Does the production mechanism of the typical Japanese utterances affect the English speech?

Some insights from a phonetic analysis perspective

Guido Izuta

Yonezawa Women's Junior College

Department of Social Information

Abstract

In this work, we dealt with two fundamental issues in the English language acquisition by Japanese learners: (1) how does the production mechanism of typical Japanese vox influence the speech of English phonics ‘a’, ‘e’, ‘i’, ‘o’ and ‘u’? (2) What is the mimicking of the North American English utterances like? The approach adopted here to tackle them was the analysis of the formants, time length of the utterances and pitch. As a result, we found that these parameters play an important role in this framework.

Keywords

(1)Utterance Analysis. (2) Formant analysis. (3) English voces ‘a’, ‘e’, ‘i’, ‘o’ and ‘u’. (4) Japanese sounds <え>, <エー>, <えい>, <い>, <イー>, <いい>, <アイ>, <お>, <オー>, <オウ>, <を>, <ゆ> and <ユー>. (5) Comparison between Japanese and English voices.

1 Introduction

Over the past decades, the developments and advances in the digital signal processing field have enormously benefited the studies of human languages and their leanings in all sorts of ways. In particular, researchers working on the phonetics and phonological analysis have witnessed remarkable changes in the repertoire of available tools to investigate the mechanisms of voice production; and this has broadened the horizons of the quantitative assessment and characterization of utterances methodologies [1] [2].

In fact, the resulting successful achievements have prompted a new kind of paradigm in the second language learning education in the sense that it brought into the classroom a computer assisted learning paraphernalia to help students with their learning tasks and drills in every single level of the native-like pronunciation training and acquisition.

In Japan, there has been some research on the evaluation of these computer assisted systems as auxiliary educational instruments in English learning classes and working out laboratories (see for example [3] [4] [5] [6] [7].) In short, the mainstream of these approaches has primarily been concerned with the measurements of the improvements accomplished in characteristics as the pronunciations, accents, rhythms and fluency as compared to the native speakers of a pre-determined

English speaking country or region.

Notwithstanding this framework has shown to be fruitful in the development of second language classes, this work, however, focuses on the relationship between the English and Japanese utterances that sound somehow close or similar to most Japanese ears. More specifically, we are interested here in figuring out how some basic characteristics as formants and pitch [8] of the Japanese sounds affect, if so, the formation of the English sound that resembles it. Therefore, the aim of this study is twofold: (a) to compare the phonics of the English ‘a’, ‘e’, ‘i’, ‘o’ and ‘u’. generated by young Japanese female students with the ones by native North American English speakers as well as benchmark these voice signals against the Japanese sounds of <え>, <エー>, <えい>, <いぃ>, <イー>, <いい>, <アイ>, <お>, <オー>, <オウ>, <を>, <ゆ> and <ユー> in order to understand how Japanese students produce these different sounds. (b) To investigate whether there is a well-defined pitch modulation strategy adopted by the Japanese and native speakers as they make the utterances.

To reach the goals proposed hitherto, we designed an experimental setup to make digital records of the sounds voiced by our volunteers as well as be able to later process these signals. The formants, pitches and the signal durations were analyzed and statistically tested between the groups in order to check for differences pair wisely. Unlike the others, the results related to the pitches presented here are, however, only qualitative graphs.

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 inference testing of the formants and the processed voice signals intensities are yielded in section 3; and outcomes are discussed in section 4.

2 Experimental procedure

This section presents the experimental setup used in this work. So, it describes the subjects, the experimental protocols and the digital data processing procedure carried out to analyze the utterances.

2.1 Data acquisition

The voice sounds were obtained from three different groups, each one composed by six subjects. The constituents of two out of three groups were young Japanese female students aged 18 through 20 whereas the remaining group was basically a set of digital utterances made by six different young ladies who were native North American English female speakers. The Japanese collegians were all attending to a two-year junior college at the time of the experiment. The details of the groups and the experimental protocol are written up as follows.

The group 1 consisted of six Japanese female college students majoring in social sciences with only ordinary background in English language just as taught in junior high and high schools of the standard Japanese educational system. None of them had ever been abroad or engaged in any extra-curricular activities as English conversation classes or private courses. As far as their places of origin are

concerned, three schoolgirls were from Yamagata Prefecture and the others from Iwate, Akita and Aomori Prefectures, respectively. After some training sessions in order to speak as possible as the standard Japanese without the accent or the rhythms peculiar to their regions, they were instructed to pronounce the Japanese sounds of the following words: <え>, <エー>, <えい>, <い>, <イー>, <いい>, <アイ>, <お>, <オー>, <オウ>, <を>,<ゆ> and<ユー>. The voices were repeated several times and recorded with a digital computer system.

The group 2 also contained six Japanese female volunteers only. They were all students from the English language department with that entire educational milieu until then and some experience studying abroad, mainly in the U.S. for at least a couple of weeks. They were selected to pair up the proveniences of their peers in group 1 in an attempt to add up the same amount of linguistic regionalism. Their experimental tasks were to utter the phonics <a>, <e>, <i>, <o> and <u>. Training sessions were performed aimed at mimicking the native English speakers' utterances as perfect as they could perform. Data acquisitions of several trials were achieved for later digital processing and assessment.

The group 3 was not exactly a party of six beings participating in the experiments, but rather a set of digital voice signals collected from a variety of internet web sites [9] - [15]. Only North American female voices were selected to take part in the experiment, and preference was given to sounds which were available with multiple copy samples available from the same speaker, allowing us to compute average voice signals.

