

**An acoustic analysis of English pronunciation systems  
of Japanese learners**

**Aya Kitagawa**

**A dissertation submitted in partial fulfilment of the requirements  
for the degree of Doctor of Philosophy in Education**

**Waseda University  
2016 January**

## Abstract

The purpose of this study was to offer practical implications for the learning and teaching of pronunciation for the field of English education in Japan. The productive aspects of learning were emphasized and two research questions were addressed: (1) which phonetic and phonological items are easy, learnable or difficult for Japanese learners of English in each element of pronunciation; and (2) whether there is any positive, supportive relationship between the elements of pronunciation in the learning process.

Eight elements of pronunciation were targeted in this study: vowel quality, vowel duration, plosives, fricatives, approximants, rhythm, intonation and connected speech phenomena. The speech samples collected from Japanese learners of English, native speakers of British English and native speakers of American English were measured for these elements, using different measurements, including F1, F2, F3, duration, pitch, intensity and four spectral moments. A cluster analysis, a multivariate analysis of variance and a discriminant analysis were carried out for each element in order to address the first research question. The presence of supportive relationships in the learning process was investigated for the second question, using descriptive profiling and Spearman's rank-order correlation coefficients, based on the results of each element.

The results for the first research question suggested easy, learnable and difficult items, as follows. The items that were found to be easy were the vowel quality of /i:, e, æ, ʌ, ɑ:, ɔ:, u:/, the durational vowel distinctions in the /i:-I, ɑ:-æ/ pairs, the nucleus placement in utterances where the nucleus commonly fell on the final word, the use of a falling tone where it was typical, the span and the level. Learnable items were the /u:-ʊ/ durational distinction, VOTs of /p, k/ and the distinction of aspirated and unaspirated voiceless plosives in the /k-sk/ contrast. Difficult items contained /I, u:, ɜ:/ for vowel quality, the /ɑ:-ʌ/ distinction for vowel duration, VOT of /t/, the distinction of aspirated and unaspirated voiceless plosives in the /t-st/ contrast, both voiceless fricatives /θ, s/, both approximants /r, l/, all four properties to produce weak vowels (i.e. pitch, intensity, duration and vowel centralization), the nucleus placement in utterances that were long or where the

nucleus fell on the non-final word, the use of non-falling tones where they were typical, and three connected speech phenomena, elision, CC linking and CV linking. These results indicate that Japanese learners of English have more difficult items to learn in order to produce, or for use in production, than easy items and learnable items.

As regards the second research question, vowel quality and approximants were found to have a supportive relationship. Vowel quality and rhythm also suggested that they might have some relationship. It follows that these elements of pronunciation are positively related to one another, where they support each other to develop a learner's pronunciation system in the learning process. Approximants and fricatives were also found to have a relationship. Referring to the articulatory behavior and the acoustic features of these elements, this relationship was interpreted as due to a similar level of difficulty. There is a need for further examination as to whether this relationship could lead them to support one another in the learning process.

Although the current study focused on Japanese learners learning English pronunciation, comparing their performances against those of native speakers, achieving a native-speaker level of proficiency is not the only goal of learners. This study thus offers practical implications for both Japanese learners who use English as a foreign language, EFL-oriented learners, and Japanese learners who use English as a lingua franca, ELF-oriented learners, as regards five potential goals of pronunciation. Comparing the findings of this study with the targets suggested by these potential goals, the items that should be learned to attain each goal were specified. When learners define the goal that they will aim for, the findings of the present study will offer them more helpful guidelines for pronunciation in learning and teaching.

## **Acknowledgements**

My many thanks go to the two professors who supervised me through this research, Prof. Yasuyo Sawaki and Prof. Michiko Nakano. Prof. Sawaki taught me how thoroughly research should be conducted. I am certain that this dissertation would not be what it is now without her. Every time I faced difficulties in working on this dissertation, she comforted and guided me, and that is exactly why I was able to continue on the right track. Prof. Nakano taught me how important it is to cultivate a broader view of looking at things, and the world. I have learned many things from her, and this was one of the most precious times in my life. The energy she devotes to her research has always stimulated me, and it pushed me to come this far. I believe that I could not have continued my study without her encouraging words.

I also would like to offer sincere gratitude to three of the teachers who brought me to this point of writing my dissertation in the field of phonetics: Prof. Yoji Tanabe, Prof. Akira Ishihara and Dr. Michael Ashby. I vividly remember the time I learned phonetics from Prof. Tanabe. He taught us phonetics in my first year at Waseda University, and was my introduction to the subject. Without the warm words of encouragement that he offered me on my graduation thesis, this academic field would have seemed like another world to me. Prof. Ishihara also had a great influence on me. I was one of the students whose pronunciation was severely corrected most often in his class. He was the person who definitely made me think about the English pronunciation I would like to learn and master. The education that I received at Waseda University guided me to University College London, where I learned the true depth of phonetics. Dr. Ashby supervised me, and taught me what a spectrogram tells us and how fun it is to consider. If I had not encountered these teachers, I would not have been so fascinated by phonetics for such a long time.

Needless to say, I am deeply grateful to the two deputy chairs of the committee for this dissertation: Prof. Hiroshi Matsusaka and Prof. Tetsuo Harada. Not only did both teachers spare much time for reading this long dissertation, but they also offered constructive comments on the various issues that this study covered, ranging over phonetic and phonological knowledge and other possible approaches to the issues. They were both helpful

and shrewd, and at the same time, encouraging and motivating. Their words will remain in my mind, leading me to the next stage of my research life.

Last but not least, no words can express my appreciation to everyone around me who supported me in every moment of my life. Thanks to them, I am able to hold the complete work that is my dissertation. All the seminar members I studied with on Thursday nights, all the staff I have worked with, all my friends and all my students – without the time that I spent with them, I could not have kept working on this one thing. Finally, let me extend my warm thanks to my family. It is because of your generous and warm support that I am here now.

# Table of Contents

<b>Chapter 1 Introduction</b> .....	<b>1</b>
1.1. The purposes of the research.....	1
1.2. Current issues.....	2
1.3. Research questions.....	7
1.4. Theoretical background .....	9
1.4.1. Models concerning learning segments of the second language .....	9
1.4.2. Models concerning learning intonation of the second language.....	11
1.4.3. The theoretical background of the first research question .....	12
1.4.4. The theoretical background of the second research question .....	16
1.5. The outline of the dissertation.....	19
<b>Chapter 2 Literature review .....</b>	<b>20</b>
2.1. Vowels.....	21
2.1.1. Contrastive phonetics and phonology between Japanese and English .....	21
2.1.2. Learning L2 vowels .....	25
2.1.3. Acoustic measurements of vowels.....	28
2.1.4. The current study and hypotheses regarding vowel quality and duration .....	30
2.2. Plosives .....	33
2.2.1. Contrastive phonetics and phonology between Japanese and English .....	33
2.2.2. Learning L2 plosives.....	37
2.2.3. Acoustic measurements of plosives .....	39
2.2.4. The current study and hypotheses regarding plosives .....	41
2.3. Fricatives.....	43
2.3.1. Contrastive phonetics and phonology between Japanese and English .....	43
2.3.2. Learning L2 fricatives.....	44
2.3.3. Acoustic measurements of fricatives .....	46
2.3.4. The current study and hypotheses regarding fricatives.....	49
2.4. Approximants.....	50
2.4.1. Contrastive phonetics and phonology between Japanese and English .....	50
2.4.2. Learning L2 approximants.....	51
2.4.3. Acoustic measurements of approximants .....	54
2.4.4. The current study and hypotheses regarding approximants.....	57
2.5. Rhythm.....	59

2.5.1.	Contrastive phonetics and phonology between Japanese and English .....	59
2.5.2.	Learning L2 rhythm .....	60
2.5.3.	Acoustic measurements of rhythm.....	64
2.5.4.	The current study and hypotheses regarding rhythm .....	67
2.6.	Intonation .....	69
2.6.1.	Contrastive phonetics and phonology between Japanese and English .....	69
2.6.2.	Learning L2 intonation .....	72
2.6.3.	Acoustic measurements of intonation.....	77
2.6.4.	The current study and hypotheses regarding intonation .....	78
2.7.	Connected speech phenomena .....	84
2.7.1.	Contrastive phonetics and phonology between Japanese and English .....	84
2.7.2.	Learning L2 connected speech phenomena .....	86
2.7.3.	Acoustic measurements of connected speech phenomena.....	90
2.7.4.	The current study and hypotheses and research question regarding connected speech phenomena .....	91
2.8.	Relationships between the elements of pronunciation.....	93
2.8.1.	Application of the theory .....	93
2.8.2.	The current study and research question regarding relationships between the elements of pronunciation.....	94
2.9.	Summary of the chapter .....	95
<b>Chapter 3</b>	<b>Methodology .....</b>	<b>97</b>
3.1.	Subjects.....	97
3.2.	Materials .....	98
3.2.1.	Monophthongal vowels: vowel quality and duration.....	100
3.2.2.	Consonants: plosives, fricatives and approximants .....	102
3.2.3.	Rhythm.....	103
3.2.4.	Intonation .....	106
3.2.5.	Connected speech phenomena .....	107
3.3.	Recording and procedure .....	109
3.4.	Acoustic measurements and analyses .....	110
3.4.1.	Monophthongal vowels: vowel quality and duration.....	111
3.4.2.	Consonants: plosives, fricatives and approximants .....	111
3.4.3.	Rhythm.....	113
3.4.4.	Intonation .....	114

3.4.5.	Connected speech phenomena .....	115
3.5.	Statistical analyses .....	116
3.5.1.	Monophthongal vowels: vowel quality and duration.....	116
3.5.2.	Plosives .....	119
3.5.3.	Fricatives.....	120
3.5.4.	Approximants.....	121
3.5.5.	Rhythm.....	122
3.5.6.	Intonation .....	124
3.5.7.	Connected speech phenomena .....	128
3.5.8.	Relationships between the elements of pronunciation.....	129
3.5.9.	Preliminary analyses .....	131
3.5.10.	Three statistical tests for the analysis of each element of pronunciation .....	135
3.5.11.	A statistical test for the analysis of the relationships between the elements of pronunciation .....	140
3.6.	Criteria of learning.....	141
<b>Chapter 4</b>	<b>Results .....</b>	<b>143</b>
4.1.	Monophthongal Vowels .....	143
4.1.1.	Vowel quality .....	143
4.1.2.	Vowel duration .....	155
4.2.	Consonants.....	165
4.2.1.	Plosives .....	165
4.2.2.	Fricatives.....	173
4.2.3.	Approximants.....	182
4.3.	Rhythm.....	189
4.4.	Intonation .....	202
4.5.	Connected speech phenomena .....	218
4.6.	Relationships between the elements of pronunciation.....	228
<b>Chapter 5</b>	<b>Discussion .....</b>	<b>240</b>
5.1.	Vowel quality .....	240
5.1.1.	Findings.....	240
5.1.2.	Hypotheses regarding vowel quality.....	241
5.2.	Vowel duration .....	244
5.2.1.	Findings.....	244
5.2.2.	Hypotheses regarding vowel duration .....	246



5.3.	Plosives .....	250
5.3.1.	Findings.....	250
5.3.2.	Hypotheses regarding plosives .....	252
5.4.	Fricatives.....	254
5.4.1.	Findings.....	254
5.4.2.	Hypotheses regarding fricatives.....	256
5.5.	Approximants.....	258
5.5.1.	Findings.....	258
5.5.2.	Hypotheses regarding approximants.....	260
5.6.	Rhythm.....	261
5.6.1.	Findings.....	261
5.6.2.	Hypotheses regarding rhythm .....	263
5.7.	Intonation .....	266
5.7.1.	Findings.....	266
5.7.2.	Hypotheses regarding intonation .....	270
5.8.	Connected speech phenomena .....	273
5.8.1.	Findings.....	273
5.8.2.	Hypotheses regarding connected speech phenomena .....	275
5.9.	Relationships between the elements of pronunciation.....	277
5.10.	General Discussion .....	279
5.10.1.	Hypotheses and models .....	280
5.10.2.	Research questions .....	285
<b>Chapter 6</b>	<b>Practical implications .....</b>	<b>288</b>
6.1.	A learning goal of English pronunciation .....	288
6.1.1.	Need to reconsider a pedagogical goal of pronunciation.....	288
6.2.	Potential pronunciation goals.....	291
6.3.	Pronunciation goals for EFL-oriented learners and ELF-oriented learners.....	292
6.3.1.	Vowels.....	294
6.3.2.	Plosives .....	298
6.3.3.	Fricatives.....	300
6.3.4.	Affricates.....	302
6.3.5.	Approximants.....	303
6.3.6.	Nasals.....	305
6.3.7.	Syllabic consonants.....	306

6.3.8.	Consonant clusters .....	306
6.3.9.	Rhythm.....	308
6.3.10.	Stress .....	310
6.3.11.	Intonation.....	311
6.3.12.	Connected speech phenomena.....	316
6.4.	Summary of the chapter .....	318
<b>Chapter 7</b>	<b>Conclusion .....</b>	<b>320</b>
7.1.	Conclusion .....	320
7.2.	Limitations .....	323
7.3.	Further studies.....	325
	References.....	328
	Appendix A: Materials.....	351
	Appendix B: Target tokens for AN subjects .....	353
	Appendix C: Dendrogram for vowel quality .....	357
	Appendix D: Correlations between the F1 variables .....	358
	Appendix E: Correlations between the F2 variables.....	359
	Appendix F: Dendrogram for vowel duration .....	360
	Appendix G: Correlations between the variables for vowel duration.....	361
	Appendix H: Dendrogram for plosives.....	362
	Appendix I: Correlations between the variables for plosives .....	363
	Appendix J: Dendrogram for fricatives .....	364
	Appendix K: Correlations between the variables for fricatives.....	365
	Appendix L: Dendrogram for approximants.....	366
	Appendix M: Correlation between the variables for approximants.....	367
	Appendix N: Correlations between the variables for rhythm .....	368
	Appendix O: Results of the pitch patterns for the target utterances .....	369
	Appendix P: Dendrogram for intonation .....	376
	Appendix Q: Correlations between the variables for intonation .....	377
	Appendix R: Dendrogram for connected speech phenomena.....	378
	Appendix S: Correlations between the variables for connected speech phenomena.....	379
	Appendix T: Results of the rate of level agreement between the two elements of pronunciation for the JL subjects .....	380

## List of Tables

Table 1.1 Postulates and hypotheses forming the SLM, cited from Flege (1995).....	14
Table 2.1 Plosives in Japanese and English .....	34
Table 2.2 Fricatives in Japanese and English.....	43
Table 2.3 Approximants in Japanese and English.....	50
Table 2.4 Rhythm in Japanese and English .....	59
Table 2.5 Summary of the Study Hypotheses.....	96
Table 3.1 Acoustic Measurements for Each Element of Pronunciation.....	110
Table 3.2 Variables for the Analysis of Vowels.....	117
Table 3.3 Variables for the Analysis of Plosives.....	120
Table 3.4 Variables for the Analysis of Fricatives .....	120
Table 3.5 Variables for the Analysis of Approximants.....	121
Table 3.6 Variables for the Analysis of Rhythm .....	123
Table 3.7 Variables for the Analysis of the Phonetic Items of Intonation.....	124
Table 3.8 Variables for the Analysis of the Phonological Items of Intonation.....	125
Table 3.9 Variables for the Analysis of Connected Speech Phenomena.....	128
Table 3.10 Variables for the Analysis of Relationships between the Elements of Pronunciation.....	129
Table 3.11 Variables Submitted to the Cluster Analysis (all variables converted to z-scores).....	137
Table 4.1 Descriptive Statistics of Vowel Quality for BN, AN and JL Groups .....	143
Table 4.2 Descriptive Statistics of Vowel Quality for Four Clusters .....	146
Table 4.3 Group Centroids for the Standardized F1 mel Values.....	150
Table 4.4 Structural Matrix for the Correlations between the Standardized F1 mel Variables and the Two Discriminant Functions .....	152
Table 4.5 Group Centroids for the Standardized F2 mel Values.....	153
Table 4.6 Structural Matrix for the Correlations between the Standardized F2 mel Variables and the Three Discriminant Functions .....	154
Table 4.7 Descriptive Statistics of Vowel Duration for BN, AN and JL Groups .....	156
Table 4.8 Descriptive Statistics of Vowel Duration for Four Clusters.....	159
Table 4.9 Group Centroids for Vowel Duration.....	162
Table 4.10 Structural Matrix for the Correlations between the Variables for Vowel Duration and the Two Discriminant Functions.....	164

Table 4.11 Descriptive Statistics of Plosives for BN, AN and JL Groups .....	165
Table 4.12 Descriptive Statistics of Plosives for Three Clusters .....	168
Table 4.13 Group Centroids for Plosives .....	171
Table 4.14 Structural Matrix for the Correlations between the Variables for Plosives and the Two Discriminant functions .....	172
Table 4.15 Descriptive Statistics of Fricatives for BN and JL Groups .....	173
Table 4.16 Descriptive Statistics of Fricatives for Four Clusters .....	175
Table 4.17 Group Centroids for Fricatives .....	179
Table 4.18 Structural Matrix for the Correlations between the Variables for Fricatives and the Three Discriminant Functions .....	180
Table 4.19 Descriptive Statistics of Approximants for BN, AN and JL Groups.....	183
Table 4.20 Descriptive Statistics of Approximants for Four Clusters .....	185
Table 4.21 Group Centroids for Approximants.....	187
Table 4.22 Structural Matrix for the Correlations between the Variables for Approximants and the Two Discriminant Functions .....	188
Table 4.23 Descriptive Statistics of Rhythm for BN, AN and JL Groups .....	189
Table 4.24 Descriptive Statistics of Rhythm for Four Clusters .....	195
Table 4.25 Group Centroids for Rhythm .....	199
Table 4.26 Structural Matrix for the Correlations between the Variables for Rhythm and the Two Discriminant Functions .....	201
Table 4.27 Typical Pitch Patterns Used by BN/AN Subjects.....	203
Table 4.28 Descriptive Statistics of Intonation for BN, AN and JL Groups.....	207
Table 4.29 Descriptive Statistics of Intonation for Four Clusters.....	211
Table 4.30 Group Centroids for Intonation.....	216
Table 4.31 Structural Matrix for the Correlations between the Variables for Intonation and the Discriminant function.....	218
Table 4.32 Descriptive Statistics of Connected Speech Phenomena for BN, AN and JL Groups.....	219
Table 4.33 Descriptive Statistics of Connected Speech Phenomena for Five Clusters ..	221
Table 4.34 Group Centroids for Connected Speech Phenomena.....	225
Table 4.35 Structural Matrix for the Correlations between the Variables for Connected Speech Phenomena and the Two Discriminant Functions.....	226
Table 4.36 Descriptive Statistics Regarding Ranks in Each Element of Pronunciation for BN, AN and JL Groups .....	232

Table 4.37 Top 5 Combinations of the Two Elements of Pronunciation for the Level Agreement.....	233
Table 4.38 Top 5 Combinations of the Two Elements of Pronunciation for the Level Disagreement .....	234
Table 4.39 Spearman Rank-Order Correlation Coefficients between the Elements of Pronunciation for the Entire Sample.....	235
Table 4.40 Spearman Rank-Order Correlation Coefficients Between the Elements of Pronunciation for the JL Subjects .....	237
Table 5.1 Summary of Items on which Associated Hypotheses were Supported.....	280
Table 5.2 Summary of Items on which Associated Hypotheses were Rejected .....	281
Table 5.3 Summary of the Findings .....	286
Table 6.1 Potential Targets for Vowels.....	295
Table 6.2 Potential Targets for Plosives .....	299
Table 6.3 Potential Targets for Fricatives .....	300
Table 6.4 Potential Targets for Affricates.....	302
Table 6.5 Potential Targets for Approximants.....	303
Table 6.6 Potential Targets for Nasals.....	305
Table 6.7 Potential Targets for Syllabic Consonants .....	306
Table 6.8 Potential Targets for Consonant Clusters .....	307
Table 6.9 Potential Targets for Rhythm .....	308
Table 6.10 Potential Targets for Vowel Weakening .....	309
Table 6.11 Potential Targets for Lexical Stress.....	311
Table 6.12 Potential Targets for Intonation.....	312
Table 6.13 Potential Targets for Connected Speech Phenomena .....	317
Table 6.14 Learnable and Difficult Items to be Learned to Attain Potential Goals .....	319

## List of Figures

Figure 2.1. Vowel diagrams for Japanese and English .....	21
Figure 2.2. F1 and F2 of [æ].....	29
Figure 2.3. VOT of [k] .....	40
Figure 2.4. FFT spectrum of [s] sliced with 40-ms Hamming window .....	48
Figure 2.5. Spectrogram of [r].....	55
Figure 2.6. Spectrogram of [l].....	56
Figure 4.1. Vowel distribution for BN, AN and JL groups .....	144
Figure 4.2. Profile of each cluster for vowel quality.....	147
Figure 4.3. Vowel distribution for four clusters .....	147
Figure 4.4. Canonical discriminant function plot for F1. ....	151
Figure 4.5. Canonical discriminant function plot for F2. ....	153
Figure 4.6. Durational values of long and short vowels for BN, AN and JL groups .....	157
Figure 4.7. Profile of each cluster for vowel duration. ....	159
Figure 4.8. Durational values of long and short vowels for four clusters.....	160
Figure 4.9. Canonical discriminant function plot for vowel duration.....	163
Figure 4.10. VOT for BN, AN and JL groups.....	166
Figure 4.11. Profile of each cluster for plosives. ....	168
Figure 4.12. VOT for three clusters .....	169
Figure 4.13. Canonical discriminant function plot for plosives.....	171
Figure 4.14. Four spectral moments for BN and JL groups.....	174
Figure 4.15. Profile of each cluster for fricatives. ....	176
Figure 4.16. Four spectral moments for four clusters.....	176
Figure 4.17. Canonical discriminant function plot for fricatives.....	179
Figure 4.18. Score for the /r/ and /l/ tokens and average number of errors for BN, AN and JL groups .....	183
Figure 4.19. Profile of each cluster for approximants. ....	185
Figure 4.20. Score for the /r/ and /l/ tokens and average number of errors for four clusters .....	186
Figure 4.21. Canonical discriminant function plot for approximants.....	187
Figure 4.22. Rhythmic values for BN, AN and JL groups.....	192
Figure 4.23. Dendrogram output for the rhythm.....	194
Figure 4.24. Profile of each cluster for rhythm.....	196

Figure 4.25. Rhythmic values for four clusters.....	197
Figure 4.26. Canonical discriminant function plot for rhythm. ....	200
Figure 4.27. Errors in the phonological representation of intonation and the span and level in the phonetic representation of intonation for BN, AN and JL groups.....	208
Figure 4.28. Dendrogram output for span and level. ....	209
Figure 4.29. Distribution of scores for three variables of intonation.....	211
Figure 4.30. Profile of each cluster for intonation. ....	213
Figure 4.31. Errors in the phonological representation of intonation and the span and level in the phonetic representation of intonation for four clusters .....	214
Figure 4.32. Canonical discriminant function plot for intonation. ....	217
Figure 4.33. Rate of CC linking and CV linking use in the target phonetic contexts for BN, AN and JL groups .....	219
Figure 4.34. Profile of each cluster for connected speech phenomena.....	222
Figure 4.35. Rate of CC linking and CV linking use in the target phonetic contexts for five clusters .....	222
Figure 4.36. Canonical discriminant function plot for connected speech phenomena. ...	225
Figure 4.37. Diagram to illustrate the presence of supportive relationships .....	239

# Chapter 1 Introduction

## 1.1. The purposes of the research

Among all linguistic components, pronunciation is unique. At whatever level of learner one is, one is not allowed to communicate, selectively using phonetic and phonological items. That is, one always has to use every property in oral communication. For instance, it is impossible to speak English without using /l/ and /r/ even if they are problematic for Japanese learners of English to produce authentically. It is unlikely that teachers would present words or sentences consisting only of vowels with no pitch movement and rhythmic features even if they would like their students to focus solely on learning vowels. One cannot gradually introduce challenging phonetic and phonological items into one's own spoken language according to one's learning stage. This is what makes pronunciation different from other linguistic components such as grammar and vocabulary.

All this suggests that pronunciation is the basis of spoken language. Because of this feature of pronunciation, there is no single, established way of teaching it and various teaching methods have been developed, as summarized in Celce-Murcia, Brinton, and Goodwin (2010). These methods began with the direct method from around the late 1800s to the early 1900s, and evolved into Audiolingualism in the United States and the Oral Approach in Britain, which were based on structuralism from the 1940s to the 1950s, the Silent Way and Community Language Learning in the 1970s, the Communicative Approach in the 1980s. In recent years, more emphasis has been placed not on teaching pronunciation by focusing merely on target sounds or sentences, but on teaching it through communication.

Along with the development of different teaching techniques, an awareness of the importance of teaching pronunciation has increased over time in Japan, as reflected in a series of guidelines for the English curriculum that the Ministry of Education Culture, Sports, Science and Technology (MEXT) has stipulated since 1947. The importance of teaching pronunciation has always been acknowledged in these guidelines. For instance, the latest course of study proposed by the MEXT (2009) points out that it is especially necessary to teach rhythm, intonation, speech rate and volume, as well as segments. This principle is



supported and promoted through the MEXT's English education reform plan. Based on the Common European Framework of Reference for Languages (CEFR; Council of Europe, 2001), the MEXT (2013) defines the goal for junior high school students as achieving either an A1 or A2 level, and the goal for high school students as achieving either a B1 or B2 level under the English education reform plan. Learning and teaching various aspects of pronunciation are required in the English educational field in order to encourage Japanese learners of English to gain better communication skills at these levels in this globalized society.

## **1.2. Current issues**

Despite the above-cited development of teaching methods and increasing awareness of teaching pronunciation, there is still a problem in the field of pronunciation learning and teaching. The reality is that it is not easy to answer the question: how pronunciation should be taught. How widely are these methods used in the classroom? Are language teachers familiar with them? Do they teach pronunciation as confidently as they do grammar? What elements of pronunciation to teach, in what order to teach them, what materials to use and how to teach them are issues that require further discussion.

Kochiyama, Arimoto and Nakanishi (2013) report that instruction in pronunciation teaching is not sufficient in English teacher-training courses in Japan. This may be an important reason why Japanese English teachers tend to be less confident in teaching pronunciation in class than they are in teaching other linguistic components such as grammar and vocabulary, and skills such as reading. They are also likely to spend less time with their students in learning and practicing phonetic and phonological aspects of English. Although the MEXT (2009) proposed the use of phonetic symbols as a supplement as a potential means of teaching pronunciation, it is not possible to teach rhythm and intonation, which the course of study considers significant, in this way.

Another serious problem in teaching pronunciation is that there is no developed and widely used teaching method, although a number of teaching methods have been proposed. While various computer-assisted pronunciation training programs have been devised,

Hismanoglu and Hismanoglu (2010) found, based on the results of questionnaires, that language teachers preferred employing traditional methods such as reading aloud, using dictionaries and dialogues to computer techniques. The researchers concluded that the teachers taught pronunciation as they had been taught it. Thomson (2013) suggested that some English language teachers recruited in Canada and the United States did not have a basic knowledge of pronunciation, although some descriptions of the questionnaire used in his study was too confusing to interpret with confidence. Derwing and Munro (2005) even argue that many teachers base their teaching on their own intuition and experience.

It cannot simply be that many pieces of research are too specialized for language teachers to access, but there is also a lack of comprehensive studies. As far as the research on phonetics and phonology is concerned, few extensive studies have been conducted to examine prosodic features of spoken languages and their learning (Mennen, 2007), for instance. This results in a failure to offer language teachers the practical and comprehensive implications of teaching certain features of pronunciation.

The elements of pronunciation are commonly divided into two, segmental features and prosodic features.<sup>1</sup> In a rough classification, the former includes vowels and consonants and the latter, rhythm and intonation. In a narrow sense, prosodic features involve loudness, tempo and rhythm (Crystal, 1985). This means that segmental features are distinguished from prosodic features by whether or not they stretch over a single segment. The smaller number of comprehensive studies on prosodic features could be partly attributed to the difficulty of analyzing features stretching over a single segment, which could range over the whole utterance.

Even though there is a complexity to defining prosodic features compared with segmental features, the importance of the function that prosodic features serve in spoken language has been emphasized. Anderson-Hsieh, Johnson and Koehler (1992) investigated

---

<sup>1</sup> Prosodic features are often interchangeably called suprasegmental features. Suprasegmental is a term originating from American structuralism in the 1940s as the linguistic concept opposite to segmental features. However, suprasegmental features such as stress and tone, are closely related to the segmental level in their representation, which blurs the division of the two features. The term prosodic features, preferred by the London School, will thus be used in this dissertation.

how the pronunciation score was correlated with three different types of pronunciation errors in prosody, segments and syllable structures. Their results showed a stronger correlation between the pronunciation score and the prosody score; hence, they concluded the overall pronunciation proficiency level was more strongly influenced by the accuracy of prosody than that of segments. The important role of prosodic features was highlighted by Munro's (1995) finding that native listeners of English could differentiate native samples from non-native samples even when they were low-pass filtered. Although the study failed to show that the listeners could rate accentedness in the same way in the filtered condition as in the unfiltered condition, the results suggested a crucial distinction in the realization of prosodic features between native speakers and non-native speakers. Derwing and Rossiter (2003), demonstrated the effectiveness of teaching pronunciation focusing on prosodic features by comparing three groups: a group which received their training in segments, one which received their training in prosodic features and one which received no specific pronunciation training. As shown in these studies, the learning of prosodic features is worth examining more extensively.

While various studies have been carried out on pronunciation by learners, there is also a methodological problem in conducting research if one is to address questions as to the nativeness and foreign-accentedness of learner pronunciation: how to assess a learner's speech samples. One of the most common methods chosen in previous research is the use of human raters. In such studies, native listeners or non-native listeners of a target language listened to the target tokens produced by the subjects and rated nativeness or foreign-accentedness on a 5-point Likert scale, for instance. However, an important problem is that there are individual differences in the judgment among human raters. Major (2007) conducted an experiment to investigate differences in the ratings of listeners, using four listener groups: Brazilian Portuguese listeners in Brazil, Brazilian Portuguese listeners in the USA, American English listeners in Brazil with Portuguese experience and American English listeners in the USA with Portuguese experience. They listened to, and rated on a 9-point Likert scale, speech samples of a short passage read by 5 native speakers and 20 non-native

speakers of Brazilian Portuguese. According to their results, where the listeners lived affected the rating of a foreign accent more than their first language. This suggests that the amount of the daily exposure to the language that they rate has an impact on foreign-accentedness judgement.

A rater's language learning background was found to be another factor that influenced the ratings. In Winke, Gass and Myford's (2012) study, 107 raters who had studied Korean, Spanish, Mandarin Chinese, German, French, Arabic or Japanese rated speech samples provided by 24 Spanish-, 24 Korean- and 24 Mandarin Chinese-speaking test takers of the Test of English as a Foreign Language internet-Based Test (TOEFL iBT®). The results showed that, for all three languages of the test takers, the experience of having learned the test takers' language would mean the raters rated leniently even after completing rater training. Their findings were consistent with those of Carey, Mannell and Dunn (2011), who revealed the effect of familiarity with a speakers' interlanguage phonology and the location of the test center on the examiner rating of the pronunciation score in International English Language Testing System (IELTS™).

Schmid and Hopp (2014) is one of the most comprehensive recent studies to address the issue of human raters from various perspectives. They first reviewed methodological differences between the studies in terms of materials, speakers, raters and procedure, and carried out three experiments to examine the effect of raters and methods on judgement of foreign accent. In the first experiment, the German-speaking raters, who had been exposed only to German in childhood, rated narrative-descriptive speech samples collected from five speaker groups: 20 monolingual speakers of German, 20 English late learners of English, 20 Dutch late learners of German, 20 native speakers of German having emigrated to English-speaking Canada and 20 native speakers of German having emigrated to the Netherlands. The results revealed that less familiarity with the language led to lower rating scores. In the second experiment, the judgment task was assigned using only 30 selected most native-like speech samples and 30 least native-like speech samples, both of which were rated in the first experiment. The former samples were rated more native-like overall, while the

latter samples were rated more broadly on the native-like scale than in the first experiment. This suggests the effect of the range within which the sample was selected. In the third experiment, the effect of instructions to the raters and the scales that they would use were tested. The raters rated all the speech samples in the first experiment, receiving different instructions. The results showed that the ratings were affected by whether they were instructed to rate foreign accentedness or nativeness. The standard of foreign accentedness was found to differ across raters. All these studies suggest a problem in the use of human raters in experiments. A rater's familiarity with the language, the population of subjects that the raters rate and the instructions that raters receive are all possible factors that undermine the reliability of the ratings.

One study reported no difference in holistic ratings as to consistency and severity, even between native English-speaking teachers and non-native English-speaker teachers, although they differed in the features that they weighted in judgement (Zhang & Elder, 2011). However, as long as various factors could affect the results of ratings, it is possible to use methods other than human raters. Above all, the effect of the subject population to be rated, which was identified by Schmid and Hopp (2014), was considered to most seriously influence the results of the present study. Birdsong (2007) also implied the effect of the population on rating, demonstrating that the nativeness ratings of low proficiency learners lacked agreement. Accordingly, acoustic analyses were employed in this study, as more objective and stable measurements. They have been commonly used in the field of speech science, as suggested in the literature reviewed in Chapter 2. Stable acoustic cues have been found to characterize each phonetic and phonological feature, as described in Kent and Read (2002) and Ladefoged (2003) in great detail. It is also possible to predict how the acoustic features measured would be perceived in the human auditory system, by transforming the acoustic scale to the auditory scales, such as Bark, mel, semitone (ST) and equivalent rectangular bandwidth (ERB). Considering the methodological issue described above, the current study thus attempted to conduct acoustic analyses in an experiment.

### **1.3. Research questions**

While the choice of materials and methods to teach or learn pronunciation of a target language could depend on what elements of pronunciation must be taught or learned, there are still two difficult questions regarding pronunciation learning. The first question is which phonetic and phonological items are easier and more difficult items to learn. Elements of pronunciation such as vowels, consonants, rhythm, intonation and connected speech phenomena are each made up of various phonetic and phonological items, but what process do Japanese learners of English go through when learning English pronunciation? The second question is whether there is any relationship between elements of pronunciation in the learning process. In other words, does learning one element enhance the learning of another element? The present study addressed these two research questions about the learning of pronunciation. Their exploration will contribute to solving some of the issues in English education in Japan. They will be discussed in more detail below.

The first research question is which phonetic and phonological items are easier or more difficult to learn to produce for Japanese learners of English who only have experience of learning English under the English curriculum in Japan. Studies to identify the difficult items for Japanese learners of English have been frequently undertaken, and so it might not be difficult to pinpoint the vowels that are problematic, for example. However, which vowels are easier or more difficult to learn? Or even if they are not easy to learn, which vowels is it possible for Japanese learners of English to learn? What are the possibilities and limitations of learning pronunciation under the English curriculum in Japan? The level of difficulty imposed by each phonetic and phonological item is one of the possible factors affecting learner achievements in learning pronunciation, and therefore, the exploration of this issue may lead to more efficient learning and teaching of pronunciation in Japan. This study thus attempted to define each of the analyzed phonetic and phonological items as either an easy item, a learnable item or a difficult item, and to identify the types of difficult items depending on the difficulty level. Learnable items and difficult items were distinguished with regard to whether there is some possibility that the majority of learners could naturally learn to produce the item concerned. When a certain item was found to be difficult for the majority of learners

to learn to produce, it was defined as a difficult item, which probably requires some treatment (special attention) for them to learn.

Motivated by the current issues described above, this study aimed to address the research question by conducting acoustic analyses of elements of pronunciation, including prosodic features. This study investigated the production of eight elements of pronunciation by Japanese learners of English: vowel quality, vowel duration, plosives, fricatives, approximants, rhythm, intonation and connected speech phenomena. Vowel quality and vowel duration were analyzed separately in the current study, and they are regarded as two separate elements. Similarly, while plosives, fricatives and approximants constitute one segmental element, consonants, they were considered to be separate elements of pronunciation because their manner of articulation is distinct from one another. These elements are known as segmental features. Rhythm, intonation and connected speech phenomena are prosodic features. Whereas connected speech phenomena such as elision and linking are features concerning the sound change of segments, this element could be defined as one of the prosodic features in the sense that they occur in connected speech stretching over a single segment. These eight pronunciation elements were studied in order to answer the first and second research questions.

The second research question asks whether there is any supportive, positive relationship between the elements of pronunciation in the process of learning. This question stems from the belief that the elements are related to each other in the development of language ability. Li and Port (2014) found that the acquisition of rhythmic features was developed along with the acquisition of the relevant individual properties. This implies that there are intimate relationships between the phonetic and phonological items in the learning process. This sort of study is in line with a dynamic systems approach (de Bot & Larsen-Freeman, 2011; de Bot, Lowie, & Verspoor, 2007), which suggests, as described in detail later in this chapter, that changes in one feature will trigger changes in another.

While a large amount of research has examined vowels and consonants, there have not been many comprehensive studies that deal with both within a single study. Few studies

have investigated multiple elements of pronunciation, and to the author's knowledge, there has been no major study addressing the issue of relationships between the elements of pronunciation. Previous studies have failed to address a research question such as whether learners who produce authentic vowels could produce authentic consonants. This is a rather new, groundbreaking issue that has not been explored thoroughly. The present study employed a dynamic systems approach to the whole system of pronunciation to explore the existence of supportive relationships between the elements of pronunciation (Verspoor & van Dijk, 2011), where, as one element of pronunciation develops, another also develops better. Identifying the difficulty in learning each phonetic and phonological item and the existence of the hypothetical relationships will reveal how target phonetic and phonological items, and each pronunciation element, can be effectively taught and learned.

#### **1.4. Theoretical background**

This section describes the theoretical background that this study was built on. Study of the first research question dealt with eight elements of pronunciation, and there is no one model for the productive aspects of pronunciation learning that could be employed with every element. Thus, this study attempted to employ a theoretical framework whose assumptions could be applied to all elements, referring to models proposed for individual elements. The concept of a dynamic systems approach was adopted for the second research question. A description of these theories will be given below.

##### **1.4.1. Models concerning learning segments of the second language**

There are currently three influential perception-based models in the field of acquisition of second language (L2) segments: the Speech Learning Model (SLM; Bohn & Flege, 1992; Flege, 1987, 1995), the Perceptual Assimilation Model (PAM; Best, 1995) and the Native Language Magnet theory (NLM; Kuhl, 1991, 2000; Kuhl & Iverson, 1995). These theoretical models have been applied in various studies, some of which compared their differences (Flege, MacKay, & Meador, 1999; Hallé, Best, & Levitt, 1999) or explored the applicability of these models to more extensive targets (Guion, Flege, Akahane-Yamada, &



Pruitt, 2000).

The SLM predicts the extent to which learners are ultimately able to learn L2. The cornerstones for the SLM are Flege (1987) and Flege and Hillenbrand (1984), who proposed that L2 learning of segments can be predicted by classifying L2 phones into three types depending on the degree of similarity to their first language (L1) counterparts: *identical*, *new* and *similar*. New L2 phones are those which have no counterpart in the L1 phonetic and phonological system, and similar L2 phones are those which have a phone that can be easily identified as a counterpart in the L1, although it is phonetically different. It is claimed that learners perceptually confuse similar L2 phones with their L1 counterpart due to a mechanism called *equivalence classification* (Flege, 1987), which prevents them from forming a distinct category for these L2 phones. Whereas identical phones are thus the easiest to learn, new phones are ultimately more easily learned than similar phones. Flege (1995) proposes four postulates and seven hypotheses of the SLM, as will be described later.

The PAM (Best, 1995) is one of the most frequently cited models in the research of L2 perception, and a great deal of research has been conducted under this theoretical framework (Best, McRoberts & Goodell, 2001). It presumes that non-native segments tend to be perceived based on their similarities and discrepancies in a learner's native phonological space, and defines the assimilation patterns and categorical goodness rating by comparing the articulatory gestures between a native language and a non-native language. Based on this prediction, the difficulty of discriminating between two non-native phones is rated and predicted to be one of the following assimilation types: two category assimilation (TC), single category assimilation (SC), a category goodness difference (CG), an uncategorized-categorized pair (UC), uncategorized speech segments (UU) and non-assimilable (NA) non-speech sounds. Best and Tyler (2007) expanded the PAM as PAM-L2, so that the model allows the prediction of perceptual learning. They refer to four possible cases of L2 contrasts that learners perceive as speech segments at the initial stage of learning, and maintain that success in perceptual discrimination will occur in the following order, depending on the assimilation type: TC, UC, CG and SC. UU probably lies somewhere

between UC and SC.

The NLM (Kuhl, 1991, 2000; Kuhl & Iverson, 1995) is the model which began with the question of why children gradually lose their ability to perceive and discriminate sounds in languages other than their L1. The NLM identified the perceptual magnet effect as the cause. This phenomenon is that exposure to a specific language makes the acoustic space underlying phonetic perception distorted with the increase in the focus on its categories. The NLM argues that there are good exemplars and poor exemplars of phonological categories, and the former types of sounds are called native-language phonetic prototypes. Because these prototypes attract other sounds in the category regardless of the perceptual distance, as does a magnet, these attracted sounds become difficult to discriminate from other sounds, as listeners lose sensitivity to identifying them. That is, the distortion of the perceptual distance between the L1 prototype and its surrounding L2 phones makes it difficult for learners to perceive and produce these L2 phones authentically. The NLM attempts to explain the reason some sounds are easy or difficult to perceive and produce based on this magnet effect. As more studies have been conducted within the framework of the NLM (Kuhl et al., 2008), this model has been revised in recent years to propose an expanded NLM (NLM-e), which suggests five principles newly added to the NLM to make specific predictions. It claims that early language experience affects future language learning, where computational and social abilities influence learning, pointing out the connection between perception and production.

#### **1.4.2. Models concerning learning intonation of the second language**

Few models have been developed and experimentally tested with respect to the learning of prosodic features, compared with the three models of segmental learning described above. However, there are two promising models currently advanced: the Perceptual Assimilation Model for Suprasegmentals (PAM-S; So & Best, 2014) and the L2 Intonation Learning Theory (LILt; Mennen, 2015; Mennen & de Leeuw, 2014).

The PAM-S (So & Best, 2014) is a model which originated from the PAM. So and Best (2014) examined the perception of Mandarin tones by native speakers of Australian English and French, which differ in whether or not lexical stress is used. According to their

results, the presence or absence of lexical stress in their L1 influenced the way they perceived non-native target tones. The authors claim, based on this finding, that the assimilation patterns of non-native tones into L1 tones and categorical goodness ratings could be defined by the phonetic similarities of prosodic categories between a target language and L1.

The LILt (Mennen, 2015; Mennen & de Leeuw, 2014) is a model to characterize similarities and differences between L1 intonation and L2 intonation in four dimensions. The first dimension is called *the phonological dimension*, which concerns the phonological items. For instance, languages differ in the type of tones and pitch accents that they allow. In other words, which type of tone falls on the nucleus in a certain context and the syllables that are prominent in utterances are different from language to language. The second dimension is called *the phonetic dimension*, which is relevant to phonetic items that implement phonological items. One example of the differences in this dimension is that even if the two languages under consideration have a falling tone, the realization of this tone differs phonetically in its alignment, slope, or height. The third dimension is called *the semantic dimension*, and involves the function of intonation. Mennen (2015) cited the example of the difference between English and Greek, which involves the tone use for yes-no questions; the former uses a rising tone, while the latter uses a falling tone. Even if the same tone is used in two languages, it may function differently. The final dimension is *the frequency dimension*, which is related to the frequency in use. This dimension is used to define the similarities and differences of two languages in terms of how frequently a certain tone is used, for example.

### **1.4.3. The theoretical background of the first research question**

The SLM, PAM and NLM share the conceptual basis that the difficulty of L2 acquisition depends on language experience of L1. However, these models vary in some respects. One of the primary differences is how they define the similarities and differences between a learner's L1 and target language, although the three models agree that similarities and differences between the languages examined need to be defined at a more precise level. The SLM is grounded on perceptual similarities and differences at an allophonic level as well as a phonemic level. The PAM attempts to predict assimilation patterns based on similarities

and differences in the articulatory gestures and phonological space. The NLM explores the possibility of the magnet effect by defining acoustic similarities and differences. Another difference is that whereas the PAM and the NLM focus more on explaining acquisition behavior at an initial stage of learning, the SLM highlights the process of life-long learning and a link between perception and production (Flege, 1995, 2003). The models also differ in that the SLM and the NLM aim to predict the learning of both perception and production, while the PAM mainly emphasizes predicting the perceptual aspect of learning.

Of these three models, the current study employed the SLM as its theoretical framework to explore the learning of the segmental features. One of the aims of this study was to examine phonetic and phonological learning with a specific focus on production and to categorize the phonetic and phonological items depending on the difficulty level. The SLM was expected to best fit the present study to address this first research question, which was the main reason the SLM was applied here. However, one problem in applying the SLM to the present study was that the model concerns the prediction of relatively experienced learners (Flege, 1995). The SLM suggests that L2 new phones are easy to learn, but this may not be applied to the learning of some new phones produced by less experienced learners. This study targeted Japanese learners of English who had learned English under the curriculum of English education in Japan, which makes the applicability of this model to the general population questionable.

This study attempted to deal with this practical problem by defining the degree of newness of the item more precisely, which was expected to make it possible to predict and investigate the learning of new phones by less experienced learners. In establishing hypotheses under the framework of the SLM, the perceived distance between Japanese and English (Flege, 1984, 1995) and acoustic differences between early and late learners (Oh et al., 2011) were considered to define L2 phones as identical, similar, or new. If new phones are easier for learners to learn to produce than similar phones, it follows that the more perceptually and acoustically dissimilar target items were to any L1 phone, the more likely they would be learned. It was predicted that new phones with a higher degree of newness

would be learned to produce even by less experienced Japanese learners of English. Thus, of the seven hypotheses the SLM put forward, which are specified with three postulates in Flege (1995), as shown in Table 1.1, the current study focused on testing four hypotheses that could apply to the category formation by less experienced learners, as well as experienced learners. These hypotheses are displayed in boldface in the table: H2, H3, H4 and H5.

Table 1.1

*Postulates and hypotheses forming the SLM, cited from Flege (1995)*

Postulates	P1	The mechanisms and processes used in learning the L1 sound system, including category formation, remain intact over the life span, and can be applied to L2 learning.
	P2	Language-specific aspects of speech sounds are specified in long-term memory representations called <i>phonetic categories</i> .
	P3	Phonetic categories established in childhood for L1 sounds evolve over life span to reflect the items of all L1 or L2 phones identified as a realization of each category. Bilinguals strive to maintain contrast between L1 and L2 phonetic categories, which exist in a common phonological space.
	P4	
Hypotheses	H1	Sounds in the L1 and L2 are related perceptually to one another at a position-sensitive allophonic level, rather than at a more-abstract phonemic level.
	<b>H2</b>	<b>A new phonetic category can be established for an L2 sound that differs phonetically from the closest L1 sound if bilinguals discern at least some of the phonetic differences between the L1 and L2 sounds.</b>
	<b>H3</b>	<b>The greater the perceived phonetic dissimilarity between an L2 sound and the closest L1 sound, the more likely it is that phonetic differences between the sounds will be discerned.</b>
	<b>H4</b>	<b>The likelihood of phonetic differences between L1 and L2 sounds, and between L2 sounds that are noncontrastive in the L1, being discerned decreases as AOL increases.</b>
	<b>H5</b>	<b>Category formation for an L2 sound may be blocked by the mechanism of equivalence classification. When this happens, a single phonetic category will be used to process perceptually linked L1 and L2 sounds (diaphones). Eventually, the diaphones will resemble one another in production.</b>
	H6	The phonetic category established for L2 sounds by a bilingual may differ from a monolingual's if: 1) the bilingual's category is "deflected" away from an L1 category to maintain phonetic contrast between categories in a common L1-L2 phonological space; or 2) the bilingual's representation is based on different features, or feature weights, than a monolingual's.
	H7	The production of a sound eventually corresponds to the items represented in its phonetic category representation.

*Note.* The hypotheses relevant to the current study are highlighted in bold.

The LILT was adopted as the theoretical framework as regards intonation, for which the PAM-S (So & Best, 2014) and the LILt (Mennen, 2015) were described in Section 1.4.2. The PAM-S is aimed at modelling the learning of perception rather than production. In this sense, the LILt was expected to be more applicable in this study. Applying the LILt created another advantage for the study, in that hypotheses of learning segments and intonation could be formulated under the shared theoretical assumptions. The LILt has theoretical assumptions in common with the SLM in the following five points: the prediction of learning, the focus on the phonetic level to be considered in the prediction, the factor affecting the learning, the possibility of life-long learning, and the interrelationship between L1 and L2.

However, there was some difficulty in developing hypotheses by following exactly what the LILt proposes. One problem is that although the LILt requires similarities and differences to be predicted in four dimensions, the phonological, phonetic, semantic and frequency dimensions, the definition of the similarities and differences in intonation between two languages is far more difficult than those for segments, as noted by Mennen (2015). It may be possible to compare Japanese and English regarding the phonetic aspects of intonation, such as pitch range or peak alignment, by considering only the phonetic dimension. However, the comparison of the phonological and semantic aspects of intonation, such as tone choice, is not as straightforward as that of the phonetic aspects of intonation.

Another difficulty in following the proposal of the LILt thoroughly is that all four dimensions need to be considered in order to predict difficulty in learning intonation (Mennen, 2015). Cabrera-Abreu, Vizcaíno-Ortega and Hernández-Flores (2013) especially highlighted the importance of the phonological, phonetic and semantic dimensions in examining intonational errors of learners. Nevertheless, there are limited studies directly comparing Japanese and English in all four dimensions. To the author's knowledge, no study has been conducted on the frequency dimension of Japanese intonation and English intonation. This suggests that the prior studies were not comprehensive enough to predict the difficulty in learning intonational items that the present study targeted.

The hypotheses of the current study were thus not completely developed based on a

consideration of similarities and differences in all four dimensions. A comparison between Japanese and English was first made only in dimensions available from prior research, mainly in the phonological and phonetic dimensions, just as in the SLM. The target intonational items were then defined as identical, similar or new in order to predict the difficulty of learning.

Although this study also dealt with the two other prosodic features, rhythm and connected speech phenomena, there was no specific model proposed to predict the learning of these elements of pronunciation. This study regarded it as reasonable to apply the theoretical assumptions shared by the SLM and LILt to predict the learning of these elements. As described above, the SLM and the LILt are similar, for example, in that both are suitable for defining similarities and differences based on the phonetic and phonological items between L1 and L2 and for predicting the difficulty of learning. Accordingly, as attempted in these models, the learning of rhythm and that of connected speech phenomena were also predicted in terms of the phonetic and phonological similarities and differences between Japanese and English. After categorizing target items as identical, similar or new, the difficulty of learning was predicted, focusing especially on how new items would be learned, with reference to previous literature.

#### **1.4.4. The theoretical background of the second research question**

As noted in Section 1.3, the second research question, which was directed at detecting supportive relationships between the elements of pronunciation within the pronunciation system, was approached within the framework of dynamic systems theory (DST; de Bot & Larsen-Freeman, 2011; de Bot et al., 2007). This theory will be described below.

Each element of pronunciation consists of various components. For example, English intonation is comprised of the skills and knowledge of using an appropriate tone type in a certain context, to place a nucleus on an appropriate syllable, to use enough pitch range, and many more. English rhythm is made up of an articulatory skill to produce schwa using an appropriate pitch, intensity, duration and vowel quality, and to place stress at a regular

interval. Under DST, the element of pronunciation that constitutes these components is called a system. In other words, intonation is one system, and the skills and knowledge are components that form the system. At the same time, there is another level of system and components. The elements of pronunciation, such as vowels and intonation, are components, and pronunciation is the whole system. This study attempted to reveal relationships between the elements at this level.

According to DST, the elements of pronunciation form *dynamic systems* of pronunciation. De Bot and Larsen-Freeman (2011) define *dynamic* and *systems* as “the changes that a system undergoes due to internal forces and to energy from outside itself” (p. 8) and “groups of entities or parts that work together as a whole” (p. 8), respectively. When applied to the development of L2 pronunciation, each element of pronunciation can be regarded as a subsystem involving a dynamic system of pronunciation and changing over time, while interacting with one another.

Thus, all components of the system are interconnected. This sounds simple, but the application of this theory requires consideration of various issues. De Bot and Larsen-Freeman (2011) mention the following basic characteristics of this theory. First, it emphasizes the importance of considering the initial conditions of learning, seeing that small differences at an initial stage of learning could affect how learners learn at a later stage (sensitive dependence on initial conditions). Secondly, how strongly components are connected varies from component to component, and which components are relevant to a system under consideration needs to be estimated carefully, based on previous studies or common sense (complete interconnectedness). Thirdly, in a dynamic system, more dynamic components and less dynamic components relate to one another over time, which produces a less linear effect in learning as more components are relevant (nonlinearity in development). These three points are very essential parts of the theory of all characteristics pointed out by de Bot and Larsen-Freeman, on which the predictions of learning were built.

Whereas the current study employed DST to detect possible relationships between the elements of pronunciation, DST aims to explain various relationships under a system as



suggested in the above-mentioned characteristics. This is why DST can be broadly applied to diverse aspects of language learning, such as lexical development and syntactic development. The aims of this study within DST need to be elaborated. The first point is that this study focused on the improved performance of components within learning. Verspoor and van Dijk (2011) stated that there were two types of development of components, “the growth or increase in level of more developmentally advanced or complex variables and the decline or decrease of less developmentally advanced variables” (p.85). For instance, in the development of the lexical system, when one difficult word is learned, another easier word might be used less frequently. This corresponds to the second type of development. However, as far as the learning of pronunciation is concerned, it is less likely that learning one element will be a factor in the decline of another element. Therefore, the second type of development was not assumed in the present study. The second point is that the selection of components to be considered within a system affects the description of relationships. As implied by the second and third characteristics above in particular, many components consist of a certain system in reality, including both internal resources within individual learners and external resources outside them (de Bot et al., 2007). However, what can be found in an experiment is restricted to the relationships between the components tested. While this study has defined eight elements of pronunciation as the components of the pronunciation system, it should be noted that there are more detailed components relevant to the system. The third point is that this study attempted to detect a supportive relationship between the elements of pronunciation. Verspoor and van Dijk (2011) note that there are other relationships, the competitive relationship and the conditional relationship. A supportive relationship, competitive relationship and conditional relationship are each defined as a relationship where two components develop better together, a relationship where one component develops better while the other changes for the worse, and a relationship where one component develops better under a certain condition of the other. The aim of the experiment in the present study was to find relationships between the elements of pronunciation in order to offer practical implications for establishing effective teaching or learning pronunciation. To satisfy this aim,

the current study looked only for supportive relationships.

### **1.5. The outline of the dissertation**

The overall structure of this dissertation is as follows. Chapter 2 will describe previous studies of each element of pronunciation: vowels, consonants, rhythm, intonation and connected speech phenomena in terms of contrastive phonetics and phonology, the learning of target phonetic and phonological items and acoustic analyses for them. Based on the literature review, the study hypotheses will be put forward at the end of respective sections. Chapters 3, 4 and 5 concern the experiment of the present study, and each of these chapters will show the methodology, the results and the discussion. Chapter 6 will offer practical implications in reference to potential pronunciation goals, including goals for EFL-oriented learners and ELF-oriented learners. The study conclusion will be provided in Chapter 7.

## Chapter 2 Literature review

This chapter will first review each element of pronunciation in the following order: vowels, plosives, fricatives, approximants, rhythm, intonation and connected speech phenomena from Sections 2.1 through 2.7. These sections will establish hypotheses with which to address the first research question: which phonetic and phonological items are easy, learnable or difficult for Japanese learners of English. Vowel quality and vowel duration, separately analyzed as elements of pronunciation in the present study, constitute one section because they have often been analyzed together within a single study. Relationships between the elements of pronunciation will be described in Section 2.8. The second research question was whether there is any supportive relationship between these elements of pronunciation. The previous literature involving dynamic systems theory (DST) will be introduced in this section, from which the idea of the relationships within the pronunciation system originated.

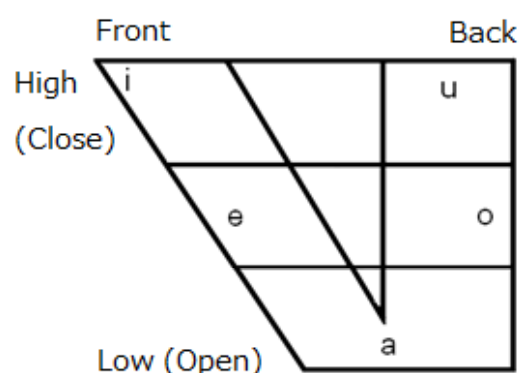
Sections 2.1 through 2.7, which concern the first research question, are structured as follows. They will begin with contrastive phonetics and phonology, where the phonetics and phonology of Japanese and English will be compared concerning the relevant elements of pronunciation. This will be followed by literature reviews of previous studies, focusing mainly on the L2 learning of the element under consideration. Acoustic measurements will also be discussed. Some elements of pronunciation have been measured using consistent methods since the experiment with the spectrogram became widely available, and other elements of pronunciation have been acoustically analyzed with different measurements in the context of finding reliable acoustic measurements. Finally, hypotheses will be formulated based on previous research, using the Speech Learning Model (SLM) for segments and L2 Intonation Learning theory (LILt) for intonation, and employing their theoretical assumptions for rhythm and connected speech phenomena, as noted in Section 1.4. These study hypotheses are summarized in Table 2.5 in Section 2.9. When an issue not tested elsewhere is raised, this study only posed a research question concerning the relevant element of pronunciation, rather than a hypothesis. In Section 2.8, the second research question will be considered after describing prior studies conducted within DST.

## 2.1. Vowels

### 2.1.1. Contrastive phonetics and phonology between Japanese and English

Vowels are articulated with air from the lungs being allowed to flow through the oral cavity freely. They are commonly defined through two dimensions: tongue height and tongue position. The vowel diagrams (a) and (b) in Figure 2.1 show the vowel system of Japanese and Received Pronunciation (RP), where the tongue height is represented vertically and the tongue position, horizontally. Some vowels, furthermore, involve a description of the lip position, such as spread, neutral or rounded. Lip rounding is another dimension with which to characterize vowels, which is not usually illustrated in the vowel diagram.

(a) Japanese vowel system



(b) English vowel system (RP)

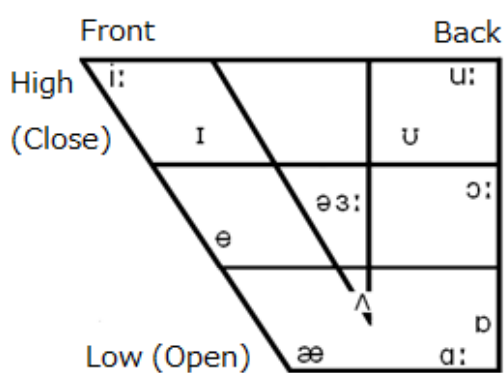


Figure 2.1. Vowel diagrams for Japanese and English.

Japanese has five vowels that have distinct qualities, as shown in Figure 2.1(a): close front /i/, close back /u/ (this vowel is also transcribed as /ɯ/, but /u/ will be used throughout this dissertation unless previous studies mentioning /ɯ/ are referred to), mid front /e/, mid back /o/ and open central /a/. In contrast, Figure 2.1(b) shows that English had a much richer vowel system, with eleven or twelve distinct vowels, the number varying slightly from accent to accent.

Similarities and differences in vowels between the two languages are experimentally investigated and identified. Strange et al. (1998) directly examined the perceptual assimilation of the American English vowels, /i:/, ɪ, eɪ, ε, æ:, α:, ʌ, ɔ:, ou, u, u:/, with the

five Japanese vowels. In this experiment, 24 native speakers of Japanese listened to stimuli in disyllable and sentence conditions, and selected the most similar Japanese vowel category to the English vowels tested, also providing the category goodness ratings on a 7-point scale from Japanese-like to not Japanese-like. Their results showed that when the experiment was oriented solely toward the quality, Japanese learners assimilated American vowels into Japanese vowel categories like /i:/ into /i/, /ɑ:/ into /a/, /ʊ, u:/ into /u/ consistently, and /ɪ/ into /i/, /ɛ/ into /e/, /ʌ, æ/ into /a/, /ɔ:/ into /o/ less consistently. The consistency of assimilating /ɪ, ɛ, æ, ʊ/ into some Japanese vowel categories varied from disyllable condition to sentence condition. These findings suggest that some English vowels sound close to some Japanese vowels, while others do not.

Nishi, Strange, Akahane-Yamada, Kubo, and Trent-Brown (2008) experimented with the opposite direction of Strange et al. (1998), the perceptual assimilation of the Japanese short vowels, /i, e, a, o, ʊ/, and long vowels, /ii, ee, aa, oo, ʊʊ/, into the American English vowel categories, /i:, ɪ, eɪ, ɛ, æ:, ɑ:, ʌ, ɔ:, ɒ, ʊ, u:/. They conducted two experiments, involving acoustic similarity and perceptual assimilation. Twelve Japanese learners of English who spoke American English fluently participated in an experiment with the perceptual assimilation, where they listened to the stimuli of a nonsense disyllable read in citation form and sentence form, and categorized the Japanese vowels into one of the American vowel categories depending on the perceptual similarity. The listeners also rated category goodness on a 7-point Likert scale, from English to foreign. Each of the Japanese vowels /i, ʊ, o/ was assimilated into /i: u: ɒ/ fairly consistently, whether the tokens were long or short vowels and whether they were given in citation or sentence forms. Two vowels, /e, a/, revealed a difference in the consistency of the assimilation, depending on the length, but not on the form. Long vowels, /ee/ and /aa/, were each assimilated into /eɪ/ and /ɑ:-ɔ:/ highly consistently—the native subjects who provided the stimuli in the experiment of acoustic similarity confused /ɑ:/ and /ɔ:/, so the results of pooled /ɑ:/ and /ɔ:/ as /ɑ:-ɔ:/ were reported. In contrast, some tokens of the short /e/ and /a/ were mainly assimilated into /eɪ/ and /ɑ:-ɔ:/, but they were also each assimilated into /ɪ, ɛ/ and /ʌ/ to a moderate degree. As

found by Strange et al. (1998), the results suggest that there are some Japanese vowels and English vowels that sound closer to each other, and some that do not.

Acoustic analyses of the spectrum also support these findings. As stated above, Nishi et al. (2008) studied similarities between Japanese vowels and American English vowels in citation and sentence forms, drawing a cross-language acoustic comparison. Four male speakers of Japanese provided stimuli of the five short and five long Japanese vowels, and four speakers of general American English, eleven English vowels. These stimuli, read in the above-mentioned two speaking conditions, were acoustically analyzed and then submitted to a discriminant analysis to examine the classification of the Japanese vowels into English vowel categories. They found that, exactly as in the perceptual assimilation, /i/, /u/ and /o/ were consistently classified into /i:/, /u:/ and /ou/ only by spectral cues, regardless of length and speaking style. However, /e/ was classified mainly into /ɪ/ except for long /ee/ in citation form into /eɪ/. Similarly, most /a/ tokens were classified into /ɑ:-ɔ:/, but short /a/ in the citation form was classified into /ʌ/.

With all these findings taken together, a greater similarity was uncovered between English /i:/ and Japanese /i/ and English /u:/ and Japanese /u/. In contrast, the other vowels have some scope left for exploration of their similarities and differences.

There are differences between English and Japanese in vowel duration as well as vowel quality. All 5 Japanese vowels have a distinction between the long vowel and short vowel, which suggests that there are 10 phonemes exist in the inventory of Japanese vowels. Hirata (2004) confirmed this using relational measures. Four native speakers of Japanese produced nonsense words and real words that contained long or short Japanese vowels in disyllabic contexts in three speaking rates, slow, normal and fast. Tokens for all 10 vowels were collected, and the durations of the words and the accented and unaccented target vowels were measured in the acoustic analysis. The results showed that the ratio of the short vowel to the long vowel tended to be stable in the experiment of nonsense words, barely affected by the speaking rate, although the same result was not obtained in the experiment of real words, where the effect of accent interacted with the effect of rate. This suggests that the temporal

cue that discriminates between long and short vowels in Japanese is rather phonological.

There is a stronger distinction of long and short vowels at a phonetic level in English than in Japanese. Cruttenden (2014) phonologically identified /i:-ɪ/, /u:-ʊ/, /ɑ:-æ/, /ɔ:-ɒ/ and /ɜ:-ə/ as the long-short pairs of vowels in English. However, the long and short versions of English vowels, which also differ phonetically in vowel quality, with one exception for /ɜ:-ə/, do not serve as minimal pairs in the same manner as those of Japanese vowels. That is to say, the vowels in the pairs above are differentiated from one another not only in duration but also in vowel quality. Thus, for example, although /i:-ɪ/ and /u:-ʊ/ may sound different only in duration to Japanese learners of English, the former vowels of each pair are categorized as tense vowels and the latter vowels are as lax vowels, which makes them different in quality as well as in duration. Also, /æ/, categorized as a short vowel, can be phonetically long with great variation in duration. In this sense, the long and short vowels take on a distinction, at both a phonetic level and at a phonological level, more in English than in Japanese.

The fact that long and short distinction in English vowels is more likely to be both phonological and phonetic than in Japanese vowels is reflected in the results of Hisagi, Nishi, and Strange (2008). They carried out an experiment of vowel quality and vowel quantity in different conditions: the consonantal condition, focus condition and speaking rate condition. Four native speakers of Japanese read nonsense words of five long and five short vowels in a carrier sentence, and four native speakers of American English did the same for seven long and four short American vowels. The target vowels appeared in various consonantal conditions, and each token was read with focus and post-focus and in a normal and a rapid speaking rate. Hisagi et al. reported that Japanese short vowels had less variation regardless of the surrounding consonants, sentence focus and speaking rate. American English vowels had more variations, but they maintain the difference between the long and short vowels more consistently except in different consonantal conditions. Japanese vowels always showed a greater distinction between long and short vowels than American English vowels. A more clear-cut durational distinction between long and short vowels, due to longer long vowels in Japanese, was also reported by Kato and Cox (2006). Kato and Cox found that Japanese had

the smallest difference in the [a-a:] pair, whereas English had the largest difference in the [ʌ-ɑ:] pair (they transcribed them as [ɛ-ɛ:]).

What should also be noted as to the difference between Japanese and English is that vowels transcribed with the same phonetic symbols in these languages are not necessarily the same in light of phonetics. English /i:/ is thus sometimes identified as the same as Japanese /i:/ in quality, but they differ phonetically. As noted above, English vowels have the tense and lax distinction, which labels the English /i:/ as a tense vowel, unlike the Japanese /i:/ (Vance, 1987). The same holds true of /u:-ʊ/.

### 2.1.2. Learning L2 vowels

There have been a number of studies regarding the learning of English vowels by Japanese learners of English, especially as regards their quality. What should be noted is that while some researchers have reported that it is difficult for Japanese learners of English to learn to discriminate English vowels according to their quality, others argued that it would be possible for them to learn some of the English vowels.

One of the most comprehensive acoustic analyses of English vowels produced by Japanese learners of English was conducted by Shimizu (1999). Shimizu measured the first formant (F1) and second formant (F2) of Japanese vowels /i, e, a, o, ʊ/ and English vowels /i:, ɪ, ɛ, æ, ɑ:, ə, ɔ:, ʊ, u:/ that seven Japanese learners of English with intermediate proficiency produced, and maintained that there was a clear negative transfer from their L1. The vowels in each of the English /i:/ and /ɪ/ pairs, /u:/ and /ʊ/ pairs and /æ/, /ɑ:/ and /ə/ groups were located closely to each other, forming one category. The three categories above corresponded to Japanese /i/, /ʊ/, and /a/, respectively. This suggests that it was difficult for the subjects to differentiate the vowels which could fall into one category in their phonological vowel space.

