An assessment of the utterances of some English vowels spoken by Japanese students

A survey at a junior college for women in Japan

Guido Izuta

Abstract
This work was aimed to carry out an assessment of how young female students pronounce some very basic vowels of the English language. The study was pursued at a junior college for women located in the northern region of Japan ('Tohoku'). Nearly all the subjects were Japanese female folks and just a few from the surrounding prefectures. To accomplish the purpose of the investigation, the formants of the utterances were analyzed and statistically compared with samples of the North American English sounds and their intensities were qualitatively examined as well. Finally, an attempt to interpret the data in terms of the tongue positioning framework as dealt in some phonetics research branches was made in order to get some insight from the results.

Keywords
(1) English vowels sounds, (2) utterance analysis, (3) English sounds pronounced by Japanese female students

1 Introduction
The English language learning has been one of the key issues in the context of the Japanese school curriculum that is revised and standardized every 10 years or so; and the recent ever growing trend has been to push the beginning year of English language training further down in the primary school work-study program. In fact, recent Japanese high school graduates come out from school with a background in English learning training that counts for one or two years in elementary, three years in junior high, and, finally, three more years in high school for the years in regular school education alone [12].

Despite all these years of language acquisition, the assimilation of the English sounds and words has been a major challenge for the students in general, so that attempts to overcome this pronunciation shortcoming have been carried out aimed either directly to assist the learners during this period (see for example [20]) or specifically to help college students improve speaking skills [11] [15] [16] [20]. On the other hand, there has been an increasing interest in assessing the English words sounds produced by Japanese speakers [18] [21].

These reports relying on the methodology of experimental phonology [9] [10] have provided a great deal of information on how Japanese people speak English. However, to the best of the author's knowledge, a picture of how students say the English vowels is not yet fully portrayed. For example,
it is not clear how close the Japanese language utterances are to the English vowels 'a', 'e', 'i', 'o' and 'u' when Japanese speakers' sounds are compared with those generated by native English speakers. Interestingly enough, as far as the spoken patterns of English speaking countries are concerned with, this kind of benchmarking is even far from being scholarly documented and/or widely available for public referencing.

Taking these facts into consideration, the purpose of this investigation is two-fold: (1) to make it clear how young Japanese female college students say the English vowels 'a', 'e', 'i', 'o' and 'u' by comparing them with samples of North American English utterances related to these vowels; (2) shed some light on the phonation strategy bearing on the formant and loudness analysis.

Finally, the remainder of the paper is organized as follows: in section 2, the experimental procedures and the data related to the beings taking part in it are presented in detail; the results of the statistical testing of the formants and the processed voice signals intensities are yielded in section 3; and outcomes are discussed in section 4.

2 Experiments

The experimental set up is presented in this section. It describes the subjects, the protocols and the data processing method. Unlike the non-native English speakers' voice signal that were computer measured on-site by the conventional digital measurement system, the native English speakers sound waves were collected from the Internet.

2.1 Method

The experimental procedure consists basically of three steps: (1) acquisition and analysis of subjects' voice signals by means of a personal computer based measurement system, (2) downloading native North American English speakers' speech signals, which are available over the Internet, and processing them, and (3) benchmarking the formants and the normalized intensities of the non-native speakers' voice signals against the native sounds. It is worth noting that the students were asked to produce the sounds of the English vowels as they were previously taught at school.

2.2 Subjects group of non-native English speakers

Twenty six college female students participated in the experiment. The age of the subjects ranged from 18 to 21; and, at the time, seven out of them were second graders majoring in English literature and language whereas 10 others were first graders, and the remaining 9 people were first graders enrolled in the social sciences course.

As far as their hometowns and the regional dialects are concerned, all the participants were born and raised in the northern region of Japan.

In addition, not only they all studied English language during the 3 years in junior high school as well as high school as part of the academic curriculum load established by the Ministry of Education,
Culture, Sports, Science and Technology of Japan [12], but they also had one or two more years of class activities aimed to give pupils very basic notions and feelings of English language in the primary school.