2.2 Data processing

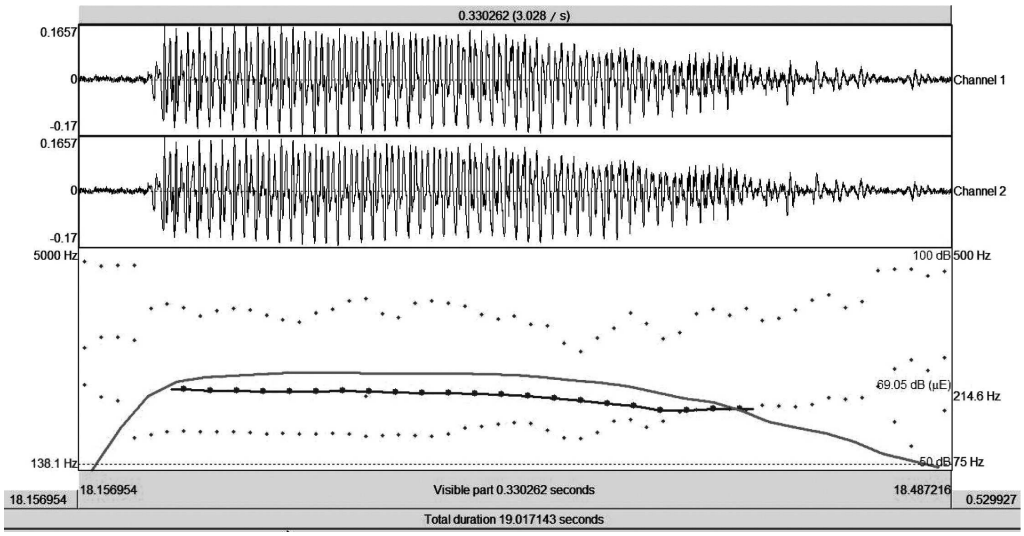

Figure 1: An overview of data processing on Praat

The data processing system consisted of a personal computer (equipped with Intel Pentium 2.5 GHz processor and running on Microsoft Windows 7) which had a commercial multi-channel sound mixer

TASCAM US-322 attached to its USB port. The subjects uttered to an electronic condenser microphone SONY ECM-PCV800 connected to the mixer. The kernel of the data acquisition environment was the software ‘Sound Engine’ (a freeware available widely on the web) set to cut off signals lower than 48 dB; and another freeware, namely ‘Praat’, was used to process the digital data as shown in figure 1. The graphics and the statistical testing were made on Microsoft Excel 2010.

3 Results

In this section we provide the outcomes of the data processing. First, the formants F1 and F2 of the utterances are plotted in a single graph for each case. Then they are tested for statistical differences with one-side confidence set to $p < 0.05$. As the duration time of many Japanese sounds is of fundamental importance in this language, this component is also analyzed here. Finally, the Japanese sounds along with the phonics ‘a’, ‘e’, ‘i’, ‘o’ and ‘u’ of both Japanese and native speakers are qualitatively compared for the pitch changes during the voice production in order to check their patterns and relationships.

3.1 Formants and durations of the utterances <え>, <エー>, <えい> and ‘a’

Figure 2 shows the plots of the formants for <え>, <エー>, <えい> and ‘a’. Note that the native sounds located between the utterances of ‘a’ made by the Japanese speakers and the sounds of <え>, <エー>, <えい>. More specifically, native sounds had frequencies around 2400 Hz and 600 Hz for F2 and F1, respectively; whereas the phonic ‘a’ uttered by Japanese students vibrated at frequencies ranging from 1700 to 2300 Hz and 600 to 800 Hz for F2 and F1. <え> and <エー> were scattered widely with <え> in a frequency region comprising the intervals of 2600 to 3000 Hz in the horizontal axis and 700 to 1100 Hz in the vertical axis. The frequency range of <エー> had a portion overlapping <え> and then extending further to higher frequencies in both axes. <えい> concentrated relatively at the bottom left of the axes.

The results of the statistical test for F1 are depicted in Figure 3. It turned out that the pair <え> and <エー> are sub sets of a common set of signals. Likewise, the couple <エー> and <えい> are elements of the same sample set. This suggests that the vertical positions of the tongue vary slightly when generating these sounds; however, if we focused on <え> and <エー>, we could not recognize clearly any differences in their vertical positions. The same rationale applies to <エー> and <えい>. It is worth noting that we cannot draw any conclusions for the pair <え> and <えい>. Actually, they came from different sample sets from the statistical point of view.

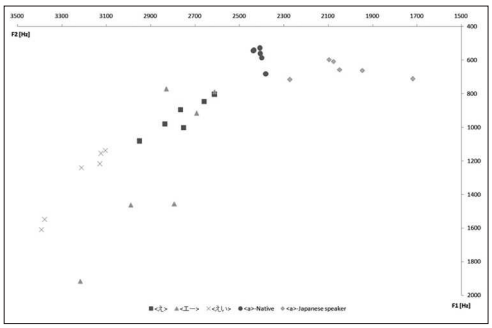

Figure 2: F2 x F1 graphs of the utterances <

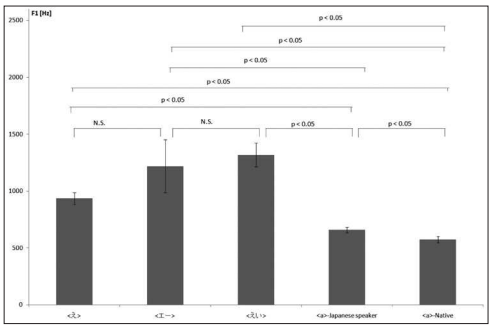

Figure 3: Results of the statistical significance tests for formants F1 of <え> 936 ± 105 Hz, <エー> 1219 ± 464 Hz, <えい> 1319 ± 207 Hz, Japanese <a> 659 ± 50 Hz, and native <a> 575 ± 56 Hz.

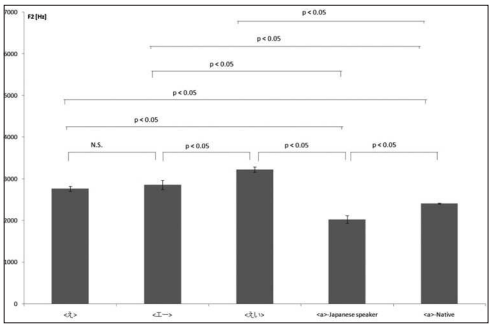

Figure 4: Results of the statistical significance tests for formants F2 of <え> 2762 ± 122 Hz, <エー> 2856 ± 464 Hz, <えい> 3223 ± 130 Hz, Japanese <a> 130 ± 184 Hz, and native <a> 2411 ± 21 Hz.

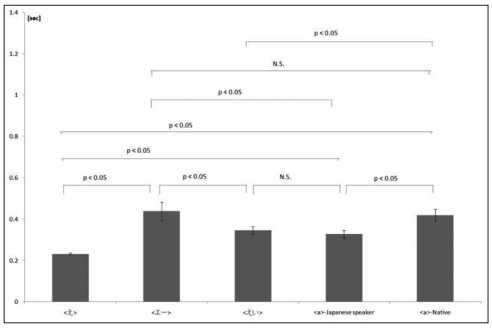

Figure 5: Results of the statistical significance tests for the duration of <え> 0.23 ± 0.01 Hz, <エー> 0.44 ± 0.09 Hz, <えい> 0.35 ± 0.04 Hz, Japanese <a> 0.33 ± 0.04 Hz, and native <a> 0.42 ±0.06 Hz.

Remark: Number of subjects = 6; each Japanese subject's data = average of 3 trials; the natives' data = average of several trials whenever possible.