In contrast, the potential to form a new L2 phonetic category was explored in some previous studies. Ingram and Park (1997) performed three experiments with native speakers of Japanese and of Korean perceiving and producing Australian English vowels, /i:, ɪ, e, æ, a:/, including an identification task, an acoustic analysis and a prototypical rating



in word form. Five less experienced Japanese learners of English, five more experienced Japanese learners of English, five less experienced Korean learners of English and five more experienced Korean learners of English participated in the production experiment, in which the acoustic analysis was employed with these subjects. While the Japanese subjects identified /æ/ as a less prototypical vowel than any other vowel at a significant level, perceptually rating it as either /e/ or /a/, they perceived and produced this vowel more accurately than the Korean participants, differentiating it from /e/. This shows that an L2 vowel that is perceptually distinct from any L1 vowel could form its own category in the L2 vowel space and did not blend into another category.

Another piece of substantial research on the formation of a new category by Japanese learners of English was conducted by Lambacher, Martens, Kakehi, Marasinghe, and Molholt (2005). They carried out experiments within the framework of the SLM, where native speakers of Japanese were assessed on how their L2 vowel categories were created through training. In their experiments, 34 participants went through a 6-week training program, all completed an identification task and 20 took a production test of American English vowels, /æ, ɑ, ʌ, ɔ, ə/, in a pretest and posttest. Their performances were compared with the control group, consisting of 20 native speakers of Japanese. The results showed the following three points. Firstly, the identification of /ə, ɔ/ greatly improved. Secondly, the ratings of the native speakers of American English confirmed the effectiveness of the training on /ə/ and the accuracy of /ɔ, æ, ʌ/ even without training. Finally, the acoustic analysis of the first three formant frequencies showed that the training promoted improvement of /ə, æ, ɔ/ production. From these findings, they concluded that while the effectiveness of the training depended on the task, the training contributed to establishing a new L2 category for /ə, æ, ɔ/ because of their auditory and phonetic distinctiveness from the Japanese /a/. On the other hand, the similarity of /ʌ, ɑ/ to the Japanese /a/ hindered them from forming their own categories: the similarities and differences influenced the formation of a new category, as predicted in the SLM.

Oh et al. (2011) also examined the production of vowels by native speakers of

Japanese under the framework of the SLM. They conducted a one-year longitudinal study targeting both Japanese adults and children, focusing on the influence of age on vowel production. Sixteen subjects, each from four different groups, Japanese adults, Japanese children, English-speaking adults and English-speaking children, participated in the experiment, where they produced sixteen English words including eight American vowels, /i:, ɪ, e, eɪ, ɑ:, ʌ, ʊ, u:/. At the first recording, the Japanese adult group outperformed the Japanese child group, while they showed some spectral differences in /ɪ, ʌ, ɑ:, ʊ/ from adult native speakers of English. The Japanese child group produced /ɪ, e, ʌ, ɑ:, ʊ/ differently from the child native speakers of English. However, at the second recording, the Japanese child group learned all these vowels with no statistical difference from the child native speakers of English. In contrast, the adult group showed no improvement. This study suggests that age plays an important role in learning vowels, also demonstrating which vowels were commonly difficult for Japanese learners of English. Additionally, Oh et al. revealed the effect of learning L2 vowels on producing L1 vowels, which supported H6, stated by the SLM, as in Table 1.1.

Previous research has studied vowel duration as well as vowel quality. Ingram and Park (1997) acoustically investigated the vowel duration production of Japanese learners of English and Korean learners of English. They measured the durational difference within the following pairs, /i:-a:/, /ɑ:-æ/, /æ-e/ and /e-ɪ/. According to their report, the Japanese learners of English produced the durational difference between /ɑ:/ and /æ/ most prominently and they tended to distinguish tense and lax vowels categorically in duration.

The difficulty of /ɑ:-ʌ/ durational distinction was pointed to by Oh et al. (2011) and Lambacher et al. (2005). Oh et al. found, as a result of the experiments described above, that the adult Japanese speakers tended to produce /ɑ:/ in a shorter duration and /i:/ in a longer duration than the adult native speakers of English. They added that /ɑ:/ was shorter than its counterpart /ʌ/, which was found for both the adult and child Japanese groups. In contrast, both of the Japanese groups attained a native-like distinction in the /i:-ɪ/, /eɪ-e/ and /u:-ʊ/ pairs.

The learning of the durational difference between long and short vowels in Japanese was investigated by Kato and Cox (2006), who examined the production by Australian learners of Japanese. This may have implications for durational differences between long and short vowels for Japanese learners of English. Four Australian learners of Japanese and three native speakers of Japanese produced four pairs of Japanese and Australian English long and short vowel contrasts. They recorded the productions of Japanese vowels collected from the Australian learners of Japanese in 3 recording sessions, 4, 8 and 16 months after they started learning Japanese. These Japanese vowels were compared with the Australian English vowels produced by the same Australian subjects and the Japanese vowels by the three native speakers of Japanese. The results showed that while the two languages differed radically in that the durational contrast between the long and short vowels was sharper in Japanese than in Australian English, the Australian subjects generally approximated the native speakers of Japanese in terms of the ratio between the long and short vowels by lengthening long vowels. This suggests that English long vowels were shorter than Japanese long vowels in duration. On the other hand, the Japanese short vowels were reported to be closer to the Australian long vowels as far as the duration is concerned. They concluded that the Australian subjects used the temporal cue of their L1 in learning L2, which supported the feature hypothesis by McAllister, Flege, and Piske (2002).

### **2.1.3. Acoustic measurements of vowels**

Vowels are acoustically measured using formants and duration (Ladefoged, 2003; Kent & Read, 2002), which are considered to be reliable acoustic measurements to quantify vowel distribution in the vowel space. First formant (F1) and second formant (F2) are acoustic cues that reflect tongue height and tongue position, respectively, and thus measuring the F1 and F2 values makes it possible to observe the distribution of vowels in the speaker's vowel space. Measurements of F1 and F2 have been used in various studies, which concern Japanese vowels (Keating & Huffman, 1984), English dialects (Mayr & Davies, 2011), cross-language comparisons (Strange et al., 2007; Strange, Bohn, Trent, & Nishi, 2004) and a learner's languages (Ingram & Park, 1997; Lambacher et al., 2005; Munro, 1993; Oh et al.,

2011), for instance. Figure 2.2 shows the spectrogram of the production of [æ]. The two horizontal lines indicated by the arrows are F1 and F2, respectively. They are measured in Hertz (Hz).

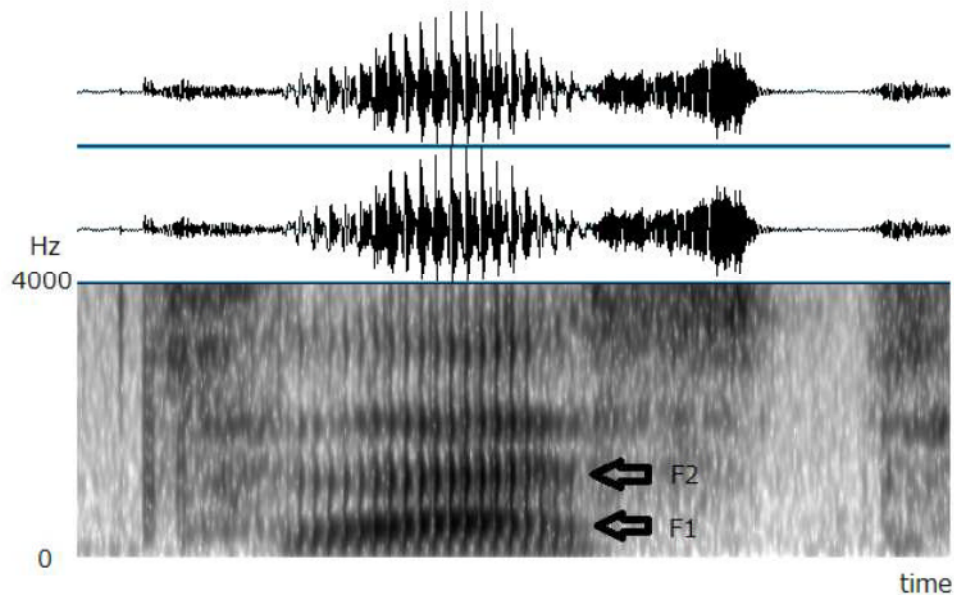


Figure 2.2. F1 and F2 of [æ].

Some researchers have claimed that vowels are not static, highlighting the need to use other measurements in addition to F1 and F2. For example, Hillenbrand, Clark, and Nearey (2001) reported that vowels were better classified by formant trajectories than static formant patterns. Following previous studies in this line, Fox and Jacewicz (2009) proposed that measuring formant movement at various points would benefit more detailed studies on vowels. In order to identify differences among three regional varieties of American English, they examined dynamic spectral changes using several acoustic measurements. They concluded that the following three acoustic measurements were effective in finding the characteristics distinct in each variety: vowel duration, trajectory length and the spectral roc. Trajectory length refers to the amount of the formant movement over the whole vowel region, and was obtained in Fox and Jacewicz by measuring the amount of change in F1 and F2 values at four sections, 20%-35%, 35%-50%, 50%-65% and 65%-80% of the vocalic portion.

The spectral roc refers to a measure to capture the speech of formant frequency changes.

Another measurement of vowels involves the overall distribution of vowels in the phonological vowel space, rather than the absolute F1 and F2 values. The technique is described by Minematsu (2004), Minematsu, Shiho, Murakami, Maruyama, and Hirose (2005), Nakamura, Suzuki, Minematsu, Hirose, and Makino (2010) and Suzuki, Qiao, Minematsu, and Hirose (2010), and was designed to characterize overall vowel distribution and to make it possible to compare the distribution of vowels between two speakers. In this technique, the speaker's distribution of vowels is first described with a distance matrix that is made up of the distance between all possible vowel pairs. By comparing a certain speaker's vowel distribution with that of another, the difference in the structure between the two speakers, the structural difference, is calculated.

#### **2.1.4. The current study and hypotheses regarding vowel quality and duration**

Studies of learning L2 vowels have been comprehensively conducted. This is related to the proposal of the currently influential models on learning L2 segments, such as the SLM, the perceptual assimilation model (PAM) and the native language magnet model (NLM). There are a countless number of studies involving Japanese learners of English learning English vowels. They have been investigated in terms of both perception and production, using various methods, which include an identification task, discrimination task, rater judgment and acoustic analyses.

As far as the methodology is concerned, the scope of these prior studies is limited in that a passage or spontaneous speech has rarely been employed as material for the analysis of vowels. The most common material is minimal pairs of both real words and nonsense words, which are read, embedded in a carrier sentence. For instance, the subjects in Ingram and Park (1997) produced vowels in the /h\_d/ frame. Lambacher et al. (2005) used 20 minimal pairs including both real and nonsense words in the /k\_d/, /k\_p/, /t\_d/, and /t\_k/ frames. Although Strange et al. (2007) reported that they analyzed the vowel production between in citation form and sentence materials, it is questionable whether their sentence materials satisfied the purpose of their study to compare formant values in different speaking styles.

Their citation utterances are disyllabic words read in isolation, while their sentence materials are trisyllabic words embedded in the sentence, *I said five \_\_\_ this time*. The words targeted in the test would be apparent in what they called sentence materials. In this sense, it would be too much to say that they investigated the effect of reading sentences on the production of vowels.

As Strange et al. (2007) argued, formant values are subject to surrounding sounds, which is the principal reason why minimal pairs have been preferred as materials for the analysis of vowels. Although subjects may exaggerate their performance, minimal pairs in well-controlled phonetic contexts are preferred as materials and expected to provide less varied formant values. However, the use of minimal pairs is highly restricted to the experimental condition. Even though native speakers articulate some vowels without attaining their target articulation in tasks such as reading a passage or giving a spontaneous speech, it is also true that these vowels are well identified in their daily conversation. Thus, in order to investigate learner's constructs of pronunciation, more studies that employ materials to measure performances in a real life are essential. This is what Strange et al. also pointed out. Therefore, the present study examined the production of English vowels by less experienced Japanese learners of English in a task involving reading a passage.

The current study targeted the following 10 monophthongal vowels in the analysis of vowel quality: /i:, ɪ, e, æ, ʌ, ɑ:, ɔ:, u:, ʊ, ɜ:/. It was hypothesized that /æ, ɔ:, ɜ:/ would be learnable items and /i:, ɪ, e, ʌ, ɑ:, u:, ʊ/ would be difficult items for Japanese learners of English to learn to produce. Within the framework of the SLM, these hypotheses were formed as follows.

Based on the degree of similarity of English vowels to those in Japanese, /i:, ɑ:, u:/ were first defined as similar and /ɪ, e, ʌ, æ, ɜ:, ɔ:, ʊ/ were as new. The classification of the three vowels into similar phones was grounded on the findings of Nishi et al. (2008), and Strange et al. (1998). The two English /i:/ and /u:/ were found to be assimilated constantly into Japanese /i:/ and /u:/, respectively. However, they are phonetically different (Igarashi, 1981; Matsusaka, 1986). They particularly differ in the lip

position: the English /i:/ and /u:/ are each spread and rounded, while the Japanese /i:/ and /u:/ are neutral. They were thus categorized as similar phones. The other vowel /ɑ:/ was defined as a similar phone, although this might be somewhat controversial. While Strange et al. claimed that Japanese /ɑ:/ was consistently assimilated into Japanese /a/, Nishi et al. found that long /a/ was assimilated into /ɑ:/ or /ɔ:/. This study followed the definition and findings of Oh et al. (2011), which presented the empirical data on the productive learning of English vowels by Japanese learners. Consequently, /ɑ:/ was defined as a similar phone. The rest of the vowels, /ɪ, e, ʌ, æ, ɜ:, ɔ:, ʊ/, were defined as new phones. This was also based on findings in the studies noted above, where these vowels were less consistently assimilated with the five Japanese vowels, or vice versa. Four vowels, /ɪ, e, ʌ, ʊ/, were also examined by Oh et al., who explicitly defined them as new.

On the basis of these definitions, the similar vowels, /i:, u:, ɑ:/, were predicted to be difficult items for Japanese learners of English to learn to produce. The hypotheses, concerning the new vowels, /ɪ, e, ʌ, æ, ɜ:, ɔ:, ʊ/, were also formulated, considering the degree of newness of each vowel to Japanese learners of English. In the present study, /æ, ɜ:, ɔ:/ were defined as vowels with a high degree of newness. According to the findings by Lambacher et al. (2005), these three vowels were found to improve both in the experiment of the spectral analysis and the perceptual rating. This led to the speculation that these vowels are phonetically prominent in quality: in other words, the newness of these vowels is high. Lambacher et al. also found some improvement of /ʌ/ even without training, implying that this vowel might have a relatively high degree of newness. However, Igarashi (1981) notes that this is the most difficult vowel for Japanese learners of English to learn to produce. Thus, /ʌ/ and the remaining vowels, /ɪ, e, ʊ/, were regarded as vowels with a low degree of newness, and were then predicted to be more difficult than /æ, ɜ:, ɔ:/. This prediction also rested on the estimation that the quality of /ɪ, ʌ, ʊ/ would be less phonetically prominent to Japanese learners of English because they have long counterparts consistently assimilated into Japanese vowels. Ingram and Park (1997) argued that Japanese learners of English tended to attend to temporal cues, which could hinder /ɪ, ʌ, ʊ/ from establishing a new L2

category. There was no report of an improvement in learning /e/, although the findings in Strange et al. (1998) imply that /e/ was new, demonstrating that Japanese /e/ was somewhere between /ɪ/ and /eɪ/ perceptually. Because the /e/ area is less dense in the English phonological vowel space, Japanese learners of English would be less sensitive to the newness of this vowel.

This study tested four long and short distinctions in vowel duration: /i:-ɪ/, /u:-ʊ/, /ɑ:-æ/ and /ɑ:-ʌ/. It was hypothesized that the distinction of long and short vowels in the /i:-ɪ/, /u:-ʊ/ and /ɑ:-æ/ pairs would be easy items for Japanese learners of English, whereas the distinction in the /ɑ:-ʌ/ pair would be an exceptionally difficult item.

The primary difference in vowel duration between English and Japanese is that the durational distinction is phonological in Japanese, whereas it is both phonological and phonetic in English. Thus, long vowels and short vowels are durationally more distinct in Japanese than in English. For example, both Hisagi et al. (2008) and Kato and Cox (2006) compared the durations of long and short vowels between Japanese and English, and found that Japanese had a greater durational difference. However, this means, at the same time, that the temporal cue itself is not unique to Japanese learners of English. This was why it was generally predicted as easy for Japanese learners of English to distinguish long and short vowels in English. In contrast, Oh et al. (2011) argued that only the /ɑ:-ʌ/ distinction was difficult for Japanese learners of English in the four pairs they tested, which was in accordance with Lambacher et al. (2005). Taken together, whereas there is a difference between Japanese and English in whether temporal cues are phonological or both phonological and phonetic, the realization of the durational distinction between long and short vowels could be defined as similar for /ɑ:-ʌ/ and identical for /i:-ɪ/, /u:-ʊ/ and /ɑ:-æ/. Therefore, it was predicted that /ɑ:-ʌ/ would be a difficult item and /i:-ɪ/, /u:-ʊ/ and /ɑ:-æ/ would be easy items.

## **2.2. Plosives**

### **2.2.1. Contrastive phonetics and phonology between Japanese and English**

Japanese and English have the same plosives in their phonological inventory:



/p, t, k, b, d, g/ as presented in Table 2.1. While there are far more plosives used in world languages, plosives are the only consonants in which Japanese and English overlap phonologically. However, at a phonetic level, some differences have been spotted. Differences in the voicing and the place of articulation can be identified by acoustically measuring voice onset time (VOT). VOT refers to a delay in the articulation between the burst of the plosive and the beginning of the voicing of the vowel that follows. As far as a within-language comparison is concerned, plosives articulated further back in the oral cavity tend to have a longer VOT, in general (Ladefoged, 2003).

Table 2.1

*Plosives in Japanese and English*

	Japanese	English
Voiceless	p t k	p t k
Voiced	b d g	b d g

One of the major differences between Japanese and English is that English plosives have a longer VOT than in many other languages, as in the previous cross-language comparisons described below. A large VOT in English plosives was reported in Lisker and Abramson (1964). They found that the duration of word-initial VOTs could be generally divided into three ranges: a voicing lead ranging from -125 ms to -75 ms, a short voicing lag from 0 ms to 25 ms and a long voicing lag from 60 ms to 100 ms. They measured VOT for 11 languages: American English, Cantonese, Dutch, Hungarian, Puerto Rican Spanish, Tamil, Korean, Eastern Armenian, Thai, Hindi and Marathi. The first six languages were classified as two-category languages with two categories of plosives differentiated by voicing and/or aspiration. As far as these two-category languages are concerned, Lisker and Abramson found that English voiced plosives generally fell into the short-lag range, while English voiceless plosives fell into the long-lag range, as with Cantonese, which was considered to have aspirated voiceless plosives and unaspirated voiceless plosives. The mean VOT values of English voiced plosives /b, d, g/ were each 1 ms, 5 ms and 21 ms, whereas those of English voiceless plosives /p, t, k/ were each 58 ms, 70 ms and 80 ms. In contrast, the rest of the

two-category languages were found to have one category of plosives with the voicing lead and the other with the long-lag range.

As shown in their findings, the word-initial VOT in English voiceless plosives tend to be long, leading to the categorization of English as a long-lag language. The VOT values of English plosives measured in previous research are comprehensively summarized in Auzou et al. (2000), who also reviewed the effect of aphasia, apraxia of speech and dysarthria on the VOT durations. All the studies they cited concerning VOT produced by normal speakers of English show that a clear separation of the range between voiced plosives and voiceless plosives. This supports a short lag for voiced plosives in English and a long lag for voiceless plosives in English. However, the VOT values for voiceless plosives seem to vary across studies. A consideration of other factors is thus needed to interpret VOT values, such as phonetic contexts and speaking rate.

One of the primary differences between Japanese and English is the range into which the word-initial VOT of each language fell. As suggested by the findings of Lisker and Abramson (1964), English voiceless plosives have a long lag, which is due to aspiration. In contrast, Japanese voiceless plosives do not involve aspiration in their articulation, and therefore, they are more precisely categorized as unaspirated plosives. In this sense, Japanese is in the same language class as Hungarian and Puerto Rican Spanish, tested in Lisker and Abramson.

Shimizu (1993) showed the above issue empirically. Just like Lisker and Abramson (1964), Shimizu compared the plosives of six Asian languages: Japanese, Mandarin Chinese, Korean, Burmese, Thai and Hindi. He analyzed these languages, using not only word-initial VOTs but also the pitch of the following vowel, the spectrum from 25 ms to 30 ms into the burst and the F1 values of the following vowel. According to his results on the VOT, the duration of VOT was divided into three ranges: a voicing lead ranging from -80 ms to -110 ms, a short voicing lag from 5 ms to 45 ms and a long voicing lag from 70 ms to 100 ms. While they are almost comparable to the ranges that Lisker and Abramson proposed, he also maintained that the short-lag range varied significantly across languages. Whereas 41 ms, 30

ms and 66 ms reported as the mean VOT values of Japanese voiceless plosives /p, t, k/, respectively, fell into the short-lag range in Shimizu, they fell into the range between the short lag and long lag in Lisker and Abramson. Shimizu also reported that, as for voiced plosives, /b, d, g/ were -89 ms, -75 ms and -75 ms long, respectively.

Shimizu (2008) reviewed his own study, where the durations of word-initial VOT were directly compared between Japanese and English. It shows that the VOTs of English voiceless plosives /p, t, k/ were 68 ms, 82 ms and 85 ms, while those of English voiced plosives /b, d, g/ were -88 ms, -74 ms and -85 ms, respectively. The values for voiced plosives were not comparable to the short-lag values that Lisker and Abramson (1964) reported for English voiced plosives. However, they added that some English speakers showed a voicing lead for voiced plosives, reporting -101 ms, -102 ms and -88 ms as the mean VOT values for /b, d, g/ of these speakers, respectively. Shimizu and Lisker and Abramson thus agree as to the word-initial VOT category in English: a long lag for voiceless plosives and a short lag or voicing lead for voiced plosives. On the other hand, according to Shimizu, Japanese voiceless plosives and voiced plosives can be categorized as a short lag and a voicing lead, respectively. Japanese plosives and English plosives are the same phonologically, but different phonetically, as regards voiceless plosives, in particular.

One thing to be noted here is that the VOT category is position-dependent. While voiceless plosives in English are aspirated with a long VOT in the word-initial position, they are unaspirated in the post-initial position, as in /sp, st, sk/. Ladefoged (2001) noted that these voiceless plosives are similar to voiced plosive counterparts. Similarly, word-medial plosives become shorter in Japanese. However, the durational difference between word-initial and word-medial VOT is not as striking as found in English. Homma (1981) reported that, in Japanese, the mean VOT value of initial voiceless plosives was 37 ms, whereas that of medial voiceless plosives was 16 ms. This result suggests that Japanese voiceless plosives /p, t, k/ are more similar to English word-initial /b, d, g/ or post-initial /p, t, k/, regardless of the position.

### 2.2.2. Learning L2 plosives

Previous studies have generally shown that the learning of VOT poses difficulty to learners. Flege and Hillenbrand (1984) and Flege (1987) are important studies, which promoted the establishment of the SLM. They carried out an experiment on learner's production of VOT, and concluded that it would be difficult for learners to achieve a monolingual native-speaker level when learning VOT at a phonetic level. By measuring the VOT value of the French /t/ produced by three groups, experienced American learners of French, inexperienced American learners of French and bilingual native speakers of French, both studies found that experienced American and bilingual native speakers of French produced /t/ in a similar duration. None of them was closer to the /t/ that the monolingual speakers of French produced in their prior study. This suggests that whereas the experienced American learners of French were able to approximate the French native speakers in the production of VOT, even the bilingual speakers of French produced /t/ differently than monolingual speakers of French. According to these findings, it would be difficult for learners to learn to produce VOT of the target language when it has a different VOT lag from their L1.

Riney and Takagi (1999) agree with these studies, indicating the difficulty of learning English plosives for Japanese learners of English. They conducted an experiment where 11 Japanese learners of English and 5 native speakers of American English read the target words, *part*, *time tub*, *cab*, *can*, and *come* and their VOTs were measured. They examined whether there was any change in VOT over an interval of 42 months. Using the global foreign accent (GFA) score for the subjects, they also investigated whether there was any correlation between the VOT measures and the GFA scores. In a previous study involving one of the authors (Riney & Flege, 1998), GFA scores were obtained for the subjects in Riney and Takagi, based on the ratings of five native-speaker listeners of five sentences produced by the subjects on a 9-point scale. The results showed that there was no improvement of VOT even with the time interval and that the VOT values were correlated with the scores of GFA.

A lack of correlation between VOT and global pronunciation ratings was also demonstrated by Birdsong (2007). However, Birdsong's study showed that late learners could

attain a native-speaker level in the production of VOT. Twenty-two native speakers of English who had started to learn French at the age of 18 or later participated in a production experiment. They read aloud target words for measuring vowel durations of five vowels and VOT durations of three voiceless plosives and passages for the rating of global pronunciation. According to the results for VOT, 9 out of 21 subjects who provided valid values were found to produce the three voiceless plosives with natively-like VOTs in comparison with the production of native speakers of French.

Although Birdsong (2007) obtained positive findings with regard to the learning of VOT by later learners, some caution is necessary in interpreting them. One issue is that the late learners in his experiment had lived in the Paris area for at least 5 years. This means that they would have had greater exposure to French. The other is that the learners' L1 is a long-lag language and their L2 is a short-lag language. As shown in Flege and Hiillenbrand (1984) and Flege (1987), English learners improved their target VOT even if they failed to attain it in a native-like manner. On the other hand, Riney and Takagi (1999) found that Japanese learners did not change their performance on VOT even with an interval of nearly four years. The subjects in Riney and Takagi were all college students, meaning that they had already learned English for a minimum of nearly 10 years, with an additional 6 years of English learning at high school. The differences in the findings of these studies might thus suggest an advantage of learners with a long-lag L1 background learning a short-lag L2 in learning VOT.

This speculation could be supported by the following study from the perceptual point of view. Stölten (2006) found that late learners with a short-lag L1 background had difficulty in perceiving different VOTs in their long-lag L2. She tested how Spanish learners of Swedish would perceive the stop continuum of Swedish. Spanish and Swedish are equivalent to Japanese and English, respectively, as far as the lag of VOT is concerned. She conducted an identification task in the experiment of perception, in which early learners of Swedish, late learners of Swedish and native speakers of Swedish participated. The results showed that while most of the early Spanish learners of Swedish in her study performed native-like in the

identification test of six plosives /p, t, k, b, d, g/, only a small number of the late Spanish learners of Swedish achieved a native-like classification in perceiving the three voicing contrast. This study of the perception of VOT suggests that learning VOT would be difficult for late foreign language learners.

Despite these difficulties, the importance of learning VOT was highlighted in terms of intelligibility by Joto, Nagase, and Funatsu (2007). In their experiment, 20 Japanese learners of English produced 39 English words containing English voiceless plosives /p, t, k/. Five native speakers of English transcribed these words and rated the intelligibility of these words and VOT on a 3-point rating scale. It was found that the intelligibility of /t/ was the lowest, while that of /k/ was the highest, and also that most voiceless plosives were transcribed as voiced plosives. Their subsequent acoustic analysis revealed that words with VOT shorter than 20 ms received low intelligibility ratings, whereas words with VOT longer than 55 ms received high intelligibility ratings for all three voiceless plosives. From these results, they claimed that /p/ pronounced with VOT longer than 30 ms, /t/ with VOT longer than 50 ms and /k/ with VOT longer than 55 ms would be fairly intelligible for the native speakers of English, adding that /t/ was the most difficult for the Japanese learners of English.

When the VOT values obtained by Joto et al. (2007) are compared with the mean VOT values in Shimizu (2008) and Riney and Takagi (1999), /t/ seems to be the most difficult English voiceless plosive for Japanese learners of English as Joto et al. argued. Shimizu showed that the mean VOT durations for the Japanese learners of English were 32 ms, 44 ms and 72 ms, respectively, for voiceless plosives /p, t, k/. These values are close to Riney and Takagi (1999), who found English /p, t, k/ produced by the Japanese learners to be 40.0 ms, 41.1 ms and 67.6 ms long, respectively, averaged across the values of the two recording sessions.

### **2.2.3. Acoustic measurements of plosives**

VOT is a major acoustic cue of plosives, as adopted as a measurement in the studies noted above. It clearly appears in the waveform and spectrogram as a gap between the burst

and the start of the formants for the following vowel. Figure 2.3 illustrates the VOT of [k], surrounded by the two vertical lines. The duration of VOT is known to be language-dependent (Ladefoged, 2003). Therefore, even if two languages are phonologically the same in that the phonological inventories in both languages contain a voiceless bilabial plosive, for instance, it does not mean this plosive is phonetically the same in duration between the two languages, as described in Section 2.2.1.

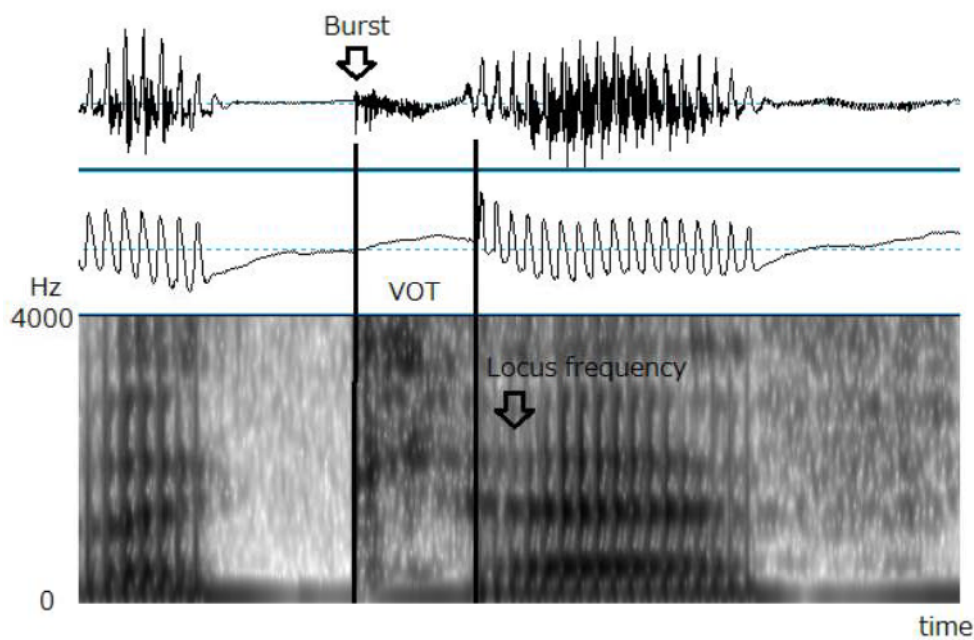


Figure 2.3. VOT of [k].

There are other acoustic cues for signaling voiced plosives, which are generally characterized with shorter VOTs than voiceless plosives. One is the presence of a voice bar, the presence of energy in low-frequency regions. A locus frequency is another possible cue, which involves the trajectory of F1, F2 and third formant (F3) of the vowels that follow voiceless and voiced plosives (Sussman, McCaffrey, & Matthews, 1991). This is shown with the arrow immediately above the formants in Figure 2.3. The examination of spectra at the release is also known to be useful in investigating the discrimination of the place of articulation (Blumstein & Stevens, 1979; Kewley-Port, 1983).

#### **2.2.4. The current study and hypotheses regarding plosives**

Whereas previous studies suggest that Japanese learners of English have some difficulty producing VOTs of voiceless plosives in English, there are issues in these studies. One is that while voiceless plosives have been well investigated so far, not many studies have been conducted regarding the learning of voiced plosives and unaspirated voiceless plosives in the post-initial position of the words. When it comes to Japanese learners of English at least, the reason for the smaller number of studies about the learning of voiced plosives would be that they are not problematic plosives for Japanese learners. Japanese voiced plosives and English voiced plosives are almost identical in that they could both be classified as a voicing lead as noted in Section 2.2.1. On the other hand, it is not clear why the learning of unaspirated voiceless plosives has rarely been studied. Another issue is that most of the studies used a list of words as material for VOT research. However, as noted in Section 2.1.4, there is a need for further investigation of learner performance in different speaking conditions. In order to address these issues, the current study used a task of reading a passage aloud to examine whether Japanese learners of English could produce the three voiceless plosives with VOTs long enough for English, and whether aspirated voiceless plosives and unaspirated voiceless plosives were well differentiated from one another in terms of VOT.

It was hypothesized that it would be difficult for Japanese learners of English to produce long VOTs for aspirated voiceless plosives in English. It was also hypothesized that it would be difficult for them to learn to differentiate in production two types of voiceless plosives, aspirated voiceless plosives and unaspirated voiceless plosives.

All aspirated voiceless plosives were defined as similar phones under the framework of the SLM. As described in Section 2.2.1, Japanese voiceless plosives and English voiceless plosives are phonologically the same, transcribed with identical phonetic symbols. However, the presence of aspiration differentiates them phonetically. Previous studies have agreed at this point by measuring VOT acoustically (Riney & Takagi, 1999; Shimizu, 2008), and thus, these similar English voiceless plosives were predicted as difficult for Japanese learners of English to produce with long VOTs.

The prediction concerning the differentiation between aspirated and unaspirated



plosives was slightly more complicated. Unaspirated voiceless plosives themselves could be defined as almost identical to those in Japanese under the framework of the SLM. According to Ladefoged (2001), English unaspirated voiceless plosives are similar to /b, d, g/. Assuming that English voiced plosives have a voicing lead, Japanese and English are in the same VOT category as far as unaspirated voiceless plosives are concerned. However, the present study did not deal with absolute VOT values of unaspirated voiceless plosives, but the differentiation of the voiceless plosives in the different positions. There is no report that Japanese has such a differentiation as aspiration and unaspiration between voiceless plosives, depending on their positions. Although Homma (1981) found that the word-initial VOT and word-medial VOT are different in duration in Japanese, the durational difference between them was too minimal to reach the separation of these VOT into different VOT categories. The discrimination of aspirated voiceless plosives and unaspirated plosives were thus defined as new under the framework of the SLM. If so, this new phonetic feature could be learned by Japanese learners of English. However, in the current study, they were predicted to have difficulty in differentiating aspirated voiceless plosives and unaspirated voiceless plosives for the following reasons. One is that, in order to realize this new phonetic discrimination, Japanese learners of English are also required to produce aspirated voiceless plosives with long VOTs. This was predicted to be difficult in the first hypothesis of learning voiceless plosives. The other is that the newness of discrimination between aspirated voiceless plosives and unaspirated voiceless plosives would be low. Evidence for this is in loan words from English. For instance, *spin* is pronounced as /supin/ in Japanese, not as /subin/. This would be the effects of both orthography and sound. Should *spin* sound [sbin] to Japanese speakers, it would serve as a piece of good evidence to show that post-initial voiceless bilabial plosives sound [b]. However, it is not the case in Japanese. For these reasons, although the discrimination between aspirated voiceless plosives and unaspirated voiceless plosives was defined as new, it was predicted to be a difficult item.

## 2.3. Fricatives

### 2.3.1. Contrastive phonetics and phonology between Japanese and English

Fricatives are produced by causing turbulence when air from the lungs goes through the stricture in the oral cavity. The number of fricatives in the phonological inventory of Japanese and English differs. Phonologically speaking, English has more fricatives as a whole. As presented in Table 2.2, English has nine fricatives, /f, v, θ, ð, s, z, ʃ, ʒ, h/, whereas Japanese has only three fricatives /s, z, h/ (IPA, 1999). Phonetically speaking, many more fricatives appear in Japanese, including [ɸ, β, ç, ʐ, ç, j, ɣ]. However, even these phones do not overlap with any English fricatives.

Table 2.2

*Fricatives in Japanese and English*

	Japanese	English
Voiceless	s h	f θ s ʃ h
Voiced	z	v ð z ʒ

This study focused on the learning of /θ/ and /s/, which were the most poorly discriminated of all fricatives by Japanese learners of English in previous studies, as will be noted below (Guion et al., 2000). They are categorized as nonsibilant and sibilant fricatives, respectively, in terms of the degree of stricture and the speed of air flow (Ladefoged, 2001). According to Cairns' (1999) description, a voiceless dental fricative, /θ/, is articulated with the tongue lowered, more relaxed and forward, even in contact with the teeth, leading to a lower intensity. Thus, /θ/, does not have any counterpart in Japanese. In contrast, a voiceless alveolar fricative, /s/, is a fricative common between the two languages. This suggests that /θ/ is not phonologically shared between the two languages but that /s/ is. However, Japanese [s] and English [s] are also known to be different at a phonetic level. Japanese /s/ has allophones, [s] and [ç], and the former does not occur when followed by a high vowel (Wells, 2000). Li, Edwards, and Beckman (2007) identified phonetic differences in [s] between Japanese and English in their acoustic analysis, suggesting that Japanese [s] is pronounced with a more dento-laminal articulation so that it would be more distinct from Japanese [ç].

Cairns also noted the difference between English /s/ and Japanese /s/ primarily from an articulatory point of view.

### **2.3.2. Learning L2 fricatives**

A comprehensive study of the learning of English fricatives by Japanese learners of English was conducted by Guion et al. (2000), who suggest that they had difficulty in discriminating between /θ/ and /s/. Guion et al. investigated the perceptual classification of 8 English consonants, /b, v, w, θ, t, s, ɹ, l/, and 7 Japanese consonants, /b, ɸ, t, d, s, r, h/, into 17 response categories of Japanese consonants. Nine native speakers of Japanese with the minimum possible exposure to English participated in the experiment, where they identified a category of the Japanese consonants to which each of the English and Japanese consonants was perceptually similar. At the same time, they rated the category goodness. According to the results, /θ/ was found to be a poor exemplar of any category of Japanese consonants, being classified to either [s] or [ɸ]. In contrast, /s/ was defined as a good exemplar of /s/. However, despite the difference in the identification and category goodness rate, these two consonants were discriminated poorly from one another in the subsequent categorical discrimination test of perception, against their prediction, which thirty native speakers of Japanese with different English learning experience completed. Because the accuracy of discrimination did not depend on the experience of English learning, the results highlighted the difficulty of learning these two consonants.

Kusumoto (2012) also conducted a production test as well as a perception test of /s-θ/ contrasts, along with six other fricative contrasts and /r-l/ contrast. Thirty-eight Japanese learners of English participated in an experiment of perception and production, and one native speaker of American English rated their production of the contrasts in minimal pairs. The results showed that the /s-θ/ contrast was the third most difficult contrast in both perception and production tests, in which the subjects perceived and produced this contrast with 72% and 59% accuracy, respectively. This highlighted the difficulty of this contrast in particular. While /f-h/ was judged as the most accurately produced contrast with 80% accuracy, the accuracy in the /s-θ/ production was just above the level of chance.

Although Kusumoto (2012) did not clarify which fricative was more difficult to produce, /s/ or /θ/, Cairns (1999) highlighted the difficulty of /θ/ in particular, reporting that the target word *youthful* produced by Japanese learners of English was frequently identified as *useful* by native speakers of English. The difficulty of learning /θ/ was also examined by Bada (2001). Based on the contrastive analysis, Bada predicted that /θ/ would be replaced by /t/ and /ts/. Japanese learners of English who had learned English for 8 years on average participated in an experiment, where they read sentences containing sounds expected to be problematic for Japanese learners of English. As hypothesized, the results showed that /θ/ was substituted by Japanese consonants such as /t/, /s/ and /z/. As a result, /θ/ was described as a major difficulty in production along with the voiced counterpart /ð/. He concluded that Japanese learners of English needed to take intensive training to articulate these consonants authentically.

The difficulty in learning /θ/ is not applicable only to Japanese learners of English. It is generally recognized that the difficulty of /θ/ is universal. It has been reported that dental fricatives of both /θ/ and /ð/ are problematic for learners of English with different L1 backgrounds (Mousa, 2014). For example, Wester, Gilbers, and Lowie (2007) assumed that the two English dental fricatives posed difficulty to Dutch speakers, and conducted an experiment of two picture descriptions with 24 native speakers of Dutch with different levels of English proficiency to investigate what they would substitute for these fricatives. They found that the Dutch speakers tended to substitute /t, d/ for them most frequently, followed by /s, z/. They replaced them with /f, v/ least frequently. Mousa (2014) also stated that substitutions of /t, d/ most frequently occurred in learners of English than those of other consonants. It would also be interesting to learn that even native speakers of English, as well as non-native speakers of English, found these sounds difficult. Wells (2000) suggested that English-speaking children acquire the dental fricatives later than the other sounds and that native speakers of English tend to replace them with /f, v/ or /t, d/ in a variety of local accents.

The difficulty in learning the English /s/ for Japanese learners of English was

identified by Sakata, Azukisawa, Shinoki, Yamada, and Wakita (1997), who acoustically measured subtle differences in the articulation of the English /s/ between Japanese learners of English and native speakers of English. They used the target tokens, *terrible storm of wind* and *people came, smiling*, produced by 16 Japanese learners of English and 5 native speakers of English. They measured the amplitude, spectrum, duration and total power. Each acoustic property revealed that there was a difference between the two groups of the subjects, indicating the difficulty of learning to produce /s/ in a native-like way. Combining the descriptions of how to articulate /s/ provided by Takebayashi (1996) with their findings, Sakata et al. advised Japanese learners to articulate /s/ longer in a louder voice with the tip of tongue held toward the alveolar ridge.

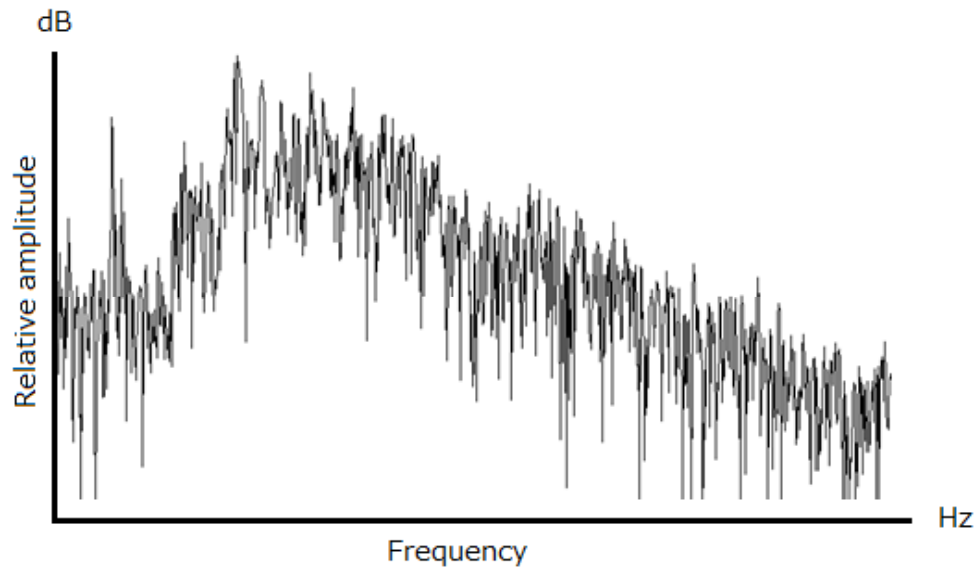
### **2.3.3. Acoustic measurements of fricatives**

There have been various acoustic measurements employed in previous studies. One of the earliest studies was conducted by Hughes and Halle (1956). They analyzed the energy density spectra of frication for the initial and non-initial /f, v, s, z, ʃ, ʒ/, and found that this acoustic cue worked to discriminate the place of articulation among the six fricatives with the region of higher frequencies, despite great individual differences. Gordon, Barthmaier, and Sands (2002) examined voiceless fricatives in seven languages, Aleut, Apache, Chickasaw, Scottish Gaelic, Hupa, Montana Salish and Toda. In a series of analyses for each language, they measured the duration, center of gravity and overall spectral shape averaged across speakers, and concluded that the overall spectral shape best characterized different fricatives. They demonstrated that the center of gravity served well to discriminate fricatives, while the duration did not.

The spectral shape is considered to be one of the most reliable acoustic measurements, as demonstrated by Gordon et al. (2002), and attempts to quantify it have been most frequently carried out. For example, de Manrique and Massone (1980) examined characteristics of the spectrum of Spanish voiced and voiceless fricatives spoken in Buenos Aires. They measured the frequency range of frication and periodic components, the frequency of the spectral maximum, the difference in intensity between the fricative and

vowel, the beginning of the frequency of the F2 and F3 transitions, and the duration of the frication. Their results showed that voiced fricatives differed from voiceless fricatives in that the former had a periodic component, a weak noise and a shorter duration. The voiceless fricatives also had two or three spectral peaks, and the main peak promoted the identification of the place of articulation among the voiceless fricatives. The results suggest that quantifying spectral shape enables the categorization of voiceless fricatives, differentiated from voiced fricatives.

Four spectral moments, mean or center of gravity (COG), standard deviation (SD), skewness and kurtosis, have been widely used to measure spectral shape. The moments are termed the first moment (M1), the second moment (M2), the third moment (M3) and the fourth moment (M4), respectively. Jongman, Wayland, and Wong (2000) maintained that these measurements clearly discriminated four places of articulation as a result of analyzing English /f, v, θ, ð, s, z, ʃ, ʒ/. Following the procedure of Forrest, Weismer, Milenkovic, and Dougall (1988), they measured the spectral moments at the four locations of spectra generated by fast Fourier transform (FFT) using 40-ms full Hamming window: onset, middle, end and offset, over which the fricatives were centered. Figure 2.4 illustrates the example of the FFT spectrum of [s] sliced with 40-ms Hamming window, which is displayed with the frequency on the x-axis and relative amplitude of each frequency on the y-axis. The major findings of Jongman et al. regarding the four spectral moments were as follows. M1 was the highest in /s, z/ and lowest in /ʃ, ʒ/. M2, differentiating the four places, was low for the sibilants and high for the non-sibilants. M3 was positive in /ʃ, ʒ/ and increased greatly in the non-sibilants at the fourth location. M4 was the highest in /s, z/. They also analyzed spectral peak location, locus equations of F2 at the following vowel onset and midpoint, overall noise amplitude, relative amplitude and noise duration.



*Figure 2.4.* FFT spectrum of [s] sliced with 40-ms Hamming window.

More comprehensive measurements were applied by Al-Tamimi and Khattab (2015). They aimed to characterize singleton and geminate fricatives in Lebanese Arabic acoustically, using the following measurements: duration, intensity at the onset, midpoint and offset of fricatives and preceded vowel, fundamental frequency (F0) at the onset, midpoint and offset of fricatives and preceded vowel, four spectral moments in medial fricatives, the peak location of the spectrum, the dynamic amplitude of spectrum, voicing patterns in medial fricatives, F1, F2 and F3 values of surrounding vowels, and the voice quality correlates of surrounding vowels. They performed a discriminant analysis using all acoustic cues as variables, and concluded that duration of the fricatives achieved the highest correct classification rate, 89%, to discriminate 10 fricatives in 4 different syllable structures. They also provide an account of how four spectral moments express actual articulatory behavior as follows. M1 and M3 have a negative correlation with the length of the front cavity, which means that sounds articulated more at the front have higher M1 and positive skewness. M2 and M4 reflect a flatness of the spectrum, which means the values become higher as it gets flatter. These two moments also show the tongue position; the articulation with the apical area is expressed by lower M2 values and higher M4 values. While the extent to which these spectral moments clearly categorize fricatives would depend on the language, they are often

used in the acoustic analysis of fricatives in different languages.

#### **2.3.4. The current study and hypotheses regarding fricatives**

The difficulty of learning to perceive and produce /s/ and /θ/ has been identified in previous studies. The characteristics of each fricative have also been described empirically with acoustic analyses, using various measurements. Nevertheless, it has been rare that a learner's production of these fricatives is acoustically investigated, compared with the research on vowels or /r/ and /l/. In this sense, Sakata et al. (1997) is one of the crucial and rare studies. However, the target words that they measured were *storm* and *smiling*, which means that /s/ in these targets was the first component of the consonant cluster. This could influence the acoustic features of pure /s/. More studies aimed at acoustically analyzing the production of fricatives by Japanese learners of English are expected to contribute to identifying the characteristics of their interlanguages, leading to more effective pronunciation teaching.

One principal reason for the lack of these studies is that researchers have not agreed on which measurements best capture the characteristics of the fricatives with which they are concerned. This is apparent in a series of attempts to quantify fricatives in different languages. As noted in Section 2.3.3, to some extent, it might be language-dependent whether a given acoustic measurement consistently characterizes all fricatives in the relevant language. As far as English fricatives are concerned, however, spectral moments could be regarded as relatively stable measurements to depict the shape of a spectrum as explored in Forrest et al. (1988) and Jongman et al. (2000). The current study thus aimed to describe the characteristics of production of /s/ and /θ/ produced by Japanese learners of English, which are recognized as one of the problematic phonemic contrasts for them.

It was hypothesized that both of the voiceless fricatives would be difficult for Japanese learners of English to learn to produce. Under the framework of the SLM, /s/ was defined as a similar phone and /θ/ was defined as a new phone. As previous studies have shown, it is clear that English [s] differs from Japanese [s] phonetically, whereas they are phonologically the same. The pattern of perceptual assimilation found by Guion et al. (2000)



for these fricatives also supports this definition. This similar phone, /s/, was thus predicted to be problematic for Japanese learners of English because of its similarity to the Japanese /s/. In contrast, there is no phonological counterpart of /θ/ in Japanese, which suggests that this fricative was new to them. However, this new phone would be still difficult for less experienced Japanese learners of English. This is because the degree of newness of this fricative's feature would be low. The feature that separates /s/ from /θ/ is the degree of stridency; the former is strident, whereas the latter is not strident. The non-stridency of /θ/ makes this phone sound close to the Japanese /s/, which is considered to have a dental feature, unlike the English /s/. Studies, such as Kusumoto (2012), Cairns (1999) and Bada (2002), also recognized the difficulty of learning /θ/. This shows that the newness of this sound is not high enough for Japanese learners to learn to produce it distinctively from a productive point of view. Thus, /θ/ was also predicted to be a difficult item.

## 2.4. Approximants

### 2.4.1. Contrastive phonetics and phonology between Japanese and English

English has four voiced approximants, whereas Japanese has only two approximants. As shown in Table 2.3, the difference between the two languages lies in the presence of a voiced alveolar approximant /r/ and a voiced alveolar lateral approximant /l/ in English (English /r/ is transcribed as /ɹ/ if based on IPA, but /r/ will be used in this dissertation following the conventional transcription), which this study targeted. Japanese does not have any corresponding liquid.

Table 2.3

*Approximants in Japanese and English*

	Japanese	English
Voiced	j w	r l j w

At the same time, it is often pointed out in previous studies that Japanese learners of English tend to perceive and produce the English syllable-initial /r/ and /l/ as Japanese /r/.

However, they are phonologically different and Japanese /r/ is not even classified as an approximant. While the transcription of Japanese /r/ is a little controversial, it is labelled as a voiced alveolar lateral flap /r/ or retroflex flap /ɻ/ (/r/ will be used for this consonant throughout this dissertation unless otherwise stated). A flap is articulated with a brief closure immediately before the following sound, which stems from a very quick contact of the tip of the tongue with the alveolar ridge (Kent & Read, 2002). Vance (1987) claims that the Japanese /r/ is similar to the English flap heard in American English, as in *better*, which suggests that this feature is rather similar to alveolar plosives, /t/ and /d/. Therefore, although English /r/ is sometimes transcribed with the same phonetic symbol as Japanese /r/, they have very different phonetic qualities.

Japanese /r/ has allophones such as [ɾ] or [ɻ], which suggests that this sound is phonetically variant. This is true of the English /r/. Some varieties in British English and those in American English primarily differ in the presence or absence of r-coloring, being classified as a non-rhotic accent and rhotic accent, respectively. Also, /l/ has well-known allophones, dark [ɫ] and clear [l]. These English approximants and Japanese /r/ are thus realized with various phonetic qualities, depending on accents, individuals and word-positions.

#### **2.4.2. Learning L2 approximants**

One of the classic studies highlighting the difficulty of English /r/ and /l/ for Japanese learners of English was conducted by Goto (1971). In his experiment, 8 Americans and 11 Japanese subjects read a list of words including /r/ and /l/ tokens, and then identified the tokens as /r/ and /l/ by listening to their own recorded samples or the other subjects.<sup>7</sup> He found that the American subjects performed well in both perception and production, that six Japanese subjects with higher English proficiency discriminated /r/ and /l/ in production fairly well but not in the perception, and that the remaining Japanese subjects performed poorly in both production and perception. Subsequently, five Americans and five Japanese, including four proficient subjects, were selected from the first test, and the perception test was performed. The results of this test revealed that the Japanese subjects failed to recognize

the difference between /r/ and /l/ pronounced successively. Taken together, these findings suggest that poor perceptual ability of Japanese learners to discriminate between /r/ and /l/ persists even if they can produce both approximants well. Hallé et al. (1999) also found that Japanese learners of English poorly identified and discriminated between /r/ and /l/. Yamada's (1995) findings were similar, adding that the experience of living in the U.S. affected accuracy in perceiving these approximants.

The difficulty of discriminating between English /r/ and /l/ was also empirically investigated in research into perception conducted by Guion et al. (2000). The results of their experiment showed that neither approximants had a good counterpart in Japanese consonants: none of the Japanese sounds, including vowels and consonants, were similar to /r/ and /l/ perceptually. They also showed that /r/ and /l/ were different in the degree of similarity to Japanese /r/. According to their report, highly experienced Japanese learners and moderately experienced Japanese learners performed better in discriminating between English /r/ and Japanese /r/ than between English /l/ and Japanese /r/, which Guion et al. interpreted as an indication that /l/ was closer to /r/, the Japanese flap.

Although Goto (1971) did not aim to reveal the difference in learning two approximants in production, the advantage of the greater perceived dissimilarity between /r/ and /r/ than /l/ and /r/ in learning is what is most relevant to the hypotheses of the SLM. This was examined by Aoyama, Flege, Guion, Akahane-Yamada, and Yamada (2004). In their study, both adult and child Japanese speakers participated in experiments discriminating between /r/, /l/ and /w/ in perception and production. The experiments were carried out twice in order to examine learning, where the second experiment took place a year later. An experiment involving perception was first performed, in which a categorical discrimination test of /l/ from /r/ and /r/ from /w/ was conducted. The adults performed better in the perceptual discrimination of the target approximants on the first test, but the children performed better on the second test. This suggests that the children showed more improvement than the adults. An experiment involving production was then performed, where the subjects read English words, and native speakers of English identified these tokens

as /r/, /l/ or /w/. They reported that the children improved the production of /r/ and /w/ more than /l/, while the adults showed only a minimal improvement in learning these approximants. Although both children and adults performed better in producing /l/ than producing /r/ and /w/ at the first session, they concluded that there was a greater improvement in the production of /r/ and /w/, emphasizing a relative improvement. A better performance of /r/ in perception was also argued by Ingram and Park (1998), who examined Japanese and Korean students in Australia, using an identification task and discrimination task.

Hazan, Sennema, Iba, and Faulkner (2005), who carried out experiments to investigate the effect of training using audiovisual stimuli, found that perceptual training with auditory stimuli improved the perception of Japanese learners' of the /l-r/ contrast with no significant difference in identification of the two approximants. The performance in the production test did not significantly differ between /l/ and /r/, either. It revealed that the perceptual training with audiovisual stimuli led their productions of both /l/ and /r/ to be better identified by native listeners. However, the results of the rating task showed that the production of /r/ was better rated as an authentic token due to the effect of the training with audiovisual stimuli. The results of the pretest and the posttest in the experiment of production suggested that there was an absolute difference in performance between the two approximants while the impact of training on /l/ and /r/ did not differ. The subjects performed better in producing /r/ than /l/ as a whole.

The data reported by Slawinski (1999) showed that the learning of /r/ and /l/ was parallel in the production, although it was not what she had attempted to examine. Slawinski carried out experiments of both perception and production of /r/ and /l/ by Japanese children and adults, pointing out that only a limited number of empirical studies had been conducted on how spoken proficiency would affect the production of /r/ and /l/. Four groups of Japanese children and three groups of adults participated in the experiments. The groups of children consisted of a 3-year-old group, a 4-year-old group, a 5-year-old group and a 7-year-old group, who had more exposure to English with age. The groups of adults were made up of groups with intensive exposure to English, with late exposure and with limited

exposure. They perceived synthesized continua of /r/ and /l/ in the experiment, whose results revealed the perceptual ability of discrimination depended on the amount of exposure to English; the adult group with greater exposure to English and the 7-year-old children group highly discriminated these target consonants. The production test was aimed at examining how the participants would use temporal and spectral cues in discriminating /r/ and /l/, the results of which also supported those of the perception test. The participants showed that they improved F2 and F3 transitions of both /r/ and /l/ with age and exposure. There was no significant difference in the duration of F1 transition among the groups, except that the adult late learners used a longer cue for /l/ at a significantly different level from the other adult groups. Flege, Takagi, and Mann (1995) also found that Japanese learners of English could learn to produce both /l/ and /r/ accurately, as they are more experienced, by comparing native speakers of American English, less experienced Japanese learners of English and experienced Japanese learners of English with experience of living in the United States for 21 years on average.

Saito and Lyster (2011) focused on investigating the effect of training for learning /r/. They tested whether 3-week sessions of training improved the production of /r/ by 65 intermediate Japanese learners divided into 3 different groups: a control group, a group which received form-focused instruction with corrective feedback and a group which received form-focused instruction without corrective feedback. Twenty-five words and five sentences were used as materials in the experiment of a pretest-posttest design. In this experiment, 100 tokens were randomly selected, and submitted for rating by five native listeners of English and to the acoustic analysis of F3. The results revealed that only the group of Japanese learners of English who took form-focused instruction with corrective feedback achieved improvement. This suggests that Japanese learners of English can learn to produce /r/ better, depending on the instruction.

### **2.4.3. Acoustic measurements of approximants**

The approximants targeted in the present study, /r/ and /l/, are similar to vowels in that they have clear formants. This provides them with an acoustic feature distinct from other

consonants, vowel-like features. The sonority of these sounds is thus higher than other consonants such as plosives, fricative and affricates. In Espy-Wilson's (1992) study, which noted these characteristics, all four English approximants were acoustically analyzed in terms of a decrease in energy at low frequencies, abrupt amplitude change, mid-frequency energy, F1, F2, F3 and fourth formant (F4).

Of all measurements, F3 is recognized as one of the primary cues used to distinguish /r/ and /l/. According to Ladefoged (2003), /r/ is characteristic of the decrease in F3, and this lowering of F3 in /r/ has been measured by many researchers (Flege, Takagi, et al., 1995; Saito & Lyster, 1999). The key role of F3 in distinguishing /r/ and /l/ was also claimed by Iverson et al. (2001). Flege, Takagi, et al. (1995) emphasized the importance of lowering F3 for English /r/, stating that a higher F3 value led to more perceived foreign accentedness. Figure 2.5 depicts this lowering feature of F3. The boxed portion in the spectrogram corresponds to the whole /r/ token. The horizontal line just below the arrow is F3 of /r/, which can be measured in Hz.

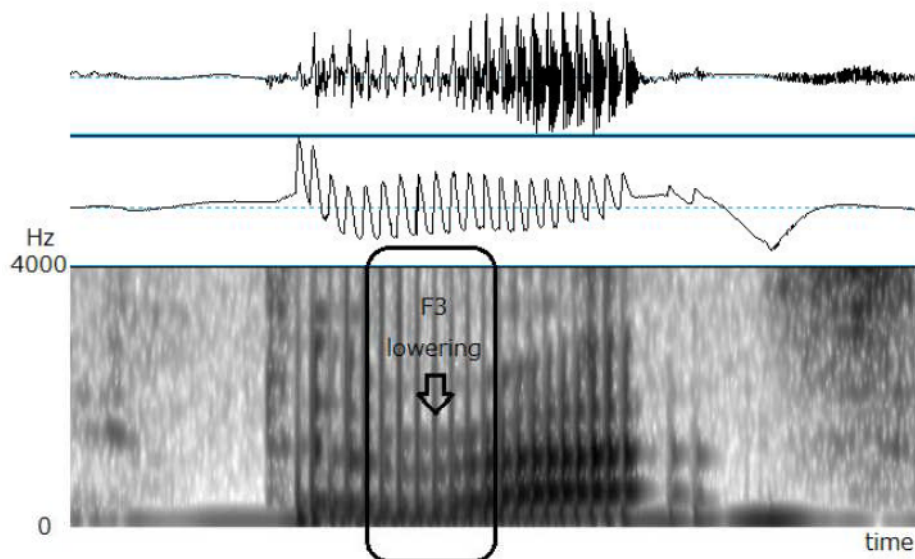


Figure 2.5. Spectrogram of [r].

There is a relatively wider range of F3 measured for English /r/, which reflects the degree of r-coloring. Ladefoged (2003) notes that some initial /r/ was produced as low as

1240 Hz, which suggests that this /r/ was articulated with the tongue curled strongly. F3 in the intervocalic /r/ is lowered only to an extent, because of the lesser degree of r-coloring, illustrated with the example of 2100 Hz for the intervocalic /r/ in *berry*. The possible F3 value for a native speaker's /r/ varies slightly across studies. For example, Saito and Lyster (2011) found that only F3 was an important cue to predict native speaker judgment, and reported that, of the /r/ tokens that the Japanese learners of English produced, those with F3 ranging from 2200 Hz to 2300 Hz were regarded as good examples of English /r/.

In contrast, /l/ does not have low F3. Figure 2.6 shows F3 of /l/, not lowered, unlike /r/. The boxed portion in the figure is the whole /l/ token. The horizontal line indicated by the arrow is F3 to be measured in Hz. As evident in a comparison with F3 of the sounds preceding and following /l/ in the figure, F3 of /l/ has no abrupt change.

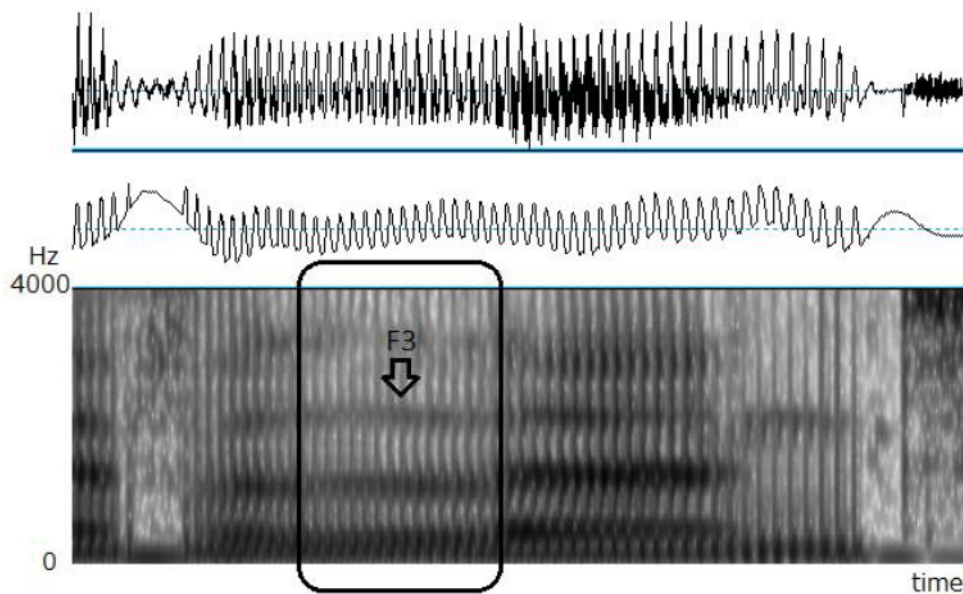


Figure 2.6. Spectrogram of [l].

F3 was one of the three measurements that Flege, Takagi, et al. (1995) analyzed, duration and F1 and F2 values at the release point being the others. They reported that the average F3 value of /l/ produced by native speakers of American English was 2854 Hz. Although Saito and Lyster (2011) did not design their experiment to measure /l/, they implied,

based on the results of the native speaker ratings of the Japanese learners' English, that authentic /l/ would be produced with F3 of 2800 Hz, while the tested tokens were judged as either /r/ or /l/ when they had F3 values of 2400 Hz to 2600 Hz.

Formants of /l/ are also generally weaker than those surrounding it. The spectrogram in Figure 2.6 shows [l] in *at last*, and the preceding [ə] in *at* and the following [æ] in *last* has clearer, darker formants. This is another acoustic cue of /l/ to be noted, called anti-formants (Kent & Read, 2002). Because /l/ is a lateral approximant, articulated with the tip of the tongue on the alveolar ridge, the air flow goes out through the side(s) of the oral cavity. This blockage causes the energy to radiate, which is reflected as anti-formants on the spectrogram, as also seen in the acoustic feature of nasals.

#### **2.4.4. The current study and hypotheses regarding approximants**

Previous studies have pointed out the difficulty of Japanese learners learning to produce English /r/ and /l/. It is generally agreed that less experienced learners had difficulty in producing these approximants in a native-like manner. There seems a lack of consensus as to which approximants are more likely to be learned, however, especially concerning production. Whereas previous studies suggest that /r/ seems to be more perceptually salient than /l/ in relation to Japanese /r/, their findings regarding how much this salience affects learning are inconclusive. Some argued that /r/ was learned with more ease, while others maintained that both were equally learned. Aoyama et al. (2004) claimed, for instance, that /r/ showed a greater degree of improvement than /l/, but their results still revealed that the production of /l/ was better judged on both the first test and the second test. The definition of learning in these studies is also a little ambiguous. Although many previous studies employed the judgments of native speakers (Aoyama et al., 2004; Goto, 1971), higher identification does not necessarily provide evidence of the subjects having learned the target approximants, as noted by Flege, Takagi, et al. (1995).

In order to address these issues, this study analyzed the production of these two approximants by less experienced Japanese learners, using acoustic analyses. Under the framework of the SLM, their production was hypothesized as follows: /r/ and /l/ would be



learnable and difficult for Japanese learners of English, respectively.

It was predicted that /r/ would be learnable, defined as a new phone. As suggested by the previous studies carried out within the framework of the SLM, /r/ phonologically and phonetically differs from any Japanese phones, and thus, it would be reasonable to claim that it is new. The major phonetic features of articulating /r/, such as retracted tongue, or lip rounding, are not prominently used in Japanese. This study therefore hypothesized that the newness of /r/ to Japanese learners of English was high. However, although previous studies found that it would be possible for Japanese learners of English to learn to produce /r/, they also suggest that training (Saito & Lyster, 2001) and exposure or age (Slawinski, 1999) would be key factors in promoting learning. These findings imply that there is no strong ground for hypothesizing that /r/ would be easy to learn. Thus, this new phone was predicted to be learnable for less experienced Japanese learners of English.

In contrast, /l/ was predicted to be difficult, and was defined as a new phone under the framework of the SLM. Phonologically speaking, /l/ is categorized as a lateral approximant whereas Japanese does not have any counterpart of this consonant pronounced in the same manner of articulation. This suggests that there is a difference in articulation between English /l/ and Japanese /r/, and thus, it was regarded as new. This definition is also based on the argument of Guion et al. (2000) that /l/ was not a good example of any Japanese consonant. However, the newness of this consonant was considered to be low. Japanese /r/ has various allophones, with greater similarity to English /l/ than English /r/. One of the articulatory similarities between Japanese /r/ and English /l/ is that both require the tip of tongue as an active articulator and the alveolar ridge, or around this region, as a passive articulator. The difference lies in the duration of the hold phase, which is reflected in the difference that the English /l/ is continuant and Japanese /r/ is not. It is not known how important this difference is. Nevertheless, it would be reasonable to assume that /l/ is less new than /r/ to Japanese learners of English. Guion et al. argued that the English /l/ and the English /r/ have a different status in the L2 phonological space of Japanese learners of English. They maintained that the difference between /r/ and /l/ was less salient than that

between /r/ and /r/. Thus, /l/ was defined as a new phone with a lesser degree of newness, unlike /r/. The results of Slawinski (1999) imply that although /r/ and /l/ might be learned in parallel, with increased exposure and age, the temporal cue for /l/ might be learned slightly more slowly. This is also further evidence that /l/ can be differentiated from /r/.