Finally, all the students have been engaged in at least one of the following extracurricular activities: going to private English lessons, training among other things speaking in order to prepare for either the 'Test in Practical English Proficiency' administered and produced by 'Eiken Foundation of Japan' [5] - which has been sort of a very common practice among the middle as well as high school students - or TOEIC or some other English proficiency tests, and studying abroad for a short period of time, which includes the three-week intensive English conversation course at a language school in the U.S. that the collegians have the opportunity to embark on as part of their optional classes.

2.3 Data acquisition from non-native speakers and processing procedure

The examinees were handed out a very simple list containing only the English vowels 'a', 'e', 'i', 'o' and 'u' without their phonetic transcriptions [8] or whatsoever. Prior to the recordings, they were given time to look up in English dictionaries and whatever audio-visual resources they wanted to check up on how to utter the vowels. However, for the sake of the experiment control and benchmark analysis, the audio samples consulted were deliberately confined to North American English sounds. Thus after enough practice time, the data acquisition sessions were actually carried out, in which the speakers were asked to (1) pronounce the vowels on the list, (2) pronounce each vowel three times for measurement, and (3) keep as much as possible their usual speaking rhythm.

In fact, a personal computer (equipped with processor Intel Pentium 2.5 GHz and running OS Microsoft Windows 7) based speech data acquisition system was used to digitally capture the voice waves. It had a commercial multi-channel sound mixer TASCAM US-322 attached to the computer, which in turn had connected to it an electronic condenser microphone SONY ECM-PCV800. The software utilized to collect the data was the free software named SoundEngine, whose signal intensity recording threshold was tuned to 48 dB to reject any speech signals lower than this level.

The data processing of the digital signals was performed on the free software Praat, which generated a bunch of text format files consisting of formants and intensities values to be later on visualized graphically and statistically analyzed on the worldwide available spreadsheet shareware Microsoft Excel 2010. As a matter of fact, the main feature of the freeware Praat that was used in the experiment is shows in figure 1. The two identical signals depicted in the upper half panel correspond to the components of the stereo sound recorded, and the superimposed signals plotted in the lower panel are the formants (spectrograms) and intensity (line path) of the speech signal. Basically, the interval to be selected for analysis is defined by inspecting carefully and coherently throughout the whole investigation the quantitative values of the voice signals, the spectrograms and intensity levels.
2.4 Collecting voice signals from native speakers and processing procedure

Unlike the experiment with non-native subjects, the native English speakers' signals were utterances recorded directly on the previously described personal computer from a variety of Internet sites [2] [3] [4] [5] [6] [7] [13] [23]. Care was taken to only single out North American female voices. Furthermore, in order to perform statistical test of different groups on mean and standard deviation, the minimum number of native subjects was set to 5 people.

Moreover, the data set, once collected, was dealt with akin to the procedure detailed in section 2.3.

3 Results

In the sequel, formant charts and normalized graphs of voice intensities, which are the cornerstones of the utterances characterization approach adopted in this investigation, are brought out and both the non-natives and natives group are reciprocally related to.

3.1 Formant charts

The graphs for each vowel were rendered in figures 2 to 6, respectively. Keeping in mind the discussions in the following sections on the characterization of the utterances based on the height as well as the
entire plot area whereas the circle points of the native subjects group (L1G) spread narrowly over a tall rectangular area with F2s cooped up in the frequency range of 2200-2500 Hz.

![Formant Charts F2 x F1 for Vowel 'a'. Square: Non-native Speakers (n=26); Circle: Native Speaker (n=7).](chart.png)

**Figure 2:** Formant charts F2 x F1 for vowel 'a'. Square: non-native speakers (n=26); circle: native speaker (n=7).

On the other hand, for the utterance of vowel 'e' as shown in figure 3, the circle points of L1G fit in a wide rectangular area whose height is the length between 300 Hz and 500 Hz and width varying from 1200 Hz and 3000 Hz; and L2G has high density of points around the center of the graph with F1 ranging from 350 Hz to 700 Hz, and F2 from 1700 Hz to 2600 Hz.