The formants F2 are yielded in figure 4. From these plots, we see that <え> and <エー> are not significantly different. Now, recalling Figure 3 and calling up the results of Figure 5, we see that the only difference between them is the duration of the utterances. The durations of <えい> and <a> are as shown in Figure 5 and they were ‘N.S.’ only for this parameter. Summing up these outcomes, we have in general that neither of these sounds are correlated in the sense that they came out of the same set of utterances.

3.2 Formants and durations of the utterances <い>, <イ ー>, <いい> and <e>

Both Japanese and native speakers emitted the English sound <e> with formants F2 in the range from 1200 to 3000 Hz and F1 from 200 to 700 Hz (Figure 6.) The native sounds, however, were clustered in the middle part of the region whereas their Japanese peers spanned over a larger area. Yet, the <い>, <イ ー> and <いい> sounds evolved at higher frequencies for both F2 and F1, where the former varied from 2400 to 4000 Hz, and the latter from 200 to 2700 Hz. Interestingly, they all dispersed along a descending diagonal line with the frequency ranges overlapping each other. In terms of tongue positioning, it tells us that these Japanese sounds are uttered with the tongue in a lower position; or to put it the other way, with the mouth slightly closed.

Statistical tests showed that Japanese and native speakers provided F2's (Figure 8) and voice durations that were ‘N.S.’ (Figure 9), but not for F1 (Figure 7.) In fact, the mean values of F2 for On the other hand, the pairs <い> and <イ ー> as well as <イ ー> and <いい> were all ‘N.S.’ for both F1 - shown in Figure 7 - and F2 (as illustrated in Figure 8), but not for the durations (Figure 9.)

These results indicate, first, that <い>and <イ ー> are uttered in the same way and they are differentiated exclusively by the their durations. This concept is reasonable considering that the Japanese people are told to practice in this way in their very early ages of language learning. Second, <イ ー> and <いい> are made placing the tongue at about the same height, but the back-front positions and durations distinguish one from the other. Finally, based on the visual inspection, the difference in F1 for ‘a’ sound generated by the schoolgirls and the native sounds relies apparently on the openness of the mouth.

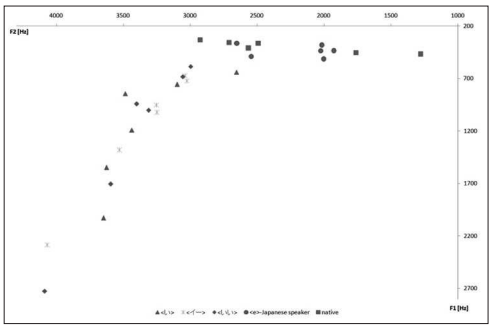

Figure 6: F2 x F1 graphics of the utterances <いぃ>, <イー>, <いぃいぃ>, Japanese-<e>, and native-<e>.

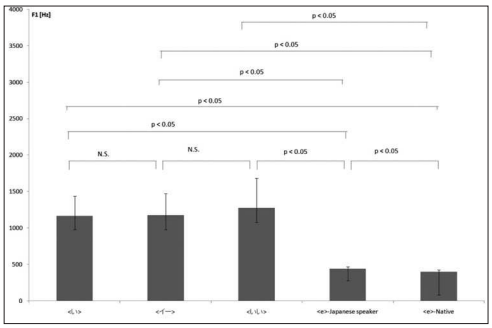

Figure 7: Results of the statistical significance tests for formants F1 of <いぃ> 1168 ± 536 Hz, <イー> 1174 ± 601 Hz, <いぃいぃ> 1275 ± 812 Hz, Japanese <e> 439 ± 59 Hz, and native <e> 399 ± 56 Hz.

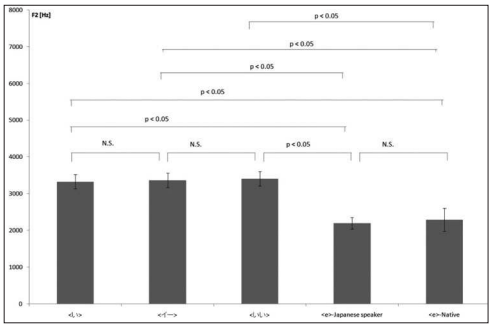

Figure 8: Results of the statistical significance tests for formants F2 of <いぃ> 3325 ± 383 Hz, <イー> 3361 ± 393 Hz, <いぃいぃ> 3407 ± 400 Hz, Japanese <e> 2195 ± 316 Hz, and native <e> 2289 ± 632 Hz

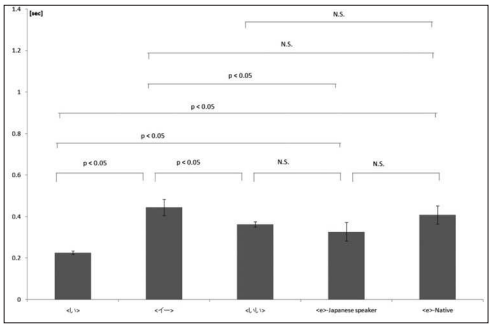

Figure 9: Statistical tests for the duration of the sounds <いぃ> 0.226 ± 0.017 Hz, <イー> 0.444 ± 0.079 Hz, <いぃいぃ> 0.363 ± 0.026 Hz, Japanese <e> 0.326 ± 0.089 Hz, and native. <e> 0.408 ± 0.088 Hz.

Remark: Number of subjects = 6; each Japanese subject's data = average of 3 trials; the natives' data = average of several trials whenever possible.

3.3 Formants and durations of the utterances <アイ> and <i>

As shown in Figure 10, the formants F2 and F1 of <アイ> had a relatively high frequency range with F2 varying from 2100 to 3200 Hz, and F1 from 600 to 1700 Hz. In contrast, the signals of <i> were concentrated in the upper right zone of the graph. Nevertheless, <i> spoken by the students and native speaker were ‘N.S.’ for only F2 (Figure 12), and statistically different for F1 (Figure 11) and the durations (Figure 13). In addition, <アイ> and natives’ <i> were ‘N.S.’ for F1 as drawn in figure 11; and <アイ> and students’ <i> were ‘N.S.’ for the durations as represented in figure 12.

As in the previous subsections, the Japanese sounds have both F1 and F2 at higher frequencies, so that as the students pronounce the English words, some of the parameters tend to be statistically different.