## 2.5. Rhythm

### 2.5.1. Contrastive phonetics and phonology between Japanese and English

English and Japanese are, respectively, categorized as a stress-timed language and a mora-timed<sup>2</sup> language, as shown in Table 2.4. Mora-timed rhythm can be regarded akin to syllable-timed rhythm since the only difference between the mora and the syllable is that geminate consonants and long vowels form a single unit as for the mora (Kubozono, 1999). This categorization, stressed-timed rhythm vs. syllable-timed or mora-timed rhythm, is grounded on the basis of the rhythmicity, which means that English and Japanese differ in that the rhythmic beat recurs only at stressed syllables or every mora (Abercrombie, 1966; Pike, 1945; Port, Al-Ani, & Maeda, 1980; Port, Dalby, & O'Dell, 1987). It is thus theoretically believed that while the duration of every mora tends to be regular in Japanese, it is that of inter-stress intervals (ISIs) that tends to be regular in English (Pike, 1945).

Table 2.4

*Rhythm in Japanese and English*

	Japanese	English
Rhythmic class	Mora-timed	Stress-timed
Phonetic cues of prominence	Pitch	Pitch Intensity Duration Vowel quality

From the phonetic perspective, the English rhythmic pattern is created by adding

<sup>2</sup> The mora is the smallest syllable-like unit of sound in Japanese; however, there are some claims to posit the existence of the mora as a rhythm unit. Beckman (1982) argued that there was no strong phonetic evidence to support the mora as a rhythmic unit. Bloch (1950) also stated that syllable is a better term to describe Japanese, regarding Japanese syllable as “a unit of duration” (p.92).

stress and weakening syllables (Celce-Murcia et al., 2010). Cruttenden (1997) suggested that, regarding the stress placed on the prominence, fundamental frequency (F0) was most important in English stress, followed by duration and intensity in this order. Fujisaki, Hirose, and Sugito (1986) found that F0 was most closely associated with English stress, as were intensity, duration and formant frequency, albeit less consistently. Vowels in unstressed syllables are reduced in the opposite way. Reduced vowels occur in weak syllables and contribute to further weakening the syllable, maintaining the English rhythm, where listeners hear the sequence of stressed syllables and unstressed syllables. English has three vowels appearing in unstressed syllables, /ə/, /u/ and /i/, and these unstressed vowels, or weak syllables, are likely to be shorter in duration, have a lower intensity and a different quality than stressed syllables (Roach, 2000). Pitch is generally, but not always, lower in unstressed syllables. It is noticeable that vowel quality becomes centralized, to sound more like a schwa.

Table 2.4 presents the phonetic cues used in English, compared with those in Japanese. Mora-timed rhythm in Japanese is simply implemented by the regular length of each mora, and there is no distinction in duration among morae. One feature in Japanese which is close to English stress is pitch accent. Every mora of a word in Japanese has an intrinsic height of pitch, low (L) or high (H), called pitch accent, and this functions lexically. As a result of the pitch difference, the two words *aya* (a figure of speech) and *Aya* (this author's first name) are differentiated in meaning, with the former recognized as having no pitch accent and the latter recognized as having the pitch accent on the first mora. However, stress in English and pitch accent in Japanese are functionally and phonetically different from one another (Venditti, 2005). Fujisaki et al. (1986) noted that only F0 expresses Japanese pitch accent, as presented in Table 2.4, which means that Japanese and English have some phonetic differences in the realization of prominence.

### **2.5.2. Learning L2 rhythm**

Because of the differences between Japanese and English in the rhythmic pattern and its phonetic realization, it is generally claimed that Japanese learners of English have difficulty in learning the rhythmality of English. Takefuta (1982) noted that rhythm is one

of the main factors causing Japanese learners' English to sound unnatural, and this has been supported by a series of empirical studies.

Lee, Guion, and Harada (2006) carried out acoustic measurements on the duration, F0, intensity and centralizing vowel quality produced by 10 early and 10 late Japanese and Korean learners of English in order to demonstrate the effect of the L1 prosodic system on L2 and the effect of language transfer depending on the age of acquisition. They analyzed the production of disyllabic or trisyllabic target words that differ in orthography, using duration, F0, intensity and vowel quality. It was reported that both the early and late Japanese learner groups produced a lower F0, a shorter duration and a weaker intensity as did the native speakers of English, although the early bilinguals tended to be more native-like for the first two items. However, a native-like centralization of vowel quality was hard to achieve for both the early and late Japanese learners.

The difference between native speakers of English and Japanese learners of English in implementing English schwa was also examined by Hatano and Kitamura (2014). Not only did they analyze F1 and F2 and duration acoustically, but they also measured articulatory behavior using an X-ray microbeam in their experiment. Three target words each for /ə, ɜ, ʌ, æ/ were produced by 16 native speakers of English and 9 native speakers of Japanese who had lived in the United States for a minimum of 8 months at the time of the experiment. It was found that the native speakers of English produced /ə/ in a shorter duration than /æ/, while the Japanese learners of English produced /ə/ in a shorter duration than /æ/ and /ʌ/. The two groups differed in vowel quality, in that the native speakers of English produced further back /ə/ than /æ/ at a significant level, reflected in a lower F2 value. The results of analysis of the tongue movement showed that, according to the production of the native speakers of English, the articulation of /ə/ was affected by the following consonants, indicating that schwa was not constantly articulated at the stable tongue height and position. However, this influence was not found in production by the Japanese learners of English. These results supported Lee et al. (2006), corroborating the findings that schwa vowel quality, but not duration, was difficult for Japanese learners of English.

Although Lee et al. (2006) and Hatano and Kitamura (2014) argued that Japanese learners of English achieved a shorter duration in unstressed syllables, Sudo and Kiritani (1991) obtained different findings. They claimed that the unnaturalness of English in the production of non-proficient Japanese learners of English probably resulted from this difference in durational patterns from native speakers of English. Measuring ISIs produced by the three groups, native speakers of American English, proficient Japanese learners of English and non-proficient Japanese learners of English, they showed that the durational increment of the ISIs was the greatest for the non-proficient Japanese learners of English, followed by the proficient Japanese learners of English and then by the native speakers of American English. Both the native speakers of American English and the proficient Japanese learners of English more remarkably shortened the unstressed vowels in the ISIs, in which they did not significantly differ. The difficulty of implementing English rhythm with duration was also found by Aoyama and Guion (2007), who measured the duration of function words. They maintained that the duration of function words was longer for the native speakers of Japanese adults and children. While these studies employed different durational measurements, both suggested the difficulty of shortening unstressed vowels for Japanese learners of English.

Satoi, Yoshimura, and Yabuuchi (2005) also suggested the difficulty in realizing the durational aspect of English rhythm. Eight Japanese undergraduates and eight native speakers of English read six sentences, which were categorized into three: a full vowel set that contains only full vowels, a reduced vowel set that contains both full and reduced vowels and a mixed vowel set that contains both full and reduced vowels and longer sentences than the reduced vowel set. The target sentences in these sets were read at three speaking rates, slow, normal and fast, in which vowel quality and duration of full vowels and reduced vowels were compared. Their results showed that the durational variability of full vowels and reduced vowels was found to be evident at all three speaking rates for the native speakers of English, whereas it was only found in slow speech for the Japanese learners of English. In a comparison between the full vowel set and the reduced vowel set, the two groups also

significantly differed in the performance of the reduced vowel set; the native speakers of English showed greater variability in this set than the Japanese learners of English. Sato et al. added that the durational variability of full vowels and reduced vowels decreased in the production by the Japanese learners of English as the sentences had more syllables. They demonstrated that whereas the distribution of full vowels did not differ significantly in vowel quality between the two groups, that of reduced vowels did differ. However, because the Japanese learners of English distinguished the tested reduced vowels from the tested full vowels in their own productions, they concluded that Japanese learners of English reduced vowels in a different manner from native speakers of English.

Sudo (2010a) pointed out both the difficulty and potential of learning to realize the durational properties of English rhythm. She demonstrated the effect of the experience of studying in the United States on the learning of English rhythmic patterns by comparing two junior high school students with a year's experience of learning English: one who studied English in the United States and the other who studied English in Japan. According to her results, a year of English study in Japan only enhanced the production of longer stressed syllables in duration, whereas that in the United States additionally improved the production of weak forms of unstressed syllables such as *the* and *to*.

These findings in Sudo (2010a) were consistent with Sudo and Kaneko (2006) and Sudo (2010b) in that these studies all showed the potential for Japanese learners to improve their English rhythmic patterns. Results from Japanese college students supported Sudo (2010a). The difference between these studies was that Sudo and Kaneko and Sudo (2010b) maintained that, through pronunciation training, Japanese learners of English would be able to improve their production of English rhythmic patterns without studying abroad. Subjects who were enrolled in an English class not focusing on pronunciation training did not show improvement in the stressed syllables of the content words nor the unstressed syllables in the functions words. Although not an empirical study, Shimada (2005) noted that some training, including explicit teaching of weak forms and strong forms, and pronunciation practice using songs, would be effective in helping Japanese learners of English to improve their production

of weak vowels.

It has thus been argued that the rhythmic features of Japanese affect the realization of the rhythm of English at a phonetic level (Ohata, 2004; Takefuta, 1982), and Lee et al. (2006), Hatano and Kitamura (2014) and Sato et al. (2005) revealed the difficulty of learning to centralize vowel quality, in particular. It is possible for Japanese learners to use other cues, such as intensity and F<sub>0</sub>, according to Lee, et al. Studies by Sato et al., Sudo and Kiritani (1991), Sudo and Kaneko (2006) and Sudo (2010a, 2010b) suggested that duration is difficult for Japanese learners of English to learn to use, but at the same time, these studies except for Sato et al. revealed that they may be able to improve their durational property to produce an English rhythm. In contrast, Lee et al. found that it was possible for even late learners of English to produce unstressed syllables shorter in duration. Therefore, it would be at least possible for Japanese learners of English to learn to realize English rhythm using the durational cue, although the level of difficulty in learning to use this property is still open to question.

### **2.5.3. Acoustic measurements of rhythm**

Pike (1945) started to use the terms, stress-timed and syllable-timed, to classify the distinctive feature of the rhythm unit, and all languages in the world are considered more or less stress-timed or syllable-timed. There have been controversies about this dichotomy, however, because no evidence has borne out the existence of these theoretically-believed rhythmic classes. With the development of acoustic techniques and instruments, the failure to capture the constant duration of ISIs raised doubts as to the existence of stress-timed rhythm and syllable-timed rhythm (Cauldwell, 2002; Roach, 1982). As it now stands, languages are auditorily categorized into one of the two rhythm classes, which could be attributed to different phonetic and phonological properties such as vowel reduction and syllable structure (Cruttenden, 1994; Roach, 1982). Complicated syllable structures and reduced vowels, as in English, could possibly lead listeners to perceive some syllables as more salient than others.

The acoustic measurement that raised such doubts is the duration of ISIs. Roach (1982) failed to confirm the difference between stress-timed languages and syllable-timed

languages, finding that English produced by far the biggest variations in ISIs across all six languages tested. Bolinger (1965) could not reveal the contribution of ISIs to the rhythmic difference, either. In contrast, Sudo and Kiritani (1991), by applying the definition of ISIs that was different from Roach's but the same as Bolinger's, succeeded in differentiating rhythmic patterns between native speakers of English and proficient and non-proficient Japanese learners of English.

While the validity of this measurement seems open to dispute, there have been ongoing attempts to acoustically capture the durational differences in rhythm that potentially underlie the different rhythmic classes. One approach was attempted by Ramus, Nespoulet, and Mehler (1999). They employed a method, not categorizing rhythmic units based on stress intervals, but based on vocalic and consonantal intervals. Vocalic intervals correspond to the portion from the onset of a vowel or vowel cluster to the offset of that vowel or that vowel cluster. Consonantal intervals are equivalent to the portion from the onset of a consonant or a consonant cluster to the offset of that consonant or that consonant cluster. Ramus et al. adopted three measurements, and found that the portion of vocalic intervals, symbolized as %V, and the standard deviation of the duration of consonantal intervals, symbolized as  $\Delta C$ , directly explained the stress-timing, syllable-timing and mora-timing of eight tested languages. In their study, the lower %V for English, which is caused by the clustering of consonants, clearly classified Japanese and English into separate rhythmic classes. Grabe and Low (2002) used these indices in their research to classify stress-timed languages, syllable-timed languages, mora-timed language and mixed-languages into stress-timed rhythm, syllable-timed rhythm or unclassified languages.

Another method of measuring durational cues was developed by Low, Grabe, and Nolan (2000). They devised a method to measure the durational variability of stressed vowels and unstressed vowels of British English and Singapore English, based on the report that vowel reduction would contribute to the categorization of the stress-timed and syllable-timed rhythmic classes. The pairwise variability index (PVI) was used in an experiment, where the durational variability of successive vowels between full vowels and



reduced vowels was calculated and normalized for each pair. Using this index, Low et al. showed that Singapore English had less variability in successive vowel durations, and the vowel reduction significantly contributed to this difference. The PVI was also employed by Torgersen and Szakay (2012) to describe the rhythmic characteristics of English spoken in inner London, Hackney, and one in outer London, Havering. Grabe and Low (2002) also used this index in their research. Sato et al. (2005) adopted this method in their study of Japanese learners of English.

Deterding (2001) also focused on calculating the durational variability that could distinguish stress-timed languages from syllable-timed languages. He proposed a variability index algorithm, a slightly different measurement from that used by Low et al. (2000), to measure the rhythmic features of conversational speech in British English and Singapore English. This method is different from Low et al. in the following points: it normalized the data based on the whole utterance, excluding the final syllable from the analysis, and the target of measurement was the duration of each syllable, not vowel. Using this procedure, they successfully showed more variability of syllable duration in British English than Singapore English. Deterding demonstrated that speaking rate did not have a substantial influence on the durational variability of syllable and vowel reduction; that is, the use of schwa had some effect on rhythm at a less significant level.

All the methods described above attempted to capture the durational variability; however, rhythmic pattern is also implemented through differences in pitch, intensity and vowel quality between stressed and unstressed syllable. Low et al. (2000) therefore dealt with not only durational variability but with spectral variability between stressed vowels and unstressed vowels. The degree of vowel centralization or dispersion was obtained by first calculating the mean F1 and F2 values, called the centroid, and then averaging the distance between the centroid and each vowel. The results suggested that reduced vowels were spectrally more centralized in British English. The degree of vowel centralization of unstressed vowels was also investigated by Lee et al. (2006), along with duration, intensity and pitch. They defined the extent of vowel centralization by calculating the Euclidean

distance between all the pairings of tested vowels and averaging them out.

#### **2.5.4. The current study and hypotheses regarding rhythm**

Research into the classification of rhythmic patterns has been conducted alongside the development of acoustic measurements, and this has contributed to devising various measurements, as described in Section 2.5.3. However, there are some methodological issues in these studies about learning rhythm. One issue is that these measurements have rarely been applied to the research on learning the rhythm of the target language. Sato et al. (2005) employed the PVI, developed by Low et al. (2000), but this is a rare study. Another issue is that in studies concerning the way Japanese learners of English implement English rhythm, the intensity and pitch of stressed syllables and unstressed syllables have been less examined than their duration and vowel quality, and there has been particular focus on durational properties.

The current study has thus attempted the following two things. The first was to apply the durational variability index to investigate the realization of rhythm in successive vowels by Japanese learners of English. Although Sato et al. (2005) employed the method proposed by Low et al. (2000) as noted above, more studies are needed to demonstrate whether it can be replicated. The applicability of their variability index also needs to be considered. The second attempt of this study was to examine the use of acoustic cues by Japanese learners of English, such as intensity and pitch, which have been investigated less often. Although Lee et al. (2006) measured all these items, they used as material a list of words consisting of disyllabic and trisyllabic words that contain stressed and unstressed vowels. The current study investigated the realization of weak vowels in weak forms in a task involving reading a passage aloud, using four phonetic items, pitch, intensity, duration and vowel quality.

It was hypothesized that it would be difficult for Japanese learners of English to realize rhythm in durational variability. As for the production of weak forms, it was hypothesized that intensity would be learnable, and pitch, duration and vowel centralization would be difficult for Japanese learners of English to learn to use.

The durational variability was defined as a similar item in terms of the realization of

rhythm, following the theoretical assumptions of the SLM. While morae in Japanese are implemented by duration (Bloch, 1950), the difference in the duration between one mora and two morae is not equal to that between stressed vowels and unstressed vowels. This prediction was also based on the study by Sato et al. (2005). They reported that Japanese learners of English lacked the durational variability of vowels, and thus, it was predicted that durational variability would be a difficult item for them to realize.

Regarding the realization of weak vowels in weak forms, duration was defined as a similar item, just like the above, under the framework of the SLM. There has been no agreement about the use of acoustic cues of duration to implement English weak vowels in previous studies, and thus, based on the theory that Japanese uses the durational distinction of vowels differently from English as noted in the previous paragraph, duration was defined as similar, leading to the hypothesis above.

The remaining three items, intensity, pitch and vowel centralization, were considered new, similar and new items, respectively, following the SLM. These definitions were based on Fujisaki et al. (1986), who claimed that only pitch contributed to the realization of Japanese pitch accent, as described in 2.5.1. It was predicted that Japanese learners of English would have difficulty using pitch as a cue to realize English rhythm, defined as similar. Pitch is a phonetic cue to produce Japanese pitch accent, which suggests that it is used phonetically. At the same time, the Japanese pitch accent is realized with H or L tones, meaning that pitch is also phonologically used in Japanese, but its function differs from the function of pitch in English. The critical difference is that the pitch height of morae is lexically determined, while this is not the case in English. Although Lee et al. (2006) argued that the Japanese learners of English succeeded in using pitch in implementing English rhythm, it was thus predicted that Japanese learners of English would have difficulty in doing so.

Intensity and vowel centralization were defined as new; it was predicted that the former would be learnable and the latter, difficult. Because they were considered new, the hypotheses were established considering the newness of these items. There is no theoretical evidence to define the degree of newness for intensity because it is a paralinguistic feature in

Japanese, rather than a linguistic feature. The prediction was thus based on the study, reported by Lee et al. (2006). Intensity was one of the items that they unexpectedly found to be well-learned by Japanese learners of English, regardless of proficiency, and thus, this item was predicted to be learnable. On the other hand, the newness of vowel centralization was regarded as low due to the ambiguity in its quality. This is supported by Hattori and Kitamura (2014), who found that schwa was not articulated with the decisive articulators, affected by the subsequent consonants. This would make it difficult for Japanese learners of English to detect its centralization. Hattori and Kitamura found that schwa produced by the Japanese subjects in their study had a clearer quality, close to that of full vowels. Given these subjects who had lived in the United States for a minimum of 8 months, less experienced learners of English in this study would have more difficulty attaining the vowel centralization of weak vowels. The hypothesis that vowel centralization would be difficult for Japanese learners of English was therefore put forward.

## **2.6. Intonation**

### **2.6.1. Contrastive phonetics and phonology between Japanese and English**

The way pitch functions differs between the two languages. In English, the sequence of rising and falling pitch shapes intonation and it functions to express the meaning of utterances in a given context. It does not change the meaning of words or sentences, but different shapes of pitch contours convey the different nuances that speakers intend (even if it is not intended). Wells (2006) notes that English intonation has the following functions: the attitudinal function, grammatical function, pragmatic function, discourse function, psychological function and indexical function.

In contrast, pitch has two major functions in Japanese; one is to implement pitch accent and the other is to implement intonation. Pitch accent involves the H or L tones that each mora holds, which function lexically, as described in Section 2.5.1. This is the function that characterizes pitch accent languages. Pitch does not function to distinguish lexical meanings in English, and this is therefore what makes Japanese different from English.

Pitch works as the basis of intonation in Japanese, as in English. Because each word

has a fixed, intrinsic height of pitches in Japanese, the basic shape of the intonation contour is highly dependent on the pitch accent. Kori (2011a) notes two factors that affect its shape other than pitch accent. The first is the phrase-final mora. Kori (2003) categorizes the standard Japanese pitch pattern on the phrase-final mora into five tones, question rise, emphatic rise, fall, rise-fall and level, claiming that they are similar to tones in English. The tone movement, which involves some rise, is also found in Japanese, and termed boundary pitch movement (BPM). Although more detailed examination might be necessary for this categorization (Kikuchi, Miyajima, & Shen, 2013), Venditti (2005) recognized five BPMs: prominence-lending rise, insisting rise, incredulity question rise, information question rise, and explanatory rise-fall. The second factor that affects pitch shape is the phrase-initial rise. The presence and degree of phrase-initial rise depend on the presence of focus and a determiner before the phrase concerned. Focus can change the pitch range (Kori, 2011b), maintaining the relative height of the pitch accent of the words.

One of the principal differences in the realization of intonation between Japanese and English comes from the presence of pitch accent in Japanese. In English, words do not each have a fixed height of pitch, and pitch height depends on the placement of accent on the stressed syllables. While the use of focus is similar between the two languages, English has greater pitch variation because of this functional difference, as Venditti (2005) and Maekawa (1999) have pointed out. The other difference is the presence of pitch movement on the end of the prosodic phrase, represented by the BPMs in Japanese. In contrast, the major highlight of pitch movement can be on any syllable in English. The syllable where pitch changes prominently in the intonation phrase (IP) is called *nucleus*. Although the basic rule of the nucleus placement is that it is located on the last accented syllable, it can be located on any syllable depending on the focus. Consequently, the last word or syllable does not necessarily hold pitch movement in English, differently from Japanese.

There do seem to be some similarities. Some tone types are shared by both Japanese and English intonation systems. For instance, O'Connor and Arnold (1973) categorized English nuclear tones, pitch patterns on the nucleus in the IP, into seven types: low fall, high

fall, rise-fall, low rise, high rise, fall-rise and level. The comparisons between Kori (2003) and O'Connor and Arnold, for example, make it clear that all tones except fall-rise are available in Japanese. However, from different points of view, Japanese and English still differ in tone types. Some pitch patterns are the same between Japanese and English, which simply implies that the two languages have typological similarities. It does not necessarily mean that the use of tones is equivalent in the other dimensions that the LILt proposed, the phonetic, semantic and frequency dimensions.

In terms of the overall realization of intonation of the IP, there are two phonetic items to be often described: span and level (Ladd, 1996). They are equivalent to key and register (Cruttenden, 1997), respectively. Span involves the range of F0, and the level, the average height of F0. Although the span has been well measured for learners' interlanguages including Japanese learners of English, studies which directly compare the span in Japanese and that in English are scarce. As for the level, Japanese females has been found to use a higher pitch. Yamazawa and Hollien (1992) measured the pitch height of 32 female native speakers of Japanese and 24 female native speakers of American English when speaking their L1, and found that the native speakers of Japanese used a higher pitch than the native speakers of American English. The same tendency was also reported by Ohara (1992). Ohara claimed that Japanese women tended to speak in a higher pitch in Japanese than in English, while Japanese men tended to use a similar level. Tsuji (2004) presented similar findings, but maintained that both female speakers of Japanese and English used a higher level in a particular situation: native speakers of Japanese tended to use a higher level when talking to their clients or customers, while native speakers of English used a higher level when talking to their friends on the phone. The use of higher pitch in a particular situation was less notable for male speakers of Japanese, but those of English showed the same tendency as female speakers of English. She added that a higher level is regarded as friendliness in English, whereas it is regarded as politeness or deference in Japanese. The level of pitch is probably related to sociocultural factors (van Bezooijen, 1995).

### 2.6.2. Learning L2 intonation

Setter, Stojanovik, and Martínez-Castilla (2010) describe English intonation as “notoriously difficult to teach” (p. 369). It is one of the last phonetic and phonological elements for learners to learn and for researchers to investigate in the field of second language learning. The acquisition of the L2 prosody is remarkably difficult for language learners (Jenkins, 2000). Ortega-Llebaria and Colantoni (2014) reported that L1 had a stronger influence on learning L2 intonation than did the level of proficiency level.

However, intonation is recognized as playing an important role in conveying messages and showing attitude and emotion. Wells (2006), for instance, emphasized the importance of learning English intonation, insisting that native speakers are less tolerant of intonational errors than segmental ones. It would therefore be important to examine learner implementations of English intonation, even if it is difficult to do so.

One of the traditional systems of analyzing intonation uses three domains called the three Ts, *tonality*, *tonicity* and *tone*, which, respectively, refer to the division of utterance into the IP, the placement of accent on syllables and the tone choice of nucleus (Halliday, 1967). Hahn (2004) emphasized the importance of learning tonicity, and investigated the effects of tonicity, that is, whether stress was correctly placed on appropriate syllables. She observed the impression that speeches by non-native speakers made to native speakers. In her experiment, native speakers of English listened to three non-native speeches that differed in the presence or accuracy of the primary stress, and their reaction time to the tone, their comprehension and their evaluation of the speeches were measured. She concluded that the text that was easiest to comprehend and process was that where primary stresses were correctly placed, following the rule that new and contrastive information receives stress.

The difficulty in Japanese learners learning tonicity was also demonstrated by Wennerstrom (1994), who examined Spanish learners of English and Thai learners of English, too. The results showed that the Japanese learners of English could not successfully produce contrastive or new information with a higher pitch, unlike the native speakers of English. It was also found that the Japanese learners of English frequently used a falling tone, indicating the difficulty of learning different tones. This was true in the free speech task in the

experiment, and she reported that the Japanese speakers used low boundary tones more often than the English speakers, 46% and 12% of the time, respectively. This same was also true of the Thai speakers, but not the Spanish speakers.

The need for Japanese learners of English to learn other tones was also noted by Arimoto, Yamamoto, Yamamoto, Kochiyama, and Makino (2008). They attempted to identify which intonational features could improve the realization of overall intonation, by rating production by Japanese learners of English. Three groups of raters, native speakers of English, Japanese learners of English and non-native speakers of English, rated the appropriateness of intonation used by Japanese learners of English in their experiment. They pointed out three features of intonation as appropriate for Japanese learners to learn. The first feature concerned the nuclear tone choice appropriate in the context, depending on the syntactic category, such as declarative sentences, yes-no questions and so on. Arimoto et al. maintained that assuming a native speaker's English was not defined as their goal of pronunciation learning, the use of a falling tone, rising tone, rise-fall tone and fall-rise tone would be recommended for learners. A fall-rise is described as a tone that can indicate that a speaker has something else to say. The second feature is a complete falling tone. Arimoto et al. argued that some utterances judged as not appropriate were characterized as having an incomplete falling tone. These two features are related to tone, but the last feature concerns tonicity. The third feature is the correct location of focus. One example of an error that they noted showed that the focus was placed on the last word of the sentence. They concluded that these intonational features would play an important role in attaining smooth communication.

Joto (1983) also investigated both phonological and phonetic aspects of intonation produced by Japanese learners of English. Six Japanese learners of English read nine dialogues, and their production of intonation was acoustically compared with that produced by a native speaker of American English. She found that Japanese learners of English had difficulty using a falling tone for wh-questions and placing the nucleus on an appropriate syllable. The results also showed that the subjects were not able to properly use a rising tone in declarative sentences to express attitudinal meaning. These findings illustrate the difficulty



of learning to implement the phonological aspects of English intonation with respect to tonicity and tone. Joto's acoustic analyses also depicted the different shape of pitch contour produced by the Japanese learners of English and native speakers of English. The differences are in that the former used a flatter pitch overall, with the highest peak in the earlier or the last part of the sentence, and the latter put the highest peak at the last sentence stress. Pitch range was also found to be smaller in the production of the Japanese learners of English. This was corroborated by Sato (1999), who reported a higher pitch at the beginning and an overall flatness in the pitch contour produced by Japanese learners of English.

These studies show the aspects of English intonation that would be problematic to Japanese learners of English. One problem concerns tonicity, in that they tend to locate the nucleus on inappropriate syllables, which was pointed out by Wennerstrom (1994), Arimoto et al. (2008) and Joto (1983). Another problem is that Japanese learners were not able to use some tone types authentically. A falling tone was reported as more commonly used by Wennerstrom, and a failure to use a rising tone in declarative sentences, substituted by a falling tone, was shown by Joto.

Kamura (2011) reported L1 transfer of tonicity on the realization of nucleus placement, conducting an experiment with four non-native groups that were different in the L1's prosodic system. She found that there was no correlation between the production and perception of nucleus placement by Japanese learners of English, and that they performed poorly in tonicity both in production and perception. A similar finding was also obtained by Saito and Ueda (2011), who also summarized four major types of errors in nucleus placement; pronouns such as *I*, interrogatives, attributive adjectives and negatives, as four syntactic categories in which the nucleus is placed in utterances produced by Japanese learners of English.

Saito and Ueda (2011) did not provide empirical data, but Maeda (2005) also realized that Japanese learners of English tended to locate the nucleus on interrogatives and pronouns. A total of 45 subjects participated in an experiment: 34 Japanese university students and 11 native speakers of English. The Japanese subjects were divided into two

groups: 19 students who had never been to North America and 15 students who had been to local schools in the United States or Canada. Their productions of a dialogue were acoustically analyzed as to F0, duration and intensity. The predicted tendency was found in the less proficient group, who put the prominence on the wh-words and pronouns with a higher F0, a greater pitch change and a longer duration. The more proficient learner group performed in a manner more similar to the native speakers. At the same time, it was also found that some subjects in the more proficient group used non-native ways of shortening the non-nuclear words in the pre-head and tail.

These claims about difficulty in learning tonicity are mainly grounded on the influence of L1. Another study which directly tested possible L1 transfer of intonation was carried out by Todaka (1994). Six target sentences were read by 20 subjects and 2 native speakers of American English, whose F0 values were measured. Although his interpretation of the pitch shape was based mainly on the visual observation of F0 contours and auditory impression, the following characteristics were found in the productions of the Japanese learners of English when compared with the native subjects: a smaller pitch range, sharp rise and fall, no deaccentuation in a contrastive sentence, more intonational boundaries and a delayed final rise for a question.

The difficulty in realizing span and level has been pointed out in previous literature (Maeda, 2005; Sato, 1999; Todaka, 1994), as noted by Joto (1983). It is agreed that Japanese learners of English use a narrower pitch range. Narita and Tanaka (2012) acoustically and statistically investigated the production of English intonation produced by 10 native speakers of English and 24 native speakers of Japanese. The former subjects produced 100 utterances and the latter Japanese speakers were divided into two, each of which produced 50 utterances. Measuring the highest pitch and the lowest pitch in a sentence to obtain the pitch range, Narita and Tanaka reported that the pitch range of the Japanese speakers was smaller than that of the native speakers of English in 88% of the utterances. Mennen (2007) measured the span and level of English speakers and German speakers, and argued that span and level are language-specific and that it would be difficult for learners to approximate native speakers

regarding the phonetic implementation of intonation. The narrower span found in Japanese learners of English would therefore be persistent.

Aoyama and Guion (2007), however, argued that Japanese learners of English produced a greater pitch range than native speakers of English. They carried out an experiment where 16 native speaking Japanese adults, 16 native speaking Japanese children, 16 native speaking English adults and 16 native speaking English children read three target utterances, *I'm fine*, *Five dollars* and *They went to school*. According to their results, the pitch range was greater for the native speaker groups than for the Japanese speaker groups only in the last target utterance. There were more pitch differences in content words produced by the Japanese speaker groups than in those by the native speaker groups. Their findings are contradictory to the general argument for a narrower span for Japanese learners of English. Their target utterances were all short, which might have affected their results. However, this at least suggests that it may be possible for Japanese learners of English to learn to use a wider pitch range.

Nagamine (2002) also had positive findings for Japanese learners of English learning intonation, suggesting that problems in implementing English intonation could be remedied by training. He investigated how the use of intonation patterns would improve after 13-week pronunciation training. Fifteen Japanese learners of English and two native speakers of English participated in the experiment. The Japanese learners of English read a passage in the pretest and posttest, and sentences, including listing, were acoustically analyzed and also rated by four native-speaker judges. According to the results, the Japanese learners of English improved their production of intonation by using a wider pitch range, the same intonation pattern as the native speakers for the listing and a complete fall to the lowest pitch. They still paused frequently and placed the highest peaks on various incorrect words. Each native-speaker rater seemed to have their own criteria or some raters even seemed to change their criteria over the rating sessions. However, it would be worth exploring which phonetic and phonological aspects the participants improved in their English intonation.

### **2.6.3. Acoustic measurements of intonation**

Research on the phonological aspect of intonation developed as various researchers proposed transcription systems of intonation, as summarized by Watanabe (1994). One well-cited system is tonetic stress marks (TSM) by O'Connor and Arnold (1973), which was originally developed to transcribe British English. Another widely used system is Tones and Break Indices (ToBI) by Beckman and Elam (1997) and Beckman, Hirschberg, and Shattuck-Hufnagel (2005) and Pierrehumbert and Hirschberg (1990). This evolved under the framework of the autosegmental-metrical theory in the field of intonation phonology. The pitch is phonologically transcribed with a sequence of tones labelled with H or L. It was originally proposed in order to transcribe Mainstream American English, but has now been applied to various languages including X-JToBI (Maekawa, Kikuchi, Igarashi, & Venditti, 2002; Maekawa, Igarashi, Kikuchi, & Yoneyama, 2004). ToBI seems to be becoming more commonly used in the field of research, along with its broader application to other languages. However, deciding which system is the most learner-friendly would be another issue. For instance, Toivanen (2005) reported the complexity of the ToBI labelling system for teaching intonation. Estebas (2013) thus attempted to establish a new ToBI model called ToBI for Teaching and Learning (TL\_ToBI).

The phonetic items of intonation have been acoustically analyzed in previous studies, including span and level. Patterson (2000) defined the span and level as the difference in F0 between all the peaks excluding the sentence-initial and post-accented valleys, and the sentence-final low, respectively. He compared several scales in measuring these items in terms of which scale would best capture the span and level. The results showed that the span was best represented in the musical scale, semitones (ST). This finding was contradictory to that of Hermes and van Gestel (1991), who found that the equivalent-rectangular-bandwidth (ERB) scale best captured the features of the pitch fitting human auditory system. While these are measurements of the overall pitch range and pitch height, pitch can be also measured locally at a phonetic level. For example, Trofimovich and Baker (2006) examined local pitch movement, such as stress timing and peak alignment. Others include pitch differences between new information and given information and the onset of the pitch, which

were employed by Kang, Rubin, and Pickering (2010) in their study of suprasegmental features. Munro (1995) investigated the overall shape of the pitch contour by locally measuring the pitch height of the vowels of key words in utterances.

A model to capture the shape of the overall F0 contour was proposed by Fujisaki and Hirose (1984) in the field of engineering. In this model, the F0 contour is decomposed into two components called the phrase and accent components, each of which reflects the declination of the intonation and the local pitch movement. Fujisaki and Hirose reported that they successfully demonstrated characteristics of intonation in Japanese declarative sentences, using this model.

#### **2.6.4. The current study and hypotheses regarding intonation**

Scientific techniques and auditory judgment have revealed the characteristics of English intonation that Japanese learners of English use. As for the tonality, Todaka (1994) showed that Japanese learners of English tended to place more boundaries in utterances. This suggests that the utterances produced by Japanese learners of English are likely to be divided into more IPs than those by native speakers of English. Regarding tonicity, studies have pointed out the wrong nucleus placement in the productions of Japanese learners of English. According to Saito and Ueda (2011), pronouns, attributive adjectives, interrogatives and negatives are major syntactic categories that the nucleus falls on in their utterances. Maeda (2005) empirically found an additional nucleus placement on pronouns and interrogatives. Narita and Tanaka (2012) reported that Japanese learners put a higher pitch on function words, which would confuse listeners in terms of the nucleus placement in their utterances. Arimoto et al. (2008) noted the inappropriate use of tone. Above all, Wennerstrom (1994) and Joto (1983) explored which tone Japanese learners of English preferred to use. Wennerstrom argued that they tended to use a falling tone more often than native speakers of English. Joto found that Japanese learners could not use a rising tone in declarative sentences, but used a falling tone instead. Many studies also have shown the tendency for Japanese learners of English to have a narrower pitch when speaking English.

These findings have clearly depicted the English intonation of Japanese learners of

English. However, they have not been comprehensive. Compared with research on other elements of pronunciation, there are more aspects to be explored in more detail. Thus, this study has attempted to examine the production of intonation in declarative, considering the following issues.

First of all, while it has been claimed that there is a higher pitch in the earlier part of an utterance (Sato, 1999), the factors that cause it have not been well investigated. According to Narita and Tanaka (2012), native-speaker subjects used a higher pitch than Japanese-speaker subjects when utterances started with content words. The opposite pattern was found when utterances started with function words. This implies that Japanese learners of English do not necessarily place an extremely higher pitch at the beginning, even though they often do so. In addition to factors such as function words, pronouns and interrogatives (Maeda, 2005), other possible factors would also need to be examined.

Secondly, the nucleus placement in the latter part of utterances has rarely been noted. The basic principle of the nucleus placement of English intonation stipulates that a nucleus is placed on the last accented word in the IP. A higher pitch in the earlier part of utterances and an extra nucleus in utterances on pronouns, attributive adjectives, interrogatives and negatives have often been discussed as problems that Japanese learners of English face in the realization of intonation. A declination, a gradual downstep toward the end of the utterance, is another intonational characteristic that has often been touched on as regards their production. However, whether the correctly accented word in the latter part of utterances bears the nucleus should also be investigated.

Thirdly, exactly which tones are difficult for Japanese learners of English in what contexts needs to be further examined in order to offer practical, pedagogical implications. Whereas assessments of the use of intonation have been conducted by human raters, according to appropriateness, naturalness, nativeness or foreign-accentedness (Munro & Derwing, 1999), some past studies do not clearly suggest which tone types should have been used in a given context. Although many studies compared Japanese learners with native speakers, it would be difficult to generalize the native-speaker performance from the results

produced in the study with a limited native-speaker sample size because more than one tone is acceptable in most contexts.

Finally, while a narrower span of Japanese learners of English has often been reported, there have not been many studies regarding their level, pitch height, particularly for Japanese male speakers. As noted in Section 2.6.1, Japanese female speakers have been the focus of investigation more commonly than Japanese male speakers. This is probably because they are widely known to speak with a higher pitch. Ohara (1992) and Tsuji (2004) are two studies that dealt with the issue, focusing both on female and male speakers, but the number of studies is still limited.

The current study aimed to investigate the production of intonation in terms of phonological items and phonetic items in order to address the issues described above. Nucleus placement and nuclear tone choice were targeted in phonological items, focusing on various utterances other than questions. Mennen (2007) noted that the phonological implementation of intonation may be learned first and that the phonetic implementation would follow. Only the span and level were therefore examined as phonetic items, and local measurements of tone realization were not employed. Hypotheses about these items were developed as follows.

As regards the nucleus placement, which concerns tonality and tonicity, it was hypothesized that it would be difficult for Japanese learners of English to place a native-like nucleus in two types of utterances: long utterances, termed *long utterances*, and utterances where the nucleus does not occur on the final word, termed *non-final utterances*. In contrast, it was hypothesized that it would be easy for them to place a native-like nucleus in utterances where the nucleus falls on the final word, termed *final utterances*. These hypotheses concerned the first two issues noted above.

This study predicted that the length of utterances would be another factor that caused the earlier high pitch that Japanese learners of English often use, as for the first issue. Maeda (2005) noted the greater difficulty in locating the nucleus correctly in long utterances. Joto (1983) also implied that a topic for further study would be whether there is some effect of

length of sentences, which suggests the difficulty of tonality and tonicity in long utterances. Todaka (1994) also found that the utterances that Japanese learners of English produced had more IPs, and thus, long utterances were predicted to be one possible factor causing difficulty in placing a native-like nucleus.

Tonality and tonicity in long utterances were then defined as similar for Japanese learners of English, which led to the hypothesis above. Japanese intonation and English intonation are different, in that the pitch accent already given to each word shapes the basis of intonation in Japanese. They are therefore different in the phonological dimension because the underlying structure of the intonation system differs between Japanese and English. However, tonality and tonicity in long utterances would seem similar to learners in terms of the phonetic dimension of the LILt. Both Japanese and English have the concept of IPs, which allows breaking utterances into IPs, depending on the grammatical structure or attitude. Intonation in both languages could also be realized by a sequence of a high tone and a low tone. Accordingly, assuming that the phonetic representation is emphasized as proposed in the SLM, the tonality and tonicity of long utterances would be regarded as similar. The current study has thus established the hypothesis that nucleus placement in long utterances would be a difficult item for Japanese learners of English.

This study predicted that whether the latter part of an utterance would correctly take the nucleus in the utterances produced by Japanese learners of English would depend on where the nucleus goes in the target utterance, regarding the second issue. That is, the performance of Japanese learners would differ depending on whether or not the nucleus occurs exactly on the final word in the utterance. This prediction is mainly based on the characteristics of Japanese intonation in which the end of the prosodic phrase holds pitch movement, such as BPM.

The tonicity of final utterances and non-final utterances were defined as identical and similar, respectively. As noted above, tonality and tonicity basically differ between English and Japanese in the phonological dimension of the LILt as a whole. The tendency for Japanese learners of English to fail to put a nucleus on appropriate syllables has also been



pointed out (Arimoto et al., 2008; Joto, 1983; Maeda, 2005; Wennerstrom, 1994). However, the phrase-final mora primarily holds pitch movement in Japanese, as described earlier (Kori, 2011a), which would make the nucleus falling on the final word in English utterances identical, even in the phonological dimension of the LILt. In contrast, this would not be applicable to non-final utterances. The nucleus can more freely occur in English. This is the same in Japanese in that a prominence could be placed anywhere. However, focus is a key factor in Japanese that places a prominence on non-final words, as Kori (2011b) described. This is not necessarily applied to English because the nucleus could be placed on non-final words in English regardless of the presence of focus, and therefore, tonicity in non-final utterances was not considered to be identical, but similar in the phonological dimension of the LILt. Thus, it was predicted that final utterances would be an easy item, while non-final utterances would be a difficult item.

Two hypotheses were formulated, focusing on the nuclear tone choice of various utterances other than questions, which concerned the third issue of which tone is specifically problematic for Japanese learners of English. First, it was hypothesized that it would be difficult for Japanese learners of English to use non-falling tones for the utterances where they are preferable, termed *non-falling utterances*. In contrast, it was also hypothesized that it would be easy for them to use a falling tone where it is appropriate, termed *falling utterances*.

These predictions were mainly based on the findings of previous studies. Wennerstrom (1994) reported that Japanese learners of English tended to use a falling tone with a higher frequency than native speakers of English. Similarly, Joto (1983) argued that a rising tone was hard for her subjects to use in declarative sentences. Taken together, it was predicted that Japanese learners would have no problem using a falling tone when required, while they would not be able to choose other tones properly. They might instead overuse a falling tone. In the phonological and semantic dimensions of the LILt, a falling tone is almost identical between Japanese and English in that it is the most common pitch pattern for various basic utterances such as declarative sentences and commands in both languages. They are different in the phonetic dimension of the LILt, as observed in an incomplete falling.

However, the nuclear tone choice itself was predicted to be an easy item because this study did not deal with the phonetic aspect of realizing pitch pattern. On the other hand, non-falling tones would be new or similar, depending on the tone types in the phonological, phonetic and semantic dimensions of the LILt. The tone defined as new was a fall-rise because Japanese does not have this tone. This suggests the complete newness of this tone to Japanese learners of English in the phonological dimension, and therefore, this type of tone would be easier than other non-falling tones. However, the newness of this tone was considered to be low because there has been no report supporting the possibility of Japanese learners of English learning to use a fall-rise in previous research, to the author's knowledge. In contrast, other non-falling tones tested in the present study were defined as similar in the three dimensions. Although there is no complete study, there is no strong evidence they are identical, especially in the semantic dimension, and thus, all non-falling tones were predicted to be difficult items in utterances other than questions.

Finally, it was hypothesized that span, pitch range, would be difficult and level, pitch height, would be easy for Japanese learners of English to learn to the level of native speakers, which concerns the final issue described above. In the phonetic dimension of the LILt, span and level were predicted to be similar and identical, respectively. The definition of span is based on a previous study (Joto, 1983), claiming that Japanese learners of English have a narrower span, although Aoyama and Guion (2007) showed the opposite results. The prediction of level was highly dependent on the subjects in the present study, who were male. As will be described in 3.1, male speakers were targeted in the experiment. There was no previous study on which the present study could directly base its prediction, but studies by Yamazawa and Hollien (1992), Ohara (1993) and Tsuji (2004) led to the prediction that level would not be an important issue for male Japanese learners of English. The level was thus defined as identical as far as male speakers are concerned, and predicted to be easy for the male subjects in the present study.

## 2.7. Connected speech phenomena

### 2.7.1. Contrastive phonetics and phonology between Japanese and English

In connected speech, each sound affects the others in a stream of sounds, and various sound changes occur. These sound changes, connected speech phenomena, could occur in any kind of speech when speakers choose to allow them, while they often occur in a casual or fast speech. All sounds are coarticulated or connected with various ways, in a sense, and this takes place beyond the segment, syllable and word boundary. Thus, the term, connected speech phenomena, involves a broad variety of phenomena. When it comes to English, commonly cited connected speech phenomena include assimilation and linking across word boundaries. Of the various types of these phenomena, three connected speech phenomena often occur at word boundaries in English: elision, linking and assimilation.

Elision is the phenomenon where a sound is deleted. Weak vowels, such as /ə/, are often elided. Also, /h/ in weak forms such as in *her* and *his* is deleted, which is known as h-dropping. Elision occurs not only in unstressed syllables, but also to a consonant preceded and followed by another consonant. A common example is the elision of /t/ and /d/ as in *next day* [neks deɪ] and *and then* [ən ðen]. Casual speech could cause weak syllable elision even more radically. Anderson-Hsieh, Riney, and Koehler (1994) cite examples of *for problem* and *for it was*, pronounced as [prəm] and [iʊz], respectively.

Linking is the phenomenon where a word-final consonant and a word-initial vowel are pronounced without inserting any pause, just like one word (CV linking). For example, in the sequence of two words, *what a*, they are not separately pronounced. They are connected, and as a result, /t/ is flapped in American English. Linking does not accompany a practical change in sounds, but words are simply connected in connected speech.

There are also the linking of a consonant and a consonant (CC linking), and that of a vowel and a vowel (VV linking). Hieke (1984) maintained that these connections of sounds in these sequences were other types of linking additional to CV linking. A primary example of CC linking concerns the change in the release of a preceding plosive followed by another consonant. These plosives do not involve audible release. When they are followed by a nasal, they are released through the nasal cavity, called nasal release. When they are followed by a

lateral, they are released through both sides or one side of the tongue, called lateral release. The release of /t/ in *hot* is thus not heard in *hot potato*. Cruttenden (2014) and Takebayashi (1996) also describe these phenomena, although they do not explicitly define them as linking. They explain that these types of linking occur in more varieties of phonetic contexts, whereas Hieke noted only no audible release before a continuant and nasal release. VV linking is also a common phenomenon in English, but it differs from CV linking and CC linking in that another sound is inserted to connect two vowels. A glide, such as /w j/, is usually inserted, as in *party is* [pa:ti 'ɪz]. The intrusive /r/ works in a similar way, too.

Assimilation is also a commonly observed connected speech phenomenon, referring to the phenomenon where the place of articulation or the manner of articulation changes due to the influence of surrounding sounds. There are three directions of assimilation: regressive assimilation, progressive assimilation and coalescence. Regressive assimilation occurs due to the influence of the following sound on the preceding sound. For example, it is found in *right back*, where /t/ assimilates to /b/, being pronounced [raɪp bæk]. Progressive assimilation involves the opposite direction of assimilation, which occurs less frequently than regressive assimilation and coalescence. Coalescence is the phenomenon where two successive sounds influence each other, leading to a sound with a mixed quality. It often happens when /t, d, s, z/ are followed by /j/, as in *could you* [kʊdʒə]. These kinds of assimilation differ from secondary articulation in that they involve a change in the phonological category of the sound; that is, even though /l/ is velarized, it remains a voiced alveolar lateral approximant, phonologically. In contrast, /t/ in *right* of the example above changes to a bilabial plosive when it assimilates to [p].

Japanese also uses elision, linking and assimilation. First, Takebayashi (1996) suggests *bokunouchi* (my house) pronounced as [bokuntɕi] as an example of elision. Elision is also seen as a more advanced effect of the devoicing of high vowels. High vowels /i, u/ are often devoiced in standard Japanese when they are surrounded by voiceless consonants (Vance, 1987), as in *aruku ka* [arukɯka]. These high vowels can be deleted. Secondly, the connected speech phenomenon similar to linking is *renjou* in Japanese. *Renjou* involves the

phenomenon where /n, m, t/ is connectedly pronounced with a vowel that follows. Because of this effect, *tannou* (proficient) consists of two components *tan* and *ou*, but it is pronounced as [tannou] with geminate. Although Yamada (2013) states that this is a kind of progressive assimilation, it is more similar to linking in that two sounds are simply connected without causing a change in the phonological category of the sound. Finally, *rendaku* is one of the most well-known phenomena of assimilation. For example, while *saka* (a hill) is pronounced as [saka], /s/ is voiced in a word such as *kagurazaka* [kagurazaka]. This voicing occurs when a voiceless consonant is surrounded by voiced sounds.

Japanese and English thus have common connected speech phenomena. However, they are more limited in Japanese regarding where these phenomena occur. *Rendaku* and *renjou* occur in compound words, which differ from English assimilation and linking at word boundaries. While the whole mora is often deleted, elision of consonants at the coda never occurs in Japanese. The variety of connected speech phenomena at word boundaries is wider in English.

One of the principal reasons for this lies in the difference in the syllable structure. The Japanese phonotactic pattern is simple. Its template is represented as  $C_0^1V$ , where C and V stand for a consonant and a vowel, and the lower number and higher number represent the number of consonants minimally required and maximally permitted, respectively. This template shows that Japanese has only two types of phoneme combinations as a syllable structure, CV or V, although there is an exception to /N/, which can consist of one syllable alone. In contrast,  $C_0^3VC_0^4$  is the template for English, which suggests that three consonants and four consonants are maximally allowed on onset and coda positions (Ashby & Maidment, 2005), as in *strong* /strɒŋ/ and *sixths* /sɪksθs/. More complicated phonotactic patterns in English lead to more types of sounds being adjacent at word boundaries, causing varied connected speech phenomena.

### **2.7.2. Learning L2 connected speech phenomena**

Some researchers insist that connected speech is not an essential element of pronunciation for learners, but a possible cause of misunderstandings. For example,

Cruttenden (2014) states that knowledge of connected speech phenomena is helpful for learners when listening to native speakers but does not strongly recommend that non-native speakers use these sound changes. Kashiwagi, Snyder, and Craig (2006) note cases of misunderstanding due to failures in using connected speech phenomena. They examined how well native speakers of English understand the productions of Japanese learners of English. One example of misunderstanding was related to linking. A failure of linking in the phrase *owns a* caused the elision of *s*, which was misheard as *owner*. Another example is that, according to their speculation, the assimilation of /d/ to /t/ in the phrase *wind takes* meant that none of the three native-speaker judges understood this phrase. As in these examples, the use of connected speech phenomena could lead to a misunderstanding of the message, and thus, the prime motive of studies concerning connected speech phenomena is to unearth the nature of language. A smaller number of studies have been conducted on the learning of these phenomena.

However, learning about these connected speech phenomena is considered helpful in terms of attaining a native-like pronunciation or improving fluency and understanding natural spoken English. As proposed by Cruttenden (2014), Arimoto (2002) claims that learning elision and linking is especially necessary in order to understand spoken English correctly and to maintain a natural English rhythm. An increasing awareness of the importance of learning these phenomena was also found in the results of Arimoto's survey, where all seven textbooks of pronunciation were found to describe them. This was corroborated by Ueda and Otsuka (2010), who compared six textbooks for middle schools. Their survey found that the textbooks analyzed covered all three connected speech phenomena on which Ueda and Otsuka focused, elision, assimilation and linking, except that one textbook did not deal with elision. They also pointed out, however, that how to implement these phenomena was not described sufficiently.

One of the studies regarding the learning of elision was conducted by Matsui (1998), who carried out both a perception experiment and a production experiment. Subjects identified synthesized target tokens as either *I can go* or *I can't go* in the former test, and they

produced the three contrast tokens of *can* and *can't* followed by a plosive in the latter test. The results showed that while native speakers consistently judged the tokens with an 80 ms or shorter hold phase between /n/ and a following vowel as *I can go*, and the tokens with a 140 ms or longer hold phase as *I can't go*. The ability to identify these tokens was inconsistent in Japanese learners of English. In the production experiment, the results of the native speakers' productions revealed that the absolute duration of the hold phase was 48.3 ms to 59.4 ms for the tokens of *can* and 142.2 ms to 328.3 ms for the tokens of *can't*. In contrast, the Japanese learners of English could not discriminate *can* and *can't*. It was also found, however, that the instruction, auditory input and repetition improved production. These findings suggest that it would be possible for Japanese learners of English to realize the /t/ elision of English phonetically.

Research into the linking produced by non-native speakers was carried out by Hieke (1984). Spontaneous, casual speech was recorded from 12 native speakers of English and 29 German learners of English with intermediate proficiency, and the frequency of using CV linking was counted. The results showed that the native speakers of English and the German learners of English used linking at a rate of 77.63% and 53.50% of the items out of all points at which linking was expected to occur. Based on this observation, Hieke also pointed out that inserting silent pauses caused a reduction in the absolute points at which linking could be used, which should be taken into account experimentally. German is a language in which linking is not used, at least not as frequently as in English. Therefore, these findings may help to predict the performance of Japanese learners of English, whose L1 also does not use linking at word boundaries in the same manner as English.

A study of the production of English linking by Japanese learners of English was conducted by Anderson-Hsieh, Riney, and Koehler (1994). They analyzed how often linking was used by five Japanese learners of English with intermediate proficiency and five Japanese learners of English with high proficiency, who were compared with five native speakers of American English. The subjects read sentences and made a spontaneous speech, for which potential and actual flapping, linking, vowel reduction and consonant cluster

simplifications were counted. The linking targeted in the study included CC linking, CV linking and VV linking. They found that the Japanese learners of English with high proficiency performed closer to the native speakers in CC linking. Both Japanese groups tended to insert a glottal stop before a vowel in CV linking, but there was no significant difference in their mean scores. The same tendency to insert a glottal stop was noted in VV linking, where the Japanese learners of English with high proficiency were not significantly different from those with intermediate proficiency or the native speakers of English. In contrast, the Japanese learners of English with intermediate proficiency were found to differ from the native speakers of English at a significant level. Their results also revealed that the rate at which linking occurs was higher in spontaneous speech than reading sentences, indicating that the style of speech could affect their performance.

Similarly, Maxwell (1997) targeted Japanese learners of English and examined how well Japanese learners of English were aware of connected speech phenomena. Thirty-eight subjects participated in an experiment. They were divided into two groups; one group wrote down the sentences that a native speaker of English read and the other group identified these sentences by selecting the correct answer from multiple choice questions. Analyzing one sentence, *Is there a cat in there?*, she found that no correct responses were given by the subjects in the sentence-writing group and only 9% of the subjects in the multiple-choice group. Observation of the sentence-writing group suggested that the linking of *cat* and *in* turned out to be problematic for the subjects. Maxwell, however, claimed that the sentences that this group transcribed contained /l/ or /r/, and that the result suggested that the subjects in the group had sensed assimilation, elision, linking and weakening. Most of the subjects in the multiple-choice group chose *Is there a calendar?* as the correct answer, which also suggests that they heard some connected speech phenomena and these phenomena made it difficult for them to comprehend the sentence. She concluded that learning connected speech phenomena would enhance a learner's listening skills, recommending exposure to these phenomena for Japanese learners.

A spectrographic comparison of assimilation was made by Sato (1999) using



spectrograms of *Please read these books*, produced by a native speaker of English and a Japanese learner of English. Because the degree of assimilation differs across even native speakers, the native-speaker subject in the study did not show complete assimilation. Nevertheless, there was an apparent difference between the native speaker of English and the Japanese learner of English. The former subject produced an intermediate sound for /ð/ between /d/ and /ð/ due to assimilation of /ð/ to /d/. Sato reported that /d/ in *read* was deleted at the same time, which could be instead interpreted as assimilation with no audible position. In contrast, no assimilation was found in the spectrogram of the Japanese learner of English, who produced a longer frication noise for /ð/. Sato argued that the Japanese learner of English failed to implement the assimilation, and did not even seem to have knowledge of assimilation.

### **2.7.3. Acoustic measurements of connected speech phenomena**

From a phonetic point of view, some connected speech phenomena involve an intricate movement of articulators. Therefore, analysis is often conducted using techniques to record movements directly, such as a three-transmitter magnetometer system and electropalatography. These techniques would particularly be required in experiments of assimilation and secondary articulation, which concern a change in the phonological category and the addition of the use of different articulators, respectively.

Conversely, in elision and linking, these connected speech phenomena are each realized by the omission and connection of sounds, to put it simply. The most common means of analyzing them is thus to count the items where these phenomena occur based on the spectrogram. The elision of /t/ and /d/ can also be analyzed using acoustic measurements, because it is recognized as involving longer hold phrases, as well as having no audible release. Matsui (1998) employed this method to observe the productions of *can* and *can* ɾ followed by a plosive.

#### **2.7.4. The current study and hypotheses and research question regarding connected speech phenomena**

As noted earlier, studies conducted on connected speech phenomena have been limited, compared with those on vowels and consonants for instance. However, the significance of learning these phenomena is recognized for the advantage of listening to spoken English accurately. Although the importance of learning to use connected speech phenomena in production has not been highlighted, it would be useful to enhance understanding of the phenomena by learning to produce them. This study therefore attempted to investigate elision, CC linking and CV linking, and to determine whether it could replicate the findings of previous studies, such as Matsui (1998) and Anderson-Hsieh et al. (1994). This study also looked into whether there was any effect of phonetic contexts on the use of CC linking and CV linking. Although elision occurs in /t/ and /d/, CC linking and CV linking are known to occur in a greater variety of phonetic contexts. This issue has not been well investigated in previous research.

Assimilation and VV linking are other common connected speech phenomena, but they were not the foci of investigation here for the following reasons. The degree and use of assimilation varies from native speaker to native speaker, and therefore there can be no decisive native-speaker reference defined. VV linking involves the insertion of vowel-like, glide sounds called semivowels, such as /j/ and /w/. Spectrographic analysis, which the current study adopted, is not necessarily the most suitable analysis for the examination of these phenomena because the boundary between a vowel and a glide is unclear on the spectrogram. These two phenomena were thus not dealt with in the current study.

A hypothesis was first developed as to the difficulty of elision, CC linking and CV linking. It was hypothesized that elision, CC linking and CV linking were all difficult items for Japanese learners of English. Following categorization by the SLM, these connected speech phenomena were defined as new phenomena. As already described in Section 2.7.1, elision and linking do occur in Japanese. However, they differ from elision and the linking occurring at word boundaries in English. While the elision of /t/ and /d/ occurs at a word-final position in English, /t/ and /d/ never appear on a syllable-final position in

Japanese. This syllable structure is not allowed in Japanese. Similarly, because a consonant never occurs in a syllable-final position, the CC linking and CV linking that are common in English are not found in Japanese. Although *renjou* was used as an example of the corresponding phenomenon in Japanese, it only concerns the formation of compound words. This is why these items were defined as new.

Because they were defined as new phenomena to Japanese learners of English, the newness of these connected speech phenomena was considered in order to predict the level of difficulty for less experienced learners, who were the target of the present study. From an articulatory point of view, elision and linking involve eliding a sound and connecting a sound. In that they do not require a new articulatory movement of the articulators for implementation, the newness of these phenomena could be regarded as low. This definition would be logical, given that elision and linking itself occur in Japanese. The previous studies cited above also failed to give strong evidence for the assumption that less experienced Japanese learners of English would be able to detect and learn to use elision, CC linking and CV linking with ease. Considering that Japanese learners of English tend to place more pauses in their utterances (Nagamine, 2002), a reasonable prediction would be that it is even more difficult for them to use connected speech phenomena in a native-like manner due to frequent pause insertion, and thus, the hypothesis above was formulated.

A research question about connected speech phenomena was addressed, concerning the possible effects on CC linking and CV linking as follows; whether there would be an effect of phonetic contexts on the use of CC linking and CV linking. Two conditions were tested for both CC linking and CV linking. The conditions for CC linking were the sequence of two consonants at the same place of articulation and in the same manner of articulation, and the sequence of two consonants at a different place of articulation or in a different manner of articulation. The two conditions for CV linking were the sequence of a voiceless consonant and a vowel, and the sequence of a voiced consonant and a vowel. These phonetic contexts were selected, considering possible articulatory difficulty.

## **2.8. Relationships between the elements of pronunciation**

### **2.8.1. Application of the theory**

DST was applied to address the second research question, as to whether there would be any supportive relationships between the elements of pronunciation examined: vowel quality, vowel duration, plosives, fricatives, approximants, rhythm, intonation and connected speech phenomena. Fewer studies have been conducted within the framework of this theory than in a model such as the SLM, as far as L2 pronunciation learning is concerned.

One of the studies conducted within the framework of DST concerned a child's lexical and syntactic development. Robinson and Mervis (1998) examined a relationship between lexical growth and grammatical growth within the framework of DST, using the longitudinal dataset for one male child from 10 months through 30 months of age. Vocal production was recorded, the words that he produced were counted and these words were also coded for lexical class, such as nouns predicates and closed class. Syntactic complexity was also investigated with respect to the mean length of utterance in morphemes and words and use of plural forms. According to their results, vocabulary size increased faster than plural forms were acquired until the lexical development reached a critical level. Both of the lexical and grammatical development then advanced together, they competed against one another, and the vocabulary size only grew once the plural form had been learned. This development could be described as follows, referring to DST. The vocabulary size and plural forms had a competitive relationship at an initial stage until the vocabulary size entered an attractor state. After that, they were related first supportively and then competitively, and reached the stage where the plural form was completely learned.

In the field of learning phonetics and phonology, Li and Post (2014) conducted a study, which implied the applicability of DST to their research. They investigated how prosodic items are related to the realization of English rhythm: accentual lengthening, which occurs on stressed syllables, and phrase-final lengthening, which occurs at the phrase final or utterance final positions. Rhythm and these two items could be considered as the system and components constituting the system, respectively. Mandarin learners of English with lower proficiency, those with higher proficiency, German learners of English with lower proficiency,

those with higher proficiency, a native English-speaking control group, a native Mandarin Chinese-speaking control group and a native German-speaking control group participated in the experiment, where they read 20 English sentences and 20 sentences in their L1. One of their findings concerning DST was that there was more accentual lengthening in L2 English by the more proficient learners only in a CV-syllable structure, but not in a more complex syllable structure. Although they did not describe this relationship, this result suggests that there was a conditional relationship between accentual lengthening and syllable structure.

### **2.8.2. The current study and research question regarding relationships between the elements of pronunciation**

DST is a very promising theory that is applicable to various fields, but no previous study has dynamically investigated the pronunciation system. One of the reasons for this may be that there are complicated variables already involved within each element of pronunciation. As in Li and Post (2014), rhythm contains various features interconnected with one another in its system. Another reason may be that longitudinal studies would better fit the study of DST. In order to establish effective learning and teaching standards of pronunciation, however, one of the most necessary perspectives is the exploration of the big picture. If DST could describe and explain the whole system of pronunciation, it would offer useful implications for learning and teaching pronunciation. To learn where a supportive relationship lies between components would be especially informative because it gives suggestions as to how pronunciation could be learned and taught effectively. Accordingly, the following research question was addressed: whether there is any supportive relationship between the elements of pronunciation in the learning process.

A hypothesis was not built for this question because of scarce research that is directly accessible on the issue. Instead, expectations were formed concerning which elements of pronunciation would have a supportive relationship with one another, based on the theory. The first pair was vowel quality and vowel duration. They were separately analyzed in the present study; however, they are usually regarded as one segmental category, vowels. Therefore, learners who learn to produce the vowel quality of English may produce

the durational distinction better, or vice versa. The second pair was vowel quality and rhythm. This study attempted to analyze English rhythm in terms of the representation of unstressed vowels and weak vowels. These two elements were thus expected to be in a close relationship. For the same reason, the third pair was vowel duration and rhythm. The fourth pair was rhythm and connected speech phenomena. Connected speech phenomena, such as elision and linking, are phenomena that facilitate maintenance of the rhythmicity of English by connecting two words. Therefore, these elements could also be supportively associated with each other. This applies to three consonants, plosives, fricatives and approximants, which could involve relationships. However, because they are apparently characterized differently from an articulatory viewpoint, it was presumed that these three consonantal elements would be less likely to have a supportive relationship in learning.

## **2.9. Summary of the chapter**

After Japanese and English were compared from a phonetic and phonological point of view, previous studies and acoustic measurements of the elements of pronunciation concerned were described. Based on the theoretical comparison and the previous literature, the aims of the current study and how the hypotheses were formed were noted. Table 2.5 presents a summary of the study hypotheses proposed in this chapter. In the table, the items of each element of pronunciation that were predicted to be easy, learnable or difficult are indicated, according to the hypotheses.

These hypotheses were developed within the framework of the SLM for segments, and the LILt for intonation, and following the theoretical assumptions for rhythm and connected speech phenomena, according to previous studies. As shown in Table 2.5, more items were predicted to be difficult than to be easy or learnable. This tendency implies that it would be generally difficult for Japanese learners of English to recognize differences between Japanese and English or to learn to produce English phonetic and phonological items, due to their lower degree of newness.

Table 2.5

*Summary of the Study Hypotheses*

Element	Easy items	Learnable items	Difficult items
Vowel quality		/æ/ /ɔ:/ /ɜ:/	/i:/ /ɪ/ /e/ /ʌ/ /ɑ:/ /ʊ/ /u:/
Vowel duration	/i:-ɪ/ /u:-ʊ/ /ɑ:-æ/		/ɑ:-ʌ/
Plosives			/p/ /t/ /k/ /st-sk/ /k-sk/
Fricatives			/f/ /θ/
Approximants		/r/	/l/
Rhythm		Intensity	Durational variability Pitch Duration of weak vowels Vowel centralization
Intonation	Final utterances <sup>a</sup> Falling utterances <sup>b</sup> Level		Long/non-final utterances <sup>a</sup> Non-falling utterances <sup>b</sup> Span
Connected speech phenomena			Elision CC linking CV linking

*Note.* <sup>a</sup>Nucleus placement was examined with these utterances. Final, long and non-final refer to utterances with a typical pattern where the nucleus fell on the final word, utterances that were long, and utterances with a typical pattern where the nucleus fell on the non-final word. <sup>b</sup>Nuclear tone choice was examined with these utterances. Falling and non-falling describe utterances where a falling tone was typical, and those where a non-falling tone was typical.

## Chapter 3 Methodology

### 3.1. Subjects

A total of 91 speakers participated in this study, where 72 were Japanese learners of English (JL), 12 were native speakers of British English (BN) and 7 were native speakers of American English (AN). As noted below, the JL's performance data were compared against those of BN/AN obtained from publicly available databases. It is known that gender affects the absolute values of acoustic items that speakers generate because of differences in the size of vocal folds and vocal tract. Females generally produce higher values for pitch and formants than males (Ashby & Maidment, 2005), and therefore, this study only collected data from male speakers.

The JL subjects who participated in the experiment were third-year high school students without previous experience of living in English-speaking countries. The subjects of this age group were selected so as to generalize the findings of this study to the broader population of Japanese learners of English. They were judged as the most suitable for the present research, considering the current situation where more than 98% of Japanese children receive high-school education (MEXT, 2011). The school that the JL subjects attended was a private boy's high school. Approximately half the students enrolled in this school come from affiliated junior high schools, and another 10% from an affiliated primary school. This means that not only had they completed a certain level of English education in Japan under the common curriculum guidelines of the MEXT (1998, 2009), but the group was also made up of potential speakers at various learning levels. This school environment was thus expected to provide data about learners who represented common features of Japanese learners of English, which would satisfy the purpose of this study.

According to the results of the Global Test of English Communication for Students (GTEC for STUDENTS) advanced-level test developed and administered by the Benesse Corporation, which most of the JL subjects had taken 2 months before the experiment, the average score among the subjects was 543 out of 810 (SD = 83.97, Max = 746, Min = 338). Although one JL subject did not provide data, it was estimated that his score would be



equivalent to the average because he was assigned to the standard level of the class in which most of the JL subjects in the present study were enrolled. Benesse Corporation (2015) reported the national average of this test to be 461 from 2012 to 2014, which suggests that the present sample's proficiency level was above that of the general high school students. However, the correspondence between the scores of this test and six levels of language ability described by the Common European Framework of Reference (CEFR; Council of Europe, 2001) shows that 543 points is comparable to the A2 level. The JL subjects in this study were thus defined as basic users of the language, and their data could be interpreted as a sample of ordinary Japanese learners of English.

