combinations and the results are shown in Table 1.

5. Results and Discussion

5 - 1 American Listeners

The results are shown in percentage in the first and the second columns of Table 1: American Group A and American Group B.

The assessment of the results leads to the following findings.

1) As to pronunciation versus prosodic features, Group A valued prosody more than pronunciation, while Group B valued pronunciation more than prosody.

2) The fundamental frequency change contributed a little more to the utterance's being judged more English-like than the duration component.

3) The fundamental frequency change contributed greatly to the utterance's being judged more English-like.

4) The duration component also contributed greatly.

5) The intensity component contributed much less to the utterance's being judged more English-like than the other two components for Group A, but a multiplier effect was observed when it was combined with either of the other two components. For Group B, on the other hand, the same effect was seen when the intensity component was combined with the duration component, but not with the fundamental frequency change.

Those Americans who extensively associated with Japanese people and speak Japanese themselves tended to value pronunciation more than prosody. All the listeners in Group B were engaged in speech therapy, and this might be the reason why they put a greater value on pronunciation.

5 - 2 Japanese Listeners

The results are shown in percentage in the third and the fourth columns of Table 1: Japanese Group A and Japanese Group B.

1) Group A valued pronunciation more than prosody. This may be

due to the instruction they had received, in which a greater emphasis was placed on the pronunciation of each English sound. Group B, on the other hand, valued prosody more than pronunciation.

2) For Group A, there was no great difference between the effect of duration and that of fundamental frequency, but the latter became a greater cue when it was combined with intensity. The picture was quite different for Group B: fundamental frequency demonstrated a far greater contribution to the utterance's being judged more English-like than duration, whether or not it was combined with intensity.

3) The fundamental frequency contributed greatly to the utterance's being judged more English-like. This may have been due to a multiplier effect. Here again, Group B put a much greater value on the fundamental frequency than Group A.

4) Duration also contributed greatly. This may also have been due to a multiplier effect.

5) Intensity showed no contribution to the utterance's being judged more English-like. No multiplier effect was observed, either, contrary to the trend in the American Groups.

The two Japanese Groups showed fairly different trends from each other. This may be ascribed to the fact that the listeners in Group A live in Tokyo and those in Group B live in the Kansai district, where the tone system of Japanese there is fairly different from that of Tokyo. But further studies are called for to determine what really caused such a great difference among the Japanese groups.

6. Additional Listening Test

An additional listening test was conducted with pairs of newly added combinations of parameters in order to investigate the following three questions. (The sentence, speakers and the process of modification were exactly the same as in the previous tests.)

1) To what degree does each of the parameters contribute to the utterance's being judged more English-like?

2) To what degree does each combination of two of the parameters contribute?

3) To what degree does each of the parameters contribute when it is added (1) to another parameter and (2) to the combination of the other two parameters.

The listeners for this additional test were 21 Americans living in Japan (American Group C), 17 Japanese students from a college in Tokyo (Japanese Group C) and 21 Japanese students

from a college in the Kansai district (Japanese Group D). The combinations of parameters and the results of the listening test are shown in Table 2.

In order to compare the degree of the contribution of each parameter, preference scores for the three combinations--D vs. P, P vs. I, D vs. I--were determined and are shown in Table 3-a. It can be observed that fundamental frequency contributed most to the utterance's being judged more English-like in all the groups, and that intensity contributed least in two groups.

In the same way, a comparison was made as to the degree of contribution in the different combinations of the parameters, DI vs. PI, DP vs. DI and DP vs. PI, and is shown in Table 3-b. In descending order of contribution: the combination of fundamental frequency and duration, the combination of fundamental frequency and intensity, and the combination of duration and intensity.

When another parameter was added to one parameter or two, as can be seen in Table 2, fundamental frequency played the major role, duration followed, with intensity playing a very minor role in the utterance's being judged more English-like.

The results of this test again show the supremacy of a fundamental frequency change in the judgement of acceptability level, or "Englishness", and a large multiplier effect when a fundamental frequency change and duration component are combined.

The Japanese Group D demonstrated scores closer to the American Group C than did the Japanese Group C. This may have been due to the fact that the listeners in the Japanese Group D were college students majoring in English and had been trained in listening to English much more extensively than the listeners in the Japanese Group C, who were college students majoring in subjects other than English.

7. Summary and Conclusions

Several basic studies were conducted on the prosodic features of the English spoken by Japanese to discover which features should be altered and how they should be changed in order for their English to be judged more English-like. By means of the PARCOR analysis and synthesis technique, various combinations of three prosodic features, i.e., the duration of each sound, the fundamental frequency change and the intensity change in an English sentence uttered by Japanese speakers in a typically Japanese fashion, were replaced with the same combinations of the prosodic features of the same English sentence read by an American. Groups of Americans and groups of Japanese listened to the recording of the modified utterances and judged their acceptability level, or "Englishness", in a pair-comparison presentation.

It was found that prosodic features were important, and that

fundamental frequency change played a particularly important role in judging the acceptability level of English spoken by Japanese speakers. The duration component was nearly as important as fundamental frequency change. The intensity component was least important. Some differences were noticed between the judgments of the Americans and those of the Japanese, and also between the two groups of Japanese, which calls for further study. The results obtained in the present study must be confirmed with sentences of different types and with various speakers. The results may be little different with different types of listeners and depending on whether they have received training in speech or phonetics, and on whether they have been exposed to English spoken by Japanese for a long time. This also needs further study.