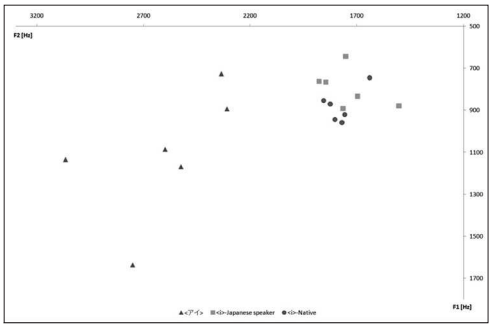

Figure 10: F2 x F1 graphs of the utterances <アイ>, Japanese <i> and native <i>.

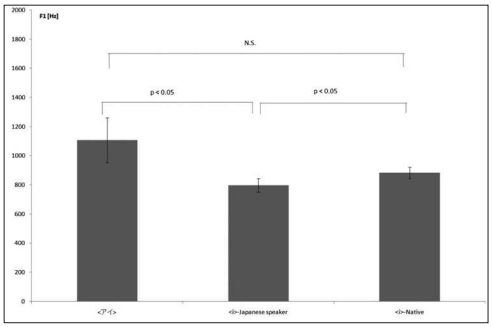

Figure 11: Results of the statistical significance tests for formants F1 of <アイ> 1108 ± 308 Hz, Japanese <i> 796 ± 93 Hz, and native <i> 882 ± 78 Hz.

Remark: Number of subjects = 6; each Japanese subject's data = average of 3 trials; the natives' data = average of several trials whenever possible.

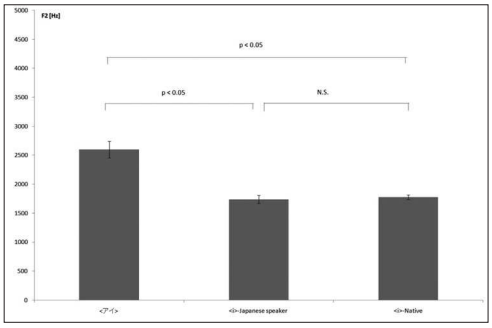

Figure 12: Results of the statistical significance tests for formants F2 of <アイ> 2599 ± 283 Hz, Japanese <i> 1740 ± 134 Hz, and native <i> 1776 ± 75 Hz.

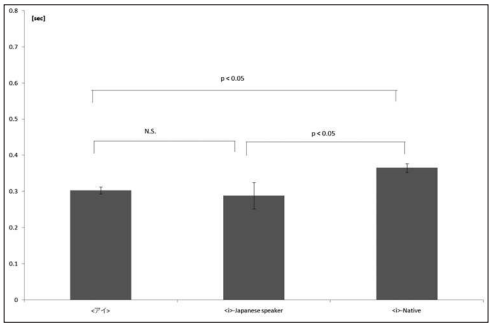

Figure 13: Statistical significance tests for the duration of the sounds <アイ> 0.303 ± 0.019 Hz, Japanese <i> 0.289 ± 0.073 Hz, and native <i> 0.365 ± 0.024 Hz.

Remark: Number of subjects = 6; each Japanese subject's data = average of 3 trials; the natives' data = average of several trials whenever possible.

3.4 Formants and durations of the utterances <お>, <オー>, <オウ>, <を> and <o>

Figure 14 displays a distinct distribution pattern with <お>, <オー>, <オウ> and <を> plots on the bottom left corner of the graph and the graph points of the phonic <o> on the upper right side. First, looking at the sound <o> voiced by Japanese students and native speakers were ‘N.S.’ for F1 as pictured in Figure 15, F2 as in Figure 16 and the duration as in Figure 17. The couple <お> and <オー> were ‘N.S.’ for F1 and F2 as in Figures 15 and 16, but statistically different for the durations (Figure 17.) Still, <オー> and <オウ> as well as <オウ> and <を> were also ‘N.S.’ as seen in Figures 15 and 16. All other combinations turned out to be statistically different for the formants.

Despite the fact that the Japanese sounds were situated relatively far from the English phonic <o> on the graph in Figure 14, the students were able to reproduce satisfactorily the native sounds in what concerns to the formants.

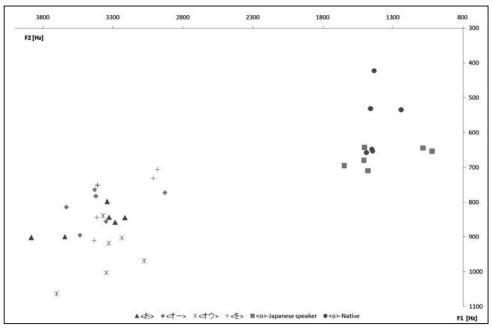

Figure 14: F2 x F1 graphs of the utterances <お>, <オー>, <オウ>, <を>, Japanese <o>, and native <o>.

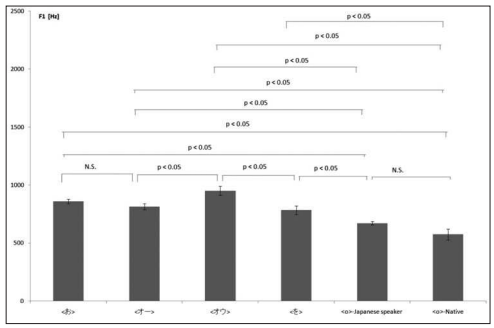

Figure 15: Results of the statistical significance tests for formants F1 of <お> 858 ± 39 Hz, <オー> 814 ± 52 Hz, <オウ> 949 ± 79 Hz, <を> 782 ± 78 Hz, Japanese <o> 671 ± 28 Hz, and native <o> 574 ± 95 Hz

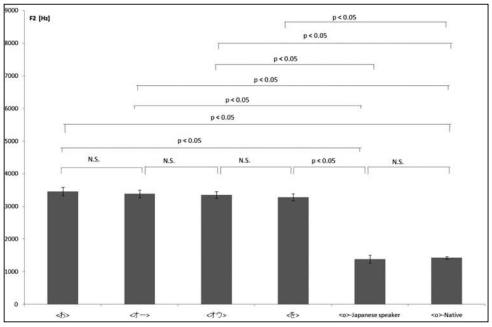

Figure 16: Results of the statistical significance tests for formants F2 of <お> 3449 ± 259 Hz, <オー> 3384 ± 244 Hz, <オウ> 3345 ± 206 Hz, <を> 3278 ± 217 Hz, Japanese <o> 1374 ± 257 Hz, and native <o> 1421 ± 90 Hz

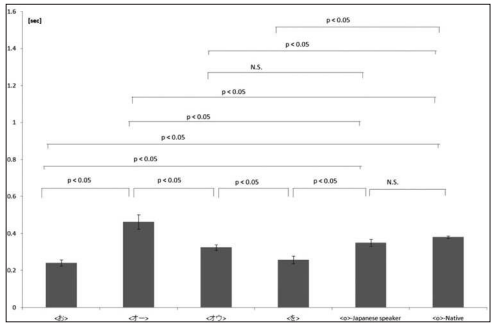

Figure 17: Statistical tests for the duration of the sounds <お> 0.242 ± 0.033 Hz, <オー> 0.463 ± 0.077 Hz, <オウ> 0.325 ± 0.031 Hz, <を> 0.257 ± 0.042 Hz, Japanese <o> 0.350 ± 0.036 Hz, and native <o> 0.380 ± 0.012 Hz.