The data for the BN subjects and the AN subjects were obtained from the UCL Speaker Database (Markham & Hazan, 2002) and the Audio Archive (Merfert, 1997), respectively, both of which are publicly available. Only speech samples for male speakers were selected as the subjects for this study. As noted above, this made it possible to remove paralinguistic features as much as possible and to compare them with the JL subjects. The former database contains the recordings of a variety of spoken data collected from female and male speakers. All speakers in this database had a neutral or mild south-eastern British English accent. The latter database, the Audio Archive, includes read speech samples by native speakers with different English accents, such as American, British, Canadian, Australian and Indian. The present study used the data obtained from native speakers of American English, including two mid-western speakers, one northeastern speaker, one southwestern speaker, one southeastern speaker, one western speaker and one speaker with influences of various American accents.

All data were designed to be used for all acoustic analyses described in Section 3.4. However, the AN data were not suitable for the spectral analyses of fricatives because the sampling rate to record at was 8 kHz, which failed to present spectra at higher regions of frequency. Only the BN and JL subjects thus underwent the analyses on fricatives.

### **3.2. Materials**

This study employed phonetically-balanced passages: *The Story of Arthur the Rat*

and *Arthur the Rat*. Data for the BN and JL subjects were collected using the former passage and those for the AN subjects, using the latter. While they were slightly different in some words used, they follow the same overall story line (see Appendix A).

One of the most well-known phonetically-balanced passages is *The North Wind and the Sun*, adopted by the International Phonetic Association (1999). Others include the *Rainbow Passage* and *The Story of Arthur the Rat*, as in the UCL Speaker Database (Markham & Hazan, 2002) and the Audio Archive (Merfert, 1997). Of these options, *The Story of Arthur the Rat* and *Arthur the Rat* were selected for the following reasons. Firstly, using these passages enabled almost the same spoken data to be obtained from both native speakers of British English and native speakers of American English. Although British English and American English are the most common accents of English taught in the classroom in Japan and spoken around the world, phonologically and phonetically speaking, they are known to have different features, especially concerning the quality of vowels and pitch patterns. Accordingly, data from native speakers of both accents were employed in this study. Secondly, these passages were phonetically-more balanced than any other passage. It was of the highest priority in selecting materials for this study that all types of phonemes occurred as frequently as possible, although it had the drawback that the more types of phonemes the passage contained, the longer it became. Finally, using these passages enabled analyses of the implementation of various pitch patterns, as noted by Markham and Hazan (2002). One aim of this study was to capture the pitch patterns used by different speakers, and thus, *The Story of Arthur the Rat* and *Arthur the Rat* were preferred to the other phonetically-balanced passages.

There are other ways to elicit spoken data from subjects, in forms such as spontaneous speech samples or citation utterances. However, they were not employed in the current study. Spontaneous speech samples would enable thorough examination of some aspects of speaking abilities, such as fluency. At the same time, however, it could make it difficult to compare the data across speech samples directly, especially in terms of segmental features. There is no guarantee that all speakers produce the same vowels and consonants and

not all target segments occur in all speech samples. On the other hand, citation utterances, the type of data that can be gathered by subjects reading aloud a list of words or sentences, allows the same spoken data to be obtained from every speaker and makes a between-subjects comparison more reliable. Another advantage is that well-balanced data can be collected because target items can be deliberately selected one by one. It is thus easy for the experimenter to manipulate the number of repetitions or phonetic environments in which target items occur. However, there is no meaningful context in these lists where the words and sentences are actually produced, which is one of the disadvantages of using citation utterances in an experiment. This research aimed to investigate the phenomena in a more natural situation where languages were used, as Strange et al. (2007) suggested. Thus, a phonetically-balanced passage, which can elicit target items produced in context, was selected as the material for the experiment.

The target items analyzed for each element of pronunciation, vowels, consonants, rhythm, intonation and connected speech phenomena, will be detailed in Sections 3.2.1 through 3.2.5. Although *The Story of Arthur the Rat* for the BN and JL data and *Arthur the Rat* for the AN data differed in the word choice, as already noted above, the target words and utterances were carefully selected so that both passages could provide comparable data. In other words, as for vowels and consonants, the target segments were shared by all groups of the subjects, even though some appeared in different phonetic contexts of different words between the two passages. Similarly, the target words and utterances for rhythm, intonation and connected speech phenomena were shared by all subjects including the AN group, even though some items were surrounded by different words between the passages. See Appendix B for the target words and utterances for the AN group.

### **3.2.1. Monophthongal vowels: vowel quality and duration**

There were 10 target monophthongal vowels, /i:/, ɪ, e, æ, ʌ, ɑ:/, ɔ:/, u:/, ʊ, ɜ:/. (A long schwa is normally transcribed as /ɜ:/ in American English, but the transcription /ɜ:/ will be used throughout this study hereafter, unless otherwise noted.) The vowel /ɒ/ was not examined because it appears only in British English, mostly replaced by /ɑ:/ in American

English. A schwa, /ə/, was also not included in the analysis of the vowel quality and duration because it was dealt with in the analysis of rhythm.

The target words for the analyses of vowel quality and duration are given below. All the words listed here were used for the analysis of vowel quality, but only the underlined words for the analysis of vowel duration. From Sections 3.2.1 through 3.2.5, the number within the square brackets in the list presents the number of repetitions, when provided:

1. /i:/: *chief*, *three*, *tree*, *immediate* and *either*
  2. /ɪ/: *this* [2], *lived*, *given*, *decision*, *bit* and *in*
  3. /e/: *never* [2], *whenever*, *ever*, *friends*, *yes*, *Helen*, *said* [6], *end*, *send*, *seven*,  
*Nelly*, *then*, *yet*, *next*, *men* and *dead*
  4. /æ/: *rat* [8], *carry*, *back* [2], *that* and *crash*
  5. /ʌ/: *young* [3], *trouble*, *up* [4], *nothing*, *coming*, *come*, *other*, *hurry*, *much* and  
*under*
  6. /ɑ:/: *Arthur* [5], *garden* and *march*
  7. /ɔ:/: *more*, *all* [2], *fallen*, *wall*, *horse*, *caught*, *board* and *saw*
  8. /u:/: *do*, *room*, *food*, *roof* [2], *too*, *ruins* and *moved*
  9. /ʊ/: *wouldn't* [2], *looked* [2], *shook* [2], *stood* [2] and *good*
  10. /ɜ:/: *learn*, *heard*, *search* and *earth*
- 4 or 6. /æ/ or /ɑ:/: *asked* [2], *answer*, *aunt*, *grass*, *last*, *can't* and *half* [2]

The last type of words categorized as /æ/ or /ɑ:/: are those that can be pronounced either way, and were therefore classified into a closer category for each speaker, judged based on the auditory impression and the values obtained and verified in a scatter graph.

The analyses of the vowels were expected to be more sensitive to the subjects' familiarity with the target words and the phonetic contexts where the target words occurred. The target words were thus selected considering these two factors. The words unfamiliar to the JL subjects were checked, referring to the results of the Range program (Nation, 2005),

which measures the level of vocabulary using word lists from the General Service List (GSL) and the Academic Word List (AWL) or British National Corpus List. In this way, the words regarded as unfamiliar to the JL subjects were discarded from the list of the target words for vowel analysis. They included *loft*, *rafters*, *rotten*, *calf*, *elm* and *angrily*. Similarly, words in the phonetic contexts that were not apt for the acoustic analysis were excluded from the subsequent analysis. Some examples of those were *once* /wʌns/ and *went* /went/, where the target vowel was preceded and followed by both /w/ and a nasal at the same time.

The above-listed words were all used for the analysis of vowel quality, but not for that of vowel duration, as noted earlier, because the extraction of duration from a stream of sounds could be greatly affected by phonetic and phonological environments. Accordingly, one-syllable words not occurring at the end of sentence and being clearly segmented from the adjacent sounds were selected for the analysis of vowel duration and are underlined in the list. The exceptions were approved for two two-syllable words, *either* and *garden*, and two sentence-final words, *lived* and *in*, so as to prioritize more well-balanced speech materials for each target vowel by setting more than one word for each vowel.

### 3.2.2. Consonants: plosives, fricatives and approximants

Plosives, fricatives and approximants were targeted in this study. They included word-initial aspirated voiceless plosives /p, t, k/, post-initial unaspirated voiceless plosives in /st, sk/, voiceless fricatives /θ, s/ and voiced approximants /r, l/. The target words used for the analysis will be described below.

Firstly, the target words for plosives were as follows:

Plosives

1. /p/: *pine*
2. /t/: *take*
3. /k/: *care*, *carry*, *kindly*, *cow*, *can't* and *came*
4. /st/: *stood* [2] and *stone*
5. /sk/: *scouts* [2]

As in Riney and Takagi (1999), the vowel that immediately follows the target voiceless plosives were controlled as much as possible. The words where /p, t, k/ were followed by a low vowel such as /a, e, æ, ɑ:/ were therefore selected as target words for the analysis to control the phonetic contexts. However, this was not applied to /st, sk/ because there was no token in the passage satisfying this condition.

Secondly, the voiceless fricatives were analyzed using the following target words:

#### Fricatives

1. /θ/: *three, Arthur* [5], *nothing* and *earth*
2. /s/: *said* [5], *send, search, seven, saw, yes, choice, grass, house, horse* and *face*

Because there were not many tokens of /θ/, compared to /s/, the position of the target items within the word was not considered. Thus, some occurred at the word-initial position and others at the word-final position for the target words of both /θ/ and /s/. The others occurred at the word-medial position for the target words of /θ/.

Finally, target words were selected for the analysis of the voiced approximants, /r/ and /l/. They occurred either at the word-initial position or the word-medial position as follows:

#### Approximants

1. /r/: *rat* [5], *rainy, room, roof* [2], *right, rode, carry* and *hurry*
2. /l/: *learn, looked, lived, last, later, lying, Helen* and *Nelly*

### 3.2.3. Rhythm

An analysis of the realization of rhythmic pattern was carried out using two rhythmic targets. One was the durational variability of successive stressed and unstressed vowels, which focused more on evaluating the overall realization of rhythm in a stream. The other was the production of weak vowels in weak forms, which involved the realization of phonetic items of weak forms.

The target utterances for the durational variability of successive stressed and unstressed vowels were as follows, where the expected stressed vowels are indicated in bold, and expected unstressed vowels are underlined:

#### Durational variability

1. *aunt Helen **said** to him*
2. *carry on like this*
3. *no more **mind** than a blade of grass*
4. *rats heard a great **noise** in the loft*
5. *send out scouts to search for a new home*
6. ***garden** with an elm tree*
7. *roof may not come down just yet*
8. *half **in** and half **out** of his hole*

It was predicted that BN/AN subjects would pronounce *for a* in the fifth utterance as one connected unstressed vowel. This token was thus regarded as one unstressed vowel.

The production of weak vowels in weak forms was analyzed regarding pitch, intensity, duration and vowel quality, which are four acoustic items known to define the English stress (Cruttenden, 1997; Roach, 2002). The target weak vowels for the analyses of pitch and intensity and those for duration and vowel quality were different, as shown:

#### Weak vowels in weak forms

1. Pitch and intensity: *a* [4], *and* [3], *at* [1], *could* [1], *he* [2], *his* [2], *of* [1], *some* [1], *them* [1] and *to* [2]
2. Duration and vowel quality: *a* [11], *an* [2], *and* [11], *the* [6], *to* [8], *them* [2], *than*, *of* [5], *some*, *just* [4], *but*, *that* and *at* [2]

The items for the duration and vowel quality were selected, expected to be weakened to one

vowel category, a schwa, /ə/. In contrast, those for pitch and intensity were expected to be weakened to any of /ə, i, u/.

As a reference for examining the weakness of these target weak vowels, the stressed vowels were also analyzed for pitch, intensity and duration. The vowel quality was not compared to the stressed vowels because more than one stressed vowel corresponding to the weak vowel was expected to be produced by the subjects. For instance, while the weak form of *of* is pronounced with a schwa, the strong form of *of* is pronounced as /ɒv/ or /ɑ:v/. Japanese learners of English could even pronounce it as [ov], influenced by its orthography. The extent of vowel centralization in the phonological space was therefore calculated, rather than comparing the vowel quality of weak vowels with that of stressed vowels, as detailed in Section 3.5.5.

The stressed vowels to be referred to are listed below, as used for the analysis of pitch and intensity. It was expected that they would be more dependent on the context, and therefore, the stressed vowels that immediately followed the target weak vowels were analyzed for the reference. They included:

Stressed vowels to compare against weak vowels

1. Against *a*: *great, stone* or *house, cow* and *garden*
2. Against *and*: *garden, watched* and *looked*
3. Against *at*: *last*
4. Against *could*: *never*
5. Against *he*: *only* and *wouldn't*
6. Against *his*: *friends* and *aunt*
7. Against *of*: *moved*
8. Against *some*: *men*
9. Against *them*: *moved*
10. Against *to*: *go* or *out* and *search*



In the nominal phrase *stone house* and the verbal phrase *go out*, the word given more stress was selected for each subject for the analysis because the word stressed in a given sentence differed from speaker to speaker. The duration averaged across stressed short vowels, /ɪ, e, ʌ, ʊ/, was used for the comparison between the stressed vowels and the weak vowels. The other stressed English short vowel, /æ/, was not added because the duration of /æ/ tended to vary greatly from speaker to speaker and from accent to accent (Cruttenden, 2014; Fox & Jacewicz, 2009). The short stressed vowels measured in the analysis of the vowel duration were used here.

#### **3.2.4. Intonation**

The current study investigated both the phonetic and phonological aspects of intonation. Span and level were analyzed for the phonetic representation of pitch, using the following three narrative sentences from the last paragraph of the story as the target:

Span and level

*That night there was a great crash that shook the earth, and down came the whole roof. Next day some men rode up and looked at the ruins. One of them moved a board, and under it they saw a young rat lying on his side, quite dead, half in and half out of his hole.*

Target utterances were selected for the analysis of the phonological representation of intonation patterns, in order to capture characteristics of nucleus placement and nuclear tone choice. There is more than one factor which could determine the most appropriate intonation pattern in a given context, as implied by the fact that intonation has several functions, as noted in Section 2.6.1 (Wells, 2006). This study focused on the syntactic and pragmatic functions in order to limit the acceptable intonation patterns as much as possible, especially for the nuclear tone choice. The following utterances were analyzed, and the abbreviations used to present the syntactic and pragmatic contexts tested are given in round brackets for convenience in reporting the results:

### Nucleus placement and nuclear tone choice

1. Antecedent modified by the relative clause (ANT): *There was once a young rat named Arthur (,who would never ...)*
2. End of the subordinate clause preceding the main clause (SD end): *go out with them*
3. Reporting clause before direct speech (bf DS): *he would only answer and said to him*
4. Short dialogue (DIA): *I don't know and This won't do*
5. Topic (TOPIC): *His aunt Helen*
6. Adverbial phrase (AdP): *one rainy day, at last, just then and that night*
7. Lists (LIST): *There was a kindly horse named Nelly, a cow and a calf*
8. Last component of closed lists (lastLIST): *and a garden with an elm tree*
9. Exclamation (EXCL): *Well*
10. Command (COM): *Right about face and March*

Two target utterances, *Well* and *March*, were not used for the analysis of the nucleus placement because they are one-word utterances. It must also be noted that although the list above is summarized according to the syntactic and pragmatic contexts for the nuclear tone choice, these contexts were not applied to the analysis of the nucleus placement. That is, 16 utterances and 10 contexts were used as the targets for nucleus placement and nuclear tone choice, respectively.

### 3.2.5. Connected speech phenomena

The three types of connected speech phenomena designed to analyze were elision, consonant-to-consonant (CC) linking and consonant-to-vowel (CV) linking. In order to examine how much these connected speech phenomena would be induced by different phonetic contexts, the target items were analyzed, classified into phonetic contexts tested for CC linking and CV linking. The target items for elision are given below. The phonetic context where elision was predicted to occur was restricted to /t/ or /d/ in the present study.

Accordingly, all target items were analyzed as one category of elision, as follows:

#### Elision

*won't do, don't know [2], named Nelly, can't wait, end with, watched the and mind than*

The target items in six phonetic contexts were tested for CC linking: the sequence of plosive and plosive where a plosive is followed by another plosive produced at the same place of articulation (PPS), that of plosive and plosive where a plosive is followed by another plosive produced at a different place of articulation (PPD), that of plosive and nasal where a plosive is followed by a nasal produced at the same place of articulation (PN), that of plosive and approximant (PA), that of plosive and fricative (PF) and that of the same consonants (CC). These six contexts were summarized into two categories: CC linking at the same place of articulation and in the same manner of articulation and CC linking at a different place of articulation or in a different manner of articulation. The following shows the classification of six phonetic contexts into these two categories and the target items analyzed, where the former category of CC linking consisted of PPS, PN and CC and the latter, of PPD, PA and PF:

#### CC linking

1. The same place of articulation and the same manner of articulation:

PPS: *said to [2]* and *quite dead*

PN: *rat named, great noise* and *that night*

CC: *with them* and *some men*

2. A different place of articulation or a different manner of articulation:

PPD: *like to* and *not come*

PA: *would like*

PF: *like this, about face* and *said the [2]*

PN was grouped into the former category, following a broad means of categorization where nasals and plosives are both stop consonants.

Four categories of CV linking were also tested: the sequence of a voiceless plosive and vowel (PV), that of a voiced alveolar plosive /d/ and vowel (DV), that of a voiced fricative and vowel (FedV) and that of a voiceless fricative and vowel (FlessV). Depending on the voicing of the preceding consonant, PV and FlessV were summarized as one target context, CV linking of a voiceless consonant, and DV and FedV, as the other target context, CV linking of a voiced consonant. These categories were formed considering the possible articulatory difficulty. The target items for each category were as follows:

#### CV linking

##### 1. A voiceless consonant:

PV: *make up [3], shook and, back and, right about, out of and looked at*

FlessV: *us all, half in and half out*

##### 2. A voiced consonant:

DV: *heard a, found a, moved a, named Arthur, would only, stood on, stood and, send out and rode up*

FedV: *friends asked, noise in, there's a, was a and with an*

### 3.3. Recording and procedure

All recordings of the JL subjects were made in a recording room, where one subject and the experimenter were alone in order to avoid background noise. At the outset of the recording session, the subjects were instructed to keep their mouth 15 cm to 20 cm away from the microphone while being recorded. Their data were recorded at a sampling rate of 44.1 kHz, 16 bit, using a digital recorder, Roland-09, and a condenser microphone, SONY ECM-MS957. The recording level was first checked and adjusted to each speaker, so that their speeches would be recorded at the volume of voice best-suited for the analyses.

The material, printed on one side of A4 paper, was distributed to each subject between 3 days and 30 minutes prior to the recording, so that they had an opportunity to

comprehend the story at their own pace and practice reading it aloud as much as they would like to. On the other side of the paper, a summary of the story was presented in Japanese, which was expected to help them grasp the outline of the story. Although the subjects were allowed to look up the pronunciation of unfamiliar words, if any, in a dictionary before participating in the recording session, no instruction was given by the experimenter as to phonetic and phonological features.

### 3.4. Acoustic measurements and analyses

Acoustic analyses were conducted using a software program for acoustic analyses, Praat (Boersma & Weenink, 2011, 2015), as a primary tool. To begin with, each of the speeches obtained was manually annotated, and then they were analyzed with the following measures: F1, F2, F3, duration, pitch, intensity and four spectral moments. Table 3.1 displays a summary of the acoustic measurements along with the element of pronunciation to which each of these were applied. More details will be given in Sections 3.4.1 through 3.4.5.

Table 3.1

*Acoustic Measurements for Each Element of Pronunciation*

Element	Measure	Target of analysis
Vowel quality	F1 and F2	Vowel quality
Vowel duration	Duration	Vowel duration
Consonants	Plosives	VOT
	Fricatives	4 spectral moments
	Approximants	F3
Rhythm	Durational variability	Degree of weakness
	Maximum pitch	
	Maximum intensity	
	Duration	
	F1 and F2	
Intonation	Pitch contour	Nucleus and tone
	Span and level	Pitch range and height
Connected speech phenomena	Spectrogram at the boundary between the two sounds <sup>a</sup>	Presence or absence of epenthesis, pause or release

<sup>a</sup>There were no specific acoustic measures for the implementation of this element because the presence or absence of an additional sound, additional pause or audible release was analyzed between the target sounds.

### **3.4.1. Monophthongal vowels: vowel quality and duration**

The values of the first formant (F1) and second formant (F2) were obtained to measure vowel quality by identifying the vocalic nuclei or the steady states that provided the most consistent F1 and F2 values on the spectrogram. Bandwidth was set at 200 Hz, the frequency range at 4 kHz and the dynamic range from 30 dB to 60 dB depending on the clearness of F1 and F2, as suggested by Ladefoged (2003). Measurements were then made using the formant track shown on the spectrogram, with the help of the LPC spectrum when the track seemed spurious and unclear. The Speech Filing System (Huckvale, 2004) was used as a supplement in cases where formants were not clearly displayed on Praat (Boersma & Weenink, 2011, 2015). The F1 and F2 values were obtained in Hz in this acoustic analysis.

The vowel duration boundary between the target vowels and surrounding sounds were first given by observing both waveform and spectrogram, and the absolute duration between the onset and offset of the target vowel was measured in units of milliseconds (ms). The target words had been meticulously selected so that the division between the target vowels and surrounding sounds were clear overall. When segmentation could not be clearly given only with the observation of waveform and spectrogram, the boundary was provided according to auditory judgment as well as through the close examination of spectrogram and waveform.

### **3.4.2. Consonants: plosives, fricatives and approximants**

The aspirated voiceless plosives /p, t, k/ and unaspirated voiceless plosives /st, sk/ were analyzed by measuring voice onset time (VOT), expressed in ms. VOT corresponds to the durational gap from the burst of the plosive to the start of the voicing of the following vowel. These locations were identified, based on the observation of both waveform and spectrogram. The release of the plosive is known to be followed by frication, and therefore, the beginning of VOT could be identified by looking at where the frication noise started. The end of VOT was defined as the location where F1 and F2 of the following vowel started on the spectrogram and the regularity of the pulse repetition on the waveform.

Regarding the voiceless fricatives /θ/ and /s/, four spectral moments of the Fast

Fourier Transform (FFT) were computed on Praat (Boersma & Weenink, 2011, 2015) to quantify the shape of spectra: center of gravity (COG), standard deviation (SD), kurtosis and skewness. This measurement was selected from among various measures described in Section 2.3.3, being regarded as one of the most stable measurements to characterize English voiceless fricatives. Previous studies have compared the spectral moments measured at more than one location, including the onset and offset, using a 20-ms Hamming window (Forrest et al., 1988) and a 40-ms Hamming window (Jongman et al., 2000). However, only the 80% window of frication noise centering the midpoint was measured in this study by manually detecting the portion where amplitude was steady, taking the differences in speaking rates among the subjects into account. Only the measurement in a linear scale was adopted in the present research. This followed the reports by Jongman et al. (2000) that there was no profound difference in the values between the linear scale and bark scale, while Forrest et al. (1988) maintained that bark-transformed frequency captured the different kurtosis of spectra between the sibilants, especially, better than linear frequency.

The third formant (F3) was measured for the voiced approximants /l/ and /r/, after the spectrogram and formant track were specified in the same way as the vowel analysis of F1 and F2. It is known that the F3 value is affected by the retraction of tongue, which abruptly lowers F3 (Ladefoged, 2003). This is why the F3 value of /r/ becomes much lower than that of /l/ when English /r/ is authentically produced (Saito & Lyster, 2011). The lowest F3 value at the beginning of the upward slope and the steady-state F3 value were thus measured for /r/ and /l/, respectively, by attentively observing the overall movement of F3 on the spectrogram. The F3 values were obtained in Hz, similar to the F1 and F2 values.

The F3 values were not measured when another sound was substituted for the target /r/ and /l/. There were two cases of this; one was the substitution of a flap-like sound and the other, that of a vowel-like sound. The flap-like pronunciation is evident from the presence of a hold phrase in most cases. However, when the presence of the hold phrase could be confused with the presence of anti-formant for /l/, a durational cue was applied to judge whether the token was /l/ or a flap-like sound, referring to the duration of a flap obtained by

Rimac and Smith (1984). For this reason, both durations and F3 values were recorded with the candidates for the flap-like tokens where the presence of a hold phrase was unclear. The articulation rate, which Praat (Boersma & Weenink, 2011, 2015) automatically calculated using the script (de Jong & Wempe, 2009), was also obtained to take into account the difference in the speaking rate between the BN/AN subjects and the JL subjects. The other type of substitution, vowel-like pronunciation, was due to an incomplete articulation of /l/. The tokens were interpreted as this type of error, when there was a clear F3 value to be measured, but no characteristics of anti-formant were visually evident in the spectrogram and waveform.

### **3.4.3. Rhythm**

The durational variability of successive vowels was calculated using the durational values of stressed vowels and unstressed vowels, and thus, durations of stressed vowel and subsequent unstressed vowels were measured on the spectrogram. In this analysis, semivowels such as /j/ and /w/ were defined as part of vowels (Grabe & Low, 2002). It must also be noted that the final syllables of the target items were excluded from the analysis, bearing in mind that the final lengthening could greatly affect duration, and that the tokens where unfilled pauses longer than 250 ms (Abe, 2011) were inserted were also discarded from the analysis, being regarded as an error.

The production of weak vowels was measured with the four acoustic items: the maximum pitch, maximum intensity, duration and F1 and F2. The first two items were measured using the F0 and intensity curves displayed on Praat (Boersma & Weenink, 2011, 2015), which are expressed in units of Hz and dB, respectively. However, the automatically-computed pitch curve sometimes depicts minor fluctuations at the onset and offset of a certain type of consonants, and is affected by voice quality such as creaky voice. In measuring F0, therefore, the waveform was also observed closely, along with the pitch curve. When the accuracy of the pitch curve could be called into question, the F0 value was manually calculated from the number of pulses on the waveforms (Ladefoged, 2003). The range of pitch displayed on Praat (Boersma & Weenink, 2011, 2015) was basically set from



50 Hz to 200 Hz unless an adjustment was needed.

The other two items, duration and F1 and F2, were measured as in the analysis of vowel duration and vowel quality, respectively, described in Section 3.4.1. There were some cases where the target vowels were difficult to segment due to blurred boundaries. This notably occurred in the phonetic contexts where the target vowels were nasalized, being influenced by the following nasal, as in *than*, *an* and *them*. In those cases, the whole nasalized portion was regarded as vowels.

#### **3.4.4. Intonation**

There were two phonetic items investigated as to intonation: span and level. These measurements were calculated by following the method suggested by Patterson (2000). The F0 values between all the peaks excluding the sentence-initial, and post-accented valleys, were measured for the span. The locations of these peaks and valleys were specified according to the F0 curve. The F0 values were also measured for the level at the sentence-final lows. As with the measurement of the maximum pitch of the stressed vowels and weak vowels, the F0 display was set from 50 Hz to 200 Hz, so that the pitch movement was clearly displayed.

Analysis on the phonological representation of intonation patterns were conducted by providing each target utterance with the tone labeling, based on the ToBI guideline (Beckman & Elam, 1997) and on the tonetic stress marks (TSM) system (O'Connor & Arnold, 1973). Under the former labelling system, the syllables that were prominent were first defined based on the observation of the peak and valley on the intensity and F0 curves with the help of the auditory judgment. They were then marked with one of the six pitch accents: H\*, L\*, !H\*, L+H\*, L\*+H or H+!H\*. One of the most appropriate phrase accents and boundary tones, L-L%, L-H%, H-L%, H-H%, !H-L% or !H-H%, was also given at the end of the intonation phrase (IP), but only the phrase accent, L-, H- or !H-, was labeled when the boundary within the utterance was identified as the break index 3. Under the latter labelling system, in contrast, after the nucleus placement was defined, primarily based on auditory judgment, one of the six tones was provided: fall, low rise, high rise, fall-rise, rise-fall and

level. Although a high fall and a low fall are distinguished in TMS, how this difference should be taken into account in ToBI varies from researcher to researcher (Ladd, 1996). This was mainly because they were phonologically almost the same, and they are difficult to discriminate from one another due to the declination, a natural lowering of the pitch toward the end of the utterance. This study thus did not distinguish between a high fall and a low fall, categorizing both as a fall.

#### **3.4.5. Connected speech phenomena**

The occurrence of elision, CC linking and CV linking was analyzed based on the observation of the spectrogram and auditory impression to record whether the target sequence was pronounced without any durational gap. All target items for the analysis of elision and CC linking had a plosive in the first element except in the CC phonetic context. When there was no audible release of the preceding plosive, it was regarded as elided. Although it is known that elision of the sound is also signaled by a longer hold phrase (Cruttenden, 2014; Matsui, 1998), this was not measured in order to focus more on the phonological realization than the phonetic realization. The target items in the CC phonetic context, which did not consist of plosives, were studied in terms of whether the same two consonants in a row were separately pronounced twice. When the second element of the sequence of the token was not articulated, it was judged to be evidence of the elision and CC linking.

CV linking was first examined as to whether an unfilled pause was inserted after the first element of a sequence was produced. When there was a pause, the token was regarded as unlinked. When there was no pause inserted between the two sounds in the sequence, it was regarded as linked as long as there was no additional vowel inserted and the following vowel was produced with modal voice. The target items where the beginning of the vowel in the target was produced with creaky voice were judged to be produced with an abnormal stream of sounds for CV linking in this study, which would be reasonable considering that creaky voice is a signal of adjusting the larynx for normal voice (Ashby & Maidment, 2005) or that creaky voice occurs at the beginning of a sentence starting with a vowel (Ladefoged, 2001).

### **3.5. Statistical analyses**

Four types of statistical analyses were performed on IBM SPSS Statistics 23 in this study: a cluster analysis, a multivariate analysis of variance (MANOVA) and a discriminant analysis and a correlation analysis with Spearman's rank-order correlation coefficient. The first research question concerned the difficulty of the target phonetic and phonological items for Japanese learners of English. The second research question was whether a supportive relationship existed among the elements of pronunciation. In order to address the first research question, the items that would be easy, learnable or difficult were hypothesized, as described in Chapter 2. The first three statistical tests, a cluster analysis, a MANOVA, and a discriminant analysis, were used to address this issue. The last statistical test was used to address the second research question.

A cluster analysis was conducted primarily to profile the JL subjects by clustering them with the BN/AN data. Using the clusters generated by this analysis as the between-subjects independent variables, a MANOVA was carried out. It revealed whether there was a statistical difference among the clusters or not. Thirdly, a discriminant analysis was performed to identify clusters that differed in the variables at a statistically significant level, which variables discriminated them and to what degree these variables contributed to the discrimination. The variables input to the test varied from element to element. Finally, a correlation analysis with Spearman's rank-order correlation coefficient was conducted in order to identify any supportive relationship underlying the elements of pronunciation. After the analyses of the elements of pronunciation were completed, the variables that could be considered to illustrate the learning of each element were selected and the correlation coefficient was obtained. Before going into the details of these statistical analyses, the variables that were submitted to the statistical tests will be first described for each element.

#### **3.5.1. Monophthongal vowels: vowel quality and duration**

Twenty-one variables were obtained for vowel quality from the acoustically-analyzed data: the standardized F1 and F2 mel values for the 10 target vowels and the score for structural difference (Table 3.2). To obtain these two kinds of variables, the

F1 and F2 values, measured in a linear scale, Hz, were first converted to an auditory scale, mel, to assess how the vowels sounded to listeners, using Equation 1 (Fant, 1968):

$$\text{Mel} = 1000/\log_2 10 \times (\log(1+F/1000)) \quad (1)$$

where F represents the frequency value. After the conversion from Hz to mel, the values were standardized by applying the procedure proposed by Lobanov (1971), z-score transformation, for a later comparison between the subjects, using Equation 2:

$$\text{Standardized } F_{(i)} = (F_{(i)} - \mu_{(i)}) / \sigma_{(i)} \quad (2)$$

where F stands for the formant value, and  $\mu_{(i)}$  and  $\sigma_{(i)}$  represent, respectively, the mean and standard deviation of the *i*th formant frequency across all vowels. This method was employed with reference to Adank, Smits, and van Hout (2004). They compared 12 normalization procedures and reported that the method proposed by Lobanov was the most workable of them.

Table 3.2  
*Variables for the Analysis of Vowels*

	Variable	No. of variables	Level of measurement	Unit
Quality	Standardized F1 mel	10	Interval	mel
	Standardized F2 mel	10	Interval	mel
	Score for structural difference among 10 vowels	1	Interval	
Duration	PVI values of long and short vowels	4	Interval	

While the F1 and F2 mel standardized values refer to the exact values of the formants, the other variable for vowel quality, the structural difference, expresses the vowel distribution in the phonological vowel space. The value for this variable was calculated using a technique described by Suzuki, Qiao, et al. (2010), where each speaker's structure of the

vowel pronunciation was first expressed with a distance matrix that was made up of the distance between all possible vowel pairs. The structural difference from each BN/AN subject was then calculated for every JL subject using Equation 3, devised by Suzuki, Qiao, et al.:

$$D(S, T) = \sqrt{\frac{1}{M} \sum_{i < j} \left( \frac{S_{ij} - T_{ij}}{S_{ij} + T_{ij}} \right)^2} \quad (3)$$

where  $D(S, T)$  represents a structural difference between a student and a teacher, simply two arbitrary speakers, and  $M$  represents the number of phonemes in the distribution. Because there were 19 BN/AN subjects with whom the JL subjects could be paired in this study, the averaged difference was defined as the score for the overall structural difference for each subject. The structural difference of the BN/AN subjects was also calculated as a reference by comparing each BN/AN subject to all BN/AN subjects other than themselves. It follows that the higher the score, the more deviant the speaker's structure from that of the native speakers.

The data analyzed concerning the vowel duration were summarized into four variables by computing the difference in the duration between the long vowels and short vowels. They were /i:-ɪ/, /u:-ʊ/, /ɑ:-æ/ and /ɑ:-ʌ/, which were paired, considering the vowels located closely in the phonological vowel space and frequently confused by Japanese learners of English (Shimizu, 1999). A mid-central vowel /ɜ:/ can be paired with /æ/ and /ʌ/, similar to /ɑ:/; however, these pairs were not tested in the current study following Jenkins' (2000) argument that the quality, not the quantity, mattered for /ɜ:/ to make it distinct from other vowels.

The idea of the normalized pairwise variability index (PVI) that Low et al. (2000) introduced in the analysis of rhythm, described in Section 2.5.3, was applied for calculating the variables in order to normalize the difference in the speaking rate between the subjects. This index was originally developed for the analysis of rhythm in order to normalize the speaking rate that varies from time to time, and it was applied throughout this study as the method to normalize the speaking rate in the production of vowels. PVI values can be obtained by dividing the durational difference between the paired vowels by the average

duration between these two vowels, using Equation 4:

$$PVI = 100 \times \left( \sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| / (m - 1) \right) \quad (4)$$

where  $m$ ,  $d$  and  $k$  represent the number of vowels to be analyzed, the duration of the vowels analysed and the  $k$ th number of vowel. One element of the formula adjusted in this study was that relative values rather than absolute values were calculated, to take into consideration cases where the target short vowels would be produced longer than the target long vowels, as shown in Equation 5:

$$\text{Adjusted PVI} = 100 \times \left( \sum_{k=1}^{m-1} \frac{d_{\text{long or stressed}} - d_{\text{short or unstressed}}}{(d_{\text{long or stressed}} + d_{\text{short or unstressed}})/2} / (m - 1) \right) \quad (5)$$

where  $m$  and  $d$  stand for the number of vowels and the duration of vowels. In the analysis of vowels, the duration of long vowels and that of short vowels were assigned to  $d_{\text{long or stressed}}$  and  $d_{\text{short or unstressed}}$ , respectively. In learners' productions, unstressed vowels or short vowels could be unexpectedly produced longer than stressed vowels or long vowels, respectively. This adjusted PVI has the advantage that these values can be expressed by negative values. The durational difference between long and short vowels can be expressed by the ratio of the duration of short vowels to that of long vowels. Applying the adjusted PVI value is therefore better in that it is clearly shown by the positive and negative sign whether or not short vowels were longer than long vowels. Use of the adjusted PVI also made it possible to avoid unnaturalness where the ratio may be over 1.0 or 100% with no upper limit, which suggests that short or unstressed vowels were longer than stressed or long vowels.

### 3.5.2. Plosives

There were five variables for the plosives: /p/, /t/, /k/, /t-st/ and /k-sk/ (Table 3.3). The first three variables were designed to assess the absolute durations of voiceless plosives /p/, /t/ and /k/. The last two variables were to assess how much the VOT differentiated between the word-initial aspirated voiceless plosives /t/ and /k/ and the post-initial

unaspirated voiceless plosives in /st/ and /sk/.

Table 3.3

*Variables for the Analysis of Plosives*

Variable	No. of variables	Level of measurement	Unit
Absolute VOT durations	3	Interval	ms
VOT differences between aspirated and unaspirated voiceless plosives	2	Ratio	

*Note.* ms = millisecond.

In order to directly examine the relative difference in the duration of VOT between the aspirated voiceless plosives and unaspirated voiceless plosives, the ratio was calculated in the /t-st/ and /k-sk/ pairs, and used as the variables. They were calculated by dividing the absolute VOT durations of aspirated /t/ and /k/ by the absolute VOT durations of unaspirated /t/ and /k/ in /st/ and /sk/.

### 3.5.3. Fricatives

The data analyzed for the target voiceless fricatives involved eight variables: COG, SD, skewness and kurtosis for /θ/ and /s/ each (Table 3.4). These were variables to quantify the spectral shape.

Table 3.4

*Variables for the Analysis of Fricatives*

Variable	No. of variables	Level of measurement	Unit
COG	2	Interval	Hz
SD	2	Interval	Hz
Skewness	2	Interval	
Kurtosis	2	Interval	

*Note.* COG = center of gravity; SD = standard deviation; Hz = Hertz.

As noted earlier, only the linear scale was used for the measurement, based on the report by Jongman et al. (2000). No transformation was therefore conducted on these values, which

were directly used for subsequent statistical analyses.

### 3.5.4. Approximants

Two variables were applied to the statistical tests on the production of the approximants /r/ and /l/: score for the /r/ and /l/ tokens, which corresponded to the number of tokens produced as the intended sound (Table 3.5). These variables were obtained by judging every target token as /r/, /l/ or other sounds and counting the number of tokens defined as intended with reference to the BN/AN data. Not only did the JL subjects tend to confuse the target approximants with one another, but they were also likely to confuse them with another sound that could not be characterized with F3. It follows that those who failed to produce the target approximants as approximants lacked the data on the F3 values. The measured F3 values were therefore not directly used as the variables for the statistical tests. The number of the correct tokens for each item, treated as a score, was employed.

Table 3.5

*Variables for the Analysis of Approximants*

Variable	No. of variables	Level of measurement	Scale
Score for the /r/ tokens	1	Interval	0-8
Score for the /l/ tokens	1	Interval	0-8

After the F3 values recorded were converted from Hz to mel using Equation 1, as in the analysis of vowel quality, scoring the /r/ and /l/ tokens started with setting the thresholds of the F3 mel value to separate /r/ and /l/ and that of the durational value to separate a flap-like sound and /l/. These threshold values were determined based on the data of BN/AN subjects. As for the F3 threshold, all tokens of initial /r/ and /l/ collected from the BN/AN subjects were each ranked according to F3, and the F3 mel values whose z-scores fell at 2 SD and -2 SD were defined as the thresholds for /r/ and /l/, respectively. The durational threshold for a flap-like sound, on the other hand, was set using the reported duration of American flaps by Rimac & Smith (1984), 33 ms, as a reference. This average duration of flaps was obtained



from the data of the adult native speakers of American English. Considering that non-native speakers are likely to speak more slowly than native speakers (Munro & Derwing, 1995), a modified threshold for the JL subjects was calculated by multiplying 33 ms by the ratio of the average articulation rate of the JL subjects to that of the BN/AN subjects.

With the threshold thus defined, every item was scored in terms of whether they were produced as intended or not, with reference to the threshold values above. First, by comparing the duration of the candidates for the flap-like tokens against the threshold value of the duration for a flap-like sound, tokens longer than the threshold were considered not to be flap-like. These tokens were submitted to the subsequent scoring process to judge whether /r/ and /l/ were produced as intended. Using the F3 threshold values, the /r/ tokens with F3 lower than threshold value for /r/ and the /l/ tokens with F3 higher than the threshold value for /l/ were judged as intended. By summing the scores for the six initial items and those of the two medial items for /r/ and /l/ each, the two variables were obtained on a scale of 0 to 8. The target word *rat*, repeated five times, was judged as intended when it was scored as intended for more than one token. Similarly, as for the target word *roof*, repeated twice, at least one token being judged as intended made it scored as intended.

### 3.5.5. Rhythm

The implementation of rhythm was summarized into five variables: the PVI values of successive stressed and unstressed vowels, the difference in the maximum pitch between the weak vowels and the stressed vowels that immediately followed, the difference in the maximum intensity between the weak vowels and the stressed vowels that immediately followed, the difference in duration between the weak vowels and the short stressed vowels and the centralization of F1 and F2 in the vowel space (Table 3.6).

The adjusted PVI was employed for the PVI values of successive stressed and unstressed vowels, using Equation 5. The duration of stressed vowels and that of the unstressed vowels were assigned to  $d_{long \text{ or } stressed}$  and  $d_{short \text{ or } unstressed}$ , respectively. This adjusted PVI equation made it possible to reflect, with a negative value, that the target unstressed vowels, supposedly shorter, were longer than the target adjacent stressed vowels.

Table 3.6

*Variables for the Analysis of Rhythm*

	Variable	No. of variables	Level of measurement	Unit
Variability	PVI values of successive stressed and unstressed vowels	1	Interval	
Weak forms	Pitch difference between stressed and weak vowels	1	Interval	ST
	Intensity difference between stressed and weak vowels	1	Interval	dB
	PVI values of stressed and weak vowels	1	Interval	
	Vowel centralization	1	Interval	mel

*Note.* PVI = pairwise variability index; ST = semitone; dB = decibel.

As regards the production of weak forms, the variables for the duration, intensity and pitch were obtained by calculating the differences between the target weak vowels and stressed vowels. While the variable of the intensity was gained by subtracting the value of weak vowels from that of stressed vowels, the maximum pitch difference was obtained by being converted from Hz to semitone (ST), a musical scale. This musical scale was used to make it possible to consider characteristic auditory sense, which is known to be less sensitive to the difference in the higher and lower pitch region (Johnson, 2003). These values were obtained by Equation 6:

$$\text{Pitch (ST)} = 12 \times \log_2\left(\frac{\text{Pitch (Hz)}_s}{\text{Pitch (Hz)}_w}\right) \quad (6)$$

where  $s$  and  $w$  stand for the stressed vowel and weak vowel, respectively. The idea of PVI was employed for the duration between the weak vowels and stressed vowels in order to take into account differences in the speaking rate among subjects. Values were calculated by pairing the duration averaged across target thirteen weak vowels and that averaged across four short stressed vowels and plugging the measured duration into Equation 5. The duration of the stressed vowels was assigned to  $d_{long \text{ or } stressed}$  and that of weak vowels was assigned to  $d_{short \text{ or } unstressed}$ .

Because there was no reference for the F1 and F2 values, the vowel centralization in the phonological vowel space was obtained and treated as the variable for the vowel quality

of the target weak vowels (Low et al., 2000). To obtain this variable, the measured F1 and F2 values were first converted to mel, and then, these F1 and F2 mel values were assigned to Equation 7:

$$\text{Vowel centralization} = \sum_i \frac{\sqrt{(\text{F1 mel}_i - \text{centroid F1 mel})^2 + (\text{F2 mel}_i - \text{centroid F2 mel})^2}}{n} \quad (7)$$

where  $n$  represents the number of the valid target items and  $i$  stands for every target item. The centroids of the F1 and F2 mel values were equal to the mean values for each. Although the equation used by Low et al. (2000) did not take the square root, this study modified it, just like transforming variance to SD, so that the values obtained were not divided by 100,000, unlike Low et al. The greater value in this variable means that the vowels produced dispersed in the vowel space, and it follows that the vowel quality becomes more deviant from a schwa as the value increases.

### 3.5.6. Intonation

Eight variables obtained for intonation: two variables were on phonetic items, span and level (Table 3.7), and the other six variables were on phonological items, intonation pattern (Table 3.8). The latter variables, those involving the phonological aspects, consisted of two major items, the scores concerning nucleus placement and the scores concerning nuclear tone choice. They were broken down into four and two variables, respectively, depending on the hypothesized difficulty of the utterance types and tone types for the JL subjects.

Table 3.7  
*Variables for the Analysis of the Phonetic Items of Intonation*

Variable	No. of variables	Level of measurement	Unit
Span	1	Interval	ST
Level	1	Interval	Hz

*Note.* ST = semitone; Hz = Hertz.

The span for the variables of the phonetic aspects of intonation was obtained by calculating the difference in the F0 values between the target peaks and valleys, and the level, by averaging the data across target utterances. While the level was expressed in Hz, the span in units of Hz was converted to that in ST using Equation 6, with reference to Patterson's (2000) report that the span was best represented in ST.

Table 3.8  
*Variables for the Analysis of the Phonological Items of Intonation*

Variable	No. of variables	Level of measurement	Scale
Score for the nucleus in the long/non-final utterances	1	Interval	0-8
Score for the non-nuclear words in the long/non-final utterances	1	Interval	0-21
Score for the nucleus in the final utterances	1	Interval	0-11
Score for the non-nuclear words in the final utterances	1	Interval	0-18
Score for the nuclear tone choice in the non-falling utterances	1	Interval	0-3
Score for the nuclear tone choice in the falling utterances	1	Interval	0-7

In order to obtain the variables of the phonological aspects of intonation, the pitch patterns labelled were quantified based on whether the nucleus placement and the nuclear tone choice were typical, referring to the performances of the BN/AN subjects. Typical performances by the BN/AN subjects were therefore specified before calculating these variables. In this study, typicality was defined as the pattern used by more than half the BN/AN subjects. When there was no pattern common to more than half these subjects, the most and the second most frequently used patterns were regarded as typical. Based on these criteria, the typical nucleus placement and typical nuclear tone choice were first defined for each target utterance. Although the tone was labelled with both ToBI and TSM, the data will be treated only with TMS because there was no critical effect of the different labelling system on the results, as far as the items in the present study were concerned.

The four variables of the nucleus placement were obtained as follows. First of all, taking into consideration the difficulty of the nucleus placement for the JL subjects, the target utterances were divided into two types: the long/non-final utterances and final utterances. The

long/non-final utterances were the utterances that were long or where a typical nucleus fell on the non-utterance-final word. The final utterances were the utterances where the utterance-final word bore a typical nucleus. As noted in Section 2.6.4, the former type of utterances was considered difficult for Japanese learners of English and the latter type easy. After the typical nucleus placement being specified from the BN/AN data, the classification of the target utterances into either type was arranged. Next, the scoring of the nucleus placement was conducted for each type, using the scoring system originating from the Levenshtein distance. This is commonly employed when the similarity between two given letter strings is calculated. Under this method, the words defined as a typical nucleus placement were scored by assigning 0 or 1. When the nucleus fell on the word where it was typically located, the target nuclear word concerned gained 1 point. When the nucleus did not fall on the word where it was typically located, the target nuclear word was not awarded a point. As will be reported in more detail in Section 4.4, one target utterance had one typical nucleus in most cases, while some target utterances had more than one typical nucleus when they were divided into more than one IP. In this case, all typical nuclei were scored dichotomously. This scoring procedure yielded the highest possible points, which determined the scale for the variables displayed in Table 3.8. The total points served as the variables of the nucleus placement in the long/non-final utterances and final utterances. Similarly, the rest of the words where the nucleus typically did not occur, the non-nuclear words, were also scored 0 or 1 point in terms of whether these words were produced without a nucleus. This score was included in order to count whether or not extra nuclei fell on the non-nuclear words. Otherwise, the subjects who located the nucleus on every word would have obtained the highest possible values. The non-nuclear words that did not bear the nucleus were assigned 1 point, whereas those that bore the nucleus were assigned no points. The scores calculated in this way correspond to the variables of the non-nuclear words in the long/non-final utterances and final utterances. The four variables were thus yielded, the scores for the nucleus in the long/non-final utterances and those in the final utterances, and the scores for the non-nuclear words in the long/non-final utterances and those in the final utterances.

The target tones for the last two variables of the nuclear tone choice were scored in terms of whether a typical nuclear tone was used or not. Firstly, the utterances were categorized into two types, non-falling utterances or falling utterances, depending on the tone typically used by the BN/AN subjects. This is to test the hypotheses that the former utterance type would be difficult and the latter would be easy in utterances other than questions, more specifically speaking, syntactic and pragmatic contexts tested, which was listed in Section 3.2.4. When the specified typical tone was used, the item concerned was awarded 1 point, and when it was not, was awarded no points. One thing to be noted here is that only the use of the typical nuclear tone had to be scored. Failure in nuclear tone choice had to be treated separately from that in nucleus placement. Thus, as long as the typical nuclear tone was used in any one word within the same syntactic phrase, the misplacement of the tone was regarded as an error of the nucleus placement, not of the nuclear tone choice, and the target tone was scored 1 point as to the nuclear tone choice. When two or more than two nuclei were erroneously placed with two or more than two different tones, however, the target tone concerned scored no points even though it took place within a single syntactic phrase. This is because both the nucleus placement and nuclear tone choice were possibly involved in this failure. Finally, the average score was calculated for each syntactic and pragmatic context presented in Section 3.2.4 because all target utterances were successfully divided into falling utterances or non-falling utterances according to the context, as detailed in Section 4.4. The scores in the falling utterances and those in the non-falling utterances were thus obtained, which served as the last two variables for nuclear tone choice, the other phonological item of intonation that was tested in the current study.

As will be described in the results in Section 4.4, nucleus placement tended to be determined by the lexical components of the utterances rather than the syntactic or pragmatic contexts of the utterances summarized in Section 3.2.4. The scores for the nucleus and the non-nuclear words were therefore calculated by regarding each target utterance as one kind. In contrast, the nuclear tone choice was affected by the syntactic and pragmatic functions, as was expected. Thus, the scores for the nuclear tone choice were obtained by averaging the

scores across the target utterances of each syntactic and pragmatic context shown in Section 3.2.4.

### 3.5.7. Connected speech phenomena

There were five variables of connected speech phenomena: one variable of elision, two variables of CC linking and two variables of CV linking (Table 3.9). The total number of potential points for the target connected speech phenomena corresponds to the highest possible value in each variable, which is presented in the scale.

Table 3.9

*Variables for the Analysis of Connected Speech Phenomena*

Variable	No. of variables	Level of measurement	Scale
Score f elision	1	Interval	0-7
Score for CC linking at SP & in SM	1	Interval	0-7
Score for CC linking at DP or in DM	1	Interval	0-6
Score for CV linking of a voiceless consonant	1	Interval	0-8
Score for CV linking of a voiced consonant	1	Interval	0-12

*Note.* CC = consonant-to-consonant; SP = the same place of articulation; SM = the same manner of articulation; DP = a different place of articulation; DM = a different manner of articulation; CV = consonant-to-vowel.

Each target token was allotted either 1 or 0 point according to whether the target connected speech phenomena occurred, and then, the total score was calculated for each variable. Repeated tokens, such as *don't know*, *said to*, *said the* and *make up*, were awarded 1 point when the target connected speech phenomena occurred at least once at one of the repeated tokens. The same applied to the tokens that represented exactly the same type of sound sequence, such as *shook and* and *back and*, and *there's a* and *was a*.

While there was only one phonetic context tested for elision, the target items of CC linking and CV linking were each summarized into two separate variables, depending on the phonetic contexts. The two variables of CC linking were the scores for CC linking at the same place of articulation and in the same manner of articulation and for CC linking at a different place of articulation or in a different manner of articulation. The former was

obtained by summing up the number of target items where CC linking occurred in PPS, PN and CC ( $2+3+2 = 7$ ), and the latter, in PPD, PA and PF ( $2+1+3=6$ ). Similarly, the scores for CV linking were summarized into two variables, according to the voicing of the preceding consonant. One variable concerning the tokens that had a voiceless consonant in the first element of the sequence was obtained from PV and FlessV ( $4+4=8$ ), and the other variable concerning those that had a voiced consonant in the first element was obtained from DV and FedV ( $8+4=12$ ).

### 3.5.8. Relationships between the elements of pronunciation

In order to address the second research question, two variables considered to clearly represent the learning process of JL subjects were selected, based on the results in a series of analyses of each element of pronunciation. This decision was made to describe relationships between the elements as clearly as possible. Table 3.10 shows the variables used for this analysis, the selection of which will be explained further in Section 4.6.

Table 3.10

*Variables for the Analysis of Relationships between the Elements of Pronunciation*

Element	Variable
Vowel quality	Standardized F1 mel value of /ɪ/ Standardized F1 mel value of /ɜ:/
Vowel duration	PVI values in the /u:-ʊ/ distinction PVI values in the /ɑ:-ʌ/ distinction
Plosives	Absolute VOT duration of /p/ Absolute VOT duration of /k/
Fricatives	SD and skewness of /θ/ SD and skewness of /s/
Approximants	Score for the /r/ tokens Score for the /l/ tokens
Rhythm	Pitch difference between stressed and weak vowels Intensity difference between stressed and weak vowels
Intonation	Score for the nucleus and non-nuclear words in the long/non-final utterances Score for the nuclear tone choice in the non-falling utterances
Connected speech phenomena	Score for CC linking at SP and in SM Score for CV linking of a voiced consonant

*Note.* PVI = pairwise variability index; CC = consonant-to-consonant; CV = consonant-to-vowel; SP = the same place of articulation; SM = the same manner of articulation.



The difficulty of learning each target item was defined as easy, learnable or difficult under the criteria in Section 3.6, based on the results described in Chapter 4 and discussed in Chapter 5. The variables tabulated in Table 3.10 were selected from the variables which were found to be learnable items or difficult items to learn. These variables were regarded as showing a sign of learning.

The values of each variable varied in units of measurement across variables. Therefore, they were transformed to the rank data in order to compare them with one another, according to the performances of the BN/AN subjects. The exact methods used to assign ranks varied across variables. When the variables were frequency data as in the variables of approximants, intonation and connected speech phenomena, the subjects who gained a higher value ranked higher. The variables of rhythm were not frequency data like this, but having a larger value means a greater difference between the stressed vowel and weak vowel. The subjects with a larger value ranked more highly for these variables. In contrast, a larger value for the remaining variables did not necessarily refer to a better performance. Rank was thus decided on each subject, using the z-scores obtained from the mean and standard deviation of the BN/AN subjects. This method of standardization was applied by Flege, Yeni-Komshian and Liu (1999). This method was adopted for this study because it made it possible to rank the JL subjects within the entire sample and within the JL subjects with reference to the BN/AN subjects. Ranks were separately given to the two variables representing each element of pronunciation, which were averaged to obtain one overall rank data for each element of pronunciation.

While most of the variables in Table 3.10 corresponded to those used for the statistical analyses for each element of pronunciation, some new variables were calculated before gaining the rank data for this specific analysis, so that the learning process of each subject could be maximally taken into consideration. These variables included SD and the skewness of fricatives and the scores for the nucleus and non-nuclear words in the long/non-final utterances. Although SD and the skewness of fricatives were dealt with as separate variables in the analysis to test the first research question, their rank data were

summarized into one rank data by averaging, which was used as variables expressing the rank of /θ/ and /s/ for the second research question. Regarding intonation, the scores for the long/non-final utterances were divided into the nucleus and non-nuclear words in the analysis of intonation for the first research question. However, they were put together as one rank data for the second research question. Because the score for the nucleus and the score for the non-nuclear words differed in the highest possible values, 8 and 21, respectively, they were summed and ranked after reducing the highest possible value.

### **3.5.9. Preliminary analyses**

Before conducting a series of statistical analyses, preliminary analyses were performed to detect the presence of outliers and to check whether the data violated the assumptions of the MANOVA and discriminant analysis to be conducted. The treatments of missing data and outliers and the assumption checks as to the normal distribution and the homogeneity of covariance matrices will be described below.

Firstly, when there were missing data because of errors or problems with the acoustic analyses, this study treated them as follows. When variables with missing values were comprised of more than one target item, these variables were obtained by averaging the values of the remaining valid data. It was estimated that these missing data would not greatly affect the variable, even if some tokens were missing in variables. However, when variables with missing data consisted of only one target sound, as with /t/ in the analysis of plosives, the missing data of this target would lead to missing the variables concerned. In this case, this subject was excluded listwise from subsequent statistical analyses of the relevant element of pronunciation.

Following this treatment, the subjects with the latter type of missing data were discarded from the statistical analyses of vowel duration and plosives. One JL subject failed to provide the /i:/ data for the vowel duration, and was excluded. There were three cases of plosives with missing data. One AN subject and one JL subject skipped the target item /t/ and one JL subject produced /st/ with a tap. These data were regarded as missing data, and these cases were discarded from the statistical analyses.

Secondly, the detection of univariate outliers was attempted based on histograms and boxplots visually and z-scores quantitatively with the criterion of 3.29, and that of the multivariate outliers, based on Mahalanobis distance at  $p < .001$  (Tabachnick & Fidell, 2007). This was done for the entire sample and for the BN, AN and JL groups, separately. When potential outliers were found, the data were acoustically checked again on Praat (Boersma & Weenink, 2011, 2015) to confirm that it was not caused by an error in the acoustic analysis. When it was confirmed that these outliers were not attributed to a failure in the acoustic analysis, the patterns of the outliers were then scrutinized. If these outliers were due to BN/AN subjects who performed differently from the JL subjects, they were not eliminated, but were considered to be legitimate data from the population. When some JL subjects were found to be outliers within the JL group because of a native-like performance, they were not excluded, either. In other cases, the subsequent statistical analyses were conducted, both including the outliers and excluding them. When there was no difference in the results between them, the results including the outliers will be reported.

Univariate outliers were identified regarding the vowel quality, when checked within the entire sample: one BN subject's F2 mel /ʌ/, one AN subject's F2 mel /æ/, one AN subject's F1 mel /ʊ/, one AN subject's F1 mel /ɜ:/ and one AN subject's F2 mel /i:/. However, because they were all the subjects of BN/AN groups, they were not discarded from the analyses. When the detection of outlier was performed for each group, two cases in the JL group were identified as outliers: one JL subject's F2 mel /ɜ:/ and F2 mel /ɔ:/ and one JL subject's F1 mel /ɜ:/ with a higher z-score as to the first outlier and a lower z-score as to the last two outliers. However, the subsequent statistical analyses conducted with and without these cases, including a cluster analysis, MANOVA and discriminant analysis, yielded the same results, whether they were included or excluded. Thus, the results including these outliers will be reported. Two cases of vowel duration were detected as univariate outliers. One JL subject's distinction between /u:/ and /ʊ/ was found to be a univariate outlier when checked within the entire sample, and another JL subject's distinction between /ɑ:/ and /ʌ/ to be a univariate outlier when checked within both the entire sample and JL subjects. Both

subjects produced much longer vowels than the other subjects and were excluded from the statistical analyses. Three cases of plosives were considered outliers. One JL subject was identified as a univariate outlier of the relative VOT difference in the /k-sk/ pair within the entire sample. Two JL subjects were detected as univariate outliers of the relative VOT differences in the /k-sk/ and /t-st/ pairs, respectively, both within the entire sample and JL subjects. One of them was also identified as a multivariate outlier within the entire sample and JL subjects with  $p < .001$ . These subjects were excluded from the analysis. One univariate outlier was detected for kurtosis of /s/ for one BN subject among the fricatives, when checked within the entire sample. However, this case was not identified as the outlier within the BN group, and therefore, it was not eliminated from the analysis. When the outlier was inspected within the JL group, two cases were identified as outliers. One JL subject was found to be a univariate outlier as to kurtosis of /s/ and one JL subject was to be a multivariate outlier. These cases were not discarded from the analysis, either, because they were distant from the other JL subjects in that they were closer to the BN/AN performances, and they were not identified as outliers when checked within the entire sample. One AN subject was identified as a univariate outlier of intensity and a multivariate outlier within the entire sample regarding the rhythm. However, this subject was not excluded from the analysis because it represented the AN group and was not defined as an outlier within the AN group. One JL subject was also found to be a univariate outlier of the maximum pitch and a multivariate outlier within the JL group. However, this case was not excluded from the analyses because this subject was detected as an outlier due to a native-like performance of this item, and because the results were not affected by the inclusion of this case. Four JL cases of intonation were excluded from the statistical analyses. One JL subject was found to be a univariate outlier in the variable of the non-nuclear words in the final utterances, when checked within the entire sample. Another JL subject was identified as a univariate outlier in the variable of the non-nuclear words in the long/non-final utterances, when checked within the JL subjects. Another JL subject was defined as a univariate outlier as to the span both within the entire sample and the JL subjects, and also as a multivariate outlier within the JL

subjects. The other JL subject was detected as a univariate outlier in the variable of the falling utterances both within the entire sample and the JL subjects. These cases were clearly deviant from the BN/AN subjects and the other JL subjects, and thus, they were discarded. On the other hand, the following two cases, which were found to be outliers, were not excluded from the analysis. One AN subject and one JL subject were identified as univariate outliers of the span within the entire sample and that of the non-falling utterances within the JL subjects, respectively. However, this AN subject was not defined as an outlier within the AN subjects, and the JL subject was defined as an outlier due to the closeness to the BN/AN subjects in performance. They were therefore included in the analysis.

Thirdly, the tests of the assumptions to be satisfied to conduct the statistical analyses involved the distribution and variances of the data. The linearity of relationships among variables was checked with scatter diagrams. Whether the data were normally distributed or not was inspected by observing histograms of the individual variables and the quantified distribution with skewness and kurtosis. Field (2009) especially suggests the way of checking the assumption of the normal distribution using z-scores of skewness and kurtosis. However, it is known that the data tend to be distributed normally as the sample size gets larger. A MANOVA is known to be robust regarding the assumption of multivariate normality. For these reasons, only when the data were distributed as bimodal, or a ceiling effect or a floor effect was found, was it regarded as a serious violation of this assumption and the relevant variable was not used for the statistical analysis.

The violation of the assumption was found for three variables for intonation. In the two variables, the score for the nucleus in the final utterances and the score for the non-nuclear words in the final utterances, the maximum values were above 1 SD of the mean value for the entire sample. In the other variable, the score for the nuclear tone choice in the non-falling utterances, the minimum value was below -1 SD of the mean value for the entire sample. They were regarded as a ceiling effect and a floor effect, respectively, as will be reported in Section 4.4. These three variables of intonation were therefore discarded from the statistical analysis.

Finally, another assumption as to the homogeneity of covariance matrices was treated with more care (Stevens, 2007; Tabachnick & Fidell, 2007; Field, 2009). First, Box's test was used to determine whether the matrices were the same among the clusters to be tested. When the difference was not significant, it follows that the data did not violate this assumption. When it was significant, the assumption could have been violated. However, because Box's test tended to produce a significant difference, this result was ignored when the sample sizes were equal (Field, 2009; Hirai, 2012). Stevens (2007), for instance, suggests that the tests are robust when the sample size in the largest group was 1.5 times smaller than that in the smallest group. The sample size was thus checked when Box's test was significant. As long as the sample size of the largest group was 1.5 times as small as that of the smallest group, statistical analyses were performed with the  $\alpha$  level set at .05. If the data did not meet any of these conditions, a MANOVA and a discriminant analysis were carried out, setting a more stringent level, .01, as suggested by Stevens. Nevertheless, as implied by Stevens, this is simply one of the practical solutions when assumptions were violated, not the ultimate solution. Thus, a non-parametric MANOVA (NPMANOVA) was also conducted using Past (Hammer, Harper, & Ryan, 2001a, 2001b) to see if the results obtained by a MANOVA were reproduced or not. However, because this is not a standard method, when both results were comparable, the results produced by a MANOVA will be reported.

### **3.5.10. Three statistical tests for the analysis of each element of pronunciation**

The ultimate aim of this study was to describe which phonetic and phonological items are easy, learnable or difficult items for Japanese learners of English in the tested elements of pronunciation, and which elements are supportively, or positively, related to one another in the learning process of pronunciation. In order to address these issues, it was necessary to classify the JL subjects depending on their process of learning or level of learning, and then to describe the characteristics of each group. A cluster analysis was therefore carried out for the classification of the subjects, a MANOVA for the detection of a significant difference among the groups and a discriminant analysis for the description of each group.

The first statistical test was a cluster analysis, where subjects were grouped into clusters, based on similarities in the input variables. This analysis was conducted using the entire sample, including the BN/AN subjects, in order to profile the subjects and to group those who were going through a similar learning process or level together. This made it possible to form the groups of JL subjects depending on similarities in their performances. The inclusion of the BN/AN subjects in this analysis also showed JL subjects that were close to the BN/AN subjects, which helped to identify the JL subjects at a higher learning level. Not only this, but the way the BN/AN subjects were clustered also allowed the investigation of similarities among the BN/AN subjects.

Because this analysis is likely to be exploratory, there was no clear rule about which method of clustering was the best for the data at hand (Adachi, 2006), or about where clustering should be cut (Shigemasu, Mori, & Yanai, 2008). Commonly used methods include the group average method, centroid method, median method and Ward's method. In the present study, Ward's method, which tended to produce less complicated clusters, was employed for all cluster analyses as the clustering method (Adachi, 2006), where the Euclidean distance was used as the distance measure. For the cutoff point, clusters were elicited, based on the theoretical hypothesis that the BN/AN subjects would be grouped together. Therefore, one of the criteria of selecting the cutoff point of the clustering process was that one of the clusters consisted of as many BN/AN subjects as possible. The other criterion, which was somewhat arbitrary, was that the JL subjects were grouped into four clusters at most, considering the balance of the sample size of each cluster for the subsequent analyses.

The variables used for the cluster analysis are shown in Table 3.11. The analysis was carried out for each element of pronunciation, which is indicated with separate rows. One exception was applied to the analysis of intonation. The phonetic and phonological variables less closely relevant to one another were mixed in this element. Therefore, the cluster analysis was separately conducted, one with the span and level and the other with the remaining six variables.

Table 3.11

*Variables Submitted to the Cluster Analysis (all variables converted to z-scores)*

Element		Variable
Vowels	Quality	Standardized F1 mel values for 10 vowels
		Standardized F2 mel values for 10 vowels
		Score for structural difference among 10 vowels
	Duration	PVI values of long and short vowels in 3 pairs
Consonants	Plosives	Absolute VOT durations of 3 voiceless plosives
		VOT differences between aspirated and unaspirated plosives in 2 pairs
	Fricatives	COG, SD, skewness and kurtosis for 2 fricatives
	Approximants	Score for the /r/ and /l/ tokens
Rhythm		Pitch difference between stressed and weak vowels
		Intensity difference between stressed and weak vowels
		PVI values of stressed and weak vowels
		Vowel centralization
Intonation	Phonological	Score for the nucleus in 2 types of utterances
		Score for the non-nuclear words in 2 types of utterances
		Score for the nuclear tone choice in 2 types of utterances
	Phonetic	Span and level
Connected speech phenomena		Score for elision
		Score for CC linking in 2 phonetic contexts
		Score for CV linking in 2 phonetic contexts

*Note.* PVI = pairwise variability index; COG = center of gravity; CC = consonant-to-consonant; CV = consonant-to-vowel.