For the vowel 'i' (figure 4), both groups have to a great extent overlapping areas as their points are connected to each other in the same group to make up a closed region. In fact, the boundaries f2 between 1500 Hz and 1900 Hz and f1 from 600 Hz to 1000 Hz define the range of L2G whereas f2 from 1650 Hz to 1900 Hz and f1 from 700 Hz to 950 Hz comprise the points of L1G.

As far as the vowel 'o' is concerned, figure 5 reveals that the F2s and F1s of points in L2G fluctuate in the interval from 1000 Hz to 1900 Hz, and from 550 Hz to 800 Hz; respectively. Furthermore, the F2s and F1s related to L1G oscillates between 1200 Hz and 1500 Hz, and 400 Hz and 700 Hz; respectively.

Finally, figure 6 exhibits that F2s and F1s of points in L2G modulate in the spectra bounded by 1750 Hz and 2150 Hz, and 350 Hz and 650 Hz; respectively. Besides, the F2s and F1s of L1G go up from 1900 Hz to 2100 Hz, and from 300 Hz to 500 Hz; respectively.
Taken these into account, it turns out that a quantitative assessment of the statistical similarity or
closeness between the formants of the groups for each vowel need to be plied in order to understand
how the L2G is generating the voice sounds.

Figure 3: Formant charts F2 x F1 for vowel 'e'. Square: non-native speakers (n=26); circle: native
speaker (n=5).

Figure 4: Formant charts F2 x F1 for vowel 'i'. Square: non-native speakers (n=26); circle: native
speaker (n=5).
Figure 5: Formant charts F2 x F1 for vowel 'o'. Square: non-native speakers (n=25); circle: native speaker (n=5).

Figure 6: Formant charts F2 x F1 for vowel 'u'. Square: non-native speakers (n=25); circle: native speaker (n=6).
3.2 Statistical test of the formants

The outcomes of the statistical test for difference of formants between L2G and L1G are exposed here to make it clear whether there were any differences between the correspondent formants of both groups not only over the full length of the signals, but also between specific portions of the processed speech signals.

To begin with, figures 7 and 8 suggest that nevertheless there was a significant difference (p<0.05) between L2G-F1 and L1G-F1 over the head and tail portions 0-16% and 84-100% of the vowel 'a' utterance, the groups (L2G-F1 mean = 630.2 Hz, L1G-F1 mean = 593.2 Hz) were not statistically different as a whole. In addition, focusing on the formant F2 part, the graphs indicate a significant difference p<0.05 between the groups over every single portion tested for, which consequently established the overall difference between the groups (L2G-F2 mean = 2050.2 Hz, L1G-F2 mean = 2330.7 Hz).

On the contrary, significant difference for F1 (L2G-F1 mean = 479.4 Hz, L1G-F1 mean = 395.8 Hz) and no difference for F2 (L2G-F2 mean = 2140.2 Hz, L1G-F2 mean = 2234.2 Hz) is the grand picture for vowel 'e' as displayed in figure 9. Interestingly, the difference of F1 is mainly due to the third portion (50%-84%) where the groups (L2G-F1 mean = 440.4 Hz, L1G-F1 mean = 363.3 Hz) differed statistically (p<0.05) from each other as illustrated in figure 10. As for F2, none of the portions demonstrated to be different with respect to their counterparts.

Unlike the previous utterance test outputs, no differences were recognized on neither F1 (L2G-F1 mean = 813.0 Hz, L1G-F1 mean = 884.3 Hz) nor F2 (L2G-F2 mean = 1766.4 Hz, L1G-F2 mean = 1766.3 Hz) for the vowel 'i' as expressed in figure 11. However, figure 12 suggests that there were portions of F1 (50% - 84%) and F2 (0% - 16%) on which the test failed to show sensitivity to difference.