Remark: Number of subjects = 6; each Japanese subject's data = average of 3 trials; the natives' data = average of several trials whenever possible.

3.5 Formants and durations of the utterances <ゆ>, <ユー> and <u>

The graph of the formants F2 and F1 exhibited in Figure 18 have two distinct pieces of points: an aggregation of points around F2=2000 Hz and F1=400 Hz, which are related to the English phonic <u> uttered by both Japanese students and native speakers; and a bunch of points along a straight line starting at around the coordinates F2=2500 Hz and F1=700Hz, and prolonging itself all the way down to the points F2=400 Hz and F1=2200 Hz. Focusing on the chunk of points, we see that the students and the native phonics were ‘N.S.’ for only F2 as easily checked in Figures 19 - 21. This result affects a bit our intuition as we see the plots of F2 and F1, because one expect expect the rightmost point in Figure 18 leading to ‘N.S.’ for F2 and not for F1. Statistically speaking, such a case is anything than a reasonable possibility. Now, <ゆ> and <ユー> differ statistically for the duration (Figure 21), but not for the formants (Figures 19 and 20) as one would obviously expect in a daily conversation.

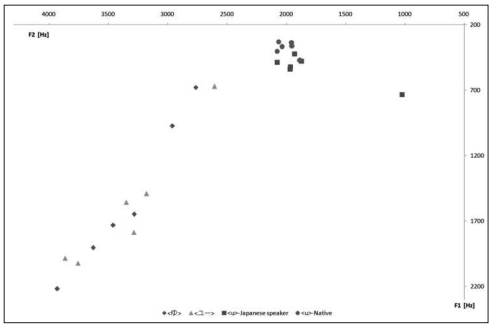

Figure 18: F2 x F1 graphs of the utterances <ゆ>, <ユー>, Japanese <u>, and native <u>.

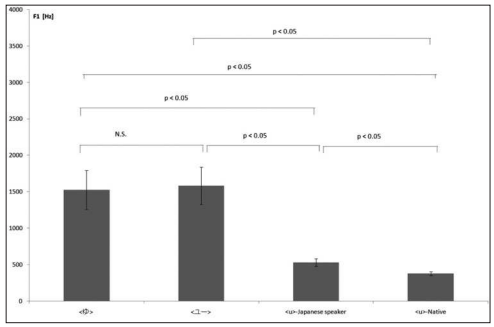

Figure 19: Results of the statistical significance tests for formants F1 of <ゆ> 3337 ± 355 Hz, <ユー> 3340 ± 414 Hz, Japanese <u> 1807 ± 430 Hz, and native <u>1995 ± 76 Hz.

Remark: Number of subjects = 6; each Japanese subject's data = average of 3 trials; the natives' data = average of several trials whenever possible.

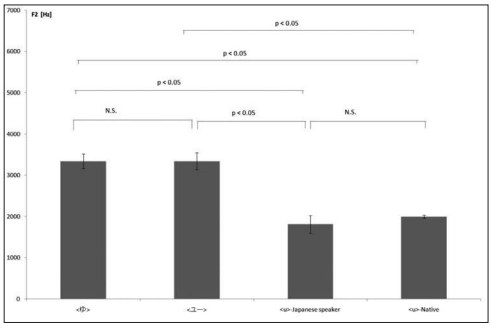

Figure 20: Results of the statistical significance tests for formants F2 of <ゆ> 3337 ± 355 Hz, <ユー> 3340 ± 414 Hz, Japanese <u> 1807 ± 430 Hz, and native <u> 1995 ± 76 Hz.

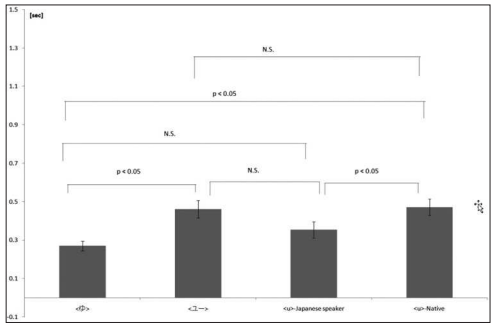

Figure 21: Statistical significance tests for the duration of the sounds <ゆ> 0.270 ± 0.050 Hz, <ユー> 0.461 ± 0.091 Hz, Japanese <u> 0.353 ± 0.085 Hz, and native <u> 0.472 ± 0.085 Hz.

Remark: Number of subjects = 6; each Japanese subject's data = average of 3 trials; the natives' data = average of several trials whenever possible.

3.6 Pitch analysis

The plots of the pitches for <え>, <エー>, <えい> and <a> are yielded in Figures 22 through 26. Ruling out the transient portions of the signals, which are considered in engineering to be the first and the last 10% of the graphic segments, <え>, <エー>, as a whole, behaved flatly almost making up curves parallel to the horizontal axis whereas the signals of the sounds <えい> waved widely up and down while moving downward as the time goes by, as shown in Figure 24. In fact, this sound production mechanism was also detected in the utterances of ‘a’ made by students (Figure 25.). Although the phonics <a> uttered by the natives kept this right-down inclination tendency, the pitches were not wave shaped as in the previous case.

Interestingly, the behavior of the plots of <い>, <イー>, <いい> delineated in Figures 27 - 29 mirrored to some extent the graphics of <え>, <エー>, <えい>: the first two paralleling the horizontal axis and the last one descending to the bottom right corner. As for the phonics <e> made by the students and the natives, all the plots presented in overall a parabolic behavior, which suggest that, the Japanese speakers modulated the pitches satisfactorily.

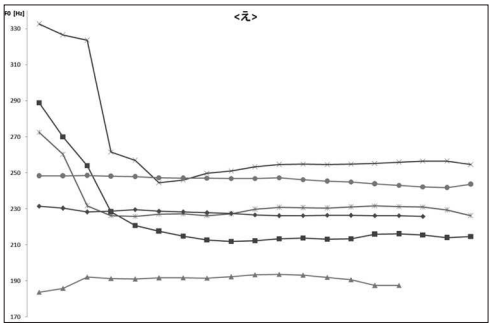

Figure 22: Pitch variations of the utterance <え>. N=6. Horizontal axis means the time evolution of the points.