Another thing to be noted is that although the variables in Table 3.11 were used for the statistical analyses both of a MANOVA and a discriminant analysis, exceptions were the variable of the score for structural differences for the vowel quality, that of the PVI values of long and short vowels for the vowel duration and that of the PVI values of successive stressed and unstressed vowels for rhythm. As regards the score for structural difference, this phonological variable was submitted to a cluster analysis with the other phonetic variables in order to take the overall phonological classification of the vowels into consideration. That is, for instance, while the variables of the F1 and F2 values made it possible to profile the subjects from the perspective of exactly where each vowel was located in the subject's vowel



space, they lacked information about how well each vowel was distributed. This variable was therefore included in the cluster analysis, although not in the other analyses. In contrast, for the PVI values in vowel duration, the variable of the /ɑ:-æ/ pair was not submitted to the cluster analysis. This was because there was another pair involving /ɑ:/, and the inclusion of the variable of /ɑ:-æ/ was expected to increase the effect of this sound on the result of the cluster analysis, as will also be noted in Section 4.1.2. The variable of the PVI values of successive stressed and unstressed vowels for rhythm was not used for the cluster analysis, either. As will be reported in Section 4.3, there were a number of JL subjects who failed to produce the target tokens by inserting pauses. This variable was thus excluded from the subsequent statistical analyses as well as the cluster analysis. It should also be noted that the values obtained were all standardized to the z-scores, using the mean and standard deviation for the entire sample, because the variables varied in unit and scale.

The second statistical analysis was a MANOVA, which was conducted in order to test whether there was a significant difference among the clusters generated by the cluster analysis. It was employed because there was more than one dependent variable for each element of pronunciation targeted in the current study. The clusters were used as independent variables and the variables presented in Table 3.11 were used as dependent variables. However, there were several exceptions, as noted above, and some variables were not used for a MANOVA: the score for structural difference among 10 vowel in the vowel quality; the score for the nucleus in the final utterances, the score for the non-nuclear words in the final utterances, the score for the nuclear tone choice in the non-falling utterances, the span and the level in intonation. The score for structural difference was the variable solely added to the cluster analysis to count the classification of vowels, as described in the paragraph above. In contrast, the reason the other variables were not submitted to the MANOVA involves the results of the analyses. The two scores for the nucleus and non-nuclear words in the final utterances had a ceiling effect and the score for the nuclear tone choice in the non-falling utterances had a floor effect, as already noted in Section 3.5.9 and to be detailed in Section 4.4. The reason why the span and level were excluded from the MANOVA has not been noted.

As will be reported in more detail in Section 4.4, the result of the cluster analysis showed that individual differences were stronger than differences between the BN/AN subjects and the JL subjects as for the span and level. These variables were thus not submitted to the MANOVA.

A one-way MANOVA was carried out for the following elements: vowel quality, vowel duration, plosives, approximants, rhythm, intonation and connected speech phenomena. For vowel quality, the F1 and F2 mel values were each submitted to a separate MANOVA. A two-way mixed-design MANOVA was conducted for fricatives. In this MANOVA, COG, SD, skewness and kurtosis were the dependent variables, the clusters were the between-subjects factors and the two voiceless fricatives were the within-subjects factors. This identified the presence of difference in the fricatives as well as the clusters, according to COG, SD, skewness and kurtosis.

The final statistical analysis was a discriminant analysis. This was carried out as a post-hoc test to supplement the MANOVA and specify the cluster that was different and the variables that contributed to discriminating between the clusters. Field (2009) proposes that a discriminant analysis should be used rather than conducting a series of univariate ANOVA to identify where there is a significant difference. This was adopted for the independent variables when a significant difference was yielded by the MANOVA. As for the repeated measure, employed in the analysis of fricatives, the within-subjects independent variables were tested using a follow-up univariate ANOVA.

In the discriminant analysis, the clusters were used as the dependent variables, while the same variables as those used in the MANOVA were the independent variables. The forced entry method was selected because this study aimed to identify the learning process of the JL subjects as well as the differences between the JL subjects and the BN/AN subjects. The study is exploratory as far as the former aim is concerned. In other words, there was no strong theoretical reason for which independent variables would contribute to discriminating the dependent variables. This method was thus considered to be preferable to the stepwise method, which is also commonly used for a discriminant analysis. The discriminant functions generated under this estimation were first omitted in terms of whether they significantly

discriminated the clusters or not. Only the functions found to differentiate the clusters significantly were considered in this study. Even though the function was significant, whether it should be interpreted or not was carefully considered when it accounted for only a small proportion of variances. The clusters that these functions discriminated and the variables that contributed to this discrimination were then examined. The former judgment was conducted based on the distance displayed in the canonical discriminant function plots and the location with reference to the group centroid indicated by positive and negative signs. The latter judgment was made, based on the structural matrix of the correlation between the variables and each of the discriminant functions. Although there is no decisive standard in the interpretation of the correlations, those higher than .33 were interpreted to suggest variables contributing to the discrimination, following the convention described by Tabachnick and Fidell (2007).

### **3.5.11. A statistical test for the analysis of the relationships between the elements of pronunciation**

In order to investigate whether there was a supportive relationship between the elements of pronunciation, a correlation analysis with Spearman's rank-order correlation coefficient was conducted. This correlation coefficient was selected to address the issue because all data were converted into rank data, ordinal scale, so as to compare the variables obtained in different units and scales. A profile of the pairwise elements of pronunciation was carried out, along with the correlation analysis. Based on the results of both analyses, the presence of a supportive relationship was comprehensively examined.

The variables presented in Table 3.10 were used for the correlation analysis, and all pairwise correlations were obtained. Field (2009) states that the correlation coefficient of  $\pm 0.1$  is a small effect, that of  $\pm 0.3$  is a medium effect and that of  $\pm 0.5$  is a large effect. These criteria are based on Cohen (1988). With reference to this, the analysis was first conducted including the BN/AN subjects in order to investigate which pairs of elements of pronunciation were highly or moderately correlated with one another and which were not correlated. However, the inclusion of the BN/AN subjects could lead to a stronger correlation

because these subjects performed much better than the JL subjects. Therefore, a correlation analysis using Spearman's rank-order correlation coefficient was also carried out excluding the BN/AN subjects to confirm whether the results of the correlation analysis with the entire sample would be replicated by an analysis with the data excluding the BN/AN subjects. This analysis was performed with the JL subjects only, and both medium effect and large effect of the correlation coefficient were equally interpreted to suggest a rather strong relationship underlying the relevant pronunciation elements. While the JL subjects in this study were expected to vary in proficiency level, as noted in Section 3.1, the extent of variation was limited because they were from the same population, and had learned English under the same guidelines of the course of study (MEXT, 1998, 2009). This could lessen a possible correlation, and therefore, the medium effect was interpreted to suggest a supportive relationship, although this was less strict than the conventional criteria.

### **3.6. Criteria of learning**

The present study aimed to reveal the learning of pronunciation by Japanese learners of English. In order to address the two research questions, each phonetic and phonological item analyzed was specified as easy item, learnable item or difficult item, as will be discussed in Chapter 5. These definitions were given by discussing the results obtained by a series of analyses, following the criteria below.

First of all, the cluster(s) that represented the performances of native speakers of English was defined depending on how the BN/AN subjects were clustered. The cluster(s) was called the BN/AN cluster(s), and the remaining clusters were called the JL clusters, respectively. The BN/AN cluster(s) could include some JL subjects who were grouped into this cluster, while the JL clusters could have some BN/AN subjects who fell outside the BN/AN cluster(s). By comparing the BN/AN cluster(s) with the JL clusters, whether the JL subjects were going through the learning of the item concerned was examined using the following criteria. The first criterion was the number of JL subjects who were classified into the BN/AN cluster(s) in the cluster analysis. The more JL subjects were clustered with them, the more learnable or easier the element of pronunciation was for Japanese learners of

English to learn to produce. This means, at the same time, that a smaller number of JL subjects in this cluster suggests that the element concerned was difficult for Japanese learners of English overall. The second criterion was which items contributed to discriminating between the JL clusters and the BN/AN cluster(s). The items that did not serve to discriminate the JL clusters from the BN/AN cluster(s) were interpreted as easy items for Japanese learners of English. The items that discriminated more than half the JL subjects from the BN/AN cluster(s) were defined as difficult items. The items that discriminated some JL subjects, but not more than half the JL subjects, were defined as learnable items.

In order to discuss difficult items in more detail, the items that made the JL clusters different from one another were also carefully observed. The difference among the JL clusters could suggest that some JL subjects were learning this item, although it may not have been fully learned. Items defined as difficult in this study were thus subdivided into the following three types. The first type are the items that discriminated more than half the JL subjects, but not all JL clusters, from the subjects in the BN/AN cluster(s). The second type are the items that discriminated all JL clusters, the majority of the JL subjects, from the BN/AN cluster(s), but where more than half the JL subjects improved toward a native-speaker level, even though these items were not fully learned. The third type are the items that discriminated between all JL clusters and the BN/AN cluster(s), and where less than half the JL subjects improved to approximate to the performances of the BN/AN cluster(s). These three types were called D1, D2 and D3, respectively. The difficult items in D1 are those that a small number of Japanese learners of English may possibly learn to produce. The difficult items in D2 are those that the majority of Japanese learners of English could approximate a native-speaker level to some extent. This suggests that although these items are generally difficult for the majority of Japanese learners to learn fully, they could possibly be learned if some treatment such as training is provided. In contrast, the items in D3 are those that it would be unlikely for Japanese learners to learn to produce or even to improve in a naturalistic learning environment. In order for them to learn these items, some radical, extensive treatment would be required.

## Chapter 4 Results

### 4.1. Monophthongal Vowels

#### 4.1.1. Vowel quality

The descriptive statistics of the BN, AN and JL groups are presented in Table 4.1 and Figure 4.1. The standardized F1 and F2 mel values and score for structural differences are summarized in the table and the former values are used for scatter diagrams. Figure 4.1(a), (b) and (c) illustrate the vowel distribution of the BN, AN and JL groups, respectively, where the x-axis corresponds to the standardized F2 mel values and the y-axis, to the standardized F1 mel values.

Table 4.1

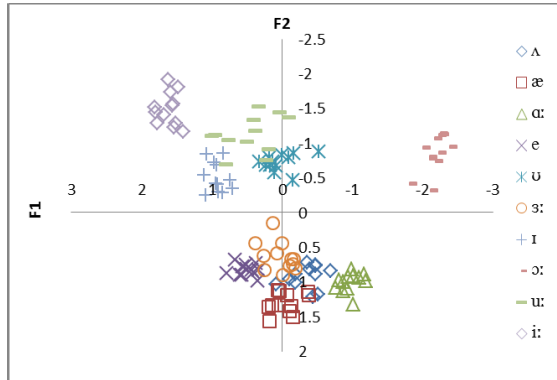
*Descriptive Statistics of Vowel Quality for BN, AN and JL Groups*

		BN ( <i>n</i> = 12)				AN ( <i>n</i> = 7)				JL ( <i>n</i> = 72)			
		<i>M</i>	<i>SD</i>	<i>Max</i>	<i>Min</i>	<i>M</i>	<i>SD</i>	<i>Max</i>	<i>Min</i>	<i>M</i>	<i>SD</i>	<i>Max</i>	<i>Min</i>
i:	F1	-1.49	0.24	-1.16	-1.91	-1.55	0.27	-1.22	-1.93	-1.33	0.17	-0.92	-1.71
	F2	1.61	0.13	1.81	1.40	1.72	0.21	2.02	1.40	1.51	0.12	1.78	1.20
ɪ	F1	-0.51	0.22	-0.25	-0.86	-0.61	0.23	-0.30	-0.92	-1.28	0.14	-0.91	-1.57
	F2	0.92	0.13	1.10	0.70	1.08	0.22	1.38	0.81	1.46	0.17	1.91	1.13
e	F1	0.83	0.09	0.98	0.68	0.42	0.16	0.57	0.10	0.23	0.22	0.70	-0.40
	F2	0.51	0.14	0.79	0.33	0.70	0.14	0.87	0.43	0.82	0.13	1.07	0.55
æ	F1	1.30	0.15	1.57	1.12	1.33	0.24	1.65	1.07	1.05	0.11	1.40	0.82
	F2	-0.05	0.19	0.19	-0.38	0.67	0.15	0.97	0.50	-0.29	0.13	0.04	-0.62
ʌ	F1	0.92	0.16	1.21	0.73	1.04	0.26	1.40	0.69	1.08	0.11	1.31	0.82
	F2	-0.35	0.21	0.08	-0.69	-0.30	0.15	-0.10	-0.54	-0.53	0.11	-0.28	-0.82
ɑ:	F1	1.01	0.14	1.33	0.81	1.42	0.12	1.55	1.26	1.13	0.14	1.47	0.81
	F2	-0.97	0.13	-0.76	-1.19	-0.88	0.17	-0.70	-1.13	-0.73	0.14	-0.26	-1.06
ɔ:	F1	-0.86	0.27	-0.31	-1.14	0.00	0.22	0.38	-0.24	-0.14	0.20	0.27	-0.62
	F2	-2.16	0.15	-1.82	-2.39	-1.77	0.10	-1.64	-1.87	-1.84	0.19	-1.32	-2.52
u:	F1	-1.13	0.26	-0.69	-1.52	-1.21	0.19	-0.99	-1.56	-0.90	0.13	-0.59	-1.23
	F2	0.46	0.35	1.00	-0.09	-0.52	0.48	0.38	-1.05	-0.09	0.30	0.58	-0.65
ʊ	F1	-0.72	0.12	-0.46	-0.88	-0.45	0.21	-0.16	-0.82	-0.89	0.13	-0.59	-1.15
	F2	0.02	0.23	0.32	-0.53	-0.39	0.20	-0.06	-0.60	0.21	0.24	0.61	-0.40

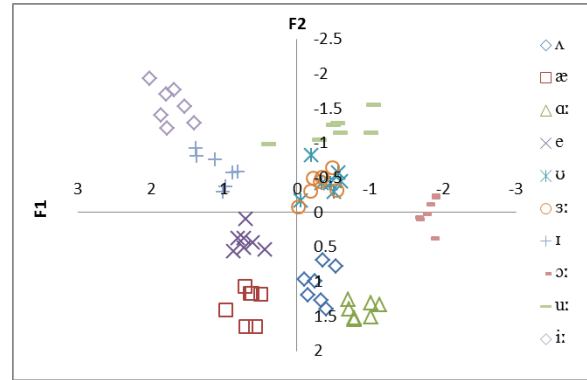
	BN				AN				JL			
ɜ: F1	0.64	0.21	0.90	0.15	-0.40	0.18	-0.08	-0.65	1.05	0.21	1.56	0.17
F2	0.03	0.19	0.38	-0.19	-0.30	0.18	-0.01	-0.54	-0.52	0.19	0.17	-0.86
Structure	0.04	0.00	0.05	0.03	0.05	0.00	0.05	0.04	0.09	0.02	0.14	0.04

Note. These values are standardized F1 and F2 mel values. Structure = Structural difference.

(a) Vowel distribution of BN group



(b) Vowel distribution of AN group



(c) Vowel distribution of JL group

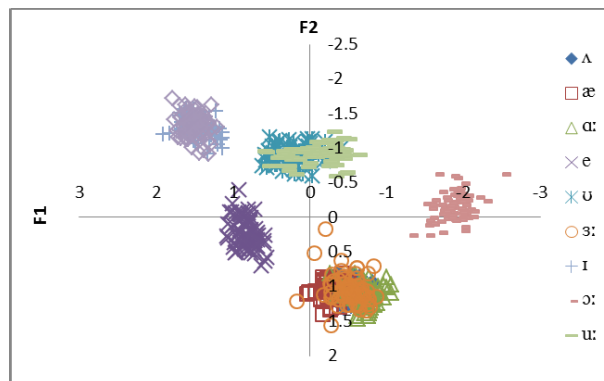


Figure 4.1. Vowel distribution for BN, AN and JL groups: (a) BN; (b) AN; and (c) JL. The standardized F1 and F2 mel values are plotted.

Figure 4.1 clearly shows the different distribution of the 10 vowels among the three groups. One apparent difference between the BN and AN groups and JL group is that the 10 vowels shrunk into five categories for the JL group as shown in Figure 4.1(c), although they were scattered for the BN and AN groups as in Figure 4.1(a) and (b). All vowels except for /ɜ:/ and /u/ of the AN group maintained their own distinct category, with each category close to one another, as shown in Figure 4.1(a) and (b). Another difference concerns the difference

between the BN group and AN group. As far as Figure 4.1(a) and (b) are compared, a mid-central back /ɜ:/ and a low-mid front vowel /e/ were articulated lower for the BN group, a mid back vowel /ɔ:/ was lower for the AN group, a low front vowel /æ/ was more front for the AN group and a high back vowel /u:/ was further back for the AN group.

In order to categorize each subject based on the overall distributions of the vowel quality, a cluster analysis was conducted using the z-scores of the following variables based on the mean and the standard deviation for the entire sample: the standardized F1 and F2 mel values and the score for structural difference. The score for structural difference was added as one of the variables for consideration of how well each vowel was classified as native-like overall.

The dendrogram generated from the analysis showed that all of the BN and AN subjects were grouped into two separate clusters, respectively, when the subjects were clustered into five (see Appendix C for the dendrogram). This suggests that this clustering successfully categorized the BN and AN subjects into different clusters depending on their accent. Cluster 1 consisted of 7 AN subjects, and Cluster 2 consisted of 12 BN subjects and 1 JL subject. These two clusters could be considered to represent the BN and AN groups, respectively. On the other hand, Clusters 3, 4 and 5 were made up only of JL subjects, representing 30, 14 and 27 subjects, respectively.

The descriptive statistics of the standardized F1 and F2 mel values and the score for structural difference are displayed for each cluster in Table 4.2. The profile of each cluster is presented in a line graph in Figure 4.2, which is based on the rank averaged across subjects in each cluster. The subjects were ranked for F1 and F2 separately, according to the absolute values of the z-scores based on the mean and standard deviation for the BN/AN subjects. These ranks were then averaged to obtain the overall rank of the vowel quality. A smaller value corresponds to a higher rank, and the mean rank was calculated for each cluster. In the subsequent sections of Chapter 4, line graphs to represent the profile of each cluster show the target items on the x-axis and the average rank on the y-axis. The vowel distribution of each cluster is visually depicted in Figure 4.3(a), (b), (c) and (d). The standardized F2 mel values



are plotted on the x-axis and the standardized F1 mel values are on the y-axis in these graphs. Cluster 1 was not added in Table 4.2 and Figure 4.3 because it corresponded exactly to the AN group, which was shown in Table 4.1 and Figure 4.1.

Table 4.2  
*Descriptive Statistics of Vowel Quality for Four Clusters*

		Cluster 2 ( <i>n</i> = 13)		Cluster 3 ( <i>n</i> = 30)		Cluster 4 ( <i>n</i> = 14)		Cluster 5 ( <i>n</i> = 27)	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
i:	F1	-1.48	0.23	-1.44	0.13	-1.29	0.14	-1.21	0.13
	F2	1.58	0.15	1.55	0.11	1.38	0.08	1.53	0.11
ɪ	F1	-0.54	0.25	-1.31	0.15	-1.28	0.10	-1.27	0.13
	F2	0.94	0.14	1.47	0.16	1.30	0.10	1.54	0.13
e	F1	0.82	0.10	0.31	0.21	0.06	0.20	0.21	0.17
	F2	0.51	0.13	0.79	0.13	0.83	0.09	0.86	0.13
æ	F1	1.29	0.15	1.07	0.13	1.01	0.08	1.04	0.10
	F2	-0.05	0.19	-0.28	0.15	-0.32	0.12	-0.30	0.09
ʌ	F1	0.93	0.15	1.08	0.12	1.07	0.12	1.08	0.10
	F2	-0.36	0.20	-0.54	0.11	-0.54	0.09	-0.50	0.12
ɑ:	F1	0.99	0.14	1.09	0.12	1.17	0.14	1.17	0.14
	F2	-0.92	0.23	-0.81	0.12	-0.79	0.06	-0.63	0.10
ɔ:	F1	-0.84	0.27	-0.08	0.19	-0.14	0.18	-0.20	0.18
	F2	-2.19	0.17	-1.78	0.17	-1.96	0.14	-1.82	0.17
ʊ	F1	-0.74	0.13	-0.83	0.14	-0.91	0.10	-0.95	0.10
	F2	0.03	0.22	0.25	0.18	0.46	0.09	0.02	0.21
u:	F1	-1.11	0.26	-0.85	0.15	-0.88	0.09	-0.97	0.10
	F2	0.42	0.37	-0.16	0.27	0.30	0.17	-0.21	0.23
ɜ:	F1	0.68	0.26	0.96	0.24	1.17	0.13	1.09	0.15
	F2	0.04	0.19	-0.50	0.19	-0.66	0.11	-0.49	0.14
Structure		0.04	0.01	0.08	0.02	0.10	0.01	0.10	0.02

*Note.* These values are standardized F1 and F2 mel values. Structure = structural difference.

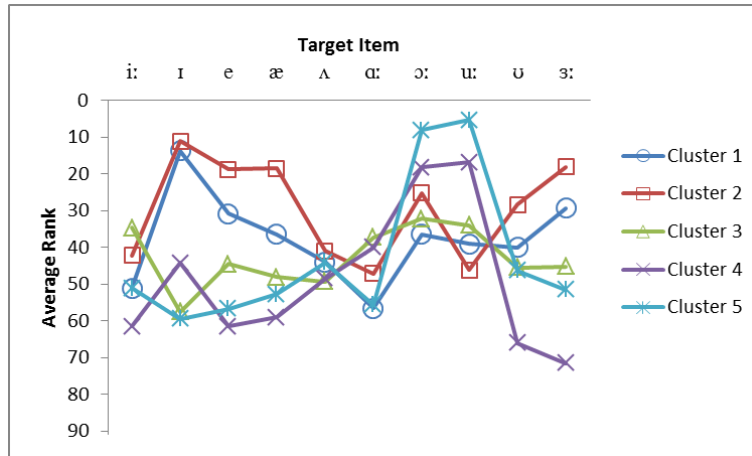
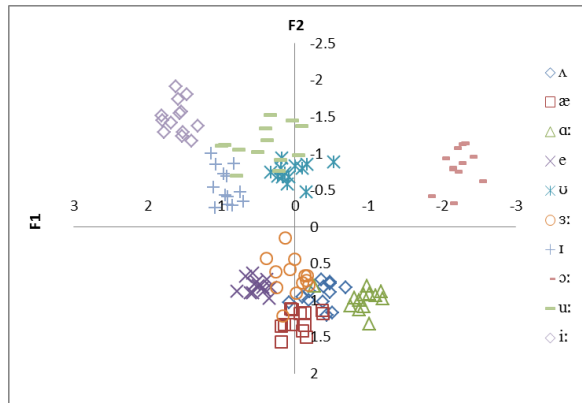
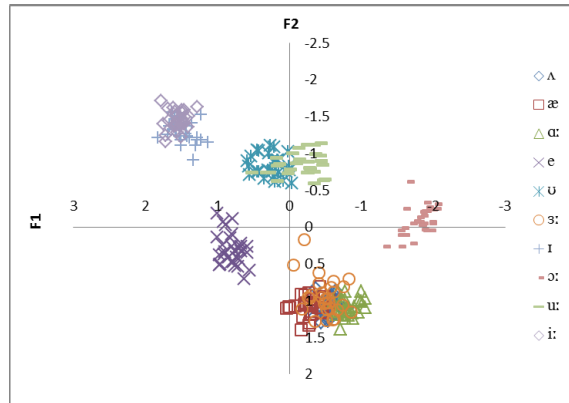


Figure 4.2. Profile of each cluster for vowel quality.

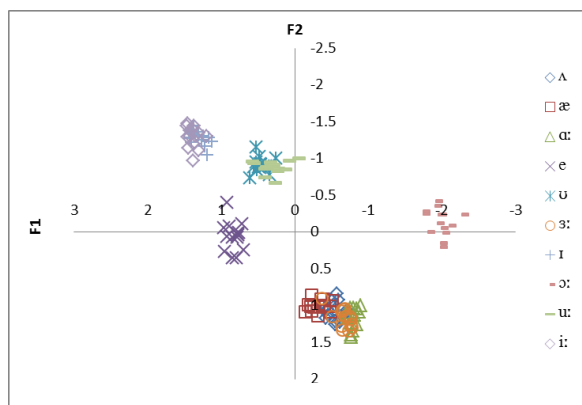
(a) Vowel distribution of Cluster 2



(b) Vowel distribution of Cluster 3



(c) Vowel distribution of Cluster 4



(d) Vowel distribution of Cluster 5

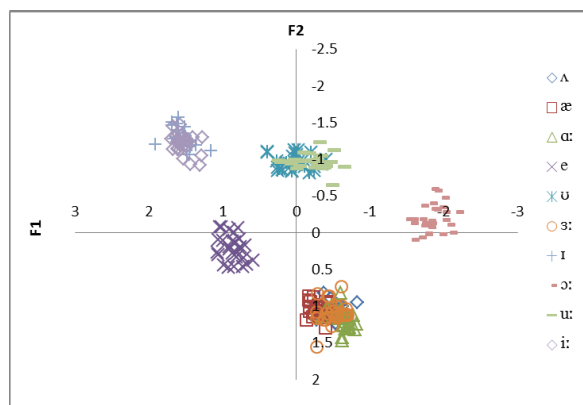


Figure 4.3. Vowel distribution for four clusters: (a) Cluster 2; (b) Cluster 3; (c) Cluster 4; and (d) Cluster 5. The standardized F1 and F2 mel values are plotted.

The profile in Figure 4.2 shows that Clusters 1 and 2, representing native speakers,

clearly performed better in articulating /ɪ, ʒ/. Obviously, they ranked higher in these two vowels. This is less noticeable, but also true of /e, æ, u/. These vowels suggest a certain extent of the difficulty learning them for the subjects in Clusters 3, 4 and 5. As for the remaining vowels, Clusters 1 and 2 did not necessarily rank higher than the JL clusters, Clusters 3, 4, and 5, indicating that there would be no critical difference in articulating these vowels among the clusters. Clusters 4 and 5 ranked even higher for /ɔ:, u:/, but the rank of these vowels could be attributed to the above-mentioned difference in the BN and AN groups.

Concerning Figure 4.3(a), (b), (c) and (d) and Figure 4.1(b), there are some similarities and differences as to the distribution of vowels between the five clusters. When Cluster 1 and Cluster 2, each representing the AN group and the BN group, were compared on Figure 4.1(b) and Figure 4.3(a), one striking difference was seen in the cluster means for /ʒ/. As in the F1 values, Cluster 1 /ʒ/ was located higher ( $M = -0.40$ ,  $SD = 0.18$ ) than Cluster 2 ( $M = 0.68$ ,  $SD = 0.26$ ) in the vowel space, and it was even more closely integrated with /u/. Their differences were also enhanced by the lower /e/, higher /ɔ:/, further back /æ/ and more front /u:/ of Cluster 2 ( $M = 0.82$ ,  $SD = 0.10$  for /e/ F1;  $M = -0.84$ ,  $SD = 0.27$  for /ɔ:/ F1;  $M = -0.05$ ,  $SD = 0.19$  for /æ/ F2;  $M = 0.42$ ,  $SD = 0.37$  for /u:/ F2) than Cluster 1 ( $M = 0.42$ ,  $SD = 0.16$  for /e/ F1;  $M = 0.00$ ,  $SD = 0.22$  for /ɔ:/ F1;  $M = 0.67$ ,  $SD = 0.15$  for /æ/ F2;  $M = -0.52$ ,  $SD = 0.48$  for /u:/ F2). In spite of these differences, Cluster 1 and Cluster 2 were analogous in that most of the vowels formed their own category in the auditory space, with a clear distinctness from the adjacent vowel(s).

On the other hand, Figure 4.3(b), (c) and (d), representing the JL clusters, show the underlying commonality of the JL subjects. The five categories in the vowel space are easily recognizable, and were most likely found in the Japanese five vowel categories. The F1 and F2 values in Table 4.2 also reflect this tendency. For instance, the subjects in Clusters 4 and 5 produced similar F1 and F2 mel values all for /æ/, /ʌ/, /ʒ/ and /ɑ:/, which suggests that these four vowels tended not to be discriminated from one another and form one category corresponding to Japanese /a/. In contrast, Cluster 3, as Figure 4.3(b), seems to show some learning, which is evident in the separation of the vowels in /i/ area, /u/ area and /a/ area.

To examine whether there was a significant difference in the articulation of the 10 vowels between the clusters, as observed above, one-way MANOVAs were carried out separately for F1 and F2 because F1 and F2 are separate acoustic items each referring to the tongue height and tongue position. In the analyses, excluding /u:/, Cluster 1 and Cluster 2 were combined to form one cluster, Cluster 1/2, for the following reasons: the small sample size in Cluster 1 could lower the statistical power; while the height of /ɜ:/, e, ɔ:/ and the position of /æ, u:/ differed between the two clusters as described above, it was estimated that the two clusters being combined would not affect the results apart from /u:/. A high back vowel /u:/ in Clusters 3, 4 and 5 was roughly located somewhere between Cluster 1 and Cluster 2 as in Figure 4.1(b) and Figure 4.3(a), (b), (c) and (d). In this case, combining Cluster 1 and Cluster 2 could hide a possible difference between the BN/AN cluster and the JL clusters. On the other hand, Clusters 3, 4 and 5 deviated from both Clusters 1 and 2 as to /æ, ɜ:/ and were very close to Cluster 1 as to /e, ɔ:/. This means that combining Cluster 1 and Cluster 2 would not affect the results greatly in the aim to detect a statistical difference between the BN/AN cluster and the JL clusters. One-way MANOVAs were thus conducted, where the nine standardized F1 values and the nine standardized F2 values were used as the dependent variables, and the classification results of the subjects into the four clusters as the independent variables. Correlations between the dependent variables are shown in Appendix D for F1 and Appendix E for F2. As regards F1, there were nearly zero correlations between the following variables: /ʌ/ and /ɑ:/, /ʌ/ and /ɜ:/, /ɑ:/ and /æ/, /i:/ and /æ/, /ɑ:/ and /ɪ/, /ɔ:/ and /ɜ:/, /i:/ and /ɔ:/ and /i:/ and /u:/. Similarly, for F2, the correlation coefficients were near zero between the following variables: /ʌ/ and /ɑ:/, /ɑ:/ and /ɜ:/, /ɑ:/ and /ɔ:/, /e/ and /ʊ/, /ʊ/ and /ɪ/, /ɔ:/ and /ʊ/ and /u:/ and /ɜ:/. In contrast, the remaining correlations were moderate, which led to the expectation that a MANOVA would perform fairly on these variables. The sample size of the largest cluster was 30, while that of the smallest cluster was 14. Because the former was more than 1.5 times larger than the latter, the  $\alpha$  level was set at .01. The results of Pillai's trace revealed that there was a significant difference in both F1 and F2 among the four clusters,  $F(27,243) = 11.96, p < .001, \eta_p^2 = .57$  for F1 and  $F(27,243) = 7.96, p$

$< .001, \eta_p^2 = .47$  for F2.

A discriminant analysis, a post-hoc test to follow up the MANOVAs, was subsequently carried out for the standardized F1 and F2 values separately. As for the F1 values, the analysis statistically found the first function, the second function and the third function, where each explained 92.2% of the variance, canonical  $R^2 = .90$ , 6.7% of the variance, canonical  $R^2 = .40$ , and 1.2% of the variance, canonical  $R^2 = .11$ . When combined, these three functions significantly discriminated between the clusters with the Wilk's lambda value of .05,  $\chi^2(27) = 246.23, p < .001$ . While the last two functions also discriminated between the clusters significantly with the Wilk's lambda value of .54,  $\chi^2(16) = 51.96, p < .001$ , the third function alone was not able to differentiate between the clusters at a significant level with the Wilk's lambda value of .90,  $\chi^2(7) = 9.30, p = .232$ . For this reason, the third function was not interpreted. The canonical discriminant function plot in Figure 4.4 and the group centroids in Table 4.3 revealed that the first function distinguished Cluster 1/2 from Clusters 3, 4 and 5. They also showed that the second function contributed to distinguishing Cluster 3 from Clusters 4 and 5. Cluster 4 and Cluster 5 were not differentiated by F1 clearly.

Table 4.3

*Group Centroids for the Standardized F1 mel Values*

Cluster	Function	
	1	2
1/2	-5.58	-0.10
3	1.31	1.08
4	2.08	-0.72
5	1.60	-0.76

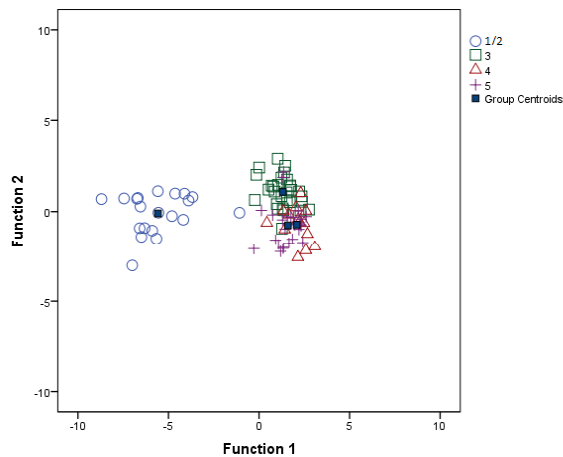


Figure 4.4. Canonical discriminant function plot for F1.

The structural matrix in Table 4.4 shows the correlations between the variables and the discriminant functions. They indicated that /ɪ, ɜ:/ loaded highly on the first function ( $r = -0.61$  and  $r = 0.33$ ), where it needs to be noted that they were correlated with the function in the opposite direction. From the standardized F1 mel value of Cluster 1, or the AN group, in Table 4.1 and that of Cluster 2 in Table 4.2, which reflects tongue height, the subjects in Cluster 1/2 produced /ɪ/ by lowering the tongue more than those in Clusters 3, 4 and 5. The subjects in the former clusters produced /ɜ:/ by setting their tongue higher than those in the latter clusters. This is more apparent in the comparison of Figure 4.1(b) and Figure 4.3(a) with Figure 4.3(b), (c) and (d). Above all, these graphs show that it is characteristic of Clusters 1 and 2 that /ɪ, ɜ:/ formed a distinct category in the phonological space.

On the other hand, /i:/ ( $r = -0.66$ ) and /ʊ/ ( $r = 0.34$ ) loaded on the second function, which worked to separate Cluster 3 from Clusters 4 and 5, as in Table 4.4. To interpret this with the standardized F1 values in Table 4.2 demonstrates that the subjects in Cluster 3 articulated /i:/, setting the tongue higher ( $M = -1.44$ ,  $SD = 0.13$ ) than Cluster 4 ( $M = -1.29$ ,  $SD = 0.14$ ) and Cluster 5 ( $M = -1.21$ ,  $SD = 0.13$ ). The opposite sign of the value suggests, on the contrary, that the subjects in Cluster 3 articulated /ʊ/, setting the tongue lower ( $M = -0.83$ ,  $SD = 0.14$ ) than Cluster 4 ( $M = -0.91$ ,  $SD = 0.10$ ) and Cluster 5 ( $M = -0.95$ ,  $SD = 0.10$ ). Figure 4.3(b), (c) and (d) also support this, although the differences between Cluster 3 and

Clusters 4 and 5 were smaller than between Cluster 1/2 and Clusters 3, 4 and 5, differentiated by the first function.

Table 4.4  
*Structural Matrix for the Correlations between the Standardized F1 mel Variables and the Two Discriminant Functions*

F1 Variable	Function	
	1	2
ɪ	<b>-.61</b>	-.26
ɜ:	<b>.33</b>	-.19
æ	-.28	.11
i:	.16	<b>-.66</b>
ɑ:	.00	-.27
ɔ:	.20	.24
e	-.32	.31
ʊ	-.25	<b>.34</b>
ʌ	.11	.04

*Note.* The variables with the absolute value of correlations with the corresponding functions of .33 and above were highlighted in bold.

Three discriminant functions were detected in the F2 values. The first function explained 86.3% of the variance, canonical  $R^2 = .90$ , the second function explained 8.9% of the variance, canonical  $R^2 = .48$ , and the third function explained 4.8% of the variance, canonical  $R^2 = .33$ . In combination, the three discriminant functions significantly discriminated between the four clusters with the Wilk's lambda value of .04,  $\chi^2(27) = 280.30$ ,  $p < .001$ . The second and third functions also distinguished between them at a significant level with the Wilk's lambda value of .35,  $\chi^2(16) = 88.34$ ,  $p < .001$ . After removing the first two functions, the third function succeeded in setting the clusters apart by itself significantly with the Wilk's lambda value of .67,  $\chi^2(7) = 33.53$ ,  $p < .001$ . The group centroids in Table 4.5 reveal that the first function distinguished Cluster 1/2 from Clusters 3, 4 and 5, which is illustrated by the discriminant function plot in Figure 4.5. The group centroids in Table 4.5 also indicate that the second function distinguished Cluster 4 from Clusters 3 and 5 and the

third function distinguished between Cluster 3 and Cluster 5.

Table 4.5

*Group Centroids for the Standardized F2 mel Values*

Cluster	Function		
	1	2	3
1/2	-5.49	0.12	-0.08
3	1.34	0.11	0.93
4	1.23	-2.07	-0.48
5	1.94	0.86	-0.72

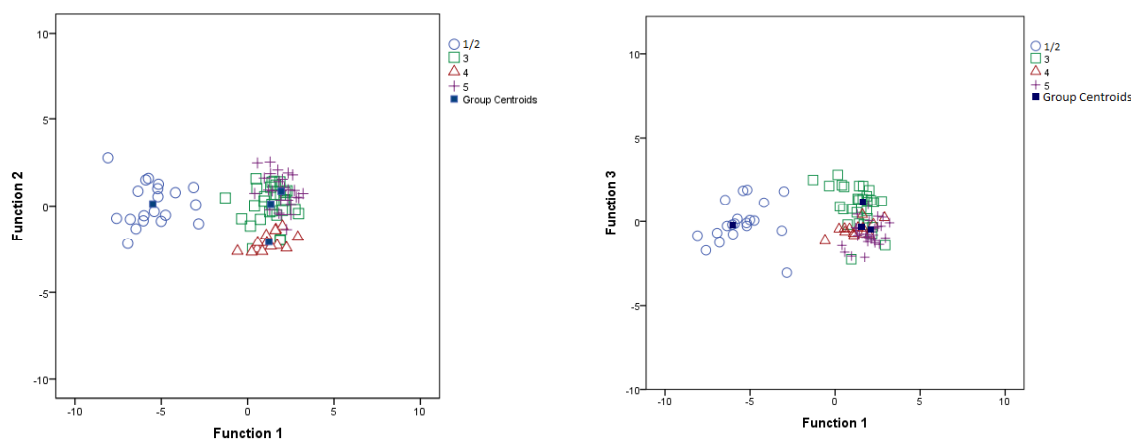


Figure 4.5. Canonical discriminant function plot for F2.

The structural matrix in Table 4.6 presents the correlations between the variables and the discriminant functions. They demonstrated that /ɪ, ɜ:/ loaded more highly than any other on the first function ( $r = .47$  and  $r = -.34$ ); that is, these vowels separated Cluster 1/2 from Clusters 3, 4 and 5. When this result being interpreted with the standardized F2 mel values in Table 4.1 and Table 4.2, which reflect tongue position, the subjects in Cluster 1/2 produced /ɜ:/ by putting their tongue at a more front position ( $M = -0.30$ ,  $SD = 0.18$  for Cluster 1;  $M = 0.04$ ,  $SD = 0.19$  for Cluster 2) and /ɪ/ by putting their tongue at a further back position ( $M = 1.08$ ,  $SD = 0.22$  for Cluster 1 and  $M = 0.94$ ,  $SD = 0.14$  for Cluster 2) than those in Clusters 3 ( $M = -0.50$ ,  $SD = 0.19$  for /ɜ:/;  $M = 1.47$ ,  $SD = 0.16$  for /ɪ/), Cluster 4 ( $M = -0.66$ ,  $SD = 0.11$  for /ɜ:/;  $M = 1.30$ ,  $SD = 0.10$  for /ɪ/) and Cluster 5 ( $M = -0.49$ ,  $SD = 0.14$  for /ɜ:/;  $M = 1.54$ ,



$SD = 0.13$  for /ɪ/. This is evident in Figure 4.1(b) and Figure 4.3(a), (b), (c) and (d), where a clearly distinct category for /ɜ:/, ɪ/ was seen in Clusters 1 and 2.

Table 4.6

*Structural Matrix for the Correlations between the Standardized F2 mel Variables and the Three Discriminant Functions*

F2 Variable	Function		
	1	2	3
ɪ	<b>.47</b>	<b>.44</b>	.09
æ	-.32	.12	-.01
e	.27	-.00	-.21
ʌ	-.20	.14	-.19
ʊ	.20	<b>-.73</b>	<b>.41</b>
i:	-.13	<b>.45</b>	.27
ɜ:	<b>-.34</b>	<b>.42</b>	.03
ɑ:	.19	.30	<b>-.61</b>
ɔ:	.16	.23	.30

*Note.* The variables with the absolute value of correlations with the corresponding functions of .33 and above were highlighted in bold.

As seen in Table 4.6, it was found that /ʊ, i:, ɪ, ɜ:/ loaded highly on the second function ( $r = -.73$ ,  $r = .45$ ,  $r = .44$  and  $r = .42$ ). This suggests that /ʊ/ and /i:, ɪ, ɜ:/ discriminated Cluster 4 from Clusters 3 and 5 in the opposite direction, which was reflected in the tongue position as follows. The subjects in Cluster 4 produced /ʊ/ with their tongue in a more front position ( $M = 0.46$ ,  $SD = 0.09$ ) than those in Cluster 3 ( $M = 0.25$ ,  $SD = 0.18$ ) and Cluster 5 ( $M = 0.02$ ,  $SD = 0.21$ ). In contrast, the subjects in Cluster 4 produced /i:, ɪ, ɜ:/ with their tongue in a further back position ( $M = 1.38$ ,  $SD = 0.08$  for /i:/;  $M = 1.30$ ,  $SD = 0.10$  for /ɪ/;  $M = -0.66$ ,  $SD = 0.11$  for /ɜ:/) than those in Clusters 3 ( $M = 1.55$ ,  $SD = 0.11$  for /i:/;  $M = 1.47$ ,  $SD = 0.16$  for /ɪ/;  $M = -0.50$ ,  $SD = 0.19$  for /ɜ:/) and Cluster 5 ( $M = 1.53$ ,  $SD = 0.11$  for /i:/;  $M = 1.54$ ,  $SD = 0.13$  for /ɪ/;  $M = -0.49$ ,  $SD = 0.14$  for /ɜ:/). A visual comparison of Figure 4.3(b), (c) and (d) also corroborated these results. According to all these graphs, these differences suggest that Cluster 4 approximated to Cluster 1/2 as for /ɪ/

and Clusters 3 and 5, to Cluster 1/2 as for /i:, ɜ:, u/.

Table 4.6 also shows that /ɑ:, ʊ/ loaded on the third function ( $r = -.61$  and  $r = .41$ ), which differentiated between Cluster 3 and Cluster 5. The subjects in Cluster 3 produced /ɑ:/ in a further back position ( $M = -0.81$ ,  $SD = 0.12$ ) and /ʊ/ in a more front position ( $M = 0.25$ ,  $SD = 0.18$ ) than those in Cluster 5 ( $M = -0.63$ ,  $SD = 0.10$  for /ɑ:/;  $M = 0.02$ ,  $SD = 0.21$  for /ʊ/). Figure 4.3(b) and (d) illustrate that Cluster 3 was closer to Cluster 1/2 in the production of /ɑ:/, whereas Cluster 5 was closer to Cluster 1/2 in the production of /ʊ/.

A high back vowel, /u:/ was excluded from the MANOVAs and discriminant analyses, and therefore, it was not statistically tested whether these vowels contributed to discriminating the clusters. As far as Figure 4.1(b) and Figure 4.3(a), (b), (c) and (d) are concerned, while Cluster 4 was closer to Cluster 2, and Clusters 3 and 5 were closer to Cluster 1, the standardized F2 values suggest that all three JL clusters seemed to be located between Cluster 1 and Cluster 2 with their /u:/ overlapping /ʊ/. The standardized F2 mel values in Table 4.1 and Table 4.2 demonstrate that Cluster 2 produced /u:/ the most front ( $M = 0.42$ ,  $SD = 0.37$ ) and Cluster 1 produced it the most back ( $M = -0.52$ ,  $SD = 0.48$ ). Cluster 3 ( $M = -0.16$ ,  $SD = 0.27$ ), Cluster 4 ( $M = 0.30$ ,  $SD = 0.17$ ) and Cluster 5 ( $M = -0.21$ ,  $SD = 0.23$ ) were between these two clusters. The subjects in Clusters 1 and 2 also placed the tongue higher, which corresponded to the lower standardized F1 mel values ( $M = -1.21$ ,  $SD = 0.19$  for Cluster 1;  $M = -1.11$ ,  $SD = 0.26$  for Cluster 2) than Cluster 3 ( $M = -0.85$ ,  $SD = 0.15$ ), Cluster 4 ( $M = -0.88$ ,  $SD = 0.09$ ) and Cluster 5 ( $M = -0.97$ ,  $SD = 0.10$ ). A comparison of the Figure 4.3(b), (c) and (d) with Figure 4.1(b) and Figure 4.3(a) shows that the three JL clusters positioned the tongue lower than the BN/AN cluster.

#### 4.1.2. Vowel duration

Table 4.7 and Figure 4.6 show the descriptive statistics of the vowel duration for the BN, AN and JL groups, respectively, based on the PVI values of the target long and short vowels and their absolute durations. The PVI values express the durational differences between the long and vowels. A greater durational difference corresponds to a greater PVI value. Three subjects of the JL group were excluded from the analyses due to missing data or

outliers. Consequently, the data of 69 JL subjects were used for the analysis. Figure 4.6(a) visually presents the PVI values of the three groups. Figure 4.6(b) and (c) display the absolute durations of the long and short vowels, respectively. In all bar graphs, the target items are presented on the x-axis and the values are on the y-axis.

Table 4.7

*Descriptive Statistics of Vowel Duration for BN, AN and JL Groups*

	BN (n = 12)				AN (n = 7)				JL (n = 69)			
	<i>M</i>	<i>S</i>	<i>Max</i>	<i>Min</i>	<i>M</i>	<i>S</i>	<i>Max</i>	<i>Min</i>	<i>M</i>	<i>S</i>	<i>Max</i>	<i>Min</i>
PVI values of long and short vowels												
i:-I	20.78	15.60	48.28	2.95	45.68	27.25	71.26	-7.07	22.11	16.82	64.41	-24.35
u:-U	37.90	19.19	63.90	2.79	35.86	27.52	67.77	-0.42	32.98	23.04	71.28	-24.27
ɑ:-æ	25.15	15.72	60.79	4.16	12.05	5.69	18.18	2.72	26.80	16.82	63.75	-14.46
ɑ:-Λ	68.12	16.52	96.91	46.14	58.54	13.98	81.67	37.22	42.72	17.31	79.06	-4.27
Absolute durations												
i:	126.50	19.36	149.00	75.00	120.86	43.16	192.00	74.00	141.49	21.66	195.50	103.50
I	102.50	16.31	128.17	68.33	73.67	19.12	97.67	45.67	113.39	18.73	153.67	73.67
u:	102.06	16.76	135.00	80.33	121.71	37.59	179.00	87.00	145.17	32.52	215.33	75.00
U	69.63	12.58	86.88	44.50	85.33	29.45	127.83	52.44	103.36	22.72	162.75	65.13
ɑ:	137.46	14.30	163.92	115.20	152.46	32.00	189.75	106.63	163.51	25.66	241.00	105.33
æ	107.25	17.23	145.00	81.25	134.55	24.96	162.86	100.21	124.86	20.76	179.38	82.30
Λ	67.82	11.91	91.50	49.33	82.98	17.69	116.33	70.17	106.77	23.29	184.33	62.33

*Note.* The PVI values are provided for the long and short vowel pairs, and the absolute durations are for each target vowel. The absolute durations are expressed in ms.

In the PVI values, the relative durational differences between the long and short vowels, the pattern to be noted in the JL group is that they were similar to the BN group. As shown in Figure 4.6(a), the JL group had similar values to the BN group. However, /ɑ:-Λ/ was exceptional. The durational difference between /ɑ:/ and /Λ/ was much smaller for the JL group ( $M = 42.72$ ,  $SD = 17.31$ ) than the BN and AN group ( $M = 68.12$ ,  $SD = 16.52$  for BN;  $M = 59.54$ ,  $SD = 13.98$  for AN). The other tendency found regarding the durational differences was that the AN group produced a greater difference for /i:-I/ and a smaller

difference for /ɑ:-æ/ than the BN and JL groups. Figure 4.6(c), which illustrates the absolute durations of the short vowels, suggests that this was attributed to a shorter /ɪ/ and a longer /æ/ of the AN group. As regards the absolute durations of the long and short vowels, the JL group produced longer durations for both long vowels and short vowels than the BN and AN groups except for AN's longer /æ/.

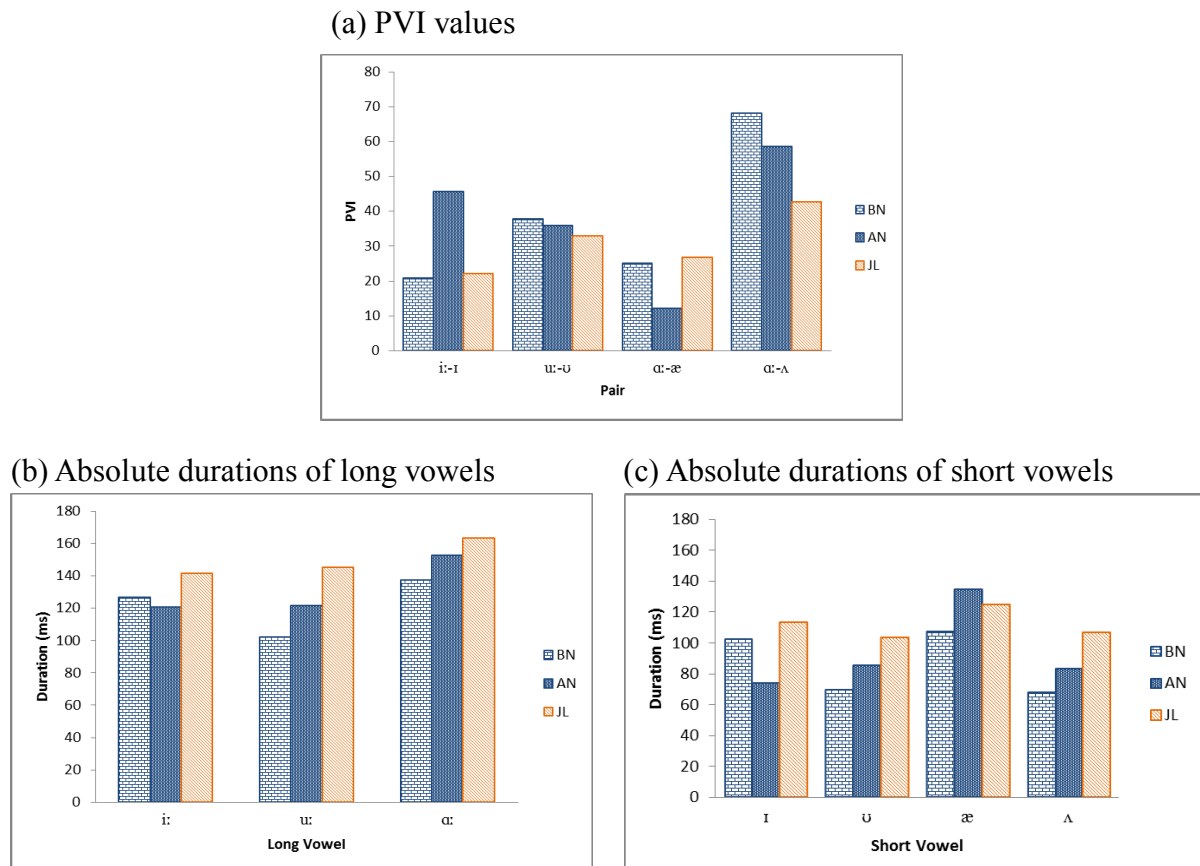


Figure 4.6. Durational values of long and short vowels for BN, AN and JL groups: (a) PVI values in the four pairs /i:-ɪ/, /u:-ʊ/, /ɑ:-ʌ/ and /ɑ:-æ/; (b) absolute durations of long vowels; and (c) absolute durations of short vowels.

In order to form a group based on the profile of the subjects, a cluster analysis was conducted using the z-scores of the three variables based on the mean and standard deviation of the entire sample, the PVI values of /i:-ɪ/, /u:-ʊ/ and /ɑ:-ʌ/. The /ɑ:-æ/ distinction was not used for the cluster analysis to avoid too much influence of /ɑ:/. It is reported that /æ/ varies individually (Cruttenden, 2014), which is the other reason why this pair was not included in

the cluster analysis. Three JL subjects were excluded from the analysis due to missing data and univariate outliers, as noted above. The analysis was therefore performed for 12 BN subjects, 7 AN subjects and 69 JL subjects in total. The dendrogram from this cluster analysis showed that 16 BN/AN subjects were clustered into two clusters at an earlier stage of clustering, regardless of their accent: eight subjects into Cluster 1 and eight subjects into Cluster 2 (see Appendix F for the dendrogram). This separation into two suggests that the pattern of the durational distinction between the long and short vowels varied among the BN and AN subjects. Cluster 1 and Cluster 2 were both comprised of most BN/AN subjects, and were thus regarded as representing native speakers. The cutoff point of clustering was set to generate two more clusters, and the remaining three BN/AN subjects were separately categorized into Clusters 3 and 4. As a result, Cluster 1 consisted of 7 BN subjects, 1 AN subject and 10 JL subjects, Cluster 2 of 3 BN subjects, 5 AN subjects and 14 JL subjects, Cluster 3 of 2 BN subjects and 18 JL subjects, and Cluster 4 of 1 AN subject and 27 JL subjects.

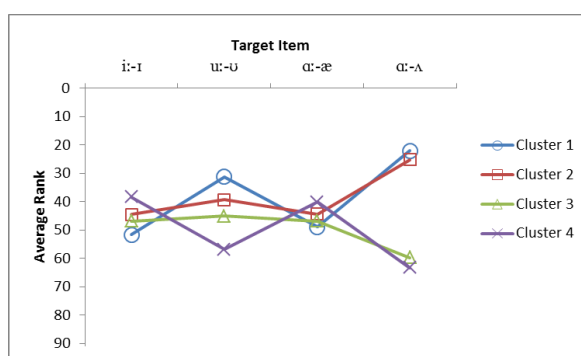
Table 4.8 presents the descriptive statistics of the four clusters, including both PVI values used for the statistical analyses and absolute durations for a reference. A line graph in Figure 4.7 shows the profile of each cluster concerning the four target items. The plot corresponds to the rank averaged across subjects in each cluster, obtained from the z-scores based on the mean and standard deviation for the BN/AN subjects. The subjects who obtained smaller absolute values ranked higher. Figure 4.8(a), (b) and (c) shows the PVI values of each target item, the absolute durations of the long vowels and the absolute durations of the short vowels, using bar graphs, where the clusters are presented on the x-axis and the values on the y-axis. Note that the PVI values refer to the durational difference between the paired vowels, just like the percentage, while the original values are expressed in ms. The higher PVI values are equal to the larger durational difference between the target long vowels and short vowels.

Table 4.8

*Descriptive Statistics of Vowel Duration for Four Clusters*

	Cluster 1 ( <i>n</i> = 18)		Cluster 2 ( <i>n</i> = 22)		Cluster 3 ( <i>n</i> = 20)		Cluster 4 ( <i>n</i> = 28)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
PVI values of long and short vowels								
i:-ɪ	12.63	8.81	45.37	13.56	17.54	17.77	18.52	13.32
u:-ʊ	30.10	15.24	46.01	17.45	55.82	7.60	11.11	14.77
ɑ:-æ	30.53	18.46	29.56	15.67	22.63	18.07	20.82	13.31
ɑ:-ʌ	65.91	10.32	61.58	13.60	34.16	15.37	33.96	10.09
Absolute durations								
i:	129.36	16.66	145.64	26.55	134.78	28.82	139.23	22.37
ɪ	113.74	13.10	91.90	18.23	112.51	21.57	116.09	21.37
u:	124.70	34.93	139.90	31.54	162.15	32.82	125.99	29.39
ʊ	91.79	24.90	89.01	27.57	91.31	18.43	111.72	22.11
ɑ:	161.59	30.52	167.15	27.70	151.17	19.86	156.77	25.85
æ	118.74	24.23	123.43	17.29	122.10	26.83	126.77	19.04
ʌ	82.72	23.40	89.43	21.36	108.40	24.69	112.07	22.88

*Note.* The PVI values are provided for the long and short vowel pairs, and the absolute durations are for each target vowel. The absolute durations are expressed in ms.

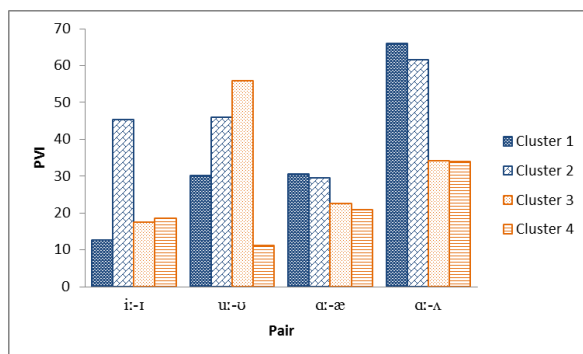


*Figure 4.7.* Profile of each cluster for vowel duration.

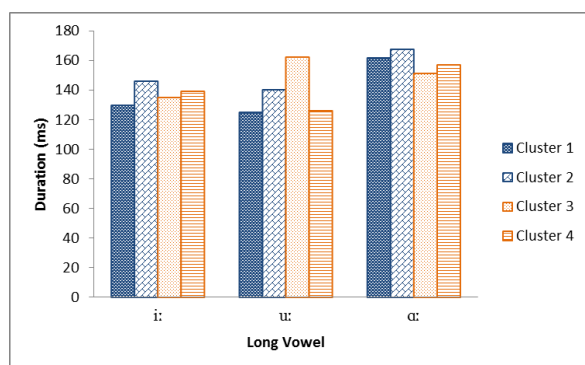
The profile of the four clusters in the line graph in Figure 4.7 shows that they did not differ in the /i:-ɪ/ and /ɑ:-æ/ distinctions. By contrast, the outperformance of Clusters 1 and 2 over Clusters 3 and 4 was found in the /u:-ʊ/ and /ɑ:-ʌ/ distinctions. This suggests that these distinctions were difficult for the subjects in Clusters 3 and 4, which was especially notable in

the /ɑ:-ʌ/ distinction.

(a) PVI values



(b) Absolute durations of long vowels



(c) Absolute durations of short vowels

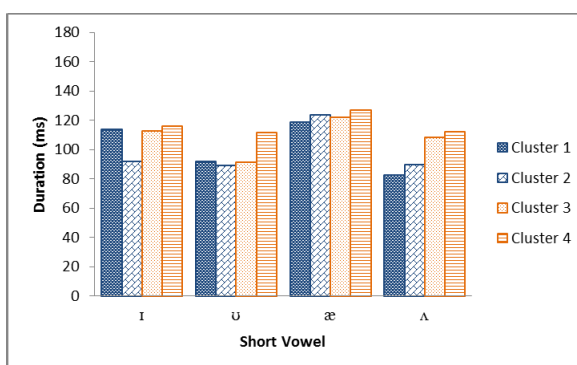


Figure 4.8. Durational values of long and short vowels for four clusters: (a) PVI values in the four pairs /i:-ɪ/, /u:-ʊ/, /ɑ:-ʌ/ and /ɑ:-æ/; (b) absolute durations of long vowels; and (c) absolute durations of short vowels.

The PVI values of Clusters 1 and 2 varied greatly in the /i:-ɪ/ distinction ( $M=12.63$ ,  $SD = 8.81$  for Cluster 1;  $M = 45.37$ ,  $SD = 13.56$  for Cluster 2), as in Figure 4.8(a). The subjects in Cluster 2 tended to produce more differences in the /i:-ɪ/ distinction. The tendency for Clusters 1 and 2 to be different was also seen in the /u:-ʊ/ distinction, although the difference is less noticeable than in the /i:-ɪ/ distinction. Figure 4.8(b) and (c) demonstrates that these differences originated from the longer long vowels and the shorter short vowels of Cluster 2.

On the other hand, the difference between Clusters 1 and 2, or the BN/AN clusters, and Clusters 3 and 4, or the JL clusters, was notable in the /ɑ:-ʌ/ distinction. Figure 4.8(a)

shows that Clusters 1 and 2 produced a greater difference for this pair ( $M = 65.91$ ,  $SD = 10.32$  for Cluster 1;  $M = 61.58$ ,  $SD = 13.60$  for Cluster 2) than Clusters 3 and 4 ( $M = 34.16$ ,  $SD = 15.37$  for Cluster 3;  $M = 33.96$ ,  $SD = 10.09$  for Cluster 4). The differences among the clusters for the absolute durations, displayed in Figure 4.8(b) and (c), were obvious for both long vowels and short vowels, but especially short vowels. As in Figure 4.8(b) and (c), Clusters 3 and 4 differed from Clusters 1 and 2 more in the short vowels than in the long vowels. The subjects in the former clusters were likely to produce short vowels in a shorter form as in /ʊ/ and /ʌ/ in particular. A smaller durational difference between a long vowel and a short vowel was also seen in the /ɑ: -æ/ distinction for Clusters 3 and 4, and the shorter /ɑ:/ seemed to affect this.

A difference between the two JL clusters was also found in the /u: -ʊ/ distinction. Cluster 3 produced a larger difference ( $M = 55.82$ ,  $SD = 7.60$ ) than both Cluster 1 ( $M = 30.10$ ,  $SD = 15.24$ ) and Cluster 2 ( $M = 46.01$ ,  $SD = 17.45$ ), while Cluster 4 ( $M = 11.11$ ,  $SD = 14.77$ ) produced a smaller difference than Clusters 1 and 2. The subjects in Cluster 4 apparently produced a longer /ʊ/ and a shorter /u:/ as in Figure 4.8(b) and (c).

In order to test whether these differences are statistically significant among the clusters, a one-way MANOVA was performed where the dependent variables were the four PVI values and the independent variables were the four clusters. Correlations between the dependent variables are summarized in Appendix G, and suggest that there were nearly zero correlations between /i: -ɪ/ and /u: -ʊ/ and between /i: -ɪ/ and /ɑ: -æ/. Although the statistical power of a MANOVA might be lowered as a result, the estimation was that it would still work well to some extent because the other correlations were not extremely high or low. The largest cluster was more than 1.5 times as large as the smallest cluster, and therefore, the  $\alpha$  level was set at .01. The results of Pillai's trace indicated a statistical difference among the clusters,  $F(12, 249) = 24.02$ ,  $p < .001$ ,  $\eta_p^2 = .54$ .

A post-hoc discriminant analysis carried out subsequently identified three discriminant functions. The first function explained 60.2% of the variance, canonical  $R^2 = .72$ , the second function explained 27.8% of the variance, canonical  $R^2 = .55$  and the third



function explained 12.0% of the variance, canonical  $R^2 = .34$ . In combination, the three discriminant functions differentiated the clusters at a significant level with the Wilk's lambda value of .08,  $\chi^2(12) = 206.47$ ,  $p < .001$ , and the last two functions also discriminated between the clusters significantly with the Wilk's lambda value of .30,  $\chi^2(6) = 100.16$ ,  $p < .001$ . After the first function and second function were removed, the third functions still differentiated between the clusters significantly with the Wilk's lambda value of .66,  $\chi^2(2) = 34.72$ ,  $p < .001$ . The group centroids in Table 4.9 and the discriminant function plot in Figure 4.9 show the clusters that were discriminated from one another by these functions. The first function contributed to separating Clusters 1, 2 and 3 from Cluster 4. Especially, Cluster 2 and Cluster 4 were maximally differentiated by this function. The second function, then, discriminated Clusters 1 and 2 from Cluster 3. The third function was found to discriminate between Cluster 1 and Cluster 2 maximally. However, this function was not interpreted, given that Cluster 1 and Cluster 2 were both regarded as representing native speakers.

Table 4.9  
*Group Centroids for Vowel Duration*

Cluster	Function		
	1	2	3
1	0.08	1.39	-1.04
2	2.17	0.34	0.70
3	0.34	-1.78	-0.55
4	-2.00	0.11	0.51

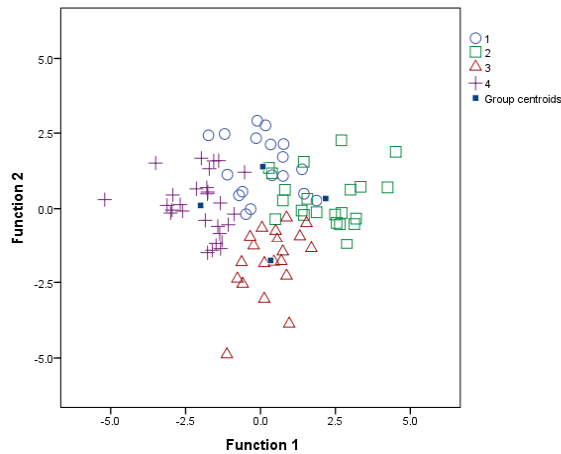


Figure 4.9. Canonical discriminant function plot for vowel duration.

The structural matrix for the PVI values is presented in Table 4.10, which depicts the correlations between the variables and the functions. For the first function, the durational difference between /u:/ and /ʊ/ ( $r = .64$ ), that between /ɑ:/ and /ʌ/ ( $r = .51$ ) and that /i:/ and /ɪ/ ( $r = .42$ ) loaded onto it, which suggests these variables were mainly related to the separation of Clusters 2 and 4. The PVI values in Table 4.8 and Figure 4.8(a) show that Cluster 2 tended to differentiate these long and short vowels with greater durational differences than Cluster 4 for the distinctions in /u:-ʊ/ ( $M = 46.01$ ,  $SD = 17.45$  for Cluster 2;  $M = 11.11$ ,  $SD = 14.77$  for Cluster 4), /ɑ:-ʌ/ ( $M = 61.58$ ,  $SD = 13.60$  for Cluster 2;  $M = 33.96$ ,  $SD = 10.09$  for Cluster 4) and /i:-ɪ/ ( $M = 45.37$ ,  $SD = 13.56$  for Cluster 2;  $M = 18.52$ ,  $SD = 13.32$  for Cluster 4). The absolute durations of /ʊ/ and /ʌ/ demonstrate that the subjects in Cluster 4 produced longer /ʊ/ and /ʌ/ ( $M = 111.72$ ,  $SD = 22.11$  for /ʊ/;  $M = 112.07$ ,  $SD = 22.88$  for /ʌ/) than those in the other clusters, Cluster 1 ( $M = 91.79$ ,  $SD = 24.90$  for /ʊ/;  $M = 82.72$ ,  $SD = 23.40$  for /ʌ/), Cluster 2 ( $M = 89.01$ ,  $SD = 27.57$  for /ʊ/;  $M = 89.43$ ,  $SD = 21.36$  for /ʌ/) and Cluster 3 ( $M = 91.31$ ,  $SD = 18.43$  for /ʊ/;  $M = 108.40$ ,  $SD = 24.69$  for /ʌ/). The first function mainly discriminated between Cluster 2 and Cluster 4, but Cluster 4 also differentiated less between /u:/ and /ʊ/ than Clusters 1 and 3 and less between /ɑ:/ and /ʌ/ than Cluster 1.

Table 4.10

*Structural Matrix for the Correlations between the Variables for Vowel Duration and the Two Discriminant Functions*

Variable	Function	
	1	2
u:-ʊ	<b>.64</b>	<b>-.59</b>
ɑ:-ʌ	<b>.51</b>	<b>.80</b>
ɑ:-æ	.13	.15
i:-ɪ	<b>.42</b>	.06

*Note.* The variables with the absolute value of correlations with the corresponding functions of .33 and above were highlighted in bold.