Now, the patterns related to the statistical tests shown in figures 13 and 15 are alike in the sense that the utterances of the vowels 'o' (L2G-F1 mean = 679.6 Hz, L1G-F1 mean = 558.6 Hz; L2G-F2 mean = 1306.2 Hz, L1G-F2 mean = 1416.0 Hz); and 'u' (L2G-F1 mean = 487.0 Hz, L1G-F1 mean = 378.1 Hz; L2G-F2 mean = 1969.1 Hz, L1G-F2 mean = 1995.3 Hz) were only different for formant F1 and not for F2. Furthermore, both vowels had similar patterns for the portions of F1 and F2 as visualized in figures 14 and 16.
Figure 7: Statistical test for vowel 'a'. Left half: L2G F1 (mean = 630.2, S.D. = 65.8) and L1G F1 (mean = 593.2, S.D. = 60.7). Right half: L2G F2 (mean = 2050.2, S.D. = 164.7) and L1G F2 (mean = 2330.7, S.D. = 170.0).

Figure 8: Statistical test for portions of vowel 'a'. Left half: 0% to 16% (L2G F1: mean = 709.2, S.D. = 100.3. L1G F1: mean = 807.9, S.D. = 154.9); 16% to 50% (L2G F1: mean = 629.1, S.D. = 77.4. L1G F1: mean = 633.5, S.D. = 51.3); 50% to 84% (L2G F1: mean = 557.1, S.D. = 90.7. L1G F1: mean = 493.7, S.D. = 52.4); and 84% to 100% (L2G F1: mean = 684.6, S.D. = 220.8. L1G F1: mean = 494.0, S.D. = 123.8). Right half: 0% to 16% (L2G F2: mean = 2048.6, S.D. = 183.1. L1G F2: mean = 2264.1, S.D. = 134.5); 16% to 50% (L2G F2: mean = 1976.2, S.D. = 243.0. L1G F2: mean = 2201.8, S.D. = 266.7); 50% to 84% (L2G F2: mean = 2068.3, S.D. = 212.1. L1G F2: mean = 2462.4, S.D. = 113.0); and 84% to 100% (L2G F2: mean = 2154.7, S.D. = 200.8. L1G F2: mean = 2410.1, S.D. = 352.9.)
Figure 9: Statistical test for vowel 'e'. Left half: L2G F1 (mean = 479.4, S.D. = 71.3) and L1G F1 (mean = 395.8, S.D. = 62.4). Right half: L2G F2 (mean = 2140.2, S.D. = 217.3), and L1G F2 (mean = 2234.2, S.D. = 690.8)

Figure 10: Statistical test for portions of vowel 'e'. Left half: 0% to 16% (L2G F1: mean = 538.3, S.D. = 113.3. L1G F1: mean = 457.4, S.D. = 95.8); 16% to 50% (L2G F1: mean = 421.6, S.D. = 54.9. L1G F1: mean = 396.0, S.D. = 82.1); 50% to 84% (L2G F1: mean = 440.4, S.D. = 76.6. L1G F1: mean = 363.3, S.D. = 53.4); and 84% to 100% (L2G F1: mean = 603.3, S.D. = 233.8. L1G F1: mean = 398.2, S.D. = 104.6). Right half: 0% to 16% (L2G F2: mean = 2175.6, S.D. = 255.0. L1G F2: mean = 2321.4, S.D. = 533.9); 16% to 50% (L2G F2: mean = 2154.0, S.D. = 277.0. L1G F2: mean = 2329.6, S.D. = 737.0); 50% to 84% (L2G F2: mean = 2129.1, S.D. = 252.9. L1G F2: mean = 2128.9, S.D. = 808.1); and 84% to 100% (L2G F2: mean = 2099.2, S.D. = 204.5. L1G F2: mean = 2155.0, S.D. = 579.2.)
Figure 11: Statistical test for vowel 'i'. Left half: L2G F1 (mean = 813.0, S.D. = 80.6) and L1G F1 (mean = 884.3, S.D. = 86.7). Right half: L2G F2 (mean = 1766.4, S.D. = 98.7), and L1G F2 (mean = 1766.3, S.D. = 79.6).