The research method used in this study can be applied to the study of the Japanese language by speakers of other languages. We wish to apply the same method to a study of Japanese sentences spoken by non-Japanese people.

Acknowledgements

This work has involved the contributions of a number of people. We would like to acknowledge the guidance of Dr. Hisashi Wakita, who initiated this study and encouraged us to go on with it, and the contribution of Dr. June E. Shoup who assisted in the interpretation of the results and in the planning of the tests. We are indebted to Ms. Kobayashi for her assistance in modifying the utterances and recording them. We also wish to acknowledge the assistance given by the listeners, who kindly and patiently judged the long series of stimuli, and the speakers who produced them.

References

1. H. Suzuki, 'An Acoustico-Phonetic Study to Improve the Prosodic Features of English Spoken by Japanese' (Oral Presentation at the Speech Study Meeting, Tokyo, March 20, 1986)

2. H. Suzuki, A. Takei, G. Ohyama and S. Kiritani, 'A Study on the Prosody of Japanese English' in "Collected Papers Presented at the Fall Meeting of the Acoustical Society of Japan, 1987" October 1987.

Table 1. Combinations of parameters and results of assessments in percentages in Test 1.
(K:spectrum, P:fundamental frequency, I:intensity, D:duration)

Combination of parameters	American Group A	American Group B	Japanese Group A	Japanese Group B
1) K vs DPI	23.8 vs 76.2	66.7 vs 33.3	54.8 vs 45.2	34.6 vs 65.4
2) D vs P	42.9 vs 57.1	40.0 vs 60.0	47.6 vs 52.4	25.9 vs 74.1
DI vs PI	47.6 vs 52.4	51.7 vs 48.3	42.9 vs 57.1	29.6 vs 70.4
3) D vs DP	7.1 vs 92.9	30.0 vs 70.0	31.0 vs 69.0	22.8 vs 77.2
4) P vs DP	28.6 vs 71.4	26.7 vs 73.3	40.5 vs 59.5	30.9 vs 69.1
5) DP vs DPI	28.6 vs 71.4	35.0 vs 65.0	54.8 vs 45.2	53.7 vs 46.3
D vs DI	38.1 vs 61.9	40.0 vs 60.0	47.6 vs 52.4	48.8 vs 51.2
P vs PI	35.7 vs 64.3	56.7 vs 43.3	50.0 vs 50.0	50.6 vs 49.4

Table 2. Combinations of replaced parameters and results of assessments in percentages in Test 2.
(K:spectrum, P:fundamental frequency, I:intensity, D:duration
N:no change)

Combination of parameters	American Group C	Japanese Group C	Japanese Group D
1) K vs DPI	41.3 vs 58.7	60.7 vs 39.3	24.0 vs 76.0
2) N vs P	32.5 vs 67.5	13.8 vs 86.2	27.3 vs 72.7
N vs D	47.6 vs 52.4	38.2 vs 61.8	32.7 vs 67.3
N vs I	47.6 vs 52.4	36.2 vs 63.8	57.3 vs 42.7
3) D vs P	43.7 vs 56.3	32.3 vs 67.7	42.9 vs 57.1
P vs I	74.6 vs 25.4	69.7 vs 30.3	68.3 vs 31.7
D vs I	56.3 vs 43.7	44.2 vs 55.8	66.7 vs 33.3
4) DI vs PI	33.3 vs 66.7	27.4 vs 72.6	40.5 vs 59.5
DP vs DI	73.8 vs 26.2	74.6 vs 25.4	73.0 vs 27.0
DP vs PI	73.8 vs 26.2	57.8 vs 42.2	71.4 vs 28.6
5) D vs DP	8.7 vs 91.3	14.8 vs 85.2	21.4 vs 78.6
I vs PI	27.0 vs 73.0	20.6 vs 79.4	32.5 vs 67.5
DI vs DPI	17.5 vs 82.5	15.7 vs 84.3	19.8 vs 80.2
P vs DP	27.0 vs 73.0	37.2 vs 62.8	27.8 vs 72.2
I vs DI	40.5 vs 59.5	49.1 vs 50.9	34.9 vs 65.1
PI vs DPI	27.0 vs 73.0	44.1 vs 55.9	25.4 vs 74.6
D vs DI	34.1 vs 65.9	48.1 vs 51.9	42.9 vs 57.1
P vs PI	55.6 vs 44.4	48.1 vs 51.9	45.2 vs 54.8
DP vs DPI	42.9 vs 57.1	45.1 vs 54.9	42.9 vs 57.1

Table 3-a. Preference scores (in percentages) for each parameter.

Parameter	American Group C	Japanese Group C	Japanese Group D
P	65.5	68.6	62.7
D	50.0	38.2	54.8
I	34.5	43.2	32.5

Table 3-b. Preference scores (in percentages) for combinations of parameters.

Parameter	American Group C	Japanese Group C	Japanese Group D
DP	73.8	66.1	72.2
PI	46.4	57.4	44.0
DI	29.8	26.5	33.7