Similarly, the durational difference between /ɑ:/ and /ʌ/ and that between /u:/ and /ʊ/ loaded on the second function ( $r = .80$  and  $r = -.59$ ). This shows that there was a difference in the production of the /u:-ʊ/ and /ɑ:-ʌ/ distinctions between Cluster 3 and Clusters 1 and 2, mainly between Cluster 1 and Cluster 3. However, unlike the first function, the sign suggests the opposite tendency between these pairs. In other words, as illustrated in Figure 4.8(a), Cluster 3 produced more durational difference for /u:-ʊ/ and fewer differences for /ɑ:-ʌ/ ( $M = 55.82$ ,  $SD = 7.60$  for /u:-ʊ/;  $M = 34.16$ ,  $SD = 15.37$  for /ɑ:-ʌ/) than Cluster 1 ( $M = 30.10$ ,  $SD = 15.24$  for /u:-ʊ/;  $M = 65.91$ ,  $SD = 10.32$  for /ɑ:-ʌ/) and Cluster 2 ( $M = 46.01$ ,  $SD = 17.45$  for /u:-ʊ/;  $M = 61.58$ ,  $SD = 13.60$  for /ɑ:-ʌ/). The degree of discrimination between Cluster 2 and Cluster 3 was smaller for the /u:-ʊ/ distinction than that between Cluster 1 and Cluster 3. The absolute durations for each target vowel in Table 4.8 suggest what underlay these differences, which are presented in Figure 4.8(b) and (c). Cluster 3 produced /u:/ and /ʌ/ longer and /ɑ:/ shorter ( $M = 162.15$ ,  $SD = 32.82$  for /u:/;  $M = 108.40$ ,  $SD = 24.69$  for /ʌ/;  $M = 151.17$ ,  $SD = 19.86$  for /ɑ:/) than Cluster 1 ( $M = 124.70$ ,  $SD = 34.93$  for /u:/;  $M = 82.72$ ,  $SD = 23.40$  for /ʌ/;  $M = 161.59$ ,  $SD = 30.52$  for /ɑ:/) and Cluster 2 ( $M = 139.90$ ,  $SD = 31.54$  for /u:/;  $M = 89.43$ ,  $SD = 21.36$  for /ʌ/;  $M = 167.15$ ,  $SD = 27.70$  for /ɑ:/). The difference between Cluster 1 and 3 was larger for /u:/ and /ʌ/ whereas that between Cluster 2 and Cluster 3 was larger for /ɑ:/.

All these results above imply that two JL clusters, Clusters 3 and 4 were not

discriminated from Cluster 1 by the /i:-ɪ/ distinction and were not discriminated from Cluster 1 nor Cluster 2 by the /ɑ:-æ/ distinction. According to visual inspection of Figure 4.8(a), Cluster 2 produced a larger durational difference between /i:/ and /ɪ/. However, this was not true of Cluster 1. Thus, this distinction did not contribute to discriminating between Cluster 1 and Clusters 3 and 4 statistically. As for the /ɑ:-æ/ distinction, Clusters 3 and 4 produced a smaller difference for this distinction than Clusters 1 and 2 as far as Figure 4.8(a) was concerned, but it did not reach a level that was statistically confirmed.

## 4.2. Consonants

### 4.2.1. Plosives

Table 4.11 and Figure 4.10 present the descriptive statistics of the plosives, showing the absolute VOT durations of /p, t, k/ and /t, k/ in /st, sk/, and the relative VOT differences in /t-st/ and /k-sk/ pairs. The latter variables concern the difference between the aspirated voiceless plosives and unaspirated voiceless plosives. One case of the AN group and five cases of the JL group were excluded from the analyses because of missing data or outlier. The total number of the subjects in the AN group was thus 6 and that in the JL group was 67.

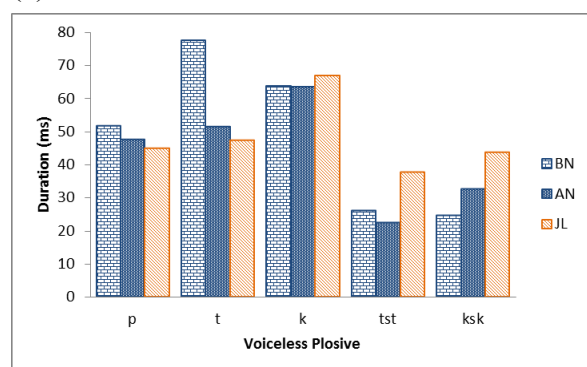
Table 4.11

*Descriptive Statistics of Plosives for BN, AN and JL Groups*

	BN (n = 12)				AN (n = 6)				JL (n = 67)			
	<i>M</i>	<i>SD</i>	<i>Max</i>	<i>Min</i>	<i>M</i>	<i>SD</i>	<i>Max</i>	<i>Min</i>	<i>M</i>	<i>SD</i>	<i>Max</i>	<i>Min</i>
Absolute VOT durations												
p	51.75	17.28	74.00	22.00	47.67	10.98	58.00	33.00	44.90	21.41	95.00	3.00
t	77.50	13.01	99.00	58.00	51.50	13.69	68.00	29.00	47.33	18.10	98.00	22.00
k	63.71	9.26	76.33	49.50	63.42	4.89	70.86	58.67	66.93	14.68	100.67	35.50
st	26.02	5.82	38.25	18.50	22.44	4.84	26.67	14.33	37.56	11.01	67.50	17.25
sk	24.54	4.40	31.00	16.00	32.50	12.17	50.00	19.00	43.61	15.34	108.00	9.00
Relative VOT differences												
t-st	0.35	0.10	0.52	0.19	0.48	0.23	0.88	0.23	0.88	0.35	1.78	0.33
k-sk	0.39	0.07	0.47	0.21	0.51	0.17	0.74	0.30	0.68	0.22	1.33	0.26

*Note.* The absolute VOT durations of /p, t, k/ and /t, k/ in /st, sk/ and the relative VOT differences in /t-st/ and /k-sk/ are expressed in ms and ratio, respectively.

(a) Absolute VOT durations



(b) Relative VOT differences

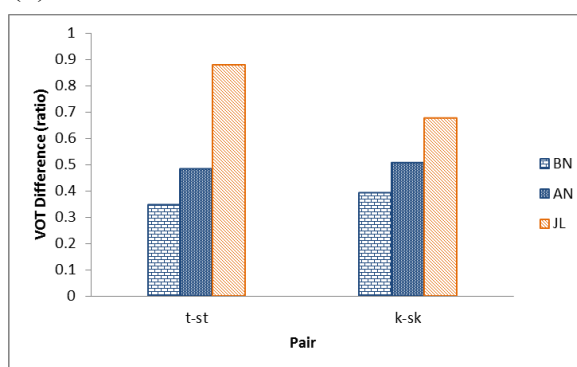


Figure 4.10. VOT for BN, AN and JL groups: (a) absolute VOT durations of /p, t, k/ and /t, k/ in /st, sk/; and (b) relative VOT differences between the aspirated and unaspirated voiceless plosives in the /t-st/ and /k-sk/ pairs.

Figure 4.10(a) and (b) illustrates the absolute durations of /p, t, k/ and /t, k/ in /st, sk/ and the relative VOT differences in /t-st/ and /k-sk/ pairs, respectively. In Figure 4.10(a), the x-axis represents the target items and the y-axis, the duration expressed in ms. Figure 4.10(b) shows the target items on the x-axis and the rate of the difference on the y-axis. In this variable, a lower value corresponds to a more durational difference.

As visual inspection of Figure 4.10(a) suggests that the JL group VOT of /k/ did not seem to differ radically from that of the BN and AN groups on average. However, they produced shorter /p/ and /t/ than the BN and AN groups. Above all, their VOT value for /t/ was much shorter than the BN group' ( $M = 47.33$ ,  $SD = 18.10$  for JL;  $M = 77.50$ ,  $SD = 13.01$  for BN). The JL group also differed markedly from BN and AN groups for the unaspirated plosives, /t/ in /st/ and /k/ in /sk/. The VOT values of both unaspirated plosives were longer for the JL group ( $M = 37.56$ ,  $SD = 11.01$  for /t/ in /st/;  $M = 43.61$ ,  $SD = 15.34$  for /k/ in /sk/) than the BN group ( $M = 26.02$ ,  $SD = 5.84$  for /t/ in /st/;  $M = 24.54$ ,  $SD = 4.40$  for /k/ in /sk/) and AN group ( $M = 22.44$ ,  $SD = 4.84$  for /st/;  $M = 32.50$ ,  $SD = 12.17$  for /k/ in /sk/). As a result, the relative differences within the pair were smaller for the JL subjects as in Figure 4.10(b). This tendency was more remarkable in the /t-st/ pair ( $M = 0.35$ ,  $SD = 0.10$  for BN;  $M = 0.48$ ,  $SD = 0.23$  for AN;  $M = 0.88$ ,  $SD = 0.35$  for JL). The maximum values of the relative VOT difference in the /t-st/ and /k-sk/ pairs were 1.78 and 1.33 for the JL group,

which suggests that the VOTs of the unaspirated plosives were even longer than those of the aspirated plosives.

Subsequently, in order to form groups of JL subjects depending on the individual performance of VOT, a cluster analysis was carried out, using the following variables: the absolute VOT durations of /p/, /t/ and /k/, and the relative VOT differences in the /t-st/ and /k-sk/ pairs. These variables were transformed to the z-scores, based on the mean and standard deviation of the entire sample. The dendrogram yielded from the cluster analysis showed that 14 BN/AN subjects out of 18 were clustered together at an earlier stage of the clustering process. When the cutoff point was selected where this BN/AN cluster was created, five JL clusters were formed (see Appendix H for the dendrogram). However, one of the five JL clusters consisted of only eight subjects, which could lower the statistical power. The cutoff point was therefore selected where these eight subjects were clustered with another JL cluster. As a result, two JL clusters, which were still distant from the BN/AN cluster, were selected. Cluster 1 was made up of 11 BN subjects, 3 AN subjects and 3 JL subjects, Cluster 2 of 1 BN subject, 3 AN subjects and 31 JL subjects, and Cluster 3 of 33 JL subjects. Cluster 1 was defined as representing native speakers.

The descriptive statistics of these clusters are shown in Table 4.12. The absolute durations of VOT are shown for /p/, /t/, /k/, /t/ in /st/ and /k/ in /sk/ and the relative differences between the aspirated and unaspirated voiceless plosives are for the /t-st/ and /k-sk/ pairs in the table. Figure 4.11 illustrates the profile of each cluster for VOT. The plot expresses the rank of each cluster averaged across subjects in each cluster, based on the z-scores of the mean and standard deviation for the BN/AN subjects. When the absolute value was smaller, it ranked higher. Figure 4.12 presents a bar graph of the absolute VOT durations for the target items, where the variables are on the x-axis and the values are on the y-axis. It also shows the a graph of the relative VOT differences between the aspirated and unaspirated voiceless plosives. The rates of the differences and the target items are shown on the y-axis and x-axis, respectively.

Table 4.12

*Descriptive Statistics of Plosives for Three Clusters*

	Cluster 1 ( <i>n</i> = 17)		Cluster 2 ( <i>n</i> = 35)		Cluster 3 ( <i>n</i> = 33)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Absolute VOT durations						
p	55.88	17.24	31.09	11.34	55.09	19.06
t	77.47	13.53	36.83	10.80	53.79	16.46
k	65.74	9.52	56.26	9.69	76.19	11.23
st	26.54	6.79	31.27	9.10	42.35	11.03
sk	26.76	7.51	40.16	13.99	46.12	13.33
Relative VOT differences						
t-st	0.35	0.09	0.94	0.41	0.83	0.26
k-sk	0.41	0.11	0.74	0.25	0.61	0.18

*Note.* The absolute VOT durations of /p, t, k/ and /t, k/ in /st, sk/ and the relative VOT differences in /t-st/ and /k-sk/ are expressed in ms and ratio, respectively.

The profile in Figure 4.11 shows that the three clusters differ less in the absolute VOT durations of /p/ and /k/. Cluster 1 and Cluster 3 overlapped for /p/, while Cluster 1 and Cluster 2 nearly overlapped for /k/. Cluster 1 performed better for the rest of the variables. The absolute VOT duration of /t/ was longer and the relative VOT differences in the /t-st/ and /k-sk/ pairs were smaller in Clusters 2 and 3. In a comparison of Cluster 2 and Cluster 3, the two JL clusters, Cluster 2 performed better in producing the VOT of /k/ and Cluster 3 was better in the remaining items.

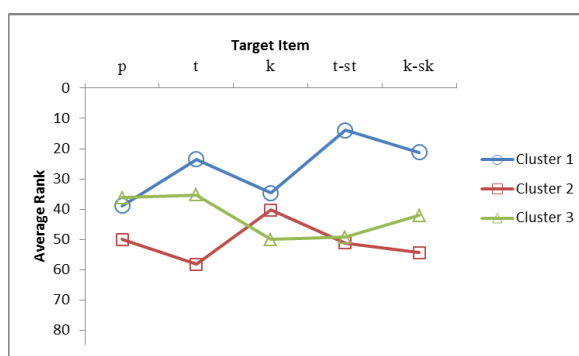
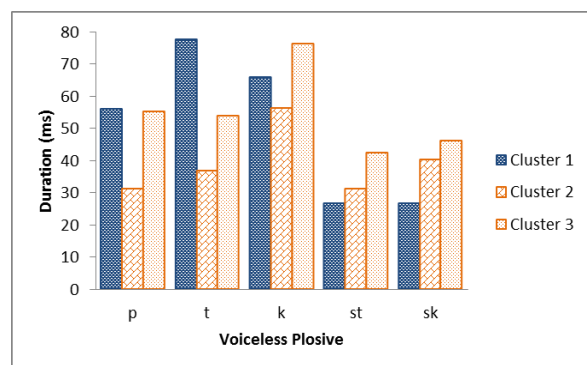


Figure 4.11. Profile of each cluster for plosives.

(a) Absolute VOT durations



(b) Relative VOT differences

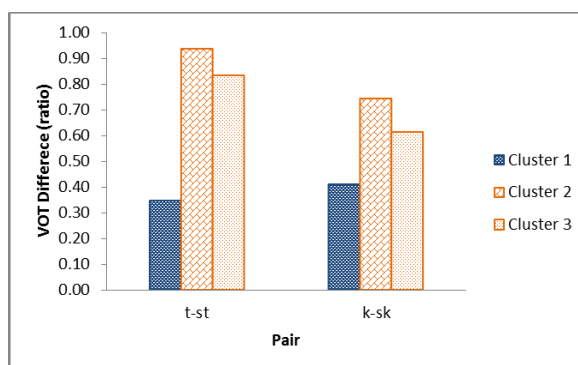


Figure 4.12. VOT for three clusters: (a) absolute VOT durations of /p, t, k/ and /t, k/ in /st, sk/; and (b) relative VOT differences between the aspirated and unaspirated voiceless plosives in the /t-st/ and /k-sk/ pairs.

Figure 4.12(a) and (b) illustrates in more detail how far these two JL clusters deviated from the BN/AN cluster. Figure 4.12(a) shows that the subjects in Cluster 2 produced the shortest VOTs for the three voiceless plosives ( $M = 31.09$ ,  $SD = 11.34$  for /p/;  $M = 36.83$ ,  $SD = 10.80$  for /t/;  $M = 56.26$ ,  $SD = 9.69$  for /k/). It is notable that this affected the poor distinction between the aspirated voiceless plosives and the unaspirated voiceless plosives. As in Figure 4.12(a) and (b), although the absolute VOT durations in /t/ in /st/ and /k/ in /sk/ were longer for Cluster 3, the other JL cluster ( $M = 42.35$ ,  $SD = 11.03$  for /t/ in /st/;  $M = 46.12$ ,  $SD = 13.33$  for /k/ in /sk/), than for Cluster 2 ( $M = 31.27$ ,  $SD = 9.10$  for /t/ in /st/;  $M = 40.16$ ,  $SD = 13.99$  for /k/ in /sk/), the relative VOT differences were greater for Cluster 3 ( $M = 0.83$ ,  $SD = 0.26$  for /t-st/;  $M = 0.61$ ,  $SD = 0.18$  for /k-sk/) than for Cluster 2 ( $M = 0.94$ ,  $SD = 0.41$  for /t-st/;  $M = 0.74$ ,  $SD = 0.25$  for /k-sk/). This suggests that the subjects in Cluster 2 produced much shorter VOT durations of /p, t, k/ considering their absolute VOT durations of /t/ in /st/ and /k/ in /sk/.

Noteworthy differences between Cluster 3 and Cluster 1 were mostly found in the relative VOT differences in the /t-st/ and /k-sk/ pairs. The differences between the aspirated plosives and the unaspirated plosives were greater for Cluster 1 ( $M = 0.35$ ,  $SD = 0.09$  for /t-st/;  $M = 0.41$ ,  $SD = 0.11$  for /k-sk/) than for Cluster 3. It is also notable that Cluster 3 ranked the lowest for the absolute VOT duration of /k/ due to the much longer VOT ( $M =$



65.74,  $SD = 9.52$  for Cluster 1;  $M = 76.19$ ,  $SD = 11.23$  for Cluster 3).

A one-way MANOVA was conducted so that these observable differences were statistically tested. The five variables served as dependent variables, the absolute VOT durations of /p, t, k/ and the relative VOT differences between the aspirated voiceless plosives and the unaspirated voiceless plosives in the /t-st/ and /k-sk/ pairs. The z-scores of these values were used, based on the mean and standard deviation of the entire sample. The independent variables were the three clusters. As in Appendix I, all variables were correlated moderately, where a MANOVA was estimated to work well. The sample size of the largest cluster, Cluster 2, was more than 1.5 times as large as of the smallest cluster, Cluster 1. Subsequent statistical tests were thus conducted at the  $\alpha$  level of .01. Pillai's trace of the MANOVA revealed that the clusters differed significantly in the production of the plosives,  $F(10, 158) = 20.11$ ,  $p < .001$ ,  $\eta_p^2 = .56$ .

A discriminant analysis was carried out as a post-hoc test following the MANOVA to detect which variables contributed to discriminating between the clusters. Two discriminant functions were found, where the first discriminant function explained 72.2% of the variance, canonical  $R^2 = .68$ , and the second function explained 27.8% of the variance, canonical  $R^2 = .44$ . When the two functions were combined, they distinguished between the clusters at a significant level with the Wilk's lambda value of .18,  $\chi^2(10) = 137.03$ ,  $p < .001$ . The second function, furthermore, was able to discriminate between the clusters significantly, when not combined with the first function, with the Wilk's lambda value of .56,  $\chi^2(4) = 47.06$ ,  $p < .001$ . The group centroids in Table 4.13 and the canonical discriminant function plot in Figure 4.13 show that the first function differentiated Clusters 1 and 3 from Cluster 2, where Cluster 1 and Cluster 2 were maximally discriminated. The second function, on the other hand, distinguished Clusters 1 and 2 from Cluster 3. Cluster 2 and Cluster 3 had been already discriminated by the first function. Thus, this function primarily differentiated Cluster 1 from Cluster 3.

Table 4.13

*Group Centroids for Plosives*

Cluster	Function	
	1	2
1	1.95	-1.28
2	-1.60	-0.34
3	0.70	1.02

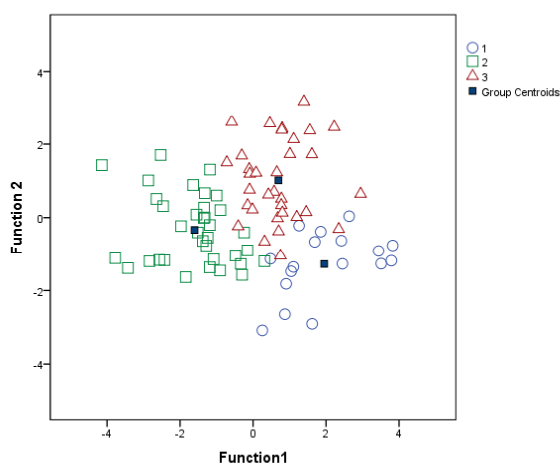


Figure 4.13. Canonical discriminant function plot for plosives.

The structural matrix of the correlations between the variables and the two functions is presented in Table 4.14. According to this matrix, the absolute VOT duration of /t/ loaded most highly on the first function ( $r = .74$ ). The other variables that loaded on it included the absolute VOT duration of /p/ ( $r = .51$ ), the absolute VOT duration of /k/ ( $r = .44$ ), the relative VOT difference in /t-st/ ( $r = -.41$ ), and the relative VOT difference in /k-sk/ ( $r = -.41$ ). These results reveal the difference between Cluster 1 and Cluster 2, and also between Cluster 3 and Cluster 2 to a lesser degree. As the positive sign of the absolute VOT duration and the negative sign of the relative VOT differences suggest, Cluster 2 tended to produce shorter VOTs for the three voiceless plosives ( $M = 31.09$ ,  $SD = 11.34$  for /p/;  $M = 36.83$ ,  $SD = 10.80$  for /t/;  $M = 56.26$ ,  $SD = 9.69$  for /k/) than Cluster 1 ( $M = 55.88$ ,  $SD = 17.24$  for /p/;  $M = 77.47$ ,  $SD = 13.53$  for /t/;  $M = 65.74$ ,  $SD = 9.52$  for /k/) and Cluster 3 ( $M = 55.09$ ,  $SD =$

19.06 for /p/;  $M = 53.79$ ,  $SD = 16.46$  for /t/;  $M = 76.19$ ,  $SD = 11.23$  for /k/), leading to smaller difference between the aspirated voiceless plosives and unaspirated voiceless plosives for Cluster 2 ( $M = 0.94$ ,  $SD = 0.41$  for /t-st/;  $M = 0.74$ ,  $SD = 0.25$  for /k-sk/). This is clearly illustrated in Figure 4.12(a) and (b).

Table 4.14

*Structural Matrix for the Correlations between the Variables for Plosives and the Two Discriminant functions*

Variable	Function	
	1	2
t	<b>.74</b>	<b>-.35</b>
p	<b>.51</b>	.26
k-sk	<b>-.41</b>	.22
k	<b>.44</b>	<b>.69</b>
t-st	<b>-.41</b>	<b>.45</b>

*Note.* The variables with the absolute value of correlations with the corresponding functions of .33 and above were highlighted in bold.

The second function was identified by three variables, the absolute VOT duration of /k/ ( $r = .69$ ), the relative VOT difference in /t-st/ ( $r = .45$ ), and the absolute VOT duration of /t/ ( $r = -.35$ ). This function contributed to discriminating Cluster 3 from Cluster 1. As the positive sign of the absolute duration of /k/ represents, Cluster 3 produced a longer VOT for /k/ ( $M = 76.19$ ,  $SD = 11.23$ ) than Cluster 1 ( $M = 65.74$ ,  $SD = 9.52$ ). In contrast, the absolute VOT duration of /t/, as shown in the negative sign, was shorter for Cluster 3 ( $M = 53.79$ ,  $SD = 16.46$ ) than Cluster 1 ( $M = 77.47$ ,  $SD = 13.53$ ). Cluster 3 produced a smaller difference ( $M = .83$ ,  $SD = .26$ ) than Cluster 1 ( $M = .35$ ,  $SD = .09$ ) for the relative VOT difference in /t-st/.

One of the JL clusters, Cluster 2, was thus discriminated from the BN/AN cluster, Cluster 1, by all variables. In contrast, the other JL cluster, Cluster 3, was not discriminated from Cluster 1 in terms of the absolute duration of /p/ or the relative VOT difference in /k-sk/. Two JL clusters were differentiated by all variables as in the first function, where the subjects in Cluster 3 performed better than those in Cluster 2 in that the former produced

longer VOTs for the aspirated voiceless plosives and larger VOT differences between the aspirated and unaspirated plosives than the latter.

#### 4.2.2. Fricatives

Table 4.15 and Figure 4.14 present the descriptive statistics of BN and JL groups based on the raw data of center of gravity (COG), standard deviation (SD), skewness and kurtosis of the two target fricatives. The AN data was not analyzed for this element of pronunciation because of the difference in the sampling rate in the digital recording. The raw values of the four spectral moments are shown in Figure 4.14(a), (b), (c) and (d), where the x-axis shows the target fricatives and the y-axis, the values. The values of COG and SD are expressed in Hz.

Table 4.15

*Descriptive Statistics of Fricatives for BN and JL Groups*

	BN (n = 12)		JL (n = 72)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
θ COG	5598.91	1137.58	8426.84	904.70
s COG	6476.91	869.80	8679.37	714.83
θ SD	5830.98	481.60	4331.24	484.37
s SD	4571.68	464.37	4175.13	464.78
θ skewness	0.92	0.29	0.45	0.24
s skewness	0.80	0.20	0.40	0.22
θ kurtosis	0.12	0.67	0.27	0.54
s kurtosis	0.80	0.80	0.30	0.52

*Note.* The values of COG and SD are expressed in Hz.

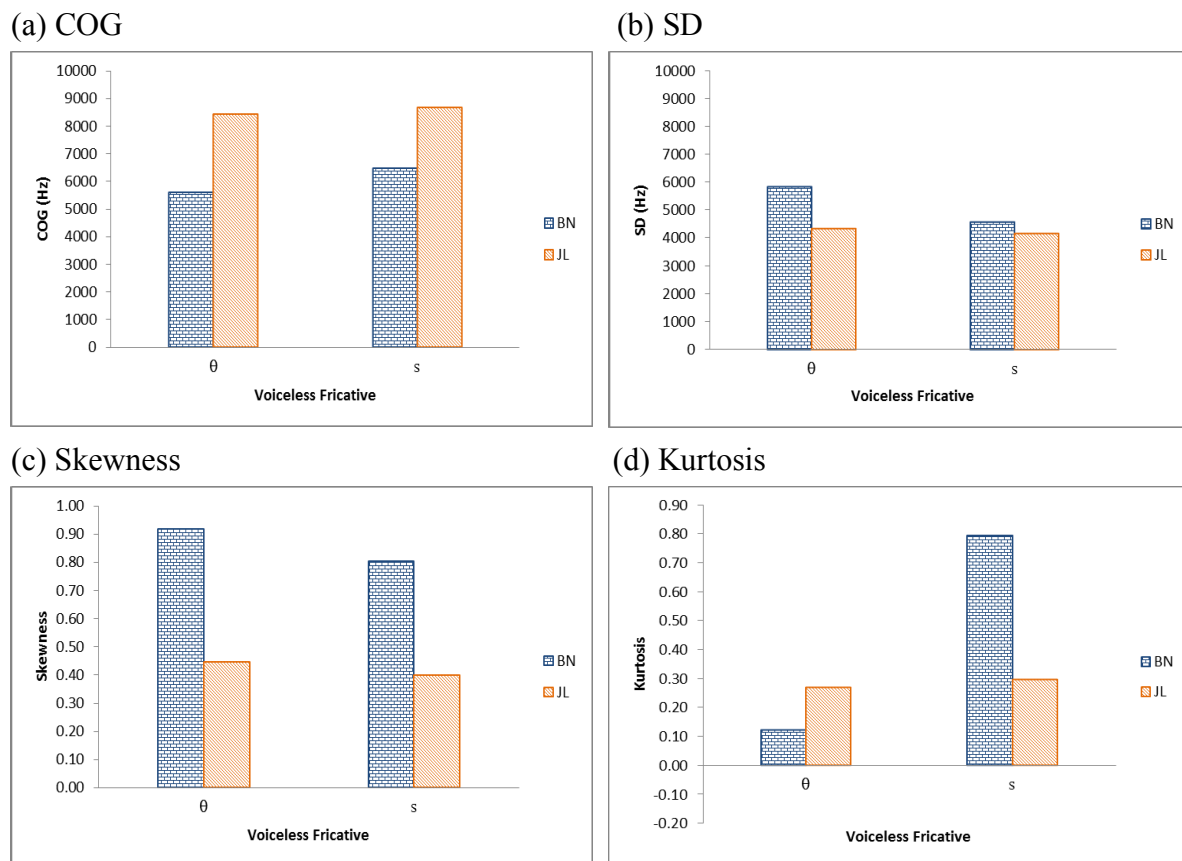


Figure 4.14. Four spectral moments for BN and JL groups: (a) COG; (b) SD; (c) kurtosis; and (d) skewness. The values of COG and SD are expressed in Hz.

A visual inspection of the bar graphs in Figure 4.14 suggests that there are two things to be noted. One is that the BN subjects clearly discriminated the two fricatives with respect to all four variables, while the JL subjects showed less clear differences between these two fricatives. These variables are not directly comparable because they differ in the scale, but this tendency holds true of all four variables. The other thing to note is that there were differences in the values between the two groups concerning all four variables. As in Figure 4.14(a), (b), (c) and (d), the BN subjects produced lower COG values for both /θ/ and /s/, a higher SD value for /θ/, higher skewness values for both /θ/ and /s/, a lower kurtosis value for /θ/ and a higher kurtosis value for /s/ than the JL subjects. Only the SD value for /s/ seemed similar between the BN group and the JL group. All these differences imply the profoundly different production of both fricatives between the two groups.

A cluster analysis was performed to profile the JL subjects, using the variables, COG

of /θ/ and /s/, SD of /θ/ and /s/, skewness of /θ/ and /s/ and kurtosis of /θ/ and /s/. These variables were transformed to the z-scores using the mean and standard deviation for the entire sample. According to the dendrogram output as the result of the cluster analysis, the BN subjects were all clustered together at an earlier stage of the clustering process (see Appendix J for the dendrogram). The cutoff point was therefore selected when they were grouped into one cluster, Cluster 1, which generated four clusters. Cluster 1 was comprised of 12 BN subjects and 3 JL subjects, being regarded as representing native speakers. Clusters 2, 3 and 4 were made up of 18 JL subjects, 27 JL subjects and 24 JL subjects, respectively.

Table 4.16 presents the descriptive statistics regarding the voiceless fricatives /θ/ and /s/ for the clusters created. A line graph in Figure 4.15 demonstrates the profile of each cluster, where the average rank across subjects is plotted. Rank was assigned based on the absolute values of z-scores, which were calculated from the mean and standard deviation for the BN subjects. A smaller value meant the subjects ranked higher. Figure 4.16(a), (b), (c) and (d) also presents the values of COG, SD, skewness and kurtosis for the clusters, respectively. In these bar graphs, the x-axis showed the clusters, and the y-axis, the measured values for each variable.

Table 4.16

*Descriptive Statistics of Fricatives for Four Clusters*

	Cluster 1		Cluster 2		Cluster 3		Cluster 4	
	(n = 15)		(n = 18)		(n = 27)		(n = 24)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
θ COG	5839.34	1125.60	8090.15	521.18	8149.21	828.68	9194.92	639.17
s COG	6623.84	833.21	8370.26	445.59	8485.42	575.88	9312.87	482.27
θ SD	5595.12	652.08	4662.12	444.73	4026.43	391.01	4385.95	437.25
s SD	4561.08	428.02	4412.65	360.78	3783.14	173.38	4395.04	488.03
θ skewness	0.91	0.26	0.48	0.14	0.54	0.24	0.25	0.11
s skewness	0.79	0.18	0.48	0.08	0.48	0.21	0.21	0.15
θ kurtosis	0.22	0.63	-0.12	0.35	0.76	0.40	-0.03	0.36
s kurtosis	0.71	0.73	0.05	0.34	0.82	0.28	-0.11	0.36

*Note.* The values of COG and SD are expressed in Hz.

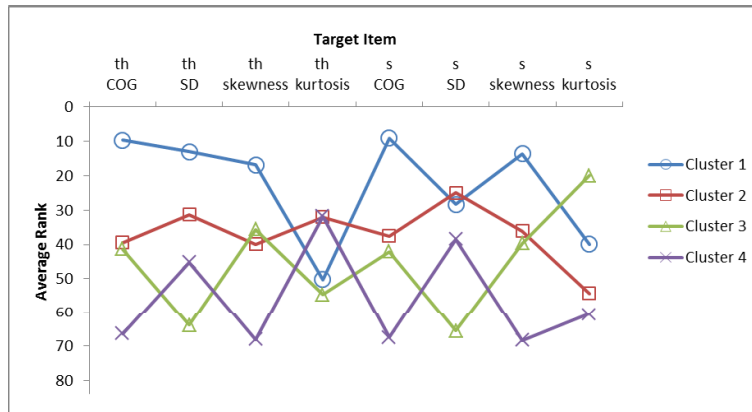
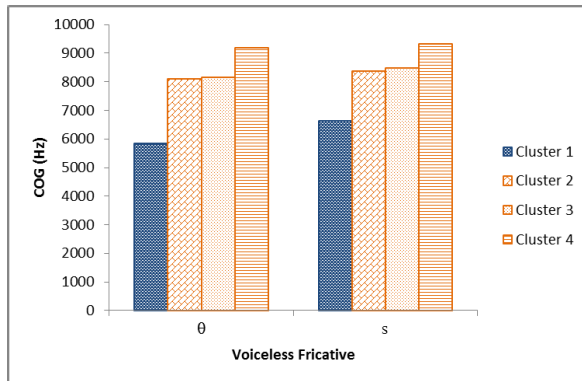
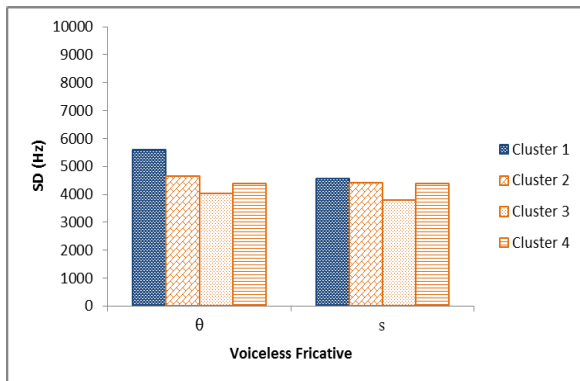


Figure 4.15. Profile of each cluster for fricatives. The alphabet in the label, *th*, demotes /θ/.

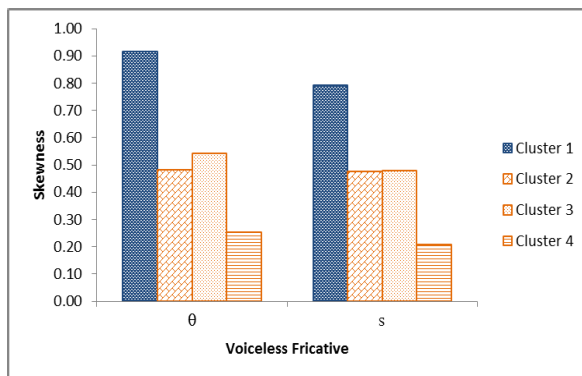
(a) COG



(b) SD



(c) Skewness



(d) Kurtosis

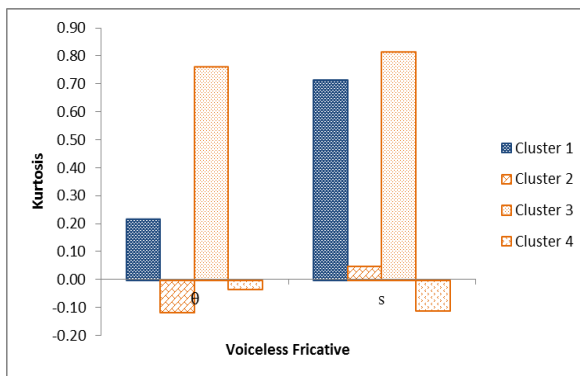


Figure 4.16. Four spectral moments for four clusters: (a) COG; (b) SD; (c) kurtosis; and (d) skewness. The values of COG and SD are expressed in Hz.

The line graph in Figure 4.15 shows that the difference was smaller in kurtosis of /θ/ and /s/ between the JL clusters, Clusters 2, 3 and 4, and the BN cluster, Cluster 1. This was also true of the SD of /s/, but Cluster 3 performed clearly more poorly than the BN cluster and the other two JL clusters. Cluster 1 ranked highest for the other items, which suggests the

better performance of the subjects in this cluster for them. Although Cluster 2 followed Cluster 1 in these variables, one exception was that Cluster 3 was closer to the BN cluster for the skewness of /θ/. Cluster 4 ranked lowest for COG and skewness of /θ/ and COG, skewness and kurtosis of /s/. Overall, the subjects in Cluster 4 performed more poorly than those in the other two JL clusters.

One of the overall patterns found in the JL clusters was that the differences between /θ/ and /s/ were smaller, as in Figure 4.16(a), (b), (c) and (d). The BN/AN cluster, Cluster 1, differentiated between the two voiceless fricatives clearly; higher COG and kurtosis values and lower SD and skewness values for /s/ than for /θ/. A pattern of this sort was less clear in the JL clusters. This tendency for the values of /θ/ and /s/ to be closer implies that the subjects in the JL clusters failed to discriminate between these target voiceless fricatives.

As noted above, of the three JL clusters, Cluster 4 seemed to deviate most from Cluster 1, especially regarding higher COG values for both /θ/ and /s/ ( $M = 9194.92$ ,  $SD = 639.17$  for /θ/;  $M = 9312.87$ ,  $SD = 482.27$  for /s/) and lower skewness values for both /θ/ and /s/ ( $M = 0.25$ ,  $SD = 0.11$  for /θ/;  $M = 0.21$ ,  $SD = 0.15$  for /s/) as in Figure 4.16(a) and (c), respectively. Cluster 3 deviated second-most from Cluster 1, with the lowest values of SD for both /θ/ and /s/ ( $M = 4026.43$ ,  $SD = 391.01$  for /θ/;  $M = 3783.14$ ,  $SD = 173.38$  for /s/), as illustrated in Figure 4.16(b). The differences in kurtosis among the JL clusters were slightly more limited compared to the other variables according to Figure 4.15, but Cluster 3 deviated the most from Cluster 1 for /θ/ ( $M = 0.76$ ,  $SD = 0.40$  for Cluster 3;  $M = 0.22$ ,  $SD = 0.63$  for Cluster 1), whereas Cluster 4 deviated the most for /s/ ( $M = -0.11$ ,  $SD = 0.36$  for Cluster 4;  $M = 0.71$ ,  $SD = 0.73$  for Cluster 1). Overall, the subjects in Cluster 2 tended to be closer to those in Cluster 1, but their productions were obviously closer to Clusters 3 and 4 for all variables than to Cluster 1.

A two-way mixed-design MANOVA was conducted with the four clusters as the between-subjects factor and the two voiceless fricatives as the within-subjects factors, where the dependent variables were COG, SD, skewness and kurtosis. Correlations among the dependent variables are provided in Appendix K. Very high correlations were found between



skewness of /θ/ and COG of /θ/ and between COG of /θ/ and COG of /s/, and nearly zero correlation was found between SD of /s/ and skewness of /s/, between kurtosis of /s/ and SD of /θ/ and skewness of /θ/ and SD of /s/. However, because the other correlations were moderate overall, it was expected that a MANOVA would work rather well. Due to an unbalanced sample size, the  $\alpha$  level was set at .01. Using Pillai's trace, there was a significant interaction effect between the fricatives and the clusters,  $F(12, 237) = 3.89$   $p < .001$ ,  $\eta_p^2 = .17$ . This suggests that the performance of the target fricatives differed depending on the variables among the clusters. A simple main effect was therefore calculated to resolve these interactions. As for the within-subjects factor, or the target fricatives, the results revealed that a significant difference was found for COG of Cluster 1 ( $p < .001$ ) and Cluster 3 ( $p = .003$ ), SD of Cluster 1 ( $p < .001$ ), skewness of Cluster 1 ( $p = .007$ ), and kurtosis of Cluster 1 ( $p = .001$ ). They suggest that the subjects in Cluster 1 significantly differentiated between the target fricatives with all four items, COG, SD, skewness and kurtosis. In contrast, the subjects in Cluster 3 discriminated between them only with COG, whereas those in Clusters 2 and 4 did not. This generally conformed to the overall pattern seen in the bar graphs in Figure 4.16, although the results unexpectedly showed that the subjects in Cluster 3 differentiated the two target fricatives by COG.

A discriminant analysis was performed on the between-subjects factor, the four clusters, in order to examine where a significant difference could be yielded. Three discriminant functions were detected. The first function explained 73.4% of the variance, canonical  $R^2 = .83$ , the second function explained 22.5% of the variance, canonical  $R^2 = .60$ , and the third function explained 4.2% of the variance, canonical  $R^2 = .22$ . When the three functions were combined, they significantly distinguished between the clusters with the Wilk's lambda value of .05,  $\chi^2(24) = 225.00$ ,  $p < .001$ . When the first function was removed, the remaining two functions also discriminated between the clusters significantly with the Wilk's lambda value of .32,  $\chi^2(14) = 88.89$ ,  $p < .001$ . Without these first two functions, the third function alone was able to discriminate the clusters at a significant level of .01 with the Wilk's lambda value of .78,  $\chi^2(6) = 18.73$ ,  $p = .005$ . Table 4.17 and Figure 4.17 each display

the group centroids and the discriminant function plot. They demonstrated that the first function discriminated Cluster 1 from Clusters 2, 3 and 4, where Cluster 1 and Cluster 4 were maximally differentiated. The second function served to discriminate Clusters 2 and 4 from Cluster 3. The third function contributed to discriminating Cluster 2 from Cluster 4.

Table 4.17

*Group Centroids for Fricatives*

Cluster	Function		
	1	2	3
1	4.34	-0.31	0.35
2	-0.23	-1.30	-0.81
3	-0.44	1.67	-0.17
4	-2.05	-0.71	0.57

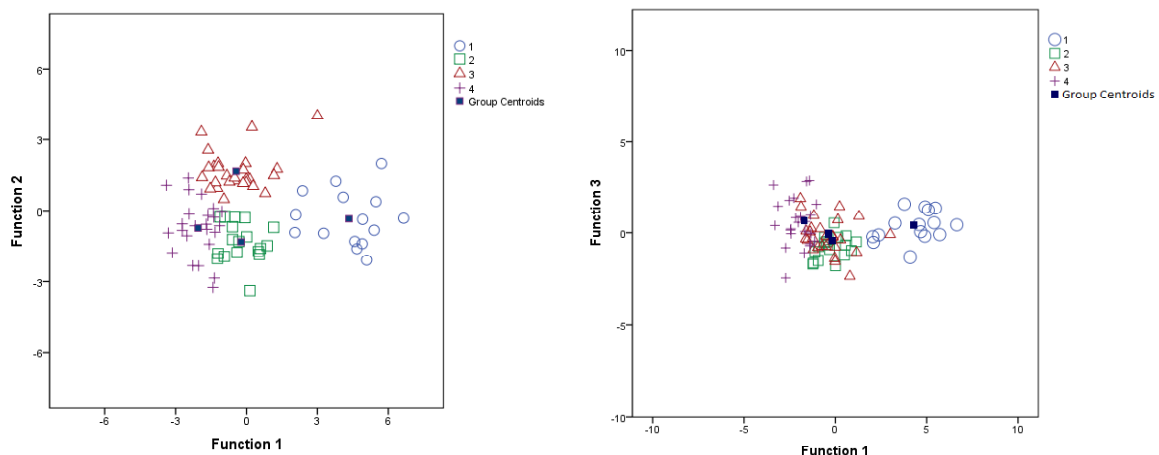


Figure 4.17. Canonical discriminant function plot for fricatives.

The structural matrix of the correlations between the eight dependent variables and the discriminant functions in Table 4.18 shows that the following variables loaded on the first function: COG of /s/ and /θ/ ( $r = -.71$  and  $r = -.66$ ), skewness of /s/ and /θ/ ( $r = .52$  and  $r = .51$ ) and SD of /θ/ ( $r = .45$ ). These variables contributed to discriminating Clusters 2, 3 and 4 from Cluster 1. Firstly, the COG values of both /θ/ and /s/ were lower for the subjects in Cluster 1 ( $M = 5839.34$ ,  $SD = 1125.60$  for /θ/;  $M = 6623.84$ ,  $SD = 833.21$  for /s/) than those

in Cluster 4 ( $M = 9194.92$ ,  $SD = 639.17$  for /θ/;  $M = 9312.87$ ,  $SD = 482.27$  for /s/), Cluster 2 ( $M = 8090.15$ ,  $SD = 521.18$  for /θ/;  $M = 8370.26$ ,  $SD = 445.59$  for /s/) and Cluster 3 ( $M = 8149.21$ ,  $SD = 828.68$  for /θ/;  $M = 8485.42$ ,  $SD = 575.88$  for /s/). Figure 4.16(a) also clearly supports all these, according to visual inspection. COG reflects where energy concentrates on average. Accordingly, the lower values for Cluster 1 suggest that the subjects in this cluster distributed energy at a lower region of frequency in producing these consonants (Jongman et al., 2000).

Table 4.18

*Structural Matrix for the Correlations between the Variables for Fricatives and the Three Discriminant Functions*

Variable	Function		
	1	2	3
/s/ COG	<b>-.71</b>	-.04	.25
/θ/ COG	<b>-.66</b>	-.06	.17
/θ/ skewnwss	<b>.51</b>	.19	-.26
/θ/ kurtosis	.04	<b>.72</b>	-.11
/s/ kurtosis	.25	<b>.67</b>	-.22
/s/ SD	.13	<b>-.65</b>	<b>.41</b>
/θ/ SD	<b>.45</b>	<b>-.51</b>	<b>.38</b>
/s/ skewness	<b>.52</b>	.16	<b>-.55</b>

*Note.* The variables with the absolute value of correlations with the corresponding functions of .33 and above were highlighted in bold.

Secondly, the opposite pattern from COG was found for the skewness of /θ/ and /s/: higher values for Cluster 1 ( $M = 0.91$ ,  $SD = 0.26$  for /θ/;  $M = 0.79$ ,  $SD = 0.18$  for /s/) than Cluster 4 ( $M = 0.25$ ,  $SD = 0.11$  for /θ/;  $M = 0.21$ ,  $SD = 0.15$  for /s/), Cluster 2 ( $M = 0.48$ ,  $SD = 0.14$  for /θ/;  $M = 0.48$ ,  $SD = 0.08$  for /s/) and Cluster 3 ( $M = 0.54$ ,  $SD = 0.24$  for /θ/;  $M = 0.48$ ,  $SD = 0.21$  for /s/). The differences are depicted in Figure 4.16(c). Positive skewness suggests that more energy concentrated in lower frequency regions (Jongman et al., 2000), and this tendency was more notable in Cluster 1.

Finally, the difference of Clusters 2, 3 and 4 from Cluster 1 was also found in the

higher SD value of /θ/ than in Clusters 2, 3 and 4, which was clear when /θ/ and /s/ were compared in Figure 4.16(b). The values shown in Table 4.16 also demonstrate the difference in this variable between Cluster 1 ( $M = 5595.12$ ,  $SD = 652.08$ ) and the three JL clusters, Cluster 4 ( $M = 4385.95$ ,  $SD = 437.25$ ), Cluster 2 ( $M = 4662.12$ ,  $SD = 444.73$ ) and Cluster 3 ( $M = 4026.43$ ,  $SD = 391.01$ ). Because a higher value of SD involves the distribution of the energy in more frequency regions (Jongman et al., 2000), these results show that this tendency was more notable in the subjects in Cluster 1 than those in Clusters 2, 3 and 4 in their production of /θ/.

The kurtosis of /θ/ loaded most highly on the second function ( $r = .72$ ), followed by kurtosis of /s/ ( $r = .67$ ) and SD of /s/ and /θ/ ( $r = -.65$  and  $r = -.51$ ). This function contributed to discriminating Cluster 3 from Clusters 2 and 4. Cluster 3 produced higher kurtosis values for both /θ/ and /s/ ( $M = 0.76$ ,  $SD = 0.40$  for /θ/;  $M = 0.82$ ,  $SD = 0.28$  for /s/) than Cluster 2 ( $M = -0.12$ ,  $SD = 0.35$  for /θ/;  $M = 0.05$ ,  $SD = 0.34$  for /s/) and Cluster 4 ( $M = -0.03$ ,  $SD = 0.36$  for /θ/;  $M = -0.11$ ,  $SD = 0.36$  for /s/). Cluster 3 produced lower SD values for both /θ/ and /s/ ( $M = 4026.43$ ,  $SD = 391.91$  for /θ/;  $M = 3783.14$ ,  $SD = 173.38$  for /s/) than Clusters 2 ( $M = 4662.12$ ,  $SD = 444.73$  for /θ/;  $M = 4412.65$ ,  $SD = 360.78$  for /s/) and Cluster 4 ( $M = 4385.95$ ,  $SD = 437.25$  for /θ/;  $M = 4395.04$ ,  $SD = 488.03$  for /s/). This suggests that the energy of Cluster 3 tended to spread more for both fricatives and the spectral peak of Clusters 2 and 4 to be sharper (Jongman et al., 2000). These differences suggest the presence of learning in the JL subject; however, the cluster that performed closer to Cluster 1, the BN cluster, depended on the variables. As illustrated in Figure 4.16(b) and (d), Cluster 3 was closer to Cluster 1 for the kurtosis of /s/, whereas Clusters 2 and 4 were closer to Cluster 1 for SD of both /s/ and /θ/ and kurtosis of /θ/.

The third function explained the smallest portion of the discrimination among the clusters, and the skewness of /s/ ( $r = -.55$ ) and SD of /θ/ and /s/ ( $r = .41$  and  $r = .38$ ) were found to load on this function. This means that Cluster 2 and Cluster 4 were distinguished by these variables. However, this function only accounted for the 4.2% of the variance. Additionally, although the positive sign of SD and the negative sign of skewness suggest the

different direction of the performance, the mean values of each cluster in Table 4.16 show that Cluster 2 had higher values for all three variables. The only variable interpreted was thus the skewness of /s/, which most strongly identified the third function. Cluster 2 achieved a higher skewness value for /s/ ( $M = .48$ ,  $SD = .08$ ) than Cluster 4 ( $M = .21$ ,  $SD = .15$ ). It follows that more energy was concentrated in lower frequency regions in the production of /s/ by the subjects in Cluster 2 than those in Cluster 4.

### 4.2.3. Approximants

The score for the /r/ tokens and the score for the /l/ tokens, the threshold value of duration for a flap-like sound and the threshold values of F3 were calculated to obtain the values for the two variables for approximants. First, the threshold value of duration was computed to separate some /l/ tokens from a flap-like sound following the procedure described in Section 3.5.4. It was found that the articulation rate of the BN and AN subjects was 4.44 syllables per second and that of the JL subjects was 3.43 syllables per second on average, which suggests that the BN and AN subjects spoke 1.29 times as fast as the JL subjects. The threshold of duration to separate /l/ from a flap-like sound was thus defined as 43 ms for the JL subjects by multiplying 33 ms (Rimac & Smith, 1984) by 1.29. The 33 ms threshold of duration was applied to the BN/AN subjects, and six tokens produced by BN/AN subjects were, as a result of this threshold, regarded as a flap-like sound.

The tokens defined as either /r/ or /l/ were then submitted to the scoring process to judge whether /r/ and /l/ were produced as intended using the threshold values of F3 for /r/ and /l/, which were obtained from the BN/AN data. The results showed that the F3 value of initial /r/ at 2 SD and that of initial /l/ at -2 SD were 1665 mel Hz and 1671 mel Hz, respectively, which were defined as the threshold value of F3 for each approximant. When these values were applied as the threshold, three tokens of /r/ out of the 113 and two tokens /l/ out of the 111 that the BN/AN subjects produced were identified as unintended.

Table 4.19 and Figure 4.18 show the descriptive statistics of the variables, obtained through the above-mentioned procedure for the BN, AN and JL groups. Figure 4.18(a) and (b) illustrates these variables, the number of the /r/ and /l/ correct tokens out of eight, and the

average number of errors, respectively. The errors were categorized into three types, which are shown in Figure 4.18(b): the substitution of /l/ for /r/ and vice versa, that of a flap-like sound for /r/ and /l/ and that of a vowel-like sound for /r/ and /l/. In Figure 4.18(a), the items are represented on the x-axis and the score for the /r/ and /l/ tokens on the y-axis. In Figure 4.18(b), the error types and the average number of errors are indicated on the x-axis and y-axis, respectively.

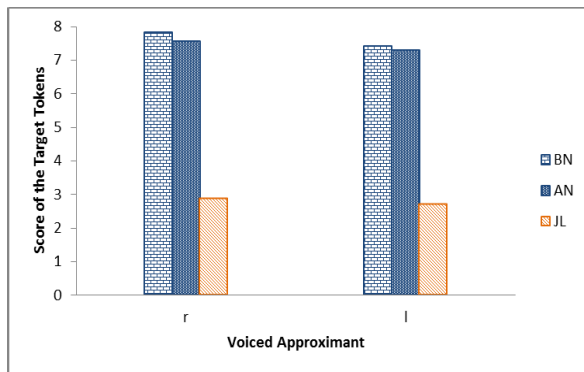
Table 4.19

*Descriptive Statistics of Approximants for BN, AN and JL Groups*

	BN (n = 12)				AN (n = 7)				JL (n = 72)			
	M	SD	Max	Min	M	SD	Max	Min	M	SD	Max	Min
r	7.83	0.39	8.00	7.00	7.57	0.79	8.00	6.00	2.86	2.62	8.00	0.00
l	7.42	0.67	8.00	6.00	7.29	0.76	8.00	6.00	2.69	2.34	8.00	0.00

Note. For /r/ and /l/, the highest possible value is 8, corresponding to the number of items.

(a) Scores for the /r/ and /l/ tokens



(b) Number of errors

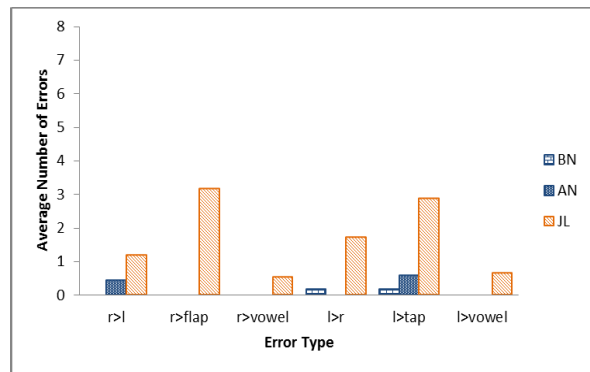


Figure 4.18. Score for the /r/ and /l/ tokens and average number of errors for BN, AN and JL groups: (a) the score for the /r/ and /l/ tokens; and (b) the number of errors for six error categories. r>l = substitution of /l/ for /r/; r>flap = substitution of a flap-like sound for /r/; r>vowel = substitution of a vowel-like sound for /r/; l>r = substitution of /r/ for /l/; l>flap = substitution of a flap-like sound for /l/; l>vowel = substitution of a vowel-like sound for /l/.

One of the differences found between the BN and AN groups and the JL group is in the number of correct tokens. The JL group achieved fewer correct tokens for both /r/ and

/l/ ( $M = 2.86$ ,  $SD = 2.62$  for /r/;  $M = 2.69$ ,  $SD = 2.34$  for /l/) than the BN group ( $M = 7.83$ ,  $SD = 0.39$  for /r/;  $M = 7.42$ ,  $SD = 0.67$  for /l/) and the AN group ( $M = 7.57$ ,  $SD = 0.79$  for /r/;  $M = 7.29$ ,  $SD = 0.76$  for /l/). Figure 4.18(a) suggests that there was no big difference in the scores between /r/ and /l/ for each group. Figure 4.18(b), furthermore, shows the errors that the BN, AN and JL groups made. It reveals that the JL group substituted a flap-like sound for both /r/ and /l/ most frequently. Substitutions of /r/ for /l/ and vice versa came next. This pattern was common between /r/ and /l/.

A cluster analysis was performed to profile the JL subjects, using the z-scores of the scores for the /r/ and /l/ tokens calculated based on the mean and standard deviation of the entire sample. The dendrogram output by the analysis is shown in Appendix L. All subjects were separated into four clusters at the earliest stage of the clustering process, which were selected for the statistical analyses that followed. Cluster 1 was comprised of 12 BN subjects, 7 AN subjects and 6 JL subjects, and was considered to represent native speakers. Clusters 2, 3 and 4 consisted of 20 JL subjects, 19 JL subjects and 27 JL subjects, respectively.

Table 4.20 shows the descriptive statistics, where the valid F3 values averaged across subjects were also presented with the scores for /r/ and /l/ tokens. Some subjects failed to provide any valid F3 value when none of their tokens were judged as either /r/ or /l/. Therefore, the number of subjects who presented those values is also shown in the table. Figure 4.19 shows the profile of each cluster. The plot corresponds to the rank averaged across subjects, which was based on the scores for /r/ and /l/ tokens. A higher score corresponded to a higher rank. Figure 4.20(a) and (b) visually presents the scores for the /r/ and /l/ tokens, and the average number of errors for each error type, respectively. In Figure 4.20(a), the items are on the x-axis and the scores on the y-axis. Figure 4.20(b) shows the error types on the x-axis and the number of errors on the y-axis.

Table 4.20

*Descriptive Statistics of Approximants for Four Clusters*

	Cluster 1 ( <i>n</i> = 25)		Cluster 2 ( <i>n</i> = 20)		Cluster 3 ( <i>n</i> = 19)		Cluster 4 ( <i>n</i> = 27)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Valid F3 <i>n</i>	25 /r/ / 25 /l/		20 /r/ / 18 /l/		13 /r/ / 19 /l/		14 /r/ / 19 /l/	
r	7.40	1.15	5.75	1.41	1.21	1.03	1.11	1.22
l	7.28	0.74	1.55	1.23	4.95	1.22	1.00	0.78
F3 Hz [r]	1735.33	156.60	1819.72	102.23	1789.36	161.11	1861.69	156.89
F3 Hz [l]	2546.50	131.71	2562.12	176.40	2497.90	140.04	2565.37	115.93
F3 mel [r]	1489.06	69.40	1491.40	52.81	1460.20	81.78	1487.68	142.50
F3 mel [l]	1824.57	52.31	1830.76	70.51	1784.35	56.42	1833.13	46.98

*Note.* The number given on the third row shows the number of subjects who provided a valid F3 value of /r/ and /l/. For /r/ and /l/, the highest possible value is 8, corresponding to the number of tokens.

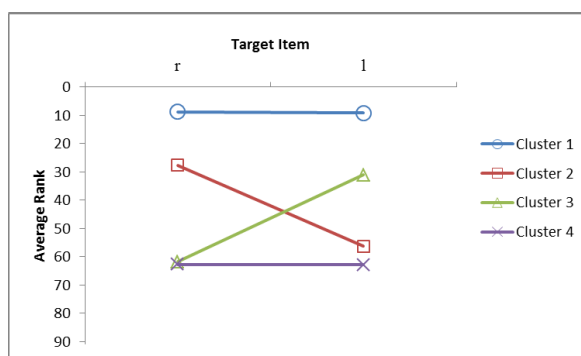


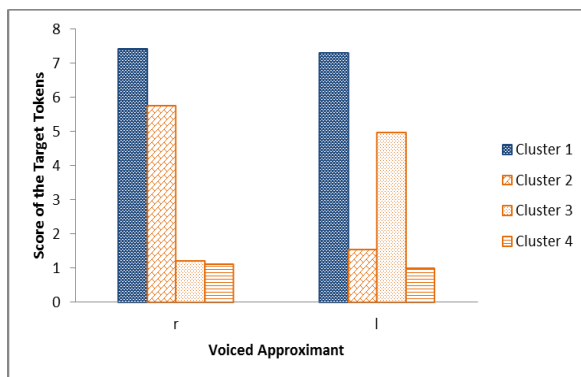
Figure 4.19. Profile of each cluster for approximants.

A notable pattern that can be found in Table 4.20, Figure 4.19 and Figure 4.20 is that the four clusters showed different performances for the production of /r/ and /l/, which was relatively clear in a comparison of Figure 4.20(a). The target tokens that the subjects in Cluster 1, the BN/AN cluster, produced were judged as intended approximants at a high rate for both /r/ and /l/ ( $M = 7.40$ ,  $SD = 1.15$  for /r/;  $M = 7.28$ ,  $SD = .74$  for /l/). In contrast, Cluster 2 performed better for /r/ than /l/ ( $M = 5.75$ ,  $SD = 1.41$  for /r/;  $M = 1.55$ ,  $SD = 1.23$  for /l/), Cluster 3 performed better for /l/ than /r/ ( $M = 1.21$ ,  $SD = 1.03$  for /r/;  $M = 4.95$ ,  $SD = 1.22$  for /l/) and Cluster 4 performed poorly for both /r/ and /l/ ( $M = 1.11$ ,  $SD = 1.22$  for /r/;  $M = 1.00$ ,  $SD = 0.78$  for /l/). These differences among the JL subjects were clearly illustrated



in the profile in Figure 4.19, and also in the pattern of errors that they made. Figure 4.20(b) shows that Cluster 2, performing better in /r/, substituted /r/ for /l/ most frequently, Cluster 3, performing better in /l/, substituted /l/ for /r/ or a flap-like sound, and Cluster 4, achieving the poorest performance in both /r/ and /l/, substituted a flap-like sound for /r/ and /l/ more often than the other clusters. These differences among the JL subjects suggest that whether each of the approximants was easy for them to learn depended on individual subjects.

(a) Scores for the /r/ and /l/ tokens



(b) Number of errors

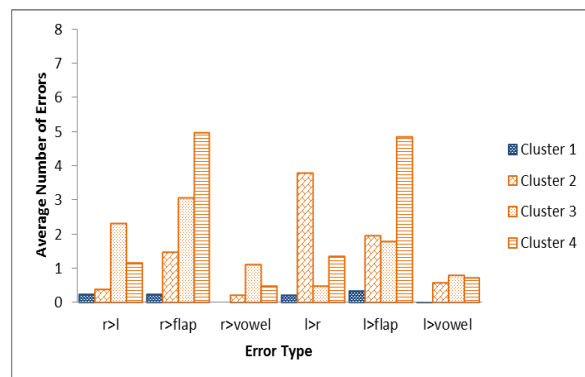


Figure 4.20. Score for the /r/ and /l/ tokens and average number of errors for four clusters: (a) the score for the /r/ and /l/ tokens; and (b) the number of errors for six error categories. r>l = substitution of /l/ for /r/; r>flap = substitution of a flap-like sound for /r/; r>vowel = substitution of a vowel-like sound for /r/; l>r = substitution of /r/ for /l/; l>flap = substitution of a flap-like sound for /l/; l>vowel = substitution of a vowel-like sound for /l/.

In order to determine whether these visually detected differences were significant or not, a one-way MANOVA was conducted with the score for the /r/ and /l/ tokens as dependent variables and the four clusters as the independent variables. Appendix M shows that there was a moderate correlation between the two variables, and therefore, a MANOVA was estimated to perform well. The sample size of the largest cluster was less than 1.5 times as large as that of the smallest cluster, so that the  $\alpha$  level was set at .05. Pillai's trace yielded a significant difference among the clusters,  $F(6, 174) = 140.91, p < .001, \eta_p^2 = .83$ .

A post-hoc discriminant analysis was performed to identify where differences existed between the BN/AN cluster and the JL clusters. Two discriminant functions were found to differentiate between the clusters. The first function explained 76.7% of the variance,

canonical  $R^2 = .91$ , and the second function explained 23.3% of the variance, canonical  $R^2 = .75$ . When these two functions were combined, the clusters were significantly discriminated from each other with the Wilk's lambda value of .02,  $\chi^2(6) = 328.64, p < .001$ . Similarly, the second function alone was able to discriminate between the clusters at a significant level with the Wilk's lambda value of .25,  $\chi^2(2) = 120.70, p < .001$ . As shown in the group centroids in Table 4.21 and the discriminant plot in Figure 4.21, the four clusters were well differentiated with the functions. The first function distinguished Clusters 2, 3 and 4 from Cluster 1, where Cluster 1 and Cluster 4 were differentiated maximally. The second function discriminated Cluster 2 from Clusters 3 and 4, where Cluster 2 and Cluster 3 were differentiated most.

Table 4.21

*Group Centroids for Approximants*

Cluster	Function	
	1	2
1	4.58	0.09
2	-0.78	2.58
3	-0.28	-2.59
4	-3.47	-0.17

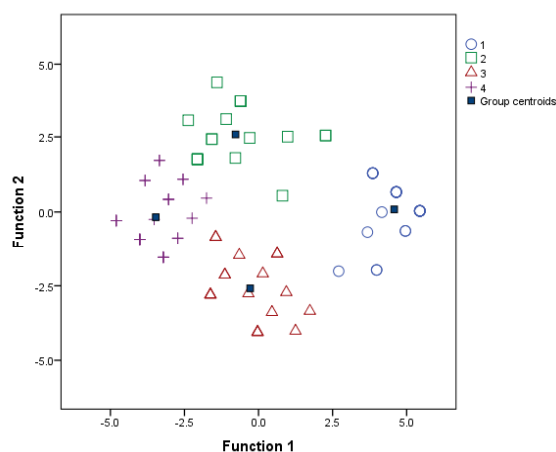


Figure 4.21. Canonical discriminant function plot for approximants.

Table 4.22 is a structural matrix to show the correlations between the variables and

each function. The results revealed that the score for the /r/ tokens most highly loaded on the first function ( $r = .81$ ), and that of /l/ also loaded on it ( $r = .62$ ). This suggests that the score for the /r/ and /l/ tokens contributed to discriminating Clusters 2, 3 and 4 from Cluster 1, especially, Cluster 4 from Cluster 1, which was clear in a comparison of the clusters in the mean values in Table 4.20 and Figure 4.20(a). The subjects in Cluster 1 achieved the highest scores for both /r/ and /l/ ( $M = 7.40$ ,  $SD = 1.15$  for /r/;  $M = 7.28$ ,  $SD = .74$  for /l/), whereas those in Cluster 4 gained the lowest scores for both /r/ and /l/ ( $M = 1.11$ ,  $SD = 1.22$  for /r/;  $M = 1.00$ ,  $SD = .78$  for /l/). Similarly, Cluster 2 attained the lower score for both targets ( $M = 5.75$ ,  $SD = 1.41$  for /r/;  $M = 1.55$ ,  $SD = 1.23$  for /l/) than Cluster 1. Cluster 3 also showed the lower score for both targets ( $M = 1.21$ ,  $SD = 1.03$  for /r/;  $M = 4.95$ ,  $SD = 1.22$  for /l/).

Table 4.22

*Structural Matrix for the Correlations between the Variables for Approximants and the Two Discriminant Functions*

Variable	Function	
	1	2
Score for the /r/ tokens	<b>.81</b>	<b>-.59</b>
Score for the /l/ tokens	<b>.62</b>	<b>.79</b>

*Note.* The variables with the absolute value of correlations with the corresponding functions of .33 and above were highlighted in bold.

The second function showed differences among the JL clusters, highlighting characteristics of Cluster 2 that differed from those of Clusters 3 and 4, particularly differences between Cluster 2 and Cluster 3. The high loadings of both the score of the /l/ tokens ( $r = .79$ ) and the score for the /r/ tokens ( $r = -.59$ ) on the second function resulted in these clusters being discriminated. As reflected in the difference in the values in Table 4.20, Cluster 3 obtained the higher score for the /l/ tokens ( $M = 4.95$ ,  $SD = 1.22$ ) than Cluster 2 ( $M = 1.55$ ,  $SD = 1.23$ ). In contrast, Cluster 2 attained the higher score for the /r/ tokens ( $M = 5.75$ ,  $SD = 1.41$ ) than Cluster 3 ( $M = 1.21$ ,  $SD = 1.03$ ). The average number of errors shown in the Figure 4.20(b) also emphasized these differences among the JL clusters, as pointed out earlier. Cluster 2, which achieved better performance in /r/, produced /r/ even for the /l/ tokens than

Clusters 3 and 4. Cluster 3, achieving the higher score for the /l/ tokens, tended to substitute /l/ for /r/ more frequently than Clusters 2 and 4. In contrast, Cluster 4 performed more poorly in both /r/ and /l/ than Cluster 2 and Cluster 3, respectively. This would be reflected to the most frequent substitution of a Japanese consonant, a flap-like sound.

### 4.3. Rhythm

Table 4.23 depicts the descriptive statistics of the four variables concerning rhythm: the maximum pitch, intensity, duration and F1 and F2 values. They are shown for the BN, AN and JL groups. The pitch and intensity are expressed in ST and dB, respectively. The PVI values of successive stressed and unstressed vowels and the durational difference between stressed vowels and weak vowels in weak forms are expressed as PVI values. A higher value of PVI reflects a greater difference. The absolute durations of the stressed vowels and weak vowels in weak forms are also shown in the table. The extent of vowel centralization is represented as one value, where a larger value means less centralization of vowels.