Figure 12: Statistical test for vowel 'i'. Left half: 0% to 16% (L2G F1: mean = 957.1, S.D. = 118.5. L1G F1: mean = 962.0, S.D. = 69.2); 16% to 50% (L2G F1: mean = 911.2, S.D. = 93.2. L1G F1: mean = 964.9, S.D. = 103.9); 50% to 84% (L2G F1: mean = 761.5, S.D. = 118.8. L1G F1: mean = 890.5, S.D. = 110.2); and 84% to 100% (L2G F1: mean = 604.7, S.D. = 123.8. L1G F1: mean = 649.7, S.D. = 85.8). Right half: 0% to 16% (L2G F2: mean = 1657.3, S.D. = 208.4. L1G F2: mean = 1510.8, S.D. = 41.5); 16% to 50% (L2G F2: mean = 1559.6, S.D. = 120.9. L1G F2: mean = 1495.4, S.D. = 144.5); 50% to 84% (L2G F2: mean = 1853.2, S.D. = 144.43. L1G F2: mean = 1932.0, S.D. = 153.6); and 84% to 100% (L2G F2: mean = 2109.0, S.D. = 151.3. L1G F2: mean = 2211.1, S.D. = 110.7).
Figure 13: Statistical test for vowel 'o'. Left half: L2G F1 (mean = 679.6, S.D. = 64.5) and L1G F1 (mean = 558.6, S.D. = 97.0). Right half: L2G F2 (mean = 1306.2, S.D. = 164.6), and L1G F2 (mean = 1416.0, S.D. = 99.5).

Figure 14: Statistical test for portions of vowel 'o'. Left half: 0% to 16% (L2G F1: mean = 779.9, S.D. = 83.7. L1G F1: mean = 680.2, S.D. = 112.5); 16% to 50% (L2G F1: mean = 673.0, S.D. = 78.0. L1G F1: mean = 565.5, S.D. = 126.7); 50% to 84% (L2G F1: mean = 643.0, S.D. = 80.6. L1G F1: mean = 464.4, S.D. = 129.9); and 84% to 100% (L2G F1: mean = 665.6, S.D. = 105.9. L1G F1: mean = 597.1, S.D. = 204.7). Right half: 0% to 16% (L2G F2: mean = 1401.9, S.D. = 238.7. L1G F2: mean = 1530.6, S.D. = 236.4); 16% to 50% (L2G F2: mean = 1121.8, S.D. = 153.0. L1G F2: mean = 1421.7, S.D. = 258.5); 50% to 84% (L2G F2: mean = 1220.9, S.D. = 291.7. L1G F2: mean = 1207.6, S.D. = 135.8); and 84% to 100% (L2G F2: mean = 1704.8, S.D. = 331.4. L1G F2: mean = 1660.1, S.D. = 614.8).
An assessment of the utterances of some English vowels spoken by Japanese students

Figure 15: Statistical test for vowel 'u'. Left half: L2G F1 (mean = 487.0, S.D. = 53.4) and L1G F1 (mean = 378.1, S.D. = 53.4). Right half: L2G F2 (mean = 1969.1, S.D. = 104.0), and L1G F2 (mean = 1995.3, S.D. = 76.3).