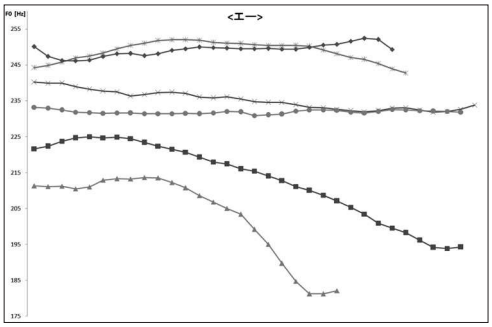

Figure 23: Pitch variations of the utterance <エー>. N=6. Horizontal axis means the time evolution of the points.

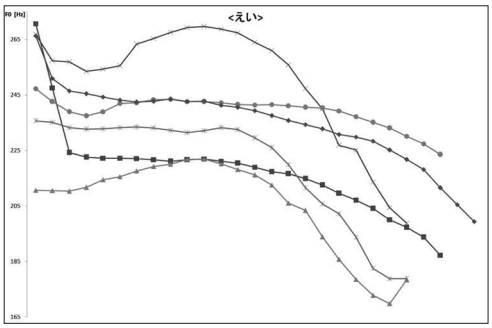

Figure 24: Pitch variations of the utterance <えい>. N=6. Horizontal axis means the time evolution of the points.

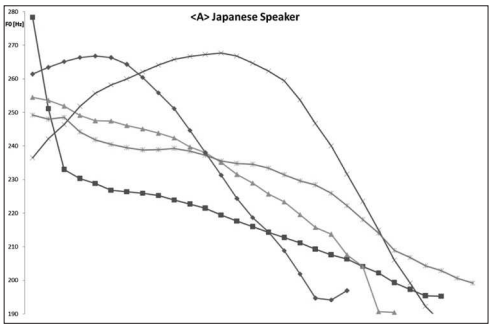

Figure 25: Pitch variations of the utterance <a> for Japanese speakers. N=6. Horizontal axis means the time evolution of the points.

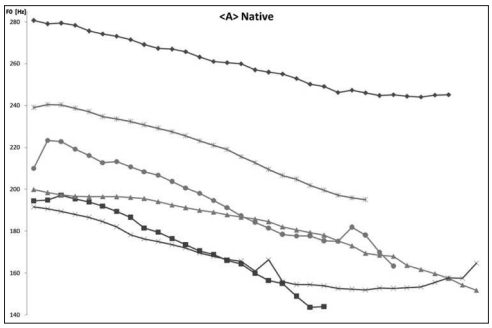

Figure 26: Pitch variations of the utterance <a> for native speakers. N=6. Horizontal axis means the time evolution of the points.

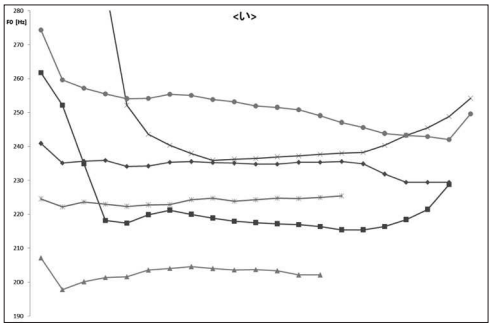

Figure 27: Pitch variations of the utterance <い>. N=6. Horizontal axis means the time evolution of the digital points.

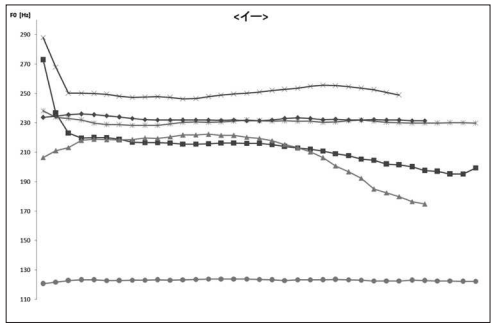

Figure 28: Pitch variations of the utterance <イー>. N=6. Horizontal axis means the time evolution of the digital points.

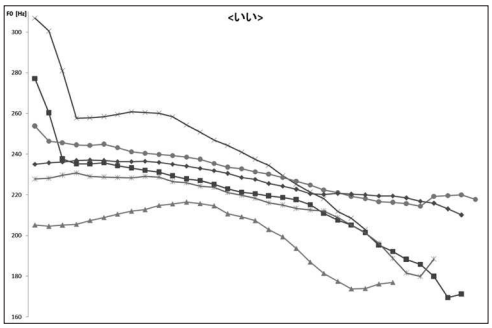

Figure 29: Pitch variations of the utterance <いい>. N=6. Horizontal axis means the time evolution of the digital points.

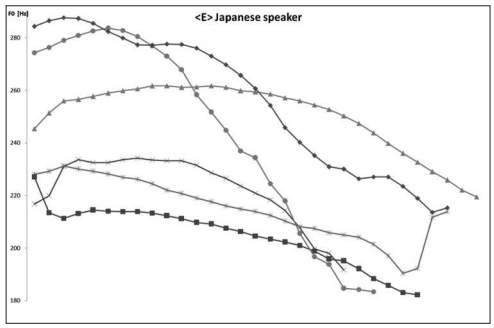

Figure 30: Pitch variations of the utterance <e> for Japanese speakers. N=6. Horizontal axis means the time evolution of the points.

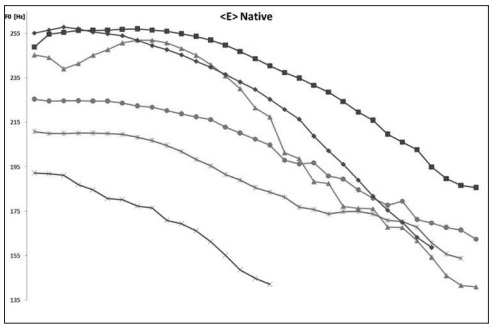

Figure 31: Pitch variations of the utterance <e> for native speakers. N=6. Horizontal axis means the time evolution of the digital points.

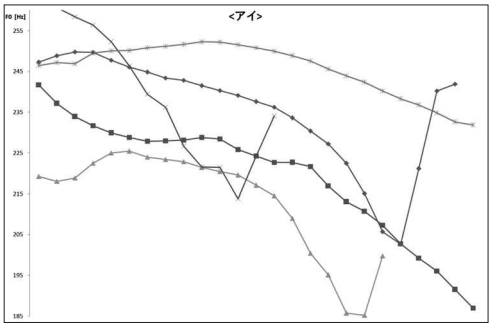

Figure 32: Pitch variations of the utterance <アイ>. N=5. Horizontal axis means the time evolution of the digital points.