Table 4.23

*Descriptive Statistics of Rhythm for BN, AN and JL Groups*

	BN (n = 12)				AN (n = 7)				JL (n = 72)			
	<i>M</i>	<i>SD</i>	<i>Max</i>	<i>Min</i>	<i>M</i>	<i>SD</i>	<i>Max</i>	<i>Min</i>	<i>M</i>	<i>SD</i>	<i>Max</i>	<i>Min</i>
Pitch	1.27	0.56	2.69	0.42	1.35	1.67	3.10	-0.78	0.12	0.67	3.00	-1.09
Intensity	2.19	0.83	3.56	0.83	3.40	1.66	6.30	0.91	0.87	1.28	3.93	-2.17
PVI weak	54.31	11.09	72.10	39.50	52.05	9.65	65.39	38.08	-11.67	13.59	15.91	-42.31
Stressed dur.	82.72	9.20	101.42	64.10	83.31	21.28	112.74	59.14	111.95	25.05	218.56	74.92
Weak dur.	47.43	6.43	58.90	35.97	48.15	8.32	59.58	40.22	125.89	26.11	223.33	69.27
Centralization	35.52	4.69	41.99	26.05	36.97	5.84	46.30	30.74	64.36	10.18	86.69	36.73
PVI successive	69.97	10.44	92.07	54.12	70.41	10.33	91.93	61.16	7.53	14.68	49.95	-27.77

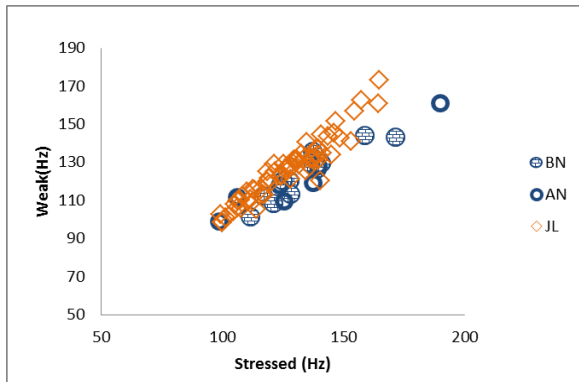
*Note.* Pitch, intensity and vowel centralization are expressed in ST, dB and mel, respectively. Absolute durations of stressed and weak vowels in weak forms are expressed in ms, on which PVI values were calculated. PVI weak = PVI values of stressed vowels and weak vowels in weak forms; dur. = duration; PVI successive = PVI values of successive stressed and unstressed vowels.

The values measured for the token *to* were excluded from the variables of pitch and

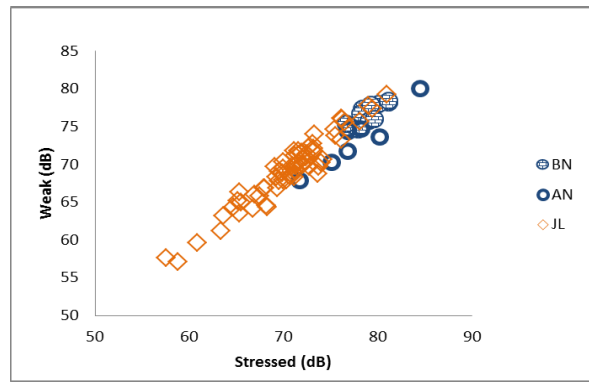
intensity because there were some subjects who weakened this word by dropping the vowel. This was another type of weakening of vowels to be considered, and it was especially characteristic of the BN subjects: of eight tokens, one BN subject dropped the vowel for six tokens, six BN subjects for five tokens, three BN subjects for four tokens, one BN subject for two tokens and one BN subject for one token. This was less likely for the AN subjects and JL subjects. One AN subject dropped the target weak vowel for three tokens and one AN subject for one token out of all five tokens, and one JL subject dropped the vowel for four tokens, two JL subjects for three tokens, three JL subjects for two tokens, eight JL subjects for one token and 57 JL subjects for zero token out of eight tokens. This suggests that all BN subjects weakened *to* by dropping its vowel, whereas the majority of the AN and JL subjects did not. This elision of the vowel was due to the influence of the VOT of the preceding /t/, which could cause the F0 and intensity curves to fluctuate. As a result, four BN subjects failed to provide data concerning the pitch and intensity of *to*. This target token was discarded from the statistical analysis of pitch and intensity for this reason.

Figure 4.22(a), (b) and (c) visually present the maximum pitch, maximum intensity and duration, respectively. The absolute values are plotted in all scatter diagrams, and the pitch, intensity and duration are each expressed in Hz, dB and ms in Figure 4.22, where the stressed vowels are on the x-axis and those of weak vowels are on the y-axis. Figure 4.22(d), (e) and (f) displays the vowel distribution of each target item, where the F2 mel values are on the x-axis and the F1 mel values are on the y-axis. Figure 4.22(g) shows the mean PVI values for each target utterance. The target utterances and the PVI values are indicated on the x-axis and the y-axis, respectively.

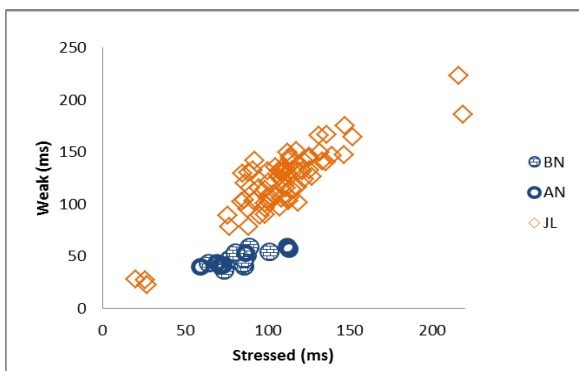
(a) Maximum pitch



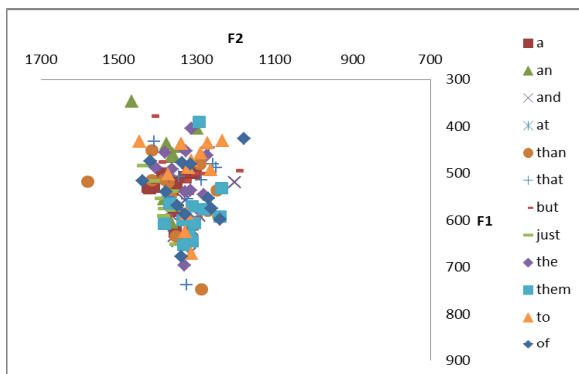
(b) Maximum intensity



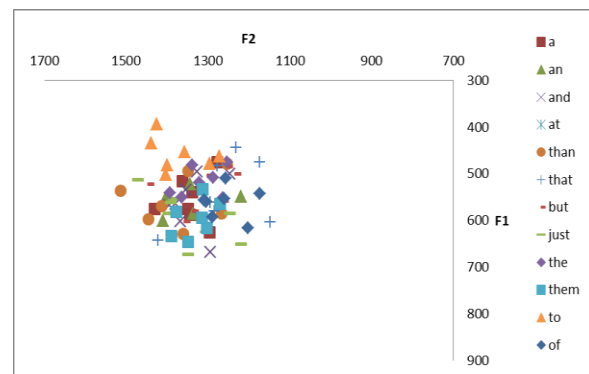
(c) Durations



(d) Vowel distribution of BN

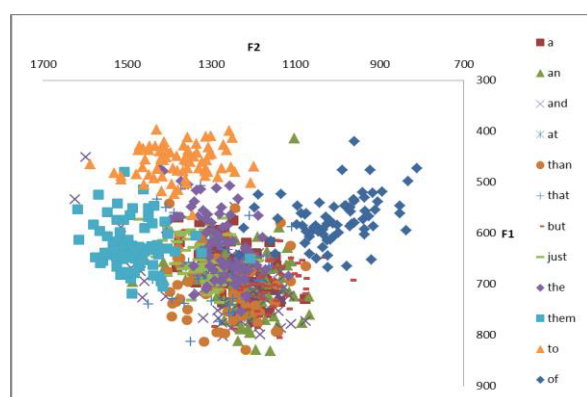


(e) Vowel distribution of AN



(Continued)

(f) Vowel distribution of JL



(g) PVI values of successive stressed and unstressed vowels

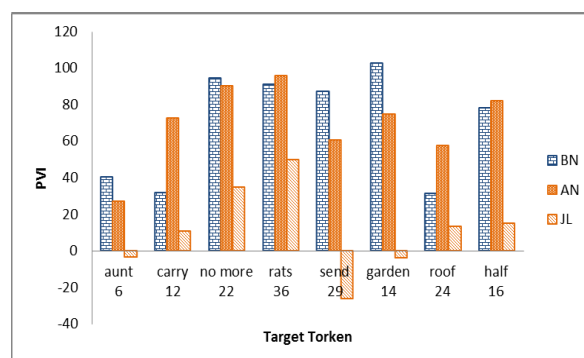


Figure 4.22. Rhythmic values for BN, AN and JL groups: (a) pitch of stressed vowels and weak vowels; (b) intensity of stressed vowels and weak vowels; (c) durations of stressed vowels and weak vowels; (d) vowel distribution of weak vowels for BN; (e) vowel distribution of weak vowels for AN; (f) vowel distribution of weak vowels for JL; and (g) PVI values of successive stressed vowels and unstressed vowels.

As far as the scatter diagrams in Figure 4.22 are concerned, the differences among the three groups are more notable for the durational difference between the stressed vowels and weak vowels in Figure 4.22(c) and the degree of vowel centralization in Figure 4.22(d), (e) and (f). As shown in Figure 4.22(c), the BN and AN groups produced much shorter weak vowels than the JL group. Similarly, a comparison between Figure 4.22(d) and (e) and Figure 4.22(f) showed that the JL group tended to produce more dispersed vowel distribution. The weak vowels in the three items, *of*, *to* and *them*, seemed to have formed their own distinct category, different from /ə/. In contrast, maximum pitch and maximum intensity seemed to be similar among the three groups, which suggests a smaller difference between the BN and AN groups and JL group. However, the values summarized in Table 4.23 also highlight some differences concerning these variables. The pitch difference and intensity difference between the stressed vowels and weak vowels were obviously smaller for the JL group ( $M = 0.12$ ,  $SD = 0.67$  for pitch;  $M = 0.87$ ,  $SD = 1.28$  for intensity) than the BN group ( $M = 1.27$ ,  $SD = 0.56$  for pitch;  $M = 2.19$ ,  $SD = 0.83$  for intensity) and AN group ( $M = 1.35$ ,  $SD = 1.67$  for pitch;  $M = 3.40$ ,  $SD = 1.66$  for intensity).

As regards the PVI values of successive stressed vowels and unstressed vowels, the number that is given under the label of each target utterance in Figure 4.22(g) presents the number of JL subjects who were excluded from the analysis due to the insertion of pauses longer than 250 ms (Abe, 2011). For instance, 36 JL subject placed pauses somewhere in the target utterance, *rats heard a great noise in the loft* and 6 JL subjects put pauses somewhere in the target utterance, *aunt Helen said to him*. As for BN/AN subjects, one token by one AN subject and three tokens by BN subjects were excluded from the analysis in total under the same criterion. This suggests that the JL subjects tended to pauses more than the BN and AN subjects. When calculated after removing all these subjects, the mean PVI values of JL subjects tended to be lower in all target sentences. This means that the JL subjects failed to differentiate the successive stressed vowels and unstressed vowels with duration. This was especially noticeable in the target utterance, *send out scouts to search for a new home*, in which the JL subjects obtained the negative mean PVI value. This suggests that they were likely to produce unstressed vowels that were longer than the adjacent stressed vowels. This was applied to two other utterances, *aunt Helen said to him* and *garden with an elm tree*. However, this pattern was not observed for the BN and AN subjects.

As described above, some JL subjects failed to present valid data for the PVI values of successive stressed and unstressed vowels because of a higher frequency of pauses. On average across tokens, 20 out of 72 tokens produced by the JL subjects were excluded from the analysis. Only 26 JL subjects successfully produced all target utterances to be measured. This variable was thus not included in the subsequent statistical analyses.

In order to focus more on the performances of individual subjects and profile the JL subjects, a cluster analysis was carried out, where the z-scores of the maximum pitch, maximum intensity, duration and F1 and F2 values based on the mean and standard deviation for the entire sample were submitted as variables. Because the clustering was slightly modified here as to the formation of one JL cluster, the resulting dendrogram is shown in Figure 4.23. All BN/AN subjects were classified into A in Figure 4.23, indicating that the native-speaker subjects in this study have similar rhythmic features. One JL subject was



categorized into this cluster. This was defined as Cluster 1. The remaining JL subjects were roughly divided into two clusters at an earlier stage of the clustering process: one was B and the other was the group consisting of C, D and E, as in Figure 4.23. The former cluster was defined as Cluster 3. The latter group was comprised of 56 subjects, and was by far larger than Clusters 1 and 3. For statistical reasons, this group was further separated into two clusters, Cluster 2 consisting of C and Cluster 4 consisting of E and D. Before the BN and AN subjects were classified into one cluster, some subjects were broken down into C, D and E at the earliest stage of the clustering. This caused the expectation of differences even between Cluster 2 and Cluster 4. Cluster 1 thus consisted of 12 BN subjects, 7 AN subjects and 1 JL subject, and was regarded as representing the native speakers' production of rhythm. Clusters 2, 3 and 4 were all made up of the JL subjects only, 27 and 15 and 29, respectively.

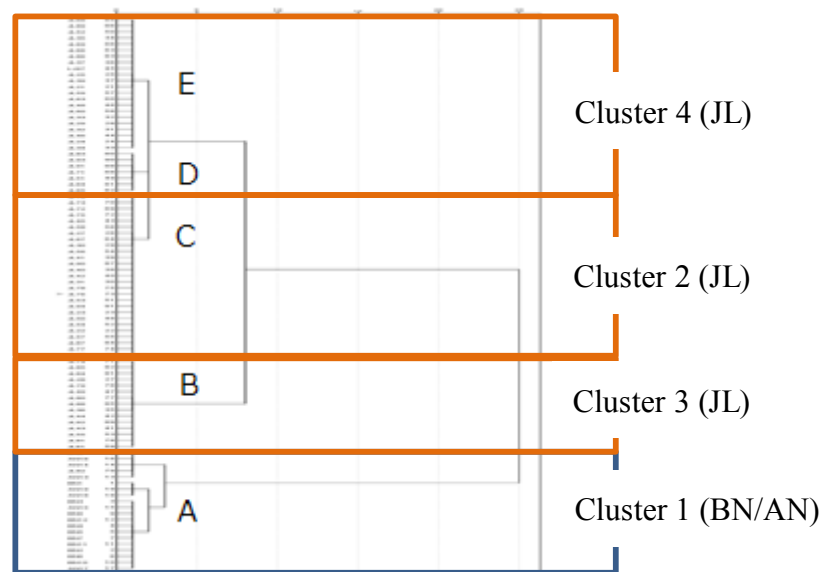


Figure 4.23. Dendrogram output for the rhythm.

The descriptive statistics are summarized in Table 4.24. As in Table 4.23, the pitch and intensity are expressed in ST and dB, respectively. The PVI values and the absolute durations are both shown as the durational values. The vowel centralization shows how much the target weak vowel was centralized, and a lower value reflects a more centralization. A line graph in Figure 4.24 illustrates the profile of each cluster based on the average of the rank

across subjects, which shows the overall level of performance of each cluster. Rank was assigned using raw values. Larger values for pitch, intensity and PVI values mean a greater difference between the stressed vowel and weak vowel, where the subjects with larger values ranked more highly for these variables. In contrast, a smaller value for vowel centralization suggests that the vowel quality was more centralized, and therefore, the subjects with a smaller value ranked more highly for this variable. The four variables are presented on the x-axis and the rank is on the y-axis. In Figure 4.25, seven scatter diagrams depict the results of each variable visually: Figure 4.25(a) for the pitch, Figure 4.25(b) for the intensity, Figure 4.25(c) for the duration and Figure 4.25(d), (e), (f) and (g) for the vowel centralization of each cluster. The pitch, intensity and duration of the target weak vowels were compared with those of the stressed vowels. Accordingly, the absolute values for the stressed vowels are on the x-axis and those for the weak vowels are on the y-axis in Figure 4.25(a), (b) and (c). Their units are Hz, dB and ms. In Figure 4.25(d), (e), (f) and (g), the centralization of the weak vowels are illustrated with the F2 values on the x-axis and the F1 values on the y-axis.

Table 4.24

*Descriptive Statistics of Rhythm for Four Clusters*

	Cluster 1 ( <i>n</i> = 20)		Cluster 2 ( <i>n</i> = 27)		Cluster 3 ( <i>n</i> = 15)		Cluster 4 ( <i>n</i> = 29)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Pitch	1.22	1.08	-0.29	0.42	1.05	0.69	0.03	0.33
Intensity	2.70	1.30	-0.20	0.64	2.61	0.68	0.85	0.69
PVI weak	51.19	14.35	-14.78	11.56	-5.19	13.42	-12.79	14.41
Stressed dur.	82.53	13.98	115.01	26.75	110.38	19.85	111.19	25.89
Weak dur.	48.78	8.32	132.23	21.94	117.51	27.47	126.28	26.71
Centralization	36.09	4.90	61.79	7.45	68.26	9.11	65.69	11.24

*Note.* Pitch, intensity and vowel centralization are expressed in ST, dB and mel, respectively. Absolute durations of stressed and weak vowels in weak forms are expressed in ms, on which PVI values were calculated. PVI weak = PVI values of stressed vowels and weak vowels in weak forms; dur. = duration.

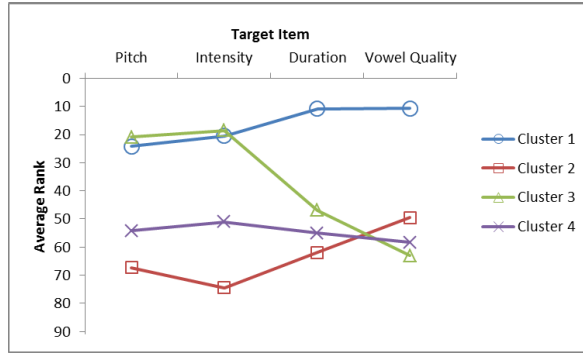
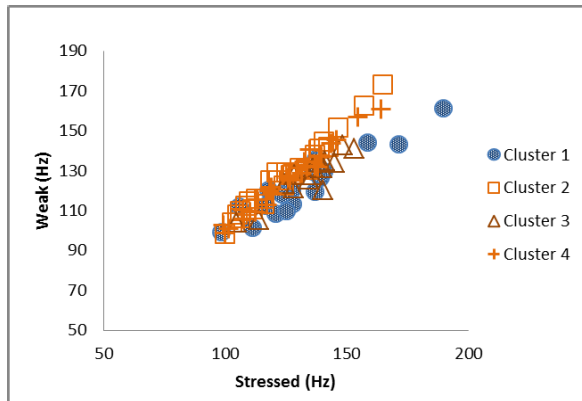
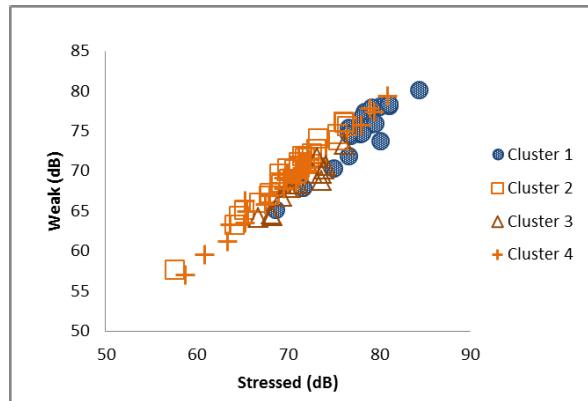


Figure 4.24. Profile of each cluster for rhythm.

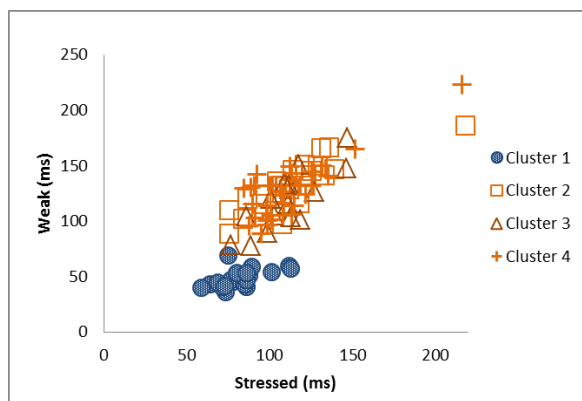
(a) Maximum pitch



(b) Maximum intensity

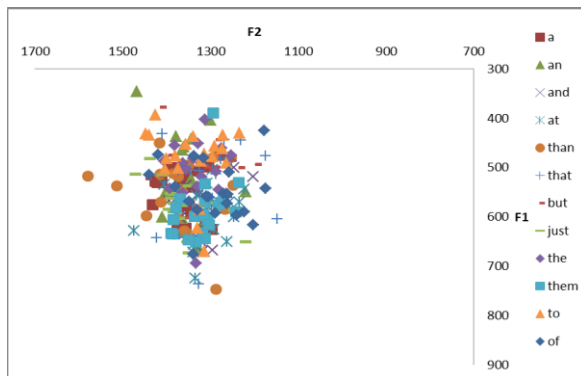


(c) Durations

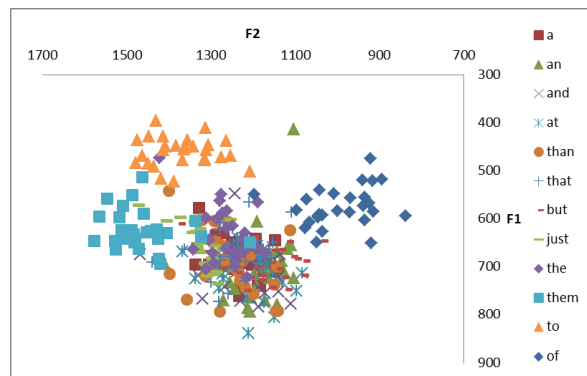


(Continued)

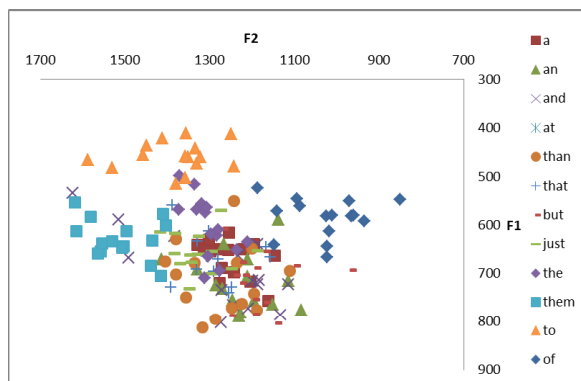
(d) Vowel distribution of Cluster 1



(e) Vowel distribution of Cluster 2



(f) Vowel distribution of Cluster 3



(g) Vowel distribution of Cluster 4

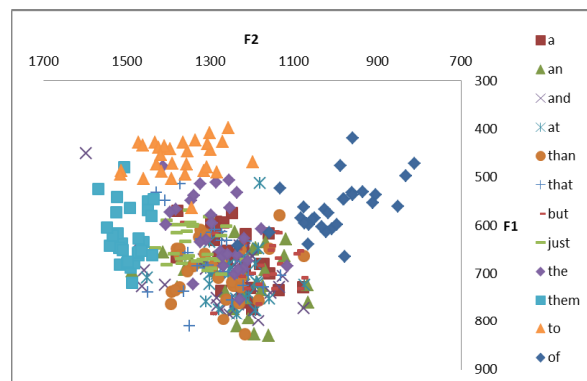


Figure 4.25. Rhythmic values for four clusters: (a) pitch of stressed vowels and weak vowels; (b) intensity of stressed vowels and weak vowels; (c) durations of stressed vowels and weak vowels; (d) vowel distribution of weak vowels for Cluster 1; (e) vowel distribution of weak vowels for Cluster 2; (f) vowel distribution of weak vowels for Cluster 3; and (g) vowel distribution of weak vowels for Cluster 4.

It can be seen from Figures 4.24 and 4.25 that the JL clusters, Clusters 2, 3 and 4, showed notable differences from the BN/AN cluster, Cluster 1, in the performance of duration and vowel centralization for the weak vowels in the weak forms. The line graph in Figure 4.24 shows that all three JL cluster deviated from the BN/AN cluster for these two variables. As shown in Figure 4.25(c), Cluster 1 produced both weak vowels and stressed vowels shorter than the other clusters, which could be interpreted as originating in the different speaking rate. The plots for the subjects in Cluster 1 were not linearly related to those for the subjects in the other clusters, suggesting that the former cluster differed from the latter clusters even though the differences in the speaking rate were counted. The PVI values of the stressed and weak vowels in Table 4.24 show that Cluster 1 produced stressed vowels

that were longer than weak vowels ( $M = 51.19$ ,  $SD = 14.35$ ), whereas all JL clusters indicated the opposite pattern ( $M = -14.78$ ,  $SD = 11.56$  for Cluster 2;  $M = -5.19$ ,  $SD = 13.42$  for Cluster 3;  $M = -12.79$ ,  $SD = 14.41$  for Cluster 4), where the weak vowels were produced even longer than the stressed vowels. For the vowel centralization, the distribution of vowels shrunk to one category for Cluster 1, as in Figure 4.25(d), regardless of the target token. Conversely, separate vowel categories could be easily recognized for the three JL clusters, and there did not seem to be much difference between them in the extent of the centralization in a comparison of Figure 4.25(e), (f), and (g). The target tokens *of*, *to* and *them* were distant from the category where the other items gathered, which was obviously due to the effect of the pronunciation of their corresponding strong form. It was also characteristic of Clusters 2, 3 and 4 that their category(ies) were generally located a little lower and further back in the vowel space than that of Cluster 1.

Seemingly, maximum pitch and intensity did not show a great difference between the clusters, as in Figure 4.25(a) and (b). According to Figure 4.25(a) and (b), there was a positive linear correlation between the stressed vowels and the weak vowels across clusters, which suggests the existence of a similar pattern among them. However, the profile given in the line graph in Figure 4.24 shows that whereas Cluster 3 did not differ from Cluster 1 in terms of the variables for the pitch and intensity, Clusters 2 and 4 deviated from Cluster 1 in both. As seen in Table 4.24, these clusters also differed in these variables from one another. Cluster 4 tended to differentiate less between the stressed vowels and weak vowels according to pitch and intensity ( $M = 0.03$ ,  $SD = 0.33$  for pitch;  $M = 0.85$ ,  $SD = 0.69$  for intensity) than Cluster 1 and Cluster 3. Cluster 2 performed a little more poorly; it did not demonstrate pitch and intensity difference between the stressed vowels and weak vowels, using an even higher pitch and stronger intensity for the weak vowels ( $M = -0.29$ ,  $SD = 0.42$  for pitch;  $M = -0.20$ ,  $SD = 0.64$  for intensity).

In order to testify to these differences statistically, a one-way MANOVA was conducted, where the pitch difference between stressed and weak vowels, the intensity difference between stressed and weak vowels, the PVI values of stressed and weak vowels

and vowel centralization were used as the dependent variables, and the four clusters as the independent variables. Correlations between the dependent variables are shown in Appendix N. This suggests that all variables were moderately correlated, and thus, it was expected that a MANOVA would work well for this analysis. The sample size of the largest cluster, Cluster 4, was nearly twice as large as that of the smallest cluster, Cluster 3, and thus, the  $\alpha$  level was set at .01 in the analysis. According to the results of Pillai's trace, the clusters were significantly different,  $F(12,258) = 21.08, p < .001, \eta_p^2 = .50$ .

A post-hoc test was carried out, using a discriminant analysis. Three discriminant functions were obtained, where the first function accounted for 79.4% of the variance, canonical  $R^2 = .86$ , the second function accounted for 20.4% of the variance, canonical  $R^2 = .61$  and the third function accounted for 0.2% of the variance, canonical  $R^2 = .15$ . When these three functions were combined, they discriminated between the four clusters significantly with the Wilk's lambda value of .05,  $\chi^2(12) = 251.29, p < .001$ . After removing the first function, the second and third functions discriminated between them significantly with the Wilk's lambda value of .38,  $\chi^2(6) = 82.47, p < .001$ . However, the third function alone failed to discriminate between the clusters at a significant level with the Wilk's lambda value of .99,  $\chi^2(2) = 1.29, p = .526$ , and thus, this function was not interpreted further. Table 4.25 and Figure 4.26 display the results of the group centroids and the discriminant function plot, respectively.

Table 4.25

*Group Centroids for Rhythm*

Variable	Function	
	1	2
1	4.24	-0.84
2	-2.11	-1.04
3	0.61	2.46
4	-1.28	0.28

The group centroids show that the first function worked to separate Clusters 1 and 3 from

Clusters 2 and 4. While Cluster 2 and Cluster 4 were discriminated maximally from the BN/AN cluster, they were also differentiated from Cluster 3, the other JL cluster. The second function differentiated between Cluster 2 and Cluster 4 and between Cluster 1 and Cluster 3. The latter clusters were maximally discriminated.

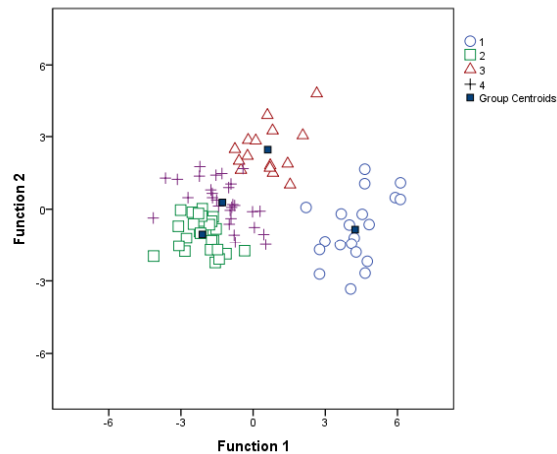


Figure 4.26. Canonical discriminant function plot for rhythm.

Table 4.26 presents the structural matrix for the correlations between the variables and the two functions, and shows that all variables highly loaded on the first function. The duration ( $r = .78$ ) loaded most highly on it, followed by the intensity ( $r = .52$ ), the vowel centralization ( $r = -.49$ ) and the pitch ( $r = .37$ ). This function discriminated Clusters 2 and 4 from Cluster 1 maximally. Thus, they were differentiated regarding all four variables. The profile in Figure 4.24 shows that the maximum pitch and intensity were primarily concerned with discrimination from Cluster 3, but to a lesser degree. As shown in Table 4.24, the subjects in Cluster 1 produced a much greater durational difference between the stressed vowels and weak vowels ( $M = 51.19$ ,  $SD = 14.35$ ) than those in Cluster 2 ( $M = -14.78$ ,  $SD = 11.56$ ) and Cluster 4 ( $M = -12.79$ ,  $SD = 14.41$ ). The negative correlation of vowel centralization with the function suggests the opposite direction of performance from the other variables. Cluster 1 had a smaller value, meaning they produced a more centralized vowel quality ( $M = 36.09$ ,  $SD = 4.90$ ) than Cluster 2 ( $M = 61.79$ ,  $SD = 7.45$ ) and Cluster 4 ( $M =$

65.69,  $SD = 11.24$ ), as presented in Table 4.24. Because the variable of the vowel quality in this study involves the centralization of the target weak vowel in the vowel space, these results suggest that the subjects in Cluster 1 produced a similar quality for each target weak vowel, while those in Clusters 2 and 4 produced different quality for each, resulting in more dispersed vowels in the space. All these results corroborated what was noted above in the visual inspection of the scatter diagrams in Figure 4.25. Figure 4.25(e), (f) and (g) demonstrate that the target weak vowels in the words *of*, *to* and *them* were articulated in a different quality from those in the other tokens in the production of the JL clusters. Similarly, Clusters 1 and 3 performed better in the differentiation of pitch and intensity between the stressed vowels and weak vowels than Clusters 2 and 4. The difference in maximum pitch and intensity was greater for Cluster 1 ( $M = 1.22$ ,  $SD = 1.08$  for pitch;  $M = 2.70$ ,  $SD = 1.30$  for intensity) and Cluster 3 ( $M = 1.05$ ,  $SD = 0.69$  for pitch;  $M = 2.61$ ,  $SD = 0.68$  for intensity) than Cluster 2 ( $M = -0.29$ ,  $SD = 0.42$  for pitch;  $M = -0.20$ ,  $SD = 0.64$  for intensity) and Cluster 4 ( $M = 0.03$ ,  $SD = 0.33$  for pitch;  $M = 0.85$ ,  $SD = 0.69$  for intensity). The negative value in these variables obtained by Cluster 2 suggests that the subjects in this JL cluster tended to place a higher pitch and stronger intensity on the weak vowels than on the stressed vowels.

Table 4.26

*Structural Matrix for the Correlations between the Variables for Rhythm and the Two Discriminant Functions*

Variable	Function	
	1	2
Duration (PVI)	<b>.78</b>	<b>-.40</b>
Vowel centralization	<b>-.49</b>	<b>.59</b>
Intensity	<b>.52</b>	<b>.55</b>
Pitch	<b>.37</b>	.30

*Note.* The variables with the absolute value of correlations with the corresponding functions of .33 and above were highlighted in bold.

The correlations between the second function and the variables are also presented in



Table 4.26, which shows that the vowel centralization, the intensity and the duration loaded on this function ( $r = .59$ ,  $r = .55$  and  $r = -.40$ ). This function contributed to differentiating primarily between Cluster 1 and Cluster 3, and also between Cluster 2 and Cluster 4. When comparing each cluster in the profile in Figure 4.24, the vowel centralization and the duration discriminated between the former clusters, and the intensity, the latter clusters. The subjects in Cluster 1 tended to use a more centralized vowel quality for the weak vowels ( $M = 36.09$ ,  $SD = 4.90$ ) than those in Cluster 3 ( $M = 68.26$ ,  $SD = 9.11$ ). This tendency is observable in a comparison between Figure 4.25(d) and Figure 4.25(f). It depicts the greater dispersion of the vowels for the subjects in Cluster 3 in spite of its smaller sample size. They were also distinguished by the duration, where the subjects in Cluster 1 produced a larger durational difference between the stressed vowels and weak vowels ( $M = 51.19$ ,  $SD = 14.35$ ) than those in Cluster 3 ( $M = -5.19$ ,  $SD = 13.42$ ). The negative value of Cluster 3 involves an even longer duration for the weak vowels than the stressed vowels. In contrast, intensity discriminated between Cluster 2 and Cluster 4. The subjects in Cluster 2 produced a smaller difference in the intensity between the stressed vowels and the weak vowels ( $M = -0.20$ ,  $SD = 0.64$ ) than those in Cluster 4 ( $M = 0.85$ ,  $SD = 0.69$ ). The negative value of this variable suggests that the subjects in Cluster 2 were likely to place a stronger intensity on the weak vowels.

#### 4.4. Intonation

The typical nucleus placement and nuclear tone choice were first identified, based on the data provided by the BN/AN subjects. Table 4.27 presents the number of the BN/AN subjects and JL subjects who used the intonation patterns that were defined as typical (see Appendix O for the summary of all results). Two words are presented when the nucleus fell on the two words within a single target utterance, which occurred to three target utterances: *There was once a young rat named Arthur*, *There was a kindly horse named Nelly* and *and a garden with an elm tree*. Nuclear tone choice was scored depending on whether the pitch movement of the typical nuclear tone occurred somewhere within the syntactic phrase containing the nucleus, as described in Section 3.5.6. Accordingly, more than one words are named for most of the tokens in Table 4.27.

Table 4.27

*Typical Pitch Patterns Used by BN/AN Subjects*

Context	Nucleus placement	Nuclear tone choice					
		BN/AN (n = 19)	JL (n = 72)			BN/AN (n = 19)	JL (n = 72)
ANT	rat & Arthur	13 <sup>a</sup>	11	named/Arthur	Fall	17	46
SD end	out	13	2	go/out/with/them	Fall-rise	6	3
					Level	6	3
bf DS 1	answer	15	69	only/answer	Level	16	18
bf DS 2	said	18	0	said/to/him	Level	9	4
					Low-rise	6	3
DIA 1	know	19	70	don't /know	Fall	14	64
DIA 2	do	18	67	this/won't/do	Fall	13	67
TOPIC	Helen	18	49	his/aunt/Helen	Fall	18	24
AdP 1	day	19	64	one/rainy/day	Low rise	8	5
					Fall	4	35
					Level	4	25
AdP 2	last	19	63	at/last	Fall	9	17
					High-rise	8	18
AdP 3	then	19	69	just/then	Fall	11	21
AdP 4	night	18	70	night	Level	7	42
					Fall	6	20
LIST 1	horse & Nelly	16	9	named/Nelly	Fall-rise	8	4
					Low rise	5	1
LIST 2	cow	19	72	cow	Low rise	15	5
LIST 3	calf	19	71	calf	Low rise	8	1
					Fall-rise	6	4
lastLIST	garden & elm	17	4	with/an/elm/tree	Fall	18	70
EXCL	well	- <sup>b</sup>	- <sup>b</sup>	well	Fall	15	35
COM 1	face	17	67	right/face	Fall	10	66
COM 2	march	- <sup>b</sup>	- <sup>b</sup>	march	Fall	19	67

*Note.* <sup>a</sup> This target utterance was slightly different between the passage for the AN subjects and that for the BN and JL subjects. The former was *Once there was a young rat named Arthur* and the latter was *There was once a young rat named Arthur*. Although all AN subjects put the nucleus on *once* due to this different word order, this nucleus was not counted here because none of the BN subjects placed the nucleus on it. <sup>b</sup> One-word utterance was not used for the analysis of nucleus placement. ANT = antecedent modified by the relative clause; SD end = end of the subordinate clause preceding the main clause; bf DS = reporting clause before direct speech; DIA = short dialogue; TOPIC = topic; AdP = adverbial phrase; LIST = lists; lastLIST = last component of closed lists; EXCL = exclamation; COM = command.

It has to be noted, regarding the nuclear tone choice, that when the target utterance had two nuclei, as in the above-mentioned three target utterances, only the latter nuclear tone was scored. For instance, Table 4.27 shows that in the target utterance, *There was once a young*

*rat named Arthur*, 13 BN/AN subjects and 11 JL subjects located the nucleus on both *rat* and *Arthur*; 17 BN/AN subjects and 46 JL subjects were judged to use a fall as the nuclear tone on *named* or *Arthur*.

Results for the BN/AN subjects showed that there was a strong tendency shared by the subjects in these groups concerning the nucleus placement as to which word of the target utterance the nucleus fell on. Although the target utterances, *There was once a young rat named Arthur* and *go out with them*, were produced with the greatest variety of nucleus placement of all, 13 out of 19 BN/AN subjects, that is, the majority of the subjects in these groups, placed the nucleus on the same word. In contrast, slightly more varieties of nuclear tone choice were found in the BN/AN subjects. No single tone type occurred in more than half the subjects for *go out with them* in the end of the subordinate clause preceding the main clause context, *said to him* in the reporting clause before direct speech context, *one rainy day*, *at last* and *that night* in the adverbial phrase context and *There was a kindly horse named Nelly* and *a calf* in the lists context. Consequently, more than two tones were defined as typical for these utterances.

On the other hand, the results of the nucleus placement for the JL subjects showed that the target utterances were divided into two, depending on whether or not almost all the subjects realized the nuclear placement typical of the BN/AN subjects. The utterances where the nucleus tended to occur on different words between the JL subjects and the BN/AN subjects included the following five target utterances, *There was once a young rat Arthur*, *go out with them*, *said to him*, *There was a kindly horse named Nelly* and *and a garden with an elm tree*. For the nuclear tone choice, the number of the JL subjects who used the typical tone varied across utterances. Nearly all the subjects succeeded in using the typical tones for the utterances, *I don't know* and *This won't do*, in the short dialogue context, *one rainy day* and *that night* in the adverbial phrase context and *a garden with an elm tree* in the last component of closed lists context and *Right about face* and *March* in the command context. In contrast, only fewer than 10 out of 72 subjects succeeded in the utterances, *go out with them* in the end of the subordinate clause preceding the main clause context, *said to him* in the reporting

clause before direct speech context and *There was a kindly horse named Nelly, a cow and a calf* in the lists context.

Based on the above-described results of the BN/AN subjects, the six variables of the phonological aspects of intonation were obtained as follows, as noted in Section 3.5.6. For the nucleus placement, the nucleus fell on the non-utterance-final word in the target utterances, *go out with them, said to him* and *and a garden with an elm tree*. The two utterances, *There was once a young rat named Arthur* and *There was a kindly horse named Nelly* were characterized as long, both starting with *there was*. These five target utterances were summarized as *long/non-final utterances* and provided two variables, the score for the nucleus in the long/non-final utterances and the score for the non-nuclear words in the long/non-final utterances. The rest of the target utterances were classified into *final utterances*, where the nucleus was on the final word. They served as the other two variables, the score for the nucleus in the final utterances and the score for the non-nuclear words in the final utterances. The scores of all four variables were given by assigning each word in the target utterance either 0 or 1 point as described in Section 3.5.6, which defined the highest possible values as follows. Three target utterances out of five in the long/non-final utterances, *and a garden with an elm tree, There was once a young rat named Arthur* and *There was a kindly horse named Nelly*, each had two nuclei. The highest possible value was thus defined as 8 points for the nucleus in the long/non-final utterances. There were 21 remaining non-nuclear words in the long/non-final utterances, which corresponded to 21 points for the score for the non-nuclear words in the long/non-final utterances. In the final utterances, all target utterances had only one nucleus, which defined the highest possible value for the nucleus in the final utterances as 11 points and that for the non-nuclear words in the final utterances as 18 points.

Similarly, as for the other two variables involving the scores for the nuclear tone choice, the target utterances were categorized as follows, based on the typical use of the nuclear tone by the BN/AN subjects. Non-falling tones were used by the BN/AN subjects in the following utterances: *go out with them, he would only answer* and *said to him, There was*

*a kindly horse named Nelly, a cow and a calf*. These six utterances were grouped together to provide one variable of the non-falling-tone type, *non-falling utterances*. The first utterance, the next two utterances and the last three utterances represent the end of the subordinate clause preceding the main clause, the reporting clause before direct speech and the lists contexts, respectively, which suggest that all utterances comprising each of the three syntactic and pragmatic context were successfully classified into the single variable. The other target utterances were categorized into the other variable, *falling utterances*. Unlike the score for the nucleus placement, it was assumed that each target utterance represented the syntactic and pragmatic target context. Therefore, the variables were calculated by averaging the scores across the utterances of each context, as referred to in Section 3.5.6. Because the score for the non-falling utterances was obtained from the three context noted above, the highest possible values were 3 points for the score for the nuclear tone choice in the non-falling utterances and 7 points for the score for the nuclear tone choice in the falling utterances.

Table 4.28 and Figure 4.27 show the descriptive statistics of all variables for the BN, AN and JL groups. Figure 4.27(a) shows the rate of common errors in the nucleus placement for each group, providing the number of subjects who put an extra nucleus on *there* or *was* for the two utterances, *There was once a young rat named Arthur* and *There was a kindly horse named Nelly*, and the number of subjects who put the nucleus on the final word of the utterances, *go out with them, said to him* and *and a garden with an elm tree*. Figure 4.27(b) depicts the rate of the subjects who used each tone type for the non-falling utterances, including a fall, a fall-rise, a level, a low-rise and a high-rise. The category error involves the tokens which could not be analyzed acoustically due to creaky voice or error speech. In both bar graphs, the number of subjects was transformed to the rate, expressed in percentages, to make it easy to compare across the groups. For Figure 4.27(b), the rate was calculated across all target utterances in the three syntactic and pragmatic contexts. Figure 4.27(c) shows the realization of the phonetic items of intonation, where the level is on the x-axis and the span is on the y-axis. The four JL subjects were identified as outliers, and removed from the subsequent statistical analyses. Consequently, the total number of JL subjects was 68.

Table 4.28

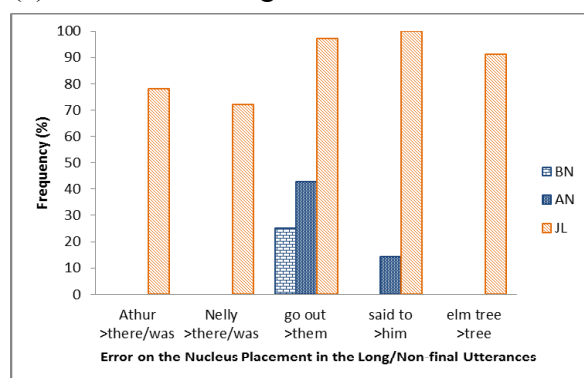
*Descriptive Statistics of Intonation for BN, AN and JL Groups*

		BN ( <i>n</i> = 12)				AN ( <i>n</i> = 7)				JL ( <i>n</i> = 68)			
		<i>M</i>	<i>SD</i>	<i>Max</i>	<i>Min</i>	<i>M</i>	<i>SD</i>	<i>Max</i>	<i>Min</i>	<i>M</i>	<i>SD</i>	<i>Max</i>	<i>Min</i>
Phonological items													
Long/ non-final	Nucleus	7.00	0.60	8.00	6.00	6.14	0.69	7.00	5.00	3.28	0.93	5.00	2.00
	Non-nuclear	20.25	0.62	21.00	19.00	19.43	0.79	20.00	18.00	14.99	1.24	18.00	12.00
Final	Nucleus	10.75	0.45	11.00	10.00	10.86	0.38	11.00	10.00	10.72	0.49	11.00	9.00
	Non-nuclear	17.50	0.67	18.00	16.00	17.86	0.38	18.00	17.00	17.63	0.60	18.00	16.00
Non-falling		2.33	0.56	3.00	1.33	1.88	0.92	3.00	0.67	0.33	0.46	1.83	0.00
Falling		5.92	0.49	7.00	5.00	5.61	1.25	7.00	3.75	4.97	1.12	7.00	2.00
Phonetic items													
Span		4.49	1.27	6.51	2.37	5.67	1.56	8.63	3.55	3.88	0.96	6.17	1.94
Level		90.26	9.36	105.10	72.50	87.00	20.03	121.60	59.70	92.44	13.11	126.90	67.07

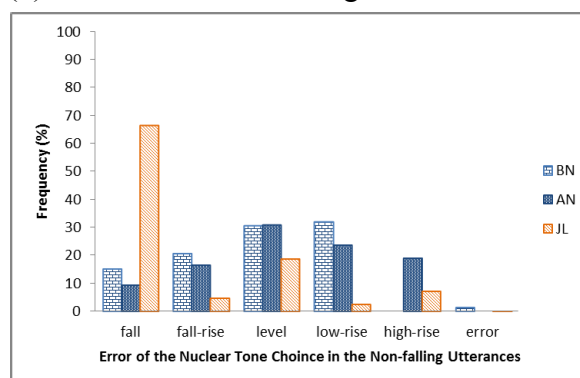
*Note.* The first six variables represent the number of utterances and contexts produced in the typical manner to the BN/AN subjects and their highest possible values were 8, 21, 11, 18, 3 and 7, respectively. The span and level are each expressed in ST and Hz.

Table 4.28 shows that major differences between the JL group and BN/AN groups lay in the scores for the nucleus and non-nuclear words in the long/non-final utterances and the score for non-falling utterances. The JL group achieved lower scores for all these variables ( $M = 3.28$ ,  $SD = 0.93$  for the score for the nucleus in the long/non-final utterances;  $M = 14.99$ ,  $SD = 1.24$  for the score for the non-nuclear words in the long/non-final utterances;  $M = .33$ ,  $SD = 0.46$  for the score for the nuclear tone choice in the non-falling utterances) than the BN group ( $M = 7.00$ ,  $SD = 0.60$  for the score for the nucleus in the long/non-final utterances;  $M = 20.25$ ,  $SD = 0.62$  for the score for the non-nuclear words in the long/non-final utterances;  $M = 2.33$ ,  $SD = 0.56$  for the score for the nuclear tone choice in the non-falling utterances) and AN group ( $M = 6.14$ ,  $SD = 0.69$  for the score for the nucleus in the long/non-final utterances;  $M = 19.43$ ,  $SD = 0.79$  for the score for the non-nuclear words in the long/non-final utterances;  $M = 1.88$ ,  $SD = 0.92$  for the score for the nuclear tone choice in the non-falling utterances).

(a) Error rate in long/non-final utterances



(b) Error rate in non-falling utterances



(c) Span and level

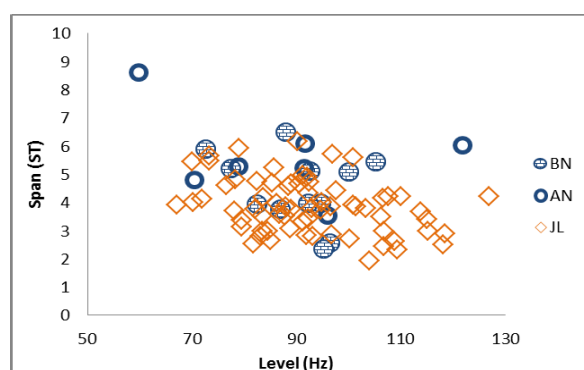


Figure 4.27. Errors in the phonological representation of intonation and the span and level in the phonetic representation of intonation for BN, AN and JL groups: (a) the rate of the extra nucleus placed on *there* or *was* for *There was once a young rat named Arthur* and *There was a kindly horse named Nelly* and the rate of the nucleus placed on the final word for *go out with them*, *said to him* and *and a garden with an elm tree*; (b) the rate of each tone type used for the utterances where a falling tone was not typical; and (c) scatter diagram of the span and level.

The tendency for low scores in these variables was, furthermore, highlighted by the rate of errors shown in Figure 4.27(a) and (b). Figure 4.27(a) illustrates errors involving the nucleus placement in long/non-final utterances: placing an extra nucleus on *there* and *was*, the beginning of two target utterances, *There was once a young rat named Arthur* and *There was a kindly horse named Nelly*, and placing the nucleus on the final words of three target utterances, *go out with them*, *said to him* and *and a garden with an elm tree*. Figure 4.27(b) shows that the JL group was likely to use a falling tone more frequently even in the non-falling utterances. The other scores concerning the phonological items of intonation did

not seem to show as great a difference as these scores, as in Table 4.28.

Another subtle difference was also found in the span and level. The JL group used a slightly narrower span and higher level than ( $M = 3.88$ ,  $SD = 0.96$  for span;  $M = 94.22$ ,  $SD = 13.11$  for level) the BN group ( $M = 4.49$ ,  $SD = 1.27$  for span;  $M = 90.26$ ,  $SD = 9.36$  for level) and the AN group ( $M = 5.67$ ,  $SD = 1.56$  for span;  $M = 87.00$ ,  $SD = 20.03$  for level) on average. The difference in the level is not apparent, but that in the span is a little more obvious in Figure 4.27(c), where the plots of the JL subjects tended to occupy the lower area of the graph.

A cluster analysis was carried out, using the z-scores of the eight variables based on the mean and standard deviation for the entire sample. Because the two variables concerned the phonetic items of intonation and the other six variables, the phonological items of intonation, the analyses were conducted separately using two variables in one and six variables in the other. The results showed that the BN/AN subjects did not form one cluster in the analysis of span and level, the two phonetic items of intonation, as in the dendrogram of Figure 4.28. This means that individual differences went over the boundary between the BN/AN subjects and JL subjects, suggesting that the span and level did not contribute to discriminating between these groups.

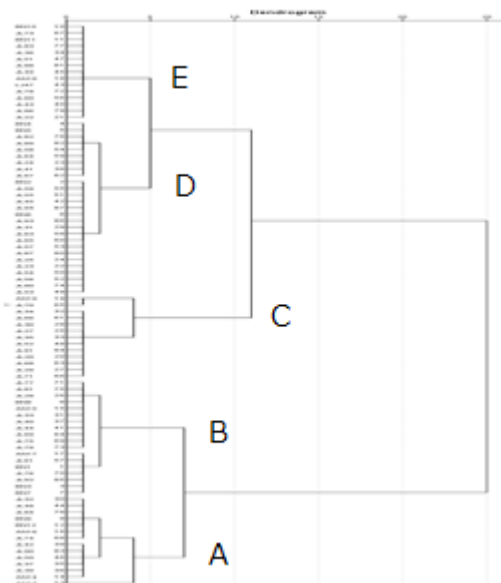


Figure 4.28. Dendrogram output for span and level.



How the BN/AN subjects distributed into separate clusters is displayed in the dendrogram in the figure. When five clusters, A, B, C, D and E, were tentatively selected as in Figure 4.28, Cluster A contained two BN subjects and one AN subject, Cluster B, four BN subjects, Cluster C, one AN subject, Cluster D four BN subjects and two AN subjects and Cluster E, two BN subjects and three AN subjects. This shows that these variables classified the BN/AN subjects into all clusters generated, which was not the case with the variables in the other elements of pronunciation. These two variables were thus regarded as not discriminating between the BN/AN subjects and the JL subjects, and were not submitted to the subsequent statistical analyses.

The other cluster analysis was carried out with the remaining six variables: the score for the nucleus in the long/non-final utterances, the score for the non-nuclear words in the long/non-final utterances, the score for the nucleus in the final utterances, the score for the non-nuclear words in the final utterances, the score for the nuclear tone choice in the falling utterances and the score for the nuclear tone choice in the non-falling utterances. According to the dendrogram output of the analysis, all BN/AN subjects were grouped in one cluster, and two JL subjects were classified into this cluster (see Appendix P for the dendrogram). All JL subjects but two also clustered together at an earlier stage of the clustering process. As a result, four clusters were selected at the point when the BN/AN subjects were grouped. Cluster 1 consisted of 12 BN subjects, 7 AN subjects and 2 JL subjects, and was regarded as representing native speakers. Clusters 2, 3 and 4 were made up of only JL subjects, 32 subjects, 16 subjects and 18 subjects, respectively.

Table 4.29 illustrates the descriptive statistics of all variables for each cluster. Not only do they include the above-mentioned phonetic variables that were excluded from the statistical analyses, but they also show the results of the following three phonological variables that were not submitted to the subsequent analyses as with the two variables for the phonetic items of intonation. Two variables, the scores for the nucleus and non-nuclear words in the final utterances, showed a ceiling effect, and one variable, the score for the nuclear tone choice in the non-falling utterances, showed a floor effect. As noted in Section 3.5.9, these

variables clearly violated the assumption of the statistical tests in terms of the normal distribution. They were thus excluded from a MANOVA and discriminant analysis. Figure 4.29(a), (b) and (c) presents the distribution of the subjects in each score for these variables, which are expressed in percentages.

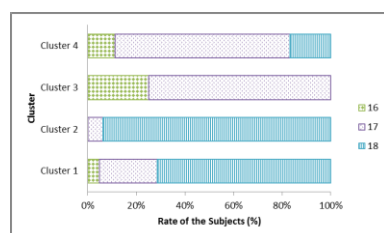
Table 4.29

*Descriptive Statistics of Intonation for Four Clusters*

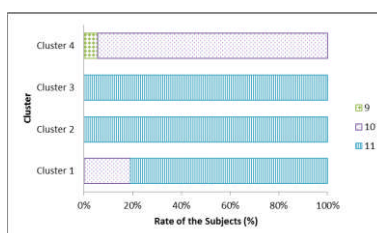
		Cluster 1 (n = 21)		Cluster 2 (n = 32)		Cluster 3 (n = 16)		Cluster 4 (n = 18)	
		M	SD	M	SD	M	SD	M	SD
Phonological items									
Long/	Nucleus	6.52	0.87	3.31	0.90	3.44	0.89	2.89	0.83
non-final	Non-nuclear	19.76	0.94	15.13	0.98	14.56	0.89	14.78	1.52
Final	Nucleus	10.81	0.40	11.00	0.00	11.00	0.00	9.94	0.24
	Nucleus	17.67	0.58	17.97	0.18	16.67	0.50	17.06	0.54
Non falling	Non-nuclear	2.05	0.83	0.33	0.39	0.25	0.46	0.35	0.46
Falling		5.77	0.85	4.80	1.17	5.49	0.91	4.75	1.10
Phonetic items									
Span		4.76	1.50	3.91	1.04	3.67	0.83	4.11	0.93
Level		90.70	14.54	92.02	12.00	89.38	14.08	94.36	13.73

*Note.* The first six variables represent the number of utterances and contexts produced in the typical manner to the BN/AN subjects and their highest possible values were 8, 21, 11, 18, 3 and 7, respectively. The span and level are each expressed in ST and Hz.

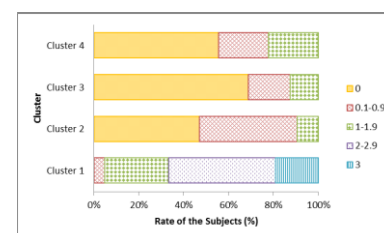
(a) Nucleus in the final utterances



(b) Non-nuclear words in the final utterances



(c) Non-falling utterances



*Figure 4.29.* Distribution of scores for three variables of intonation: (a) the score for the nucleus in the final utterances; (b) the score for the non-nuclear words in the final utterances; and (c) the score for the nuclear tone choice in the non-falling utterances.

Figure 4.29(a) shows that the majority of the JL subjects in Clusters 3 and 4 performed one point worse than those in Clusters 1 and 2, as a whole, in the score for the

nucleus in the final utterances. Figure 4.29(b) shows that nearly all JL subjects in Cluster 4 performed one point worse than those in the other clusters on average for the score for the non-nuclear words in the final utterances. This suggests that almost all subjects in all clusters achieved 17 or 18 points out of 18, the highest possible value, for the nucleus in the final utterances, and 10 or 11 points out of 11, the highest possible value, for the non-nuclear words in the final utterances. This produced a ceiling effect. All subjects in Clusters 2, 3 and 4 were thus regarded as achieving performances like those in Cluster 1. In contrast, Figure 4.29(c) represents a floor effect for the non-falling utterances as to the JL clusters. Apparently, Cluster 2 performed slightly better because the number of JL subjects who gained 0 points, which involves the failure to use a typical tone in all target non-falling utterances, was smaller in Cluster 2 than in Clusters 3 and 4. However, the ultimate tendency was that the majority of the JL subjects in all three JL clusters, Clusters 2, 3 and 4, obtained a score ranging 0 to 1 point in the score for the nuclear tone choice in the non-falling utterances, corresponding to 90.6%, 87.5% and 77.8%, respectively. This means that they failed to use a typical tone type even in one target syntactic and pragmatic context out of three. This created a floor effect, and the subjects in Clusters 2, 3, and 4 were therefore regarded as failing to perform in the same way as those in Cluster 1. These variables were excluded from the subsequent statistical analyses.

The profile in Figure 4.30 was based on the rank averaged across subjects, with the three phonological variables and two phonetic variables being excluded. The subjects who obtained higher scores were ranked higher. Figure 4.31 displays two bar graphs (a) and (b), which show more detail about the performances. Figure 4.31(a) and (b) depicts exactly the same kinds of errors in each cluster as in Figure 4.27(a) and (b). Figure 4.31 also shows scatter diagram (c) to show the results of the phonetic items of intonation, where the level is plotted on the x-axis and the span is plotted on the y-axis.

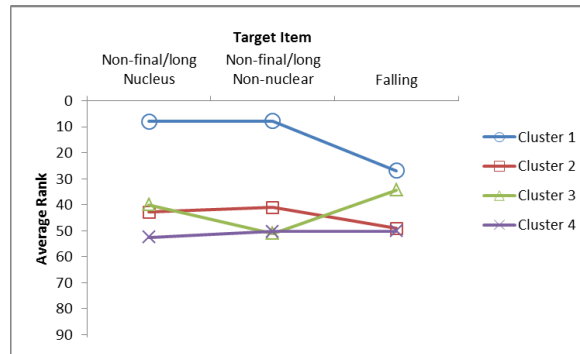
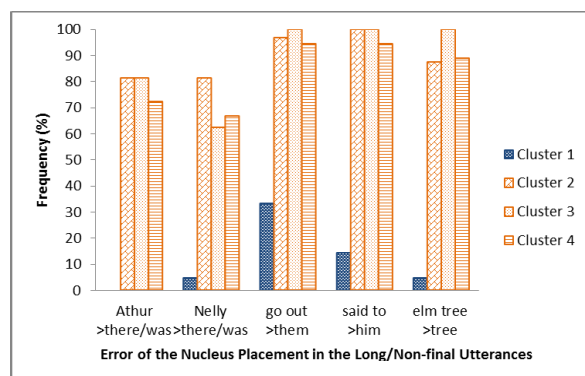


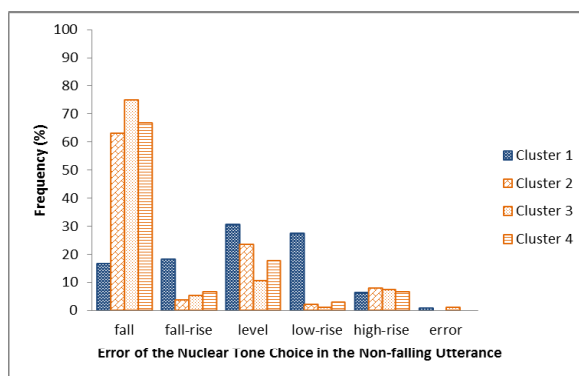
Figure 4.30. Profile of each cluster for intonation.

When focusing on the three variables not yet described, it is noticeable that, as seen in Table 4.29 and Figure 4.31, all JL clusters, Clusters 2, 3 and 4, achieved lower scores than the BN/AN cluster for two variables: the score for the nucleus in the long/non-final utterances and the score for the non-nuclear words in the long/non-final utterances. Cluster 1 clearly performed better for the nucleus in the long/non-final utterances ( $M = 6.52$ ,  $SD = 0.87$ ) than Cluster 2 ( $M = 3.31$ ,  $SD = 0.90$ ), Cluster 3 ( $M = 3.44$ ,  $SD = 0.89$ ) and Cluster 4 ( $M = 2.89$ ,  $SD = 0.83$ ). Although Cluster 4 performed worst in the three JL clusters, the difference between them was very subtle, according to the profile in Figure 4.30 and the scores in Table 4.29. Figure 4.31(a) shows that nearly 100% of the JL subjects located the nucleus on the final word in the utterances, *go out with them*, *said to him* and *and a garden with an elm tree*. The same pattern was applied to the score for the non-nuclear words in the long/non-final utterances. Cluster 1 obtained a higher score ( $M = 19.79$ ,  $SD = 0.94$ ) than Cluster 2 ( $M = 15.13$ ,  $SD = 0.98$ ), Cluster 3 ( $M = 14.56$ ,  $SD = 0.89$ ) and Cluster 4 ( $M = 14.78$ ,  $SD = 1.52$ ). Figure 4.31(a) emphasized that they tended to place an extra nucleus on *there* or *was* on the long utterances, which was one of the common errors in nucleus placement.

(a) Error rate in long/non-final utterances



(b) Error rate in non-falling utterances



(c) Span and level

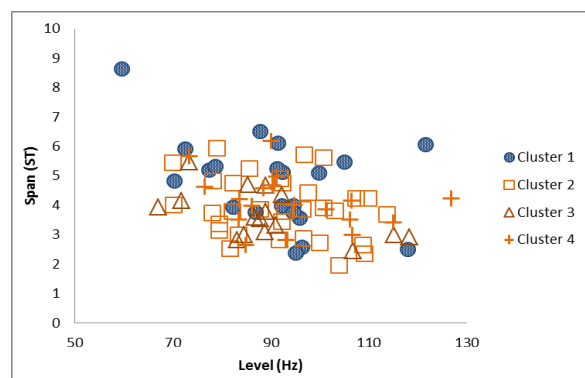


Figure 4.31. Errors in the phonological representation of intonation and the span and level in the phonetic representation of intonation for four clusters: (a) the rate of the extra nucleus placed on *there* or *was* for *There was once a young rat named Arthur* and *There was a kindly horse named Nelly* and the rate of the nucleus placed on the final word for *go out with them*, *said to him* and *and a garden with an elm tree*; (b) the rate of each tone type used for the utterances where a falling tone was not typical; and (c) scatter diagram of the span and level.

The difference in the score for the nuclear tone choice in the falling tone between the clusters seemed less notable. The profile in Figure 4.30 demonstrates that Cluster 3 performed close to Cluster 1, both of which were a little distant from Clusters 2 and 4. However, the JL clusters clearly approximated the BN/AN cluster more closely than the other two variables for the long/non-final utterances. The values provided in Table 4.29 represent the minimal differences in the score for the nuclear tone choice in the falling utterances wherein Clusters 1 and 3 gained higher scores ( $M = 5.77$ ,  $SD = 0.85$  for Cluster 1;  $M = 5.49$ ,  $SD = 0.91$  for Cluster 3) than Clusters 2 and 4 ( $M = 4.80$ ,  $SD = 1.17$  for Cluster 2;  $M = 4.75$ ,  $SD = 1.10$  for Cluster 4).

One thing to be noted for the variables that have already been described is the pattern of the tone type used by the subjects in each cluster in the non-falling utterances. The frequent use of a falling tone even for non-falling utterances is shown by the subjects in the three JL clusters in Figure 4.31(b). The results showed that using a falling tone for the non-falling utterances accounted for 63.02%, 75.00% and 66.67% of the errors in Clusters 2, 3 and 4, respectively. In contrast, the subjects in Cluster 1 were likely to use each tone type in a more balanced manner, 16.54%, 18.11%, 30.71% and 27.56% for a fall, a fall-rise, a level and a low-rise, respectively. The variables of the span and level depicted in Figure 4.31(c) show that the plots of the four clusters overlapping with one another, and there did not seem to be a clear boundary between the clusters. The result that the BN/AN cluster was not well discriminated from the JL clusters by these variables supports the argument given earlier that the span and level failed to differentiate the BN/AN subjects from the JL subjects in this study.

In order to examine whether there was a statistical difference among the clusters, a one-way MANOVA was conducted, using the three variables, the score for the nucleus in the long/non-final utterances, the score for the non-nuclear words in the long/non-final utterances and the score for the nuclear tone choice in the falling utterances as the dependent variables, and the four clusters as the independent variables. Correlations between the dependent variables are given in Appendix Q, which shows that the score for the nucleus in the long/non-final utterances and the score for the non-nuclear words in the long/non-final utterances are highly correlated, but not extremely. A MANOVA was therefore expected to perform fairly. The  $\alpha$  level was set at .01 because the sample size of the largest cluster, 32 subjects, was twice as large as that of the smallest cluster, 16 subjects. Pillai's trace revealed that there was a significant difference among the four clusters for both items,  $F(9.249) = 12.69, p < .001, \eta_p^2 = .31$ .

A post-hoc test was carried out, using a discriminant analysis, in order to investigate which variables contributed to discriminating between the four clusters. Three discriminant functions were found: the first function accounted for 97.3% of the variance, canonical  $R^2$

= .83, the second function accounted for 2.6% of the variance, canonical  $R^2 = .11$  and the third function accounted for 0.1% of the variance, canonical  $R^2 = .00$ . When combined, these three functions significantly discriminated between the four clusters with the Wilk's lambda value of .15,  $\chi^2(9) = 154.56, p < .001$ . After the first function was removed, the second and third functions discriminated between the clusters with the Wilk's lambda value of .88,  $\chi^2(4) = 10.25, p = .036$ . The  $\alpha$  level was set at .01 in this statistical test, and this function failed to differentiate the cluster significantly. The third function, by itself, also failed to differentiate between the clusters significantly with the Wilk's lambda of 1.00,  $\chi^2(1) = .34, p = .561$ . The second and third functions were thus not interpreted. How each function contributed to discriminating between the clusters is depicted by the group centroids in Table 4.30 and the discriminant plot in Figure 4.32, which shows that the first function separated Cluster 1 from the other clusters to a similar extent. Three JL clusters were not differentiated by the variables for intonation tested in the current study.

Table 4.30

*Group Centroids for Intonation*

Cluster	Function
	1
1	3.76
2	-1.01
3	-1.28
4	-1.46

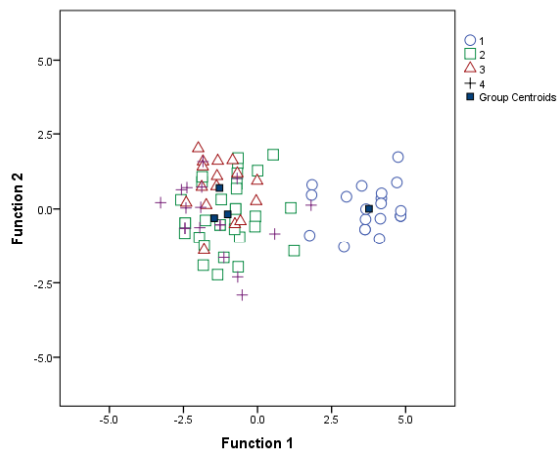


Figure 4.32. Canonical discriminant plot for intonation.