Figure 16: Statistical test for portions of vowel 'u'. Left half: 0% to 16% (L2G F1: mean = 540.3, S.D. = 132.0, L1G F1: mean = 303.1, S.D. = 54.5); 16% to 50% (L2G F1: mean = 450.1, S.D. = 47.3, L1G F1: mean = 332.2, S.D. = 27.1); 50% to 84% (L2G F1: mean = 447.7, S.D. = 36.8, L1G F1: mean = 379.1, S.D. = 50.6); and 84% to 100% (L2G F1: mean = 569.0, S.D. = 205.4, L1G F1: mean = 532.0, S.D. = 166.3). Right half: 0% to 16% (L2G F2: mean = 2158.0, S.D. = 20.03, L1G F2: mean = 2296.7, S.D. = 81.4); 16% to 50% (L2G F2: mean = 2133.3, S.D. = 151.4, L1G F2: mean = 2368.4, S.D. = 248.5); 50% to 84% (L2G F2: mean = 1767.9, S.D. = 208.4, L1G F2: mean = 1708.3, S.D. = 259.3); and 84% to 100% (L2G F2: mean = 1861.7, S.D. = 214.2, L1G F2: mean = 1549.0).
3.3 Normalized graphs of sound signals for intensity and duration

In this experiment, speech loudness graphs were investigated in order to provide some insight into the utterance strategy adopted by the non-native speakers. In what follows, the graphs of intensities were doubly normalized: firstly, the signal values were recalculate to shift the horizontal axis of the graph right in the middle position between the maximum value at the top and the minimum value at the valley of the graph plot; then the intensities were divided by the difference between these extremes.

As shown in figure 17, nevertheless some L2G subjects started uttering the vowel 'a' loudly without an evident fade-in interval, the stack of L2G and L1G graphs behaved homogeneously in the interval from 0% to 15%. However, the majority of L2G students kept the voice intensity at a relatively constant and high volume during the recording leading to a ‘plateau’ shape graph before fading out, whose breaking down point was at around 70% of the utterance length against the 50% estimated for L1G. Note there was not a well-defined behavior in the decreasing part of L1G.

![Normalized graphs of sound signals for intensity and duration](image)

Figure 17: Sound signals normalized for length and intensity - vowel 'a'.

Although a plateau of approximately 20% to 70% length range was also present in the bunch of L2G graphs as depicted in figure 18, it resembled the L1G bundle behavior, which came out in the interval 0% to 20% and went essentially down from 40% point onward.
On focusing on the vowel 'i', figure 19 reveals that the clusters of L2G and L1G plots overlapped over most of the time spectrum, except for the transition band comprising 0% to 15% period.

Figure 18: Sound signals normalized for length and intensity - vowel 'e'.

Figure 19: Sound signals normalized for length and intensity - vowel 'i'.

- 77 -
Now, figures 20 and 21 mainly indicate that vowels 'o' and 'u' had similar plateaux as their L2G and L1G activities were reciprocally compared. In addition, apart from the oscillations on the voice waves generated by L1G individuals, the L2G and L1G graphs can be assumed to be close for the same utterance.

Figure 20: Sound signals normalized for length and intensity - vowel 'o'.

Figure 21: Sound signals normalized for length and intensity - vowel 'u'.

4 Discussions

The statistical tests of F1 and F2 for difference between L2G and L1G are summarized in the following table:

<table>
<thead>
<tr>
<th>Vowel</th>
<th>high/low/same</th>
<th>front/back</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>same</td>
<td>front</td>
</tr>
<tr>
<td>E</td>
<td>low</td>
<td>same</td>
</tr>
<tr>
<td>I</td>
<td>same</td>
<td>same</td>
</tr>
<tr>
<td>O</td>
<td>low</td>
<td>same</td>
</tr>
<tr>
<td>U</td>
<td>low</td>
<td>same</td>
</tr>
</tbody>
</table>

Table 1: Tongue position of L2G compared to L1G. Here 'same' means that there is no statistical difference between the groups.