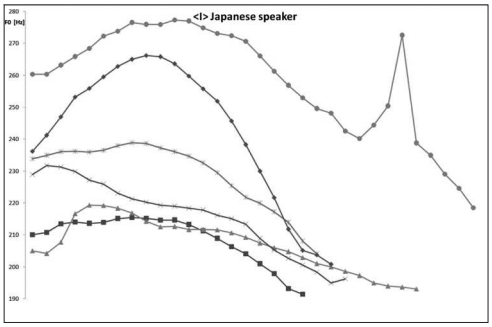

Figure 33: Pitch variations of the utterance <i> for Japanese speakers N=6. Horizontal axis means the time evolution of the digital points.

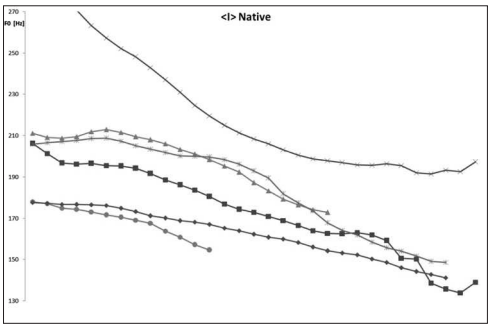

Figure 34: Pitch variations of the utterance <i> for native speakers. N=6. Horizontal axis means the time evolution of the digital points.

The pitches of <アイ> and <i> are presented in Figures 32 to 34. For <アイ>, ruling out the transient fractions and number of points, the curves could be compared to the phonic <i> voiced by the natives. However, the utterances <i> spoken by the students varied largely their pitches. Thus, it may happen that the pitch graphs of the utterances <i> articulated by both groups were a by-product of a different voice production strategy.

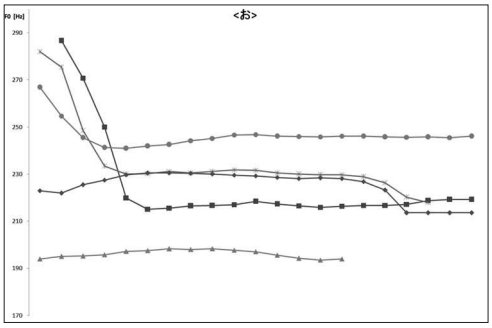

Figure 35: Pitch variations of the utterance <お>. N=6. Horizontal axis means the time evolution of the points.

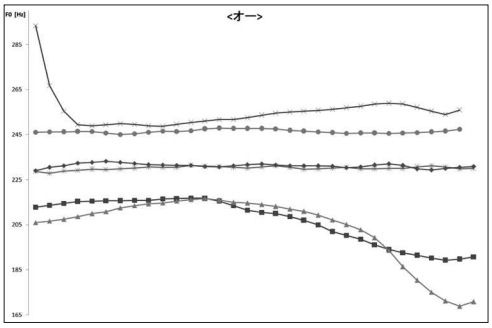

Figure 36: Pitch variations of the utterance <オー>. N=6. Horizontal axis means the time evolution of the points.

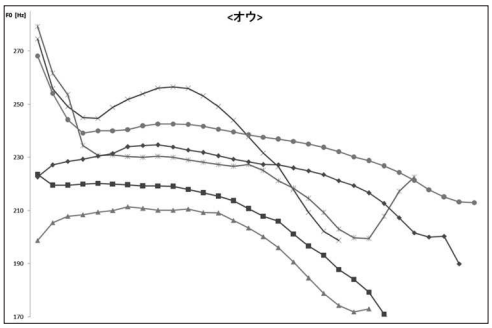

Figure 37: Pitch variations of the utterance <オウ>. N=6. Horizontal axis means the time evolution of the points.

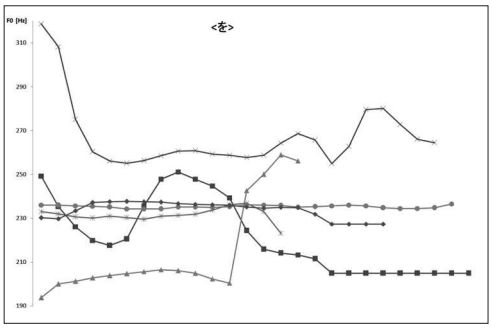

Figure 38: Pitch variations of the utterance <を>. N=6. Horizontal axis means the time evolution of the points.

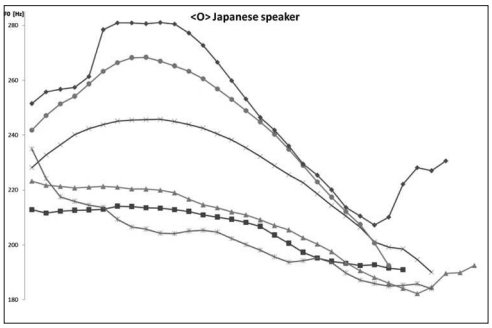

Figure 39: Pitch variations of the utterance <o> for Japanese speakers. N=6. Horizontal axis means the time evolution of the points.

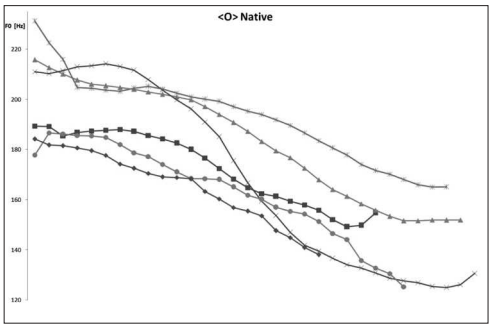

Figure 40: Pitch variations of the utterance <o> for native speakers. N= 6. Horizontal axis means the time evolution of the points.

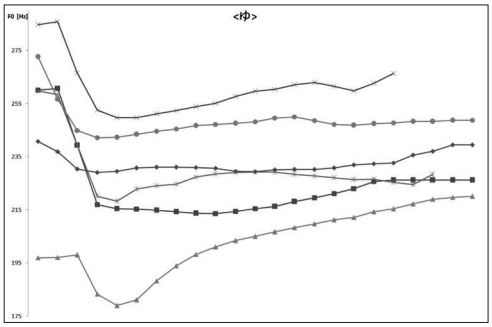

Figure 41: Pitch variations of the utterance <ゆ>. N=6. Horizontal axis means the time evolution of the digital points.