From the phonetics analysis point of view [9] [19], these results suggest that the L2G individuals tended to keep their tongues in a lower position than L1G peers for vowels 'e', 'o' and 'u'. As a matter of fact, vowels 'a' and 'i' showed no statistical difference in terms of tongue high/low positions. However, unlike the vowel 'i', which was indifferent to the frontness/backness, the vowel 'a' did indicate that L2G uttered placing the tongue in a frontal position. Note that the English vowel 'a' is pronounced with the tongue in a relatively low position, and L2G did not show sensitive difference. These show that there was a tendency for L2G students to lower the tongue as they uttered the vowels in focus. It is worth noting that the sound of vowel 'i' is equivalent to a Japanese diphthong sound.

Now, looking qualitatively at the length of the plateaux in the plots of the intensities shown in figures 17 through 21, they lead to the results drawn in table 2 below.

<table>
<thead>
<tr>
<th>Vowel</th>
<th>length of the plateau</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>not close</td>
</tr>
<tr>
<td>E</td>
<td>close</td>
</tr>
<tr>
<td>I</td>
<td>close</td>
</tr>
<tr>
<td>O</td>
<td>close</td>
</tr>
<tr>
<td>U</td>
<td>close</td>
</tr>
</tbody>
</table>

Table 2: Qualitative comparison of the plateau patterns in the intensities graphs.

Comparing the third column of table 1 with the second column of table 2, they hint to the possibility that, in general, the L2G subjects lengthened the steady state portion of the voice.
in order to articulate the formant F2. However, the price paid embracing this strategy was a statistical
difference of the formant F1, which roughly speaking on the grounds of phonetics field means the
tongue in a lower position. Further investigations are required to make it clear whether this approach
is correlated to the regional Japanese of the Tohoku region [14] or it is collectively peculiar to
Japanese female students.

5 Final remarks
Recent investigations on the English speaking skills acquired and developed by Japanese students
have focused mainly on the analysis of the word pronunciations and phrase speaking intonations as
well [1] [17] [24] [22]. They are very interesting and important on their own rights; pretty much so
that, taking for grant some of the data presented there, we can say that the result suggest on one hand
that, nevertheless English teachers are very well trained to teach in schools, their
speaking/pronunciation abilities are not correspondingly at the same level as their writing/grammar
comprehension counterparts. On the other hand, we see that middle and high school students reserve
just only a little, if any, time to practice English pronunciations and spellings. One of the reasons for
this is that English language has not been taught at schools as a second language acquisition aimed at
communication or broadening the cultural perspectives in a global framework; rather, it has likely
been handled just as a curricular discipline which is part of both the standard high school and college
entrance exams, which heavily emphasize the grammar, reading and writing abilities.

Despite in this work we focused on the utterance patterns of the five English vowels by young female
students with a little more experience than their average pears, the overall picture of the results
showed to be quite distinct from that yielded by the North American sounds group. The experimental
settings and conditions under which the experiment took place make it unreasonable to simply
interpret the results here bearing on the models mentioned in the previous paragraph. Indeed, the
results revealing the differences in tongue positioning when non-native and native English speakers
were compared suggests that a better understanding of how the tongue movement and positioning
patterns of Japanese speech are related to the English patterns is needed.

Obviously, due to the fact the Japanese educational system has neither adopted nor favored a
particular English speaking country as a reference for word spelling and intonation as well; one may
argue that such a study would depend on the English speaking country taken for reference. In
addition, arguments concerning the influence of different Japanese dialects, which shape more or less
the tongue movement, would be anything than correctly posed in this context. Fair enough and in fact,
these have been two of the crucial points in designing an adequate experimental set up and working
hypothesis in order to go further with our ongoing complementary investigations; and hopefully, the
results will be published elsewhere in a near future.
Acknowledgements

The authors would like to thanks the students of Yonezawa Women's College who provided us with the voice signals and helped us with the experiments. Last but not least, thanks also go to the staff members, colleagues and friends for their valuable support and cooperation.

References