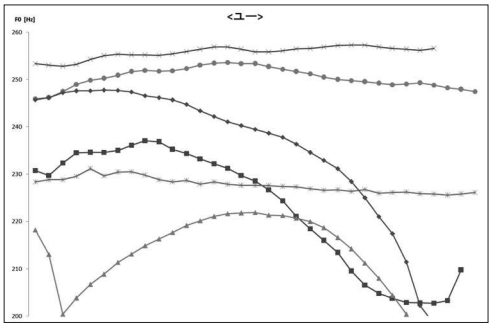

Figure 42: Pitch variations of the utterance <ユー>. N=6. Horizontal axis means the time evolution of the digital points.

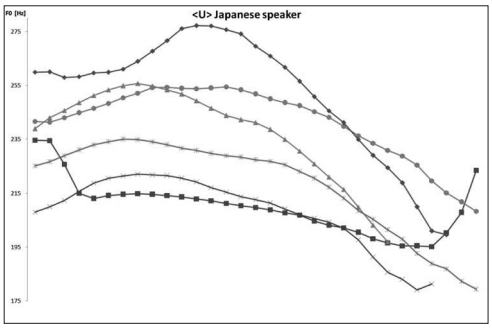

Figure 43: Pitch variations of the utterance <u> for Japanese speakers. N=6. Horizontal axis means the time evolution of the digital points.

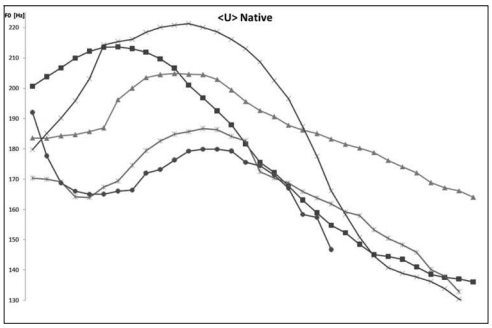

Figure 44: Pitch variations of the utterance <u> for native speakers. N=6. Horizontal axis means the time evolution of the digital points.

A close look at the pitches of <お>, <オー>, <オウ>, <を> and <o> in Figures 35 - 40 as well as <ゆ>, <ユー> and <u> in Figures 41 - 44, one concludes that the discussions so far also apply to these cases.

Taking all the results into account, we reason out that the students tried to modulate the pitch as they made English utterances; however, since the Japanese words, despite their supposed closeness, are not provided with similar pitch changes, make it difficult to voice the English sounds as the native speakers.

4 Discussions

The analysis of the formants revealed that the Japanese sounds and the English utterances located far apart from each other on the graphs of F2 and F1 with the batch of English phonics being relatively closer to the proximities of the graph origins. In terms of the tongue positioning, the native speakers tended to voice the sounds not only with their glossae in a slightly higher position but also whispered them out from the back of their mouths. On the other hand, the Japanese utterances were characterized by lower and somewhat flat planar tongue posture. Thus comparing the tongue positions of these groups, we infer that the speech of the native sounds were brought forth keeping the tongue a bitter rolled up.

As a consequence, due to the peculiarities in the tongue placement inherent to each language, when the Japanese collegians vocalize the English phonics, either F2 or F1 or both do not match statistically up the sounds made by the native people, i.e. the North American English speakers as suggested by the graphs related to the statistical tests.

As far as the pitches are concerned, the Japanese sounds were in general uniform in the sense that the up and down variations were small enough to make the graph curves look almost parallel to their horizontal axes. In contrast, the English phonics started in general at a high pitch and decreasing seamlessly towards the end.

These distinct strategies to modulate the pitches, which are intrinsic to the languages, affected the utterances spoken by the students as they attempted to echo the English sounds in such a way that the students intentionally surged the pitch gradually from a normal level and then brought them down from middle way towards the end, forcing the sound graphs to have a parabolic shape.

Finally, these results suggest that the when learning English language, the Japanese students should be aware, in addition to the formants (equivalently, the positioning of the tongue), of the pitch changes innate to the English utterances.

Acknowledgements

The authors would like to thank the students of Yonezawa Women’s College for providing the voice signals and helping with the experiments. Last but not least, thanks also go to the staff members and colleagues for their valuable support and cooperation.

5 References

[1] Ladefoged, P. (1993). “*A Course in Phonetics*”. Harcourt Brace Jonanovich, Orlando.

[2] Ladefoged, P. (2007). “*Phonetic Data Analysis*”. Blakcwell Publishing, Carlton.

[3] Minematsu N., Asakawa S. and Hirose K. (2006). “S*tructural representation of the pronunciation and its use for CALL*”. Proc. Int. Workshop on Spoken Language Technology, 126-129, Palm Beach, Aruba.

[4] Otsuka S. (2012). “*Utilizing Sound Spectrograms for Teaching Students Vowel Pronunciation*”.

Tokyo Josi Daigaku Kiyo Ronshu, 62(2):131-156. (in Japanese).

[5] Tsubota Y., Dantsuji M. and Kawahara T. (1999). "*English Pronunciation Instruction System for Japanese using Formant Structure Estimation*". Spoken Language Information Processing of Information Processing Society of Japan, 99(64):77-84. (in Japanese).

[6] Taguchi K. (2010). "*How can we assess intelligibility of English pronunciation more efficiently?*" Toyo Daigaku Keizai Kenkyukai, 35(2):221-226.

[7] Yoshimaru H. and Yamada J. (2010). "*An Acoustic Comparison of English Phonemes Between English Males and Male Japanese English Learners*". CASELE research bulletin, 40:41-50. (in Japanese).

[8] Titze R. (1994). "I. Principles of Voice Production". Prentice-Hall, Englewood Cliffs, NJ.Ministry of Education, Culture, Sports, Science and Technology-Japan. Accessed June of 2013 at"*http://www.mext.go.jp/*".

[9] Cambridge Dictionaries Online. Accessed January of 2013 at "*http://dictionary.cambridge.org*".

[10] Collins Dictionary. Accessed January of 2013 at "*http://www.collinsdictionary.com*".

[11] dictionary.com. Accessed January of 2013 at "*http://dictionary.reference.com/*".

[12] forvo. Accessed January of 2013 at "*http://ja.forvo.com/word/villages/*".

[13] howjsay.com. Accessed January of 2013 at "*http://www.howjsay.com/*".

[14] Merriam-Webster Dictionary. Accessed January of 2013 at "*http://www.merriam-webster.com/dictionary/*".

[15] WEBLIO. Accessed January of 2013 at "http://ejje.weblio.jp/".