The structural matrix in Table 4.31 reflects the correlations between the variables and each of the functions. The results show that the two scores in the long/non-final utterances, the nucleus and the non-nuclear words, loaded most highly on the first function ( $r = .90$  and  $r = .76$ ). The score for the falling tone statistically tested did not contribute to discriminating Cluster 1 from Clusters 2, 3 and 4. In the long/non-final utterances, the subjects in Clusters 2, 3 and 4 performed more poorly for nucleus ( $M = 3.31$ ,  $SD = 0.90$  for Cluster 2;  $M = 3.44$ ,  $SD = 0.89$  for Cluster 3;  $M = 2.89$ ,  $SD = 0.83$  for Cluster 4) and non-nuclear words ( $M = 15.13$ ,  $SD = 0.98$  for Cluster 2;  $M = 14.56$ ,  $SD = 0.89$  for Cluster 3;  $M = 14.78$ ,  $SD = 1.52$  for Cluster 4) than those in Cluster 1 ( $M = 6.52$ ,  $SD = 0.87$  for nucleus;  $M = 19.76$ ,  $SD = 0.94$  for non-nuclear words). As noted above, Figure 4.31(a) revealed that, in their production, an extra nucleus tended to fall on *there* or *was* in the *there was* utterances, and the nucleus tended to take place on the final word even for the utterances where the BN/AN subjects preferably put the nucleus on the non-final words in the utterances.



Table 4.31

*Structural Matrix for the Correlations between the Variables for Intonation and the Discriminant function*

Variable	Function
	1
Long/non-final: Nucleus	<b>.90</b>
Long/non-final: Non-nuclear words	<b>.76</b>
Falling	.16

*Note.* The variables with the absolute value of correlations with the corresponding functions of .33 and above were highlighted in bold.

The other five variables were not tested in the one-way MANOVA or the discriminant analysis above, as discussed earlier. To sum up the results on these variables, the span, level, score for the nucleus in the final utterances and score for the non-nuclear words in the final utterances did not contribute to discriminating the JL clusters from the BN/AN clusters. The results of the cluster analysis suggested that there was a stronger effect of individual differences for the span and level. A ceiling effect was found in the two scores for the final utterances, which suggests that the subjects in the JL clusters performed in the same manner as the BN/AN subjects. In contrast, a floor effect was detected regarding the score for the nuclear tone choice in the non-falling utterances. The majority of the subjects in the JL clusters gained less than 1 point, while the mean of the BN/AN cluster was 2.05 out of 3. Taken together, this variable discriminated between the JL clusters and the BN/AN cluster.

#### **4.5. Connected speech phenomena**

Table 4.32 and Figure 4.33 show the descriptive statistics of connected speech phenomena regarding the five variables of the BN, AN and JL groups. Figure 4.33(a) and (b) illustrates the frequency of using three types of connected speech phenomena, elision, consonant-to-consonant (CC) linking, and consonant-to-vowel (CV) linking. The mean frequency use is expressed in percentages, so that each category could be compared. In these bar graphs, the phonetic contexts and the rate of use are presented on the x-axis and y-axis, respectively.

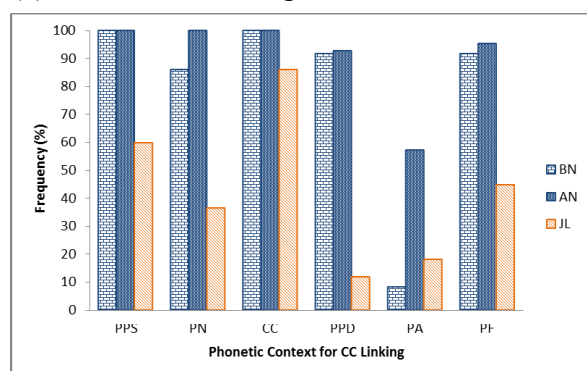
Table 4.32

*Descriptive Statistics of Connected Speech Phenomena for BN, AN and JL Groups*

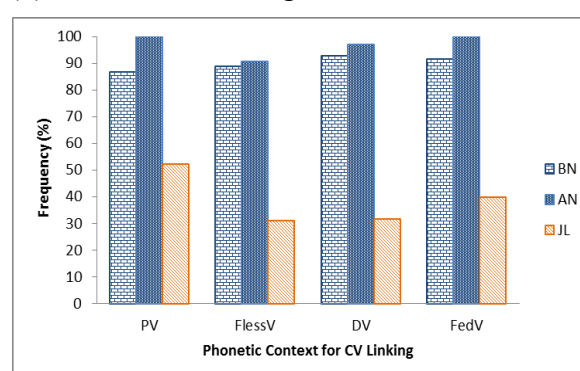
	BN ( <i>n</i> = 12)				AN ( <i>n</i> = 7)				JL ( <i>n</i> = 72)				
	<i>M</i>	<i>SD</i>	<i>Max</i>	<i>Min</i>	<i>M</i>	<i>SD</i>	<i>Max</i>	<i>Min</i>	<i>M</i>	<i>SD</i>	<i>Max</i>	<i>Min</i>	
Elision	6.08	0.90	7.00	5.00	6.71	0.76	7.00	5.00	3.10	1.63	7.00	0.00	
CC linking	Same	6.58	0.52	7.00	6.00	7.00	0.00	7.00	7.00	4.01	1.60	7.00	1.00
	Different	4.67	1.16	5.00	1.00	5.29	0.76	6.00	4.00	1.76	1.18	5.00	0.00
CV linking	Voiceless	7.00	0.95	8.00	5.00	7.71	0.49	8.00	7.00	3.54	2.02	7.00	0.00
	Voiced	12.00	1.13	13.00	9.00	12.71	0.49	13.00	12.00	4.43	2.73	12.00	0.00

*Note.* The highest possible value of each variable was 7 for elision, 7 for the score for CC linking at the same place of articulation and in the same manner of articulation, 6 for the score for CC linking at a different place of articulation or in a different manner of articulation, 8 for the score for CV linking of a voiceless consonant and 13 for the score for CV linking of a voiced consonant. CC = consonant-to-consonant; CV = consonant-to-vowel.

(a) Rate of CC linking use



(b) Rate of CV linking use



*Figure 4.33.* Rate of CC linking and CV linking use in the target phonetic contexts for BN, AN and JL groups: (a) CC linking; and (b) CV linking. PPS = CC linking where a plosive is followed by another plosive produced at the same place of articulation; PN = CC linking where a plosive is followed by a nasal produced at the same place of articulation; CC = CC linking where a consonant is followed by the same consonant; PPD = where a plosive is followed by another plosive produced at a different place of articulation; PA = CC linking where a plosive is followed by a approximant; PF = CC linking where a plosive is followed by a fricative; PV = CV linking of a voiceless plosive; FlessV = CV linking of a voiceless fricative; DV = CV linking of a voiced alveolar plosive, /d/; FedV = CV linking of a voiced fricative.

It is apparent that the JL group was less likely to use elision, CC linking and CV

linking overall, as shown in Table 4.32. As far as Figure 4.33(a) and (b) is concerned, the BN/AN groups used CC linking and CV linking in almost all target contexts, except CC linking where a plosive is followed by an approximant, abbreviated as PA. There was only one target token in this context, *would like*, and 1 subject out of 12 in the BN group and 4 subjects out of 7 in the AN group used CC linking for this.

According to Figure 4.33(a), the phonetic context where the JL group tended to use CC linking more frequently was where a consonant was followed by the same consonant, labelled CC (86.1%). By contrast, they used CC linking much less frequently than the BN/AN groups in the phonetic context where a plosive was followed by another plosive produced at a different place of articulation, abbreviated as PPD (11.8%). CC linking was also less often used when a plosive was followed by an approximant, abbreviated as PA (18.1%), but it was more frequent than in the BN group. As in Figure 4.33(b), the JL group was more likely to use CV linking of a voiceless plosive (52.2%) than that in the other phonetic contexts for CV linking. However, the effect of the phonetic context on the use of CV linking seemed less notable, compared to CC linking.

A cluster analysis was carried out to profile the subjects based on the performances of connected speech phenomena, using the five variables for the frequency of the use of these phenomena: the score for elision, the score for CC linking at the same place of articulation and in the same manner of articulation, the score for CC linking at a different place of articulation or in a different manner of articulation, the score for CV linking of a voiceless consonant and the score for CV linking of a voiced consonant. After these variables were standardized to the z-scores based on the mean and standard deviation for the entire sample, they were submitted to the cluster analysis. The result of clustering is shown in the dendrogram in Appendix R. Two BN subjects were grouped with JL subjects, but the rest of the BN subjects and all AN subjects were grouped together in Cluster 1 at the earliest stage of the clustering process. Although three separate clusters were formed for the two BN subjects and the JL subjects, the sample sizes of these clusters were much larger than that of Cluster 1. Therefore, the clusters with a smaller sample size, which was created at the earliest stage of

the clustering process, were selected. The dendrogram in Appendix R shows that there were five smaller clusters. Because the second cluster from the top in the dendrogram was made up of only seven JL subjects, they were combined with the next closest cluster, the third cluster from the top, following the results of the cluster analysis, in order to avoid this cluster lowering the statistical power of the subsequent analysis. Clusters 1, 2, 3, 4 and 5 were thus each made up of 17, 17, 28, 14 and 15 subjects. Cluster 1 consisted of 10 BN subjects and 7 AN subjects, and was defined as the BN/AN cluster, regarded as representing native speakers. Cluster 2 was comprised of 2 BN subjects and 15 JL subjects, while the remaining JL clusters had only JL subjects.

The descriptive statistics are shown in Table 4.33. A line graph in Figure 4.34 presents the profile of each cluster based on the rank averaged across subjects. Rank was allotted to each subject based on the frequency of using the target connected speech phenomena. The subjects who showed a higher frequency, represented by the higher score, ranked higher. The target connected speech phenomena are indicated on the x-axis and the rank is on the y-axis.

Table 4.33

*Descriptive Statistics of Connected Speech Phenomena for Five Clusters*

		Cluster 1 ( <i>n</i> = 17)		Cluster 2 ( <i>n</i> = 17)		Cluster 3 ( <i>n</i> = 28)		Cluster 4 ( <i>n</i> = 14)		Cluster 5 ( <i>n</i> = 15)	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Elision		6.47	0.80	4.59	1.33	3.43	1.37	2.64	1.45	1.47	0.64
CC linking	Same	6.76	0.44	5.41	1.06	4.75	1.40	2.93	0.92	2.40	0.74
	Different	5.12	0.49	2.88	1.32	1.96	1.04	1.29	0.73	0.73	0.59
CV linking	Voiceless	7.47	0.62	5.88	0.99	2.46	1.71	4.57	1.02	2.20	1.26
	Voiced	12.53	0.51	7.41	2.09	2.86	1.56	7.14	1.70	2.20	1.26

*Note.* The highest possible value of each variable was 7 for elision, 7 for the score for CC linking at the same place of articulation and in the same manner of articulation, 6 for the score for CC linking at a different place of articulation or in a different manner of articulation, 8 for the score for CV linking of a voiceless consonant and 13 for the score for CV linking of a voiced consonant. CC = consonant-to-consonant; CV = consonant-to-vowel.

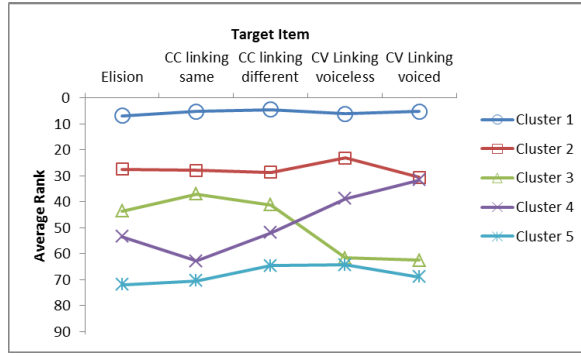
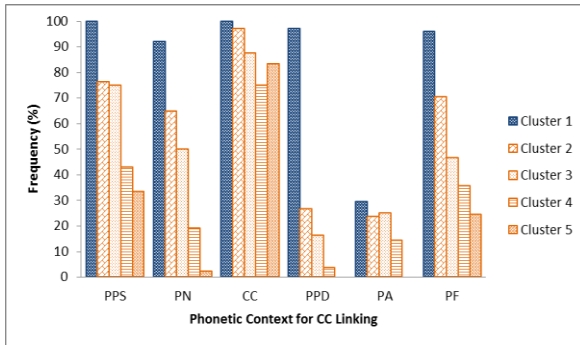


Figure 4.34. Profile of each cluster for connected speech phenomena.

Figure 4.35(a) and (b) displays the results of the CC linking and CV linking in the target phonetic contexts in more detail, where the target phonetic context is shown on the x-axis and the rate of use, on the y-axis. To compare the results of each context, the values were transformed to the mean frequency of the use in percentages by averaging the number of items across subjects in each cluster.

(a) Rate of CC linking use



(b) Rate of CV linking use

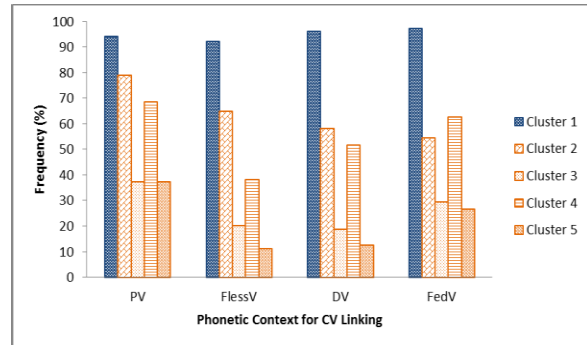


Figure 4.35. Rate of CC linking and CV linking use in the target phonetic contexts for five clusters: (a) CC linking; and (b) CV linking. PPS = CC linking where a plosive is followed by another plosive produced at the same place of articulation; PN = CC linking where a plosive is followed by a nasal produced at the same place of articulation; CC = CC linking where a consonant is followed by the same consonant; PPD = where a plosive is followed by another plosive produced at a different place of articulation; PA = CC linking where a plosive is followed by a approximant; PF = CC linking where a plosive is followed by a fricative; PV = CV linking of a voiceless plosive; FlessV = CV linking of a voiceless fricative; DV = CV linking of a voiced alveolar plosive, /d/; FedV = CV linking of a voiced fricative.

According to the line graph in Figure 4.34, which shows the profile of each cluster, Cluster 1 clearly used elision, CC linking and CV linking in the target context most frequently in all the clusters, and it is followed by Cluster 2. Cluster 3 and Cluster 4 showed the opposite tendency in performance; the former used elision and CC linking more often but not CV linking, and the latter used CV linking more often but not elision and CC linking. Cluster 5 tended to perform most poorly for elision, CC linking and CV linking. These patterns are reflected in the scores for each connected speech phenomena presented in Table 4.33.

What should be noted from Figure 4.35 is the more detailed pattern in the use of CC linking and CV linking in each phonetic context. Cluster 1 almost always used CC linking and CV linking in the target contexts except for CC linking where a plosive was followed by an approximant, PA. By contrast, all four JL clusters used CC linking and CV linking less frequently than Cluster 1, which was especially noticeable for CC linking where a plosive was followed by a plosive at a different place of articulation, PPD, as in Figure 4.35(a) and (b). Even Cluster 2, which most frequently used the connected speech phenomena of the JL clusters as a whole, achieved 26.5% of the mean frequency for this context. None of the subjects in Cluster 5 used CC linking in this phonetic context. In contrast, although the JL clusters were less likely to use CC linking in the phonetic context where a plosive was followed by an approximant, PA, this was applied to Cluster 1, too, as in Figure 4.35(a). As for CV linking, there did not seem as much influence from the phonetic context as in CC linking, in a comparison between the clusters. Clusters 2 and 4 used CV linking more frequently than Clusters 3 and 5 in any phonetic context. A within-cluster difference was that while all JL clusters used CV linking of a voiceless plosive most often, the phonetic context where they used CV linking least frequently varied: CV linking of a voiced fricative for Cluster 2, CV linking of /d/ for Cluster 3 and CV linking of a voiceless fricative for Clusters 4 and 5.

In order to examine the presence of the significant difference between the clusters, a one-way MANOVA was conducted with the five variables of elision, CC linking and CV

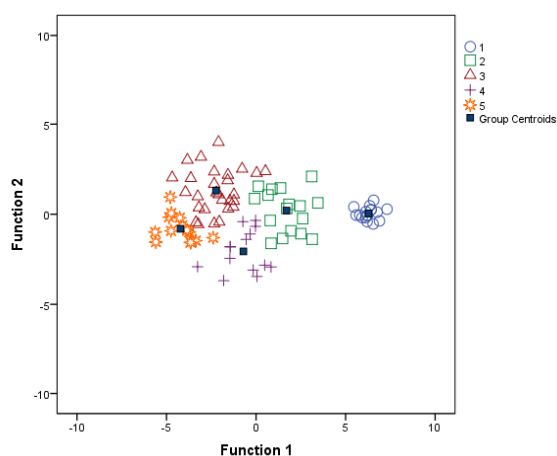
linking being the dependent variables and the three clusters being independent variables. Correlations between the five dependent variables are presented in Appendix S, which shows that whereas CV linking of a voiced consonant and CV linking of a voiceless consonant had a higher correlation, the other variables were moderately correlated. This suggests a MANOVA would work well on these variables. The  $\alpha$  level was set at .01 because the largest cluster, Cluster 3, had a sample size more than 1.5 times as large as the smallest cluster, Cluster 4. There was a significant difference between the clusters,  $F(20,340) = 11.89, p < .00, \eta_p^2 = .41$ , according to Pillai's trace.

To follow up the results of the MANOVA and identify the variables that discriminated between the clusters, a post-hoc analysis was carried out, using a discriminant analysis. Four discriminant functions were found, where the first function explained 89.6% of the variance, canonical  $R^2 = .93$ , the second function explained 9.3% of the variance, canonical  $R^2 = .58$ , the third function explained 0.8% of the variance, canonical  $R^2 = .11$  and the fourth function explained 0.2% of the variance, canonical  $R^2 = .03$ . In combination, these functions discriminated between the clusters at a significant level with the Wilk's lambda value of .03,  $\chi^2(20) = 311.24, p < .001$ . Similarly, the second function, third function and fourth function combined were able to discriminate between the clusters significantly with the Wilk's lambda value of .37,  $\chi^2(12) = 85.76, p < .001$ . After the removal of the second function, however, the third function and the fourth function failed to differentiate between the clusters significantly with the Wilk's lambda value of .87,  $\chi^2(6) = 12.23, p = .057$ . The fourth function on its own did not discriminate between the clusters at a significant level, either, with the Wilk's lambda value of 1.00,  $\chi^2(2) = 2.77, p = .250$ . The third function and fourth function were therefore not interpreted. According to the results of the group centroids in Table 4.34 and the canonical discriminant function plot in Figure 4.36, the first function contributed to separating Clusters 1 and 2 from Clusters 3, 4 and 5. Cluster 5 was maximally distinguished from Cluster 1 above all. The second function contributed to distinguishing between three JL clusters, Clusters 4 and 5 from Cluster 3.

Table 4.34

*Group Centroids for Connected Speech Phenomena*

Cluster	Function	
	1	2
1	6.29	0.05
2	1.70	0.20
3	-2.24	1.31
4	-0.71	-2.06
5	-4.22	-0.81

*Figure 4.36.* Canonical discriminant function plot for connected speech phenomena.

According to the structural matrix in Table 4.35, which shows the correlations between the variables and the discriminant functions, the first function was identified by all variables. The score for CV linking of a voiced consonant most highly loaded on the first function ( $r = .67$ ), which was followed by the score for CV linking of a voiceless consonant ( $r = .44$ ), the score for CC linking at a different place of articulation or in a different manner of articulation ( $r = .43$ ), the score for elision ( $r = .36$ ) and the score for CC linking at the same place of articulation and in the same manner of articulation ( $r = .34$ ). All five variables therefore discriminated Clusters 1 and 2 from Clusters 3, 4 and 5. As shown in Table 4.33 and Figure 4.35, the former clusters showed a higher frequency of CV linking of a voiced consonant ( $M = 12.53$ ,  $SD = 0.51$  for Cluster 1;  $M = 7.41$ ,  $SD = 2.09$  for Cluster 2) than the



latter clusters ( $M = 2.86$ ,  $SD = 1.56$  for Cluster 3;  $M = 7.14$ ,  $SD = 1.70$  for Cluster 4;  $M = 2.20$ ,  $SD = 1.26$  for Cluster 5). However, as reflected in the group centroids in Table 4.34, Cluster 4 differed more minimally from Cluster 2. Figure 4.35(b) supports this, showing that the subjects in Cluster 4 even used CV linking of a voiced fricative more often than those in Cluster 2. Similarly, Clusters 1 and 2 used CV linking of a voiceless consonant more frequently ( $M = 7.47$ ,  $SD = 0.62$  for Cluster 1;  $M = 5.88$ ,  $SD = 0.99$  for Cluster 2) than Cluster 3 ( $M = 2.46$ ,  $SD = 1.71$ ), Cluster 4 ( $M = 4.57$ ,  $SD = 1.02$ ) and Cluster 5 ( $M = 2.20$ ,  $SD = 1.27$ ). Figure 4.35(b) shows that all clusters used CV linking of a voiceless fricative, FlessV, slightly less frequently than CV linking of a voiceless plosive, PV.

Table 4.35

*Structural Matrix for the Correlations between the Variables for Connected Speech Phenomena and the Two Discriminant Functions*

Variable	Function	
	1	2
CV linking voiced	<b>.67</b>	<b>-.51</b>
CC linking same	<b>.34</b>	<b>.69</b>
Elision	<b>.36</b>	<b>.38</b>
CV linking voiceless	<b>.44</b>	-.28
CC linking different	<b>.43</b>	<b>.37</b>

*Note.* The variables with the absolute value of correlations with the corresponding functions of .33 and above were highlighted in bold.

Clusters 1 and 2 also achieved a higher frequency of CC linking. They used CC linking at a different place of articulation or in a different manner of articulation ( $M = 5.12$ ,  $SD = 0.49$  for Cluster 1;  $M = 2.88$ ,  $SD = 1.32$  for Cluster 2) than Cluster 3 ( $M = 1.96$ ,  $SD = 1.04$ ), Cluster 4 ( $M = 1.29$ ,  $SD = 0.73$ ) and Cluster 5 ( $M = 0.73$ ,  $SD = 0.59$ ). Figure 4.35(a) shows that the subjects in Clusters 3, 4 and 5 used CC linking where a plosive was followed by another plosive produced at a different place of articulation, PPD, and CC linking where a plosive was followed by an approximant, PA, less frequently than CC linking where a plosive was followed by a fricative, PF. This suggests that their performances were affected by the

phonetic context. As for CC linking at the same place of articulation and in the same manner of articulation, Clusters 1 and 2 also had higher scores ( $M = 6.76$ ,  $SD = 0.44$  for Cluster 1;  $M = 5.41$ ,  $SD = 1.06$  for Cluster 2) than Clusters 3, 4, and 5 ( $M = 4.75$ ,  $SD = 1.40$  for Cluster 3;  $M = 2.93$ ,  $SD = 0.92$  for Cluster 4;  $M = 2.40$ ,  $SD = 0.74$  for Cluster 5). According to Figure 4.35(a), the difference between Clusters 1 and 2 and Clusters 3, 4 and 5 was smaller in CC linking where a consonant was followed by the same consonant than in CC linking where a plosive was followed by a plosive at the same place of articulation, PPS, and in CC linking where a plosive was followed by a nasal at the same place of articulation, PN. Even the subjects in Clusters 3, 4 and 5 used CC linking where a consonant was followed by the same consonant as often as those in Clusters 1 and 2. It follows that the other two phonetic contexts of CC linking at the same place of articulation and in the same manner of articulation above all discriminated between Clusters 1 and 2 and Clusters 3, 4 and 5.

The score for elision also discriminated Clusters 1 and 2 from Clusters 3, 4 and 5. The former clusters used elision more often ( $M = 6.47$ ,  $SD = 0.80$  for Cluster 1;  $M = 4.59$ ,  $SD = 1.33$  for Cluster 2) than Cluster 3 ( $M = 3.43$ ,  $SD = 1.37$ ), Cluster 4 ( $M = 2.64$ ,  $SD = 1.45$ ) and Cluster 5 ( $M = 1.47$ ,  $SD = 0.64$ ).

The three JL clusters, Clusters 3, 4 and 5, were discriminated by the second function, where Cluster 3 was separated from Clusters 4 and 5. According to the structural matrix in Table 4.35, the following four variables loaded on this function: the score for CC linking at the same place of articulation and in the same manner of articulation ( $r = .69$ ), the score for CV linking of a voiced consonant ( $r = -.51$ ), the score for elision ( $r = .38$ ) and the score for CC linking at a different place of articulation or in a different manner of articulation ( $r = .37$ ). In a comparison of the mean score for each cluster, Cluster 3 more frequently used CC linking at the same place of articulation and in the same manner of articulation ( $M = 4.75$ ,  $SD = 1.40$ ), CC linking at a different place of articulation or in a different manner of articulation ( $M = 1.96$ ,  $SD = 1.04$ ) and elision ( $M = 3.43$ ,  $SD = 1.37$ ) than Cluster 4 ( $M = 2.93$ ,  $SD = 0.92$  for the score for CC linking at the same place of articulation and in the same manner of articulation;  $M = 1.29$ ,  $SD = 0.73$  for the score for CC linking at a different place of

articulation or in a different manner of articulation;  $M = 2.64$ ,  $SD = 1.45$  for the score for elision) and Cluster 5 ( $M = 2.40$ ,  $SD = 0.74$  for the score for CC linking at the same place of articulation and in the same manner of articulation;  $M = 0.73$ ,  $SD = 0.59$  for the score for CC linking at a different place of articulation or in a different manner of articulation;  $M = 1.47$ ,  $SD = 0.64$  for the score for elision). This was also visually apparent in the profile of each cluster presented in Figure 4.34. Figure 4.35(a) illustrates that the clusters more remarkably differed in the phonetic contexts of CC linking where a plosive was followed by a plosive at the same place of articulation, PPS, and CC linking where a plosive was followed by a nasal at the same place of articulation, PN, than in that of CC linking where a consonant was followed by the same consonant, CC.

As shown by the negative correlation of CV linking of a voiced consonant, the subjects in Cluster 3 used CV linking of a voiced consonant less frequently ( $M = 2.86$ ,  $SD = 1.56$ ) than those in Cluster 4 ( $M = 7.14$ ,  $SD = 1.70$ ). The subjects in Cluster 5 used this CV linking least frequently of all the clusters ( $M = 2.20$ ,  $SD = 1.26$ ), and one interpretation could therefore be that the second function differentiated between Cluster 3 and Cluster 4 for CV linking of a voiced consonant. Figure 4.35(b) shows that Cluster 4 clearly used this CV linking in all contexts more often than Clusters 3 and 5, including CV linking of a voiceless consonant, which was not found to discriminate between Cluster 3 and Cluster 4.

#### **4.6. Relationships between the elements of pronunciation**

In order to examine supportive relationships between the elements of pronunciation, the profile of each subject was investigated and correlation analyses were conducted. As noted in Section 3.5.8, the following variables were selected, based on the results in a series of analyses reported above:

1. Vowel quality:
  - Standardized F1 mel value of /ɪ/
  - Standardized F1 mel value of /ɜ:/

2. Vowel duration:
  - PVI value in the /u:-ʊ/ distinction
  - PVI value in the /ɑ:-ʌ/ distinction
3. Plosives:
  - Absolute VOT duration of /p/
  - Absolute VOT duration of /k/
4. Fricatives:
  - SD and skewness of /θ/
  - SD and skewness of /s/
5. Approximants
  - Score for the /r/tokens
  - Score for the /l/ tokens
6. Rhythm:
  - Pitch differences between stressed and weak vowels
  - Intensity differences between stressed and weak vowels
7. Intonation
  - Score for the nucleus and non-nuclear words in the long/non-final utterances
  - Score for the nuclear tone choice in the non-falling utterances
8. Connected speech phenomena:
  - Score for CC linking at the same place of articulation and in the same manner of articulation
  - Score for CV linking of a voiced consonant

The criterion for selection these variables was whether learning was found to some extent in the JL subjects. As will be discussed in detail in Chapter 6, each tested item was defined as easy, learnable or difficult, based on the criteria of learning noted in Section 3.6. Difficult items were, furthermore, categorized into three types, D1, D2 or D3, according to the difficulty level within the difficult items. The difficult items in D1 are those that less than half,

but some, JL subjects learned to produce. The difficult items in D2 are those that the majority of JL subjects did not fully learn, but more than half were learning toward a native-speaker level. The difficult items in D3 are those that less than half or none of the JL subjects were learning to approximate a native-speaker level. With reference to this categorization, the variables were selected in the following order of priority: a learnable item, a difficult item in D1, a difficult item in D2 and a difficult item in D3. When the only candidates of the element were more than two difficult items in D3, which item to be selected was carefully considered in terms of where learning could be observed in the JL subjects.

Difficult items in D3, /ɜ:/ and /ɪ/, were selected for vowel quality. The other difficult item in D3, /u:/, was not selected because the difference in its quality between the BN and AN groups would make it difficult to define the level of learning precisely. The definition of /ɜ:/ and /ɪ/ as the difficult items in D3 was based on both F1 and F2 mel values; however, only the results of F1 mel values were used, considering that a comparison of the vowel distribution depicted in the scatter diagrams showed that tongue height counted more regarding the learning of their articulation.

The /u:-ʊ/ and /ɑ:-ʌ/ distinctions were selected concerning vowel duration. The /u:-ʊ/ distinction was found to be a learnable item, which was expected to show a clear learning difference among the JL subjects. In contrast, fewer than half the JL subjects produced the /ɑ:-ʌ/ distinction in a native-like way. This led the item to be defined as a difficult item in D1. This suggests that while the degree of learning by the JL subjects would be more limited than in the /u:-ʊ/ distinction, some learning was observed.

The absolute VOT durations of /p/ and /k/ were found to be learnable items, and thus, they were selected for plosives. The VOT difference /k/ in the /k-sk/ pair was also defined as a learnable item. However, the VOT durations of /p/ and /k/ were selected because the variable /k-sk/ was originally obtained based on the absolute duration of /k/.

The kurtosis of /θ/ and /s/, and SD of /s/ were found to be easy, and the remaining items were difficult as regards fricatives. These difficult items were, furthermore, defined as D3. Of these difficult items, only the SD of /θ/ and skewness of /s/ showed some differences

among the JL subjects. In contrast, the COG of /s/ and /θ/, the difficult item in D3, did not show enough evidence of learning by the JL subjects. This was why the combined values of SD and skewness were obtained by averaging the rank, as described in Section 3.5.8, and were applied to this analysis. Although the SD of /s/ was defined as easy, the above result for the SD of /θ/ was more strongly emphasized and the combined value of SD and skewness was used as one of the variables here, considering that the learning of fricatives depends on the overall spectral shape.

Although some JL subjects learned the approximants /r/ or /l/ well, they were not more than half in number. This suggests that both approximants were difficult items in D3 for the JL subjects to learn. Only two variables were used in the analysis of approximants, the scores for the /r/ and /l/ tokens. These variables were thus the only candidates to be used for this analysis.

Intensity and pitch were only defined as difficult items in D1 for rhythm. This means that these items had the most potential to be learned by the JL subjects, although they were difficult to learn to the level of native speakers. The remaining two items were defined as difficult items in D3, and thus, the variables of pitch and intensity were selected.

All variables regarding intonation were found to be either easy or difficult. The only options were therefore the difficult items in D3, the score for the nucleus in the long/non-final utterances, the score for the non-nuclear words in the long-non-final utterances and the score for the nuclear tone choice in the non-falling utterances. The first two items were related to one another in that both scores concerned tonality and tonicity. Their combined scores were thus calculated as noted in Section 3.5.8, and selected for analysis along with the score for the nuclear tone choice in the non-falling utterances.

All variables including elision, CC linking and CV linking were defined as D1 for connected speech phenomena. In order to specify which variables better demonstrate the learning of the JL subjects, therefore, the variables that highly loaded on the second function in the discriminant analysis and mainly discriminated the JL clusters from one another were identified. Consequently, the following two variables, by which the second function was

identified, were selected for the present analysis: CC linking at the same place of articulation and in the same manner of articulation and CV linking after a voiced consonant.

The variables noted above were converted to ranks in the entire sample, which made it possible to make a comparison of the elements of pronunciation. The values were ranked according to the performances of the BN/AN subjects, using the raw value or the z-scores based on the mean and standard deviation for the BN/AN subjects, as detailed in Section 3.5.8. Table 3.10 presents the rank data that were used for this analysis. Table 4.36 shows the descriptive statistics regarding the rankings for the BN, AN and JL groups.

Table 4.36

*Descriptive Statistics Regarding Ranks in Each Element of Pronunciation for BN, AN and JL Groups*

	BN				AN				JL			
	(n = 12)				(n = 7)				(n = 72)			
	MDN	SD	High	Low	MDN	SD	High	Low	MDN	SD	High	Low
Vowel quality	11	6	2	23	22	9	12	39	55	16	17	87
Vowel duration	36	16	16	67	39	24	10	71	47	19	7	86
Plosives	36	13	23	65	28	10	13	37	46	18	10	82
Fricatives	18	9	6	34	-	-	-	-	48	15	12	81
Approximants	7	5	1	17	7	7	1	21	50	16	7	78
Rhythm	22	7	6	33	21	19	5	54	55	21	5	91
Intonation	8	3	1	12	14	7	4	22	46	12	17	66
Connected speech phenomena	8	8	1	15	1	4	1	10	48	14	19	80

*Note.* The values correspond to the rank, and therefore, the minimum value is 1 and the maximum value is the number of subjects. Some cases were excluded from the analysis due to missing data or outliers, which resulted in the different number of subjects analyzed in each element of pronunciation. The number of subjects included in the analyses of vowel quality, vowel duration, approximant, rhythm and connected speech phenomena was 91 and those in the analysis of vowel duration, plosives and intonation were 89, 85 and 87, respectively. The data from AN subjects were not used for the analysis of fricatives, whose number of subjects was 84. The values are rounded.

A general pattern is that JL subjects tended to be lower in rank and the BN and AN subjects to be higher in rank. As indicated in the median value of the JL subjects in Table 4.36, they ranked in the 40s or 50s on average. The BN/AN subjects ranked much higher, but

except in vowel duration ( $MDN = 36$ ,  $SD = 16$  for BN;  $MDN = 39$ ,  $SD = 24$  for AN) and plosives ( $MDN = 36$ ,  $SD = 13$  for BN;  $MDN = 28$ ,  $SD = 10$  for AN). This suggests some JL subjects ranked higher than the BN/AN subjects in these elements of pronunciation.

In order to profile the JL subjects, their learning level for each element of pronunciation was represented with a multicategory scheme, 1, 2 and 3, each of which stands for a high level, a middle level and a low level. The first one third, the next one third and the last one third in the ranking were categorized as 1, 2 and 3, respectively. Agreement of the learning level was examined for all pairwise elements of pronunciation. Appendix T illustrates all results for the profiles of learning in a comparison of the two elements of the pronunciation. The profiles where the JL subjects showed a higher percentage of the level agreement were selected and summarized in Table 4.37. Profiles, 11, 22 and 33, suggest that the level in the learning process completely agreed between paired elements, and each of them means that the subject was at a high level, a middle level and a low level in both elements concerned. Table 4.37 shows the five combinations of the elements that had more subjects with profiles of 11, 22 or 33, whose percentages are highlighted in bold.

Table 4.37

*Top 5 Combinations of the Two Elements of Pronunciation for the Level Agreement*

		Top 5 combinations with the highest rate of level agreement				
	Profile	AP-FR	VQ-R	VQ-AP	VD-AP	VD-FR
Levels agreed	11	<b>18.84</b>	<b>15.79</b>	<b>18.92</b>	<b>18.92</b>	<b>13.04</b>
	22	<b>11.59</b>	<b>17.11</b>	<b>13.51</b>	<b>13.51</b>	<b>13.04</b>
	33	<b>15.94</b>	<b>13.16</b>	<b>13.51</b>	<b>13.51</b>	<b>15.94</b>
1-level gap	12	5.80	10.53	8.11	8.11	10.15
	21	10.14	7.90	9.46	9.46	11.59
	23	13.04	7.90	10.81	10.81	8.70
	32	13.04	9.21	13.51	13.51	7.25
2-level gap	13	5.80	10.53	8.11	8.11	10.15
	31	5.80	7.90	4.05	4.05	10.15

*Note.* The values are expressed in percentages. The figures showing the percentage of the subjects with the 11, 22, or 33 profiles are highlighted in bold. AP = approximants; FR = fricatives; VQ = vowel quality; R = rhythm; VD = vowel duration.



As can be seen in Table 4.37, there were five combinations of the two elements of pronunciation where the subjects had profiles with the highest agreement of the level: approximants and fricatives, vowel quality and rhythm, vowel quality and approximants, vowel duration and approximants, and vowel duration and fricatives. What the two elements of pronunciation have in common in each combination seemed to differ from one another, which implies that there was more than a simple pattern concerning the relationships between the elements. At the same time, some of these combinations suggest the possibility that there are supportive relationships between them in the learning process, where two elements are learned, positively supporting one another.

Table 4.38 presents the results of the profiles for the JL subjects, as in Table 4.37, focusing on a higher percentage of the level disagreement in the combinations of the two elements of pronunciation. Profiles 13 and 31 both mean that the subjects were at a high level in one element of pronunciation, but at a low level in the other. These profiles were considered to suggest the lack of a supportive relationship between the two elements. Table 4.38 shows the top five combinations of the two elements where more JL subjects were profiled as a 13 or 31 type, whose percentage is highlighted in bold.

Table 4.38

*Top 5 Combinations of the Two Elements of Pronunciation for the Level Disagreement*

		Top 5 combinations with the highest rated of level disagreement				
	Profile	AP-CO	PL-CO	PL-AP	FR-R	VD-PL
Levels agreed	11	6.49	14.49	8.45	9.33	10.77
	22	7.79	14.49	9.86	14.67	7.69
	33	3.90	8.70	9.86	10.67	9.23
1-level gap	12	15.58	11.59	15.49	13.3	12.31
	21	12.99	4.35	11.27	9.33	12.31
	23	10.39	10.14	7.04	8	10.77
	32	12.99	8.70	11.27	8	10.77
2-level gap	13	<b>15.58</b>	<b>11.59</b>	<b>15.49</b>	<b>13.33</b>	<b>12.31</b>
	31	<b>14.29</b>	<b>15.94</b>	<b>11.27</b>	<b>13.33</b>	<b>13.85</b>

*Note.* The values are expressed in percentages. The figures showing the percentage of the subjects with the 13 or 31 profiles are highlighted in bold. AP = approximants; CO = connected speech phenomena; PL = plosives; FR

= fricatives; R = rhythm; VD = vowel duration.

As shown in Table 4.38, the highest percentage of the two-level gap was found in the profile of the JL subjects for the following combinations: approximants and connected speech phenomena, plosives and connected speech phenomena, plosives and approximants, fricatives and rhythm, and vowel duration and plosives. Of these five combinations, plosives were included in four combinations. This is a notable pattern, suggesting that this element of pronunciation tended to be isolated from other elements of pronunciation.

In order to quantify the relationships between the elements of pronunciation, two correlation analyses were performed, using the entire sample and only the JL subjects. Table 4.39 shows Spearman rank-order correlation coefficients of all pairwise elements when the analyses were carried out within the entire sample. The rankings of all subjects including both BN/AN subjects and JL subjects were used for this correlation analysis, and a two-tailed test was conducted.

Table 4.39

*Spearman Rank-Order Correlation Coefficients between the Elements of Pronunciation for the Entire Sample*

	Vowel quality	Vowel duration	Plosives	Fricatives	Approximants	Rhythm	Intonation
Vowel quality	—						
Vowel duration	.21*	—					
Plosives	.16	.04	—				
Fricatives	.48**	.22*	.16	—			
Approximants	<b>.66**</b>	.09	.18	<b>.54**</b>	—		
Rhythm	.44**	.11	.12	.14	.38**	—	
Intonation	.49**	.17	.17	.30**	<b>.54**</b>	.40**	—
Connected speech Phenomena	<b>.53**</b>	.08	.18	.29**	.44**	.42**	<b>.50**</b>

*Note.* A two-tailed test was conducted. A high correlation is highlighted in bold.

\* $p < .05$ . \*\* $p < .01$ .

When the correlation coefficients in Table 4.39 were interpreted under the criteria that the magnitudes of .3 and .5 suggest a moderate correlation and a high correlation (Cohen,

1988), the following results were indicated. Of the top five combinations of the two elements of pronunciation with a high agreement of the learning level as in Table 4.37, the two combinations were found to be highly correlated: vowel quality and approximants,  $\rho = .66, p < .001$  and approximants and fricatives,  $\rho = .54, p < .001$ . Vowel quality and rhythm were moderately correlated,  $\rho = .44, p < .001$ . In contrast, there was no significant correlation found for vowel duration and approximants,  $\rho = .09, p = .423$ , and vowel duration and fricatives,  $\rho = .22, p = .049$ .

When the results in Table 4.39 were compared with the five combinations where more JL subjects showed had profiles with a disagreement in the learning level as presented in Table 4.38, four combinations were found not to correlate with other elements. The results revealed that there was no correlation between plosives and connected speech phenomena,  $\rho = .18, p = .104$ , between plosives and approximants,  $\rho = .18, p = .103$ , between fricatives and rhythm,  $\rho = .14, p = .200$ , and between vowel duration and plosives,  $\rho = .04, p = .718$ . This correlation analysis was carried out including the BN/AN subjects, which could have resulted in a high correlation. Taking this into consideration, this lack of correlation would confirm that there was no meaningful relationship between these elements. In contrast, approximants and connected speech phenomena were moderately correlated with each other,  $\rho = .44, p < .001$ . However, these combinations had only 18.18% of agreement for the learning level when focusing on the number of JL subjects who were profiled as 11, 22 or 33. Therefore, this moderate correlation could have been estimated to be caused by the effect of the BN/AN subjects.

If the results summarized in Table 4.39 are looked at in terms of the other combinations of two elements with a high correlation, they shows that between intonation and approximants,  $\rho = .54, p < .001$ , between vowel quality and connected speech phenomena,  $\rho = .53, p < .001$ , and between intonation and connected speech phenomena,  $\rho = .50, p < .001$ . The rate of the JL subjects who agreed in their level of learning between these combinations above, regarded as having profiles of 11, 22 or 33, was as follows. According to the order of the higher rate, 32.79% of the JL subjects had an agreed learning

level between intonation and connected speech phenomena, 31.82% of the JL subjects, between the vowel quality and connected speech and 37.68% of the JL subjects, between approximants and intonation. Given that 33.33% was set as the level of chance for being categorized into 11, 22 or 33, based on nine different profiles, the percentage of the level agreement in these combinations was only slightly higher than it. These figures are also much lower than the 46.37% and 45.94% of the level agreement found in the combination of vowel quality and approximants and that of approximants and fricatives, respectively, which had a high correlation and fell within the top five combinations of two elements with the complete agreement of the learning level. Taken together, while these pairwise elements were also found to be highly correlated, this was not clearly supported by the results of the profile concerning the JL subjects.

Although the results of this correlation analysis suggest where there are possible relationships between the elements of pronunciation in this way, there was a possibility that the correlation coefficient was likely to be affected by the performances of the BN/AN subjects in this analysis. The other correlation analysis was therefore conducted excluding the BN/AN subjects. Table 4.40 showed the results of the correlation coefficients for this analysis.

Table 4.40

*Spearman Rank-Order Correlation Coefficients Between the Elements of Pronunciation for the JL Subjects*

	Vowel quality	Vowel duration	Plosives	Fricatives	Approximants	Rhythm	Intonation
Vowel quality	—						
Vowel duration	.07	—					
Plosives	-.03	-.04	—				
Fricatives	.24*	.12	.09	—			
Approximants	<b>.38**</b>	-.11	-.03	<b>.34**</b>	—		
Rhythm	.11	-.05	-.03	-.13	.03	—	
Intonation	-.01	.05	-.02	-.04	.09	.01	—
Connected speech phenomena	.12	-.09	-.05	-.06	-.09	.12	-.01

*Note.* A two-tailed test was conducted. A high correlation is highlighted in bold.

\* $p < .05$ . \*\* $p < .01$ .

Under the criteria of the medium effect and the large effect being emphasized in the interpretation of Table 4.40, a moderate correlation was found only between vowel quality and approximants,  $\rho = .38$ ,  $p = .001$ , and between approximants and fricatives,  $\rho = .34$ ,  $p = .003$ . This correlation analysis was carried out only with the JL subjects, and a high correlation would be unlikely considering that the JL subjects in this study were likely to be homogeneous in their overall proficiency level and language learning background. The moderate correlation found here should thus be seen as meaningful.

The presence of a correlation between vowel quality and approximants and between approximants and fricatives confirmed the results of the profile in Table 4.37 and the correlation analysis including the BN/AN subjects in Table 4.39. On the other hand, vowel quality and rhythm were not found to be correlated even moderately,  $\rho = .11$ ,  $p = .362$ . These two elements suggested some relationship with one another, according to the analyses of the profile and the preceding correlation analysis. However, this was not confirmed by the results of the correlation analysis excluding the BN/AN subjects.

The results of the correlation analysis excluding the BN/AN subjects also confirmed the absence of a supportive relationship between approximants and connected speech phenomena. The correlation analysis including the BN/AN subjects yielded a moderate correlation between them and the profile of this combination was where two elements most disagreed as to the JL subjects' learning level. This moderate correlation was assumed to be due to the effect of the BN/AN subjects, which was confirmed in the correlation analysis excluding the BN/AN subjects.

A summary of the results concerning the relationships between the elements of pronunciation is presented in Figure 4.37. The figure includes the percentage of the JL subjects with a profile of 11, 22 or 33 and the correlation coefficients obtained in the correlation analyses, both including and excluding BN/AN subjects.

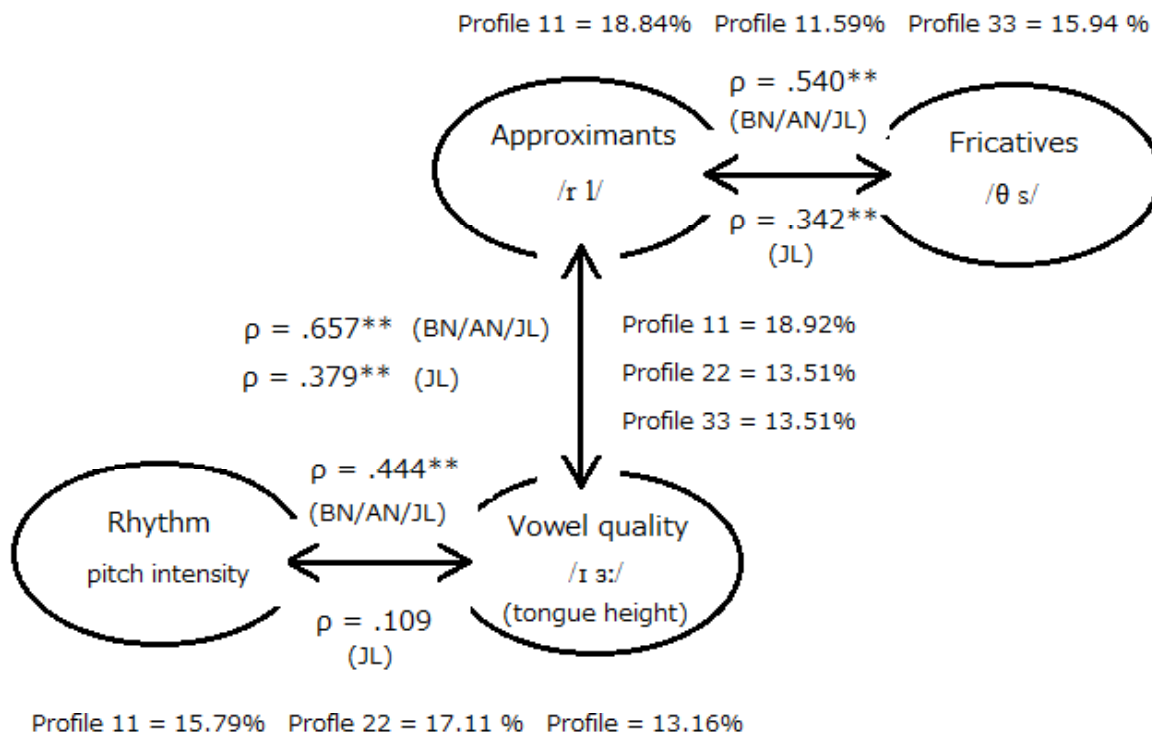


Figure 4.37. Diagram to illustrate the presence of supportive relationships

Figure 4.37 shows that four elements of pronunciation were found to have some supportive relationship with another element of pronunciation, where elements are positively correlated. The correlation analysis conducted only with the JL subjects did not confirm the relationship between vowel quality and rhythm. However, their relationship was also illustrated in the figure, based on the correlation analysis including BN/AN subjects and the results of the JL subject profiles. To summarize, the results showed that there were supportive relationships between approximants and vowel quality and between approximants and fricatives in the learning process. Vowel quality and rhythm also potentially had some weak supportive relationship.

## Chapter 5 Discussion

In each section of this chapter, the results described in Chapter 4 will first be summarized, focusing on the most relevant results for the hypotheses in the present study. It was then determined whether the hypotheses in Chapter 2 were supported or rejected. In order to determine whether the phonetic and phonological items analyzed were easy, learnable or difficult for Japanese learners of English, the rate or amount of learning was defined following the criteria described in Section 3.6.

### 5.1. Vowel quality

#### 5.1.1. Findings

It was found that there were two clusters representing the performance of native speakers of English: 12 BN subjects formed one cluster, and 7 AN subjects formed the other. Each was regarded as representing native speakers of British English or American English, and thus, the results suggest that there were clear phonetic and phonological features characterizing vowel quality of each accent. However, they were combined for further statistical analyses because it was estimated that the combined clusters would still reveal differences from or similarities to the JL clusters.

Only one JL subject out of 72 was grouped into the BN cluster. This suggests that there was a clear-cut difference between the JL subjects and the BN/AN subjects in terms of vowel quality. In other words, it was demanding for the JL subjects to learn vowel quality of all English monophthongal vowels, although it might not be impossible.

The vowels that did not contribute to discriminating between the BN/AN cluster and any of the JL clusters were /i:, e, æ, ʌ, ɑ:, ɔ:, u:, ʊ/. This suggests that these vowels were easy items for the JL subjects to learn to produce. In contrast, major differences between the BN/AN cluster and the JL clusters were found in /I/ and /ɜ:/ for both F1 and F2. The BN/AN subjects produced /I/ with the tongue placed lower and further back and /ɜ:/ with the tongue placed higher and more front. All JL clusters were discriminated from the BN/AN cluster by these vowels. These vowels were thus difficult items for the JL subjects to learn.

However, some degree of learning these difficult vowels was implied by differences within the JL clusters. It was found that the JL clusters differed in F1 of /i:, u/ and F2 of /i:, u, ɪ, ɜ:, ɑ:/. Of these, /ɪ, ɜ:/ were identified as the items discriminating the BN/AN cluster from the JL clusters. The differences in the F2 values in these items among the JL clusters would therefore suggest the rate or amount of learning, even though it did not reach the level of the BN/AN clusters. One JL cluster of 14 subjects was closer to the production of F2 of /ɪ/ to the BN/AN cluster and the other JL clusters of 57 subjects, to the production of F2 of /ɜ:/ to the BN/AN cluster. When one JL subject in the BN/AN cluster was added, /ɪ/ was defined as the third type of the difficult item, D3, under the criterion of requiring more than half the JL subjects to show learning. More than half the JL subjects approximated the subjects in the BN/AN cluster for /ɜ:/, but it was also defined as D3, considering none of the JL clusters showed any sign of learning this sound as to F1, the tongue height. In contrast, the differences in the other items among the JL clusters, F1 of /i:, u/ and F2 of /i:, u, ɑ:/, could only involve individual differences that did not affect the discrimination between the BN/AN cluster and the JL clusters. The three JL clusters were very close to the BN cluster for /u, ɑ:/ and were also closer to both BN cluster and AN cluster for /i:/. Therefore, these differences were considered to be individual differences rather than evidence of learning.

A high back vowel /u:/ was not statistically tested because the two clusters representing native speakers could not be combined for this vowel. The subjects in the three JL clusters produced /u:/ setting their tongue further back than those in the BN cluster and more front than those in the AN cluster. They also placed their tongue lower than both BN and AN clusters. Their /u:/ also overlapped /ʊ/, while /ʊ/ did not contribute to discriminating the JL clusters from the BN/AN cluster. Considering all these results, /u:/ was tentatively defined as a difficult item for the JL subjects to learn to produce. The difference among the JL clusters was not statistically tested, and it was therefore defined as D3.

### 5.1.2. Hypotheses regarding vowel quality

It was hypothesized that similar vowels, /i:, ɑ:, u:/, and new vowels, /ɪ, e, ʌ, ʊ/, would be difficult items, and new vowels, /æ, ɜ:, ɔ:/, would be learnable items for Japanese



learners of English in production. The results showed that /i:, e, æ, ʌ, ɑ:, ɔ:, ʊ/ were easy items, and /ɪ, u:, ɜ:/ were difficult items. Among all 10 vowels, those that supported the hypotheses were /ɪ, u:/. The results in this study rejected the hypotheses for the rest of the vowels. However, the hypotheses for /æ, ɔ:/ were partially supported. These vowels were predicted to be learnable items in the hypotheses, and found to be easy items to learn. In other words, while the prediction of the level of easiness was not supported, the potential for learning these vowels was supported.

Support for the hypothesis regarding /ɪ/, defined as D3, suggests the following things. Although /ɪ/ was defined as a new phone, it was hypothesized that this vowel would be a difficult item for the JL subjects in the present study to learn to produce. This result supported Shimizu (1999). One of the possible explanations for the difficulty of learning the quality of this vowel is that it was possible for Japanese learners of English to produce this vowel by shortening /i:/, located close to /ɪ/ in the phonological vowel space. These two vowels, /i:/ and /ɪ/, could be differentiated using the temporal cue. This could hinder the establishment of a distinct category of /ɪ/ even though it was a new phone, as found in Ingram and Park (1997).

A high back vowel /u:/ was not statistically analyzed, but it was identified as a difficult item as hypothesized by inspecting the scatter diagrams and comparing both F1 and F2 values among the clusters. One possible reason for the difficulty of this vowel is that it requires a clear lip rounding, which is not used for any of the Japanese vowels. Lip protrusion is very evident in this English vowel, known to make a different vowel quality from Japanese /u:/. Another reason could be that the BN cluster and the AN cluster had a different quality. Apparently, /u:/ produced by the BN cluster was much more front than /u:/ produced by the AN cluster. This phenomenon wherein /u:/ is likely to be produced more toward the front in British English is one of the pronunciation changes recently pointed out (Kleber, Harrington, & Reubold, 2011). The results in this study showed that the JL subjects failed to follow this phenomenon and to articulate it further back as found in the AN subjects. The clear difference in the quality between the two accents could make Japanese learners of English fail to capture

the authentic quality of this vowel.

Three vowels, /æ, ɔ:, ɜ:/, for which the hypotheses were rejected, could be discussed together. Although the hypotheses were not upheld, those for /æ, ɔ:/ were partially supported in that they were not learnable items but easy items. The hypothesis for /ɜ:/ that this vowel would be a learnable item was rejected, and it was defined as D3. A higher tongue position is required for this vowel, but the JL clusters showed no evidence of learning for F1. In a sense, these results agreed with Lambacher et al. (2005), which suggested the potential for learning these vowels. The difference in the difficulty of learning that /ɜ:/ was more difficult than /æ, ɔ:/ might be relevant to the findings in Lambacher et al. that the latter vowels improved even without the training. The difference between these vowels, /ɜ:/ and /æ, ɔ:/, could also be discussed from the articulatory point of view: the tongue must be set at a resting position for /ɜ:/, whereas the tongue is placed in a low front position for /æ/ and in a mid back position for /ɔ:/. The latter vowels clearly take effort to move the articulators, which Japanese learners of English might find easier. The former vowel does not require the articulators to move greatly. This articulation might be rather difficult for Japanese learners of English, who are more used to pronouncing full vowels with clear quality.

As regards the remaining five vowels, the hypotheses for /i:, e, ʌ, ɑ:, ʊ/ were not supported, although it was predicted that these vowels would be difficult items. These vowels can be divided into two categories, new phones and similar phones. Three vowels, /e, ʌ, ʊ/, were defined as new phones, and found to be easy items. The other two vowels, /i:, ɑ:/, were defined as similar phones, and found to be easy items.

New phones, /e, ʌ, ʊ/, were supposed to be easy items under the framework of the SLM because they are new. However, taking into consideration that the JL subjects in this study were less experienced, it was predicted that these vowels would still be difficult items for them to learn to produce. The results suggest that these new phones were easy for even less experienced learners to learn. In contrast, /i/ was found to be a difficult item to learn, although it was also defined as a new phone that was difficult to learn, along with /e, ʌ, ʊ/. These mixed results make it difficult to interpret why the hypotheses for /e, ʌ, ʊ/ were

rejected. Above all, /ʌ, ʊ/ and /ɪ/ are very akin to one another in that they all have a long counterpart located close to them in the phonological vowel space, /ɑ:, u:, i:/, respectively. It is therefore not straightforward to explain why /ʌ, ʊ/ were easy items, while /ɪ/ was a difficult item. Which new phones were more likely to be learned by less experienced learners might depend on the tongue position. For some reason, vowels in a low position and those in a high back position might be easier than those in a high front position. Further studies are required to explore the issue.

The remaining vowels /i:, ɑ:/, hypothesized to be difficult, were found to be easy items. The reason a low back /ɑ:/ and a high front vowel /i:/ were found to be easy to learn is not clear. The result of /ɑ:/ did not agree with the findings in Lambacher et al. (2005), in particular. Possibly, due to the phenomenon called undershoot (Lindblom, 1963), more varieties of vowels were allowed for these vowels, located at the edge of the vowel space. When undershoot occurs in a stream of sounds, the target articulation is incompletely reached. A less wide jaw opening for /ɑ:/ and a less front position of the tongue or less spread of lips for /i:/ might have been allowed in the context where the passage was read. Clearly produced sounds tended to show clearer spectral features (Maniwa, Jongman, & Wade, 2008), which implies that the task like reading a passage could cause more unclear articulations to be allowed, such as undershoot. Thus, some differences between the BN/AN cluster and the JL clusters might have been canceled out. Further studies are needed to conclude the status of these vowels.

## **5.2. Vowel duration**

### **5.2.1. Findings**

The BN/AN subjects were divided into four different clusters: seven BN subjects and one AN subject were grouped into Cluster 1, three BN subjects and five AN subjects into Cluster 2, two BN subjects into Cluster 3 and one AN subject into Cluster 4. The first two clusters where 16 BN/AN subjects were put together in total were regarded as representing native speakers. This dispersion of the BN/AN subjects suggests that there were some individual differences among them, but as discussed below, there were also critical

differences between the BN/AN clusters and the JL clusters.

Cluster 1 and Cluster 2 were considered to be BN/AN clusters, and 24 JL subjects were classified into either cluster. This number of subjects was greater than that of the subjects who were grouped in the BN/AN cluster(s) for other elements of pronunciation, which suggests that it was generally easier for the JL subjects to learn the durational aspect of vowels. At the same time, this result might have been related to the broader variation observed in the BN/AN subjects for this element of pronunciation. Unlike other pronunciation elements, not all BN/AN subjects were clustered into one group, as noted above. This suggests that there were some variations in the performance of the vowel duration within the BN/AN subjects, which might have allowed more JL subjects to achieve a native-like level.

Two JL clusters consisting of 18 JL subjects and 27 JL subjects were created, and there was one item that did not discriminate between the BN/AN clusters and the JL clusters: the /ɑ:-æ/ distinction. The distinction of this pair was not found to contribute to discriminating between any clusters statistically. This suggests that /ɑ:/ and /æ/ was an easy item for the JL subjects to learn to discriminate in production using a temporal cue.

In contrast, the durational distinction of /ɑ:-ʌ/ was found to differentiate between the BN/AN clusters and both JL clusters, with the difference between the two vowels smaller for the two JL clusters. Based on the criterion of more than half the JL subjects, it was identified as a difficult item for the JL subjects to learn. Considering 24 JL subjects who were classified into the BN/AN clusters, it was defined as D1.

Similarly, there was also a difference in the production of /i:-ɪ/ and /u:-ʊ/ between the BN/AN clusters and the JL clusters. For the /i:-ɪ/ distinction, one JL cluster of 27 JL subjects was discriminated from one of the BN/AN clusters. However, this JL cluster was not differentiated from the other BN/AN clusters as to the /i:-ɪ/ distinction, and the other JL cluster of 18 JL subjects was not differentiated from either of the BN/AN clusters. It follows that neither of the JL clusters differed statistically from both BN/AN clusters or either of them. This pair was therefore defined as an easy item for the JL subjects to learn to

distinguish with a temporal cue. In the distinction of /u:-ʊ/, one JL cluster of 27 subjects produced a smaller durational difference than at least one of the BN/AN clusters, while the other JL cluster of 18 subjects produced a larger durational difference than the other BN/AN clusters. A larger durational difference could have been regarded as evidence of learning or a lack of learning. However, the JL cluster of 18 subjects was less differentiated from one BN/AN cluster by the /u:-ʊ/ pair, and a larger durational difference led to the clearer separation of long and short vowels. A greater durational difference found in the JL cluster of 18 subjects was therefore regarded as the evidence of learning. More than half the JL subjects learned the /u:-ʊ/ distinction close to the BN/AN subject including 24 JL subjects in the BN/AN clusters, which defined this pair as a learnable item for the JL subjects.

### 5.2.2. Hypotheses regarding vowel duration

It was hypothesized that the durational difference between the long vowels and short vowels in the /i:-ɪ/, /ɑ:-æ/ and /u:-ʊ/ pairs would be easy items and that in the /ɑ:-ʌ/ pair would be a difficult item for Japanese learners of English to learn. The results showed that the difficulty depended on the pair: /i:-ɪ/ and /ɑ:-æ/ were easy items, /u:-ʊ/ was a learnable item, and /ɑ:-ʌ/ was a difficult item. The hypotheses were therefore supported for /i:-ɪ/, /ɑ:-æ/ and /ɑ:-ʌ/, whereas that for /u:-ʊ/ was rejected.

One issue to be discussed here is why no uniform results were found across pairs: /i:-ɪ/ and /ɑ:-æ/ were easy items, /u:-ʊ/ was a learnable item and /ɑ:-ʌ/ was a difficult item. The easiness of /ɑ:-æ/ would be partially due to an allowance of longer variation of this short vowel (Cruttenden, 2014). This would lead to the distinction of this pair as easy for JL subjects, who tended to produce vowels longer than those in the BN/AN clusters, as pointed out in Section 4.1.2. This also gives some clue to address the question of the remaining pairs. Similar to /æ/, one of the BN/AN clusters, which was not discriminated from the JL clusters by the /i:-ɪ/ distinction, produced /ɪ/ longer than /ʊ/ and /ʌ/. That is, the longer the subjects in the BN/AN clusters produced the short vowels, the more easily those in the JL clusters achieved the BN/AN level of distinction. This was evident in the results where the subjects in two JL clusters produced a relatively longer /ʌ/ and one JL cluster produced a longer /ʊ/ than

those in the two BN/AN clusters, compared with /ɪ, æ/. This agreed perfectly with the findings that /ɑ:-ʌ/ was a difficult item and /u:-ʊ/ was a learnable item, whereas /i:-ɪ/ and /ɑ:-æ/ were easy items. Producing short vowels longer than the subjects in the BN/AN clusters could lead to the failure of a native-like durational difference between the long and short vowels.

All this suggests that if the JL subjects follow the pattern in the variation of the short vowels shown by the BN/AN clusters, they can achieve a native-like distinction in the long and short vowel pairs. The subjects in the BN/AN clusters produced /æ/ longest, which was much longer than their shortest short vowel, either /ʌ/ or /ʊ/. This pattern was followed exactly by the subjects in the JL clusters. In their production, the longest short vowel was /æ/, the second longest was /ɪ/, and either /ʊ/ or /ʌ/ was shortest. Even if short vowels tended to be produced longer by the subjects in the JL clusters, it would not cause a problem for the relative durational difference as long as the pattern of the intrinsic duration was maintained.

The subjects in the BN/AN clusters thus did not always maintain the same amount of durational difference for each pair, which is one of the causes of the variation in difficulty across the pairs. However, the variation in the short vowels was not the only cause. The reason that some pairs were more difficult to learn also involves durational variations in long vowels. The subjects in the BN/AN clusters produced /ɑ:/ by far the longest, and /i:/ and /u:/ were both shorter with a similar duration to each other. Unlike the case of the short vowels, the subjects in the JL clusters showed a greater variety of patterns, which made the two JL groups different from one another as well as different from the BN/AN groups. One JL cluster produced /u:/ longest, which was followed by the slightly shorter vowel /ɑ:/, and /i:/ was the shortest of all. This means that the subjects in the JL cluster failed to maintain a relative durational difference between the long vowels and short vowels the way the subjects in the BN/AN clusters did. In contrast, the other JL cluster produced /ɑ:/ longest, followed by /i:/ and then by /u:/. Seemingly, this agreed with the pattern the BN/AN clusters showed, but the difference in the absolute duration between /i:/ and /u:/ was more considerable for the JL cluster.

The subjects in the JL clusters thus produced some short vowels longer than other short vowels, and the duration of the long vowels varied in different patterns from those in the BN/AN clusters. These two things were intertwined, which would have resulted in no uniform result across pairs. Longer short vowels and durational variations of long vowels lead to a shorter lag, which makes it more difficult for Japanese learners of English to produce a native-like distinction between a short vowel and long vowel for some pairs.

Hisagi et al. (2008) argued that there were great variations in duration in English vowels, depending on different contexts. This could be related to the different distinction across pairs. Although they did not discuss durational variation across vowels, Kato and Cox (2006) found that each vowel had its intrinsic durational feature. This is natural, considering that English vowels are phonetic as well as phonological. This would also lead to the variations of short and long lag found in this study. On the other hand, the results that the subjects in the JL clusters had difficulty in attaining a longer lag between the long vowels and short vowels contradict the findings in Kato and Cox and Hisagi et al., if the transfer from their L1 could explain learning of vowel duration. They reported that the durational contrast between long vowels and short vowels was greater in Japanese than in English. In order to identify this contradiction, more studies would be required, but one possible cause of poor performance might be connected with the effect of the stress on vowel duration. The target tokens of short and long vowels in the present study all received a sentence stress, which required the subjects to place a stress on the target words. If the degree of stress could affect vowel duration, there is a possibility that it promoted a larger durational difference between the long and short vowels in the production by the subjects in the BN/AN clusters. The variations in vowel duration, depending on the context, would need to be examined in further studies.

The result that /ɑ:-ʌ/ was the most difficult pair agreed with the findings of Oh et al. (2011). In the present study, the subjects in the JL clusters tended to produce /ʌ/ longer and /ɑ:/ shorter than those in the BN/AN clusters. Overall, the subjects in the JL clusters produced short vowels longer than the BN/AN subjects did, which is understandable because

learners tend to speak more slowly than native speakers of English (Munro, 1995; Munro & Derwing, 1995). However, their long vowels were relatively shorter for their speaking rate. This was true of /ɑ:/, which was produced longest by the BN/AN subjects. This could have been attributed to the difficulty in maintaining the relative durational difference in the /ɑ:-ʌ/ distinction.

At the same time, however, defining this difficult pair as D1 may suggest that the temporal cue is overall accessible to Japanese learners of English even though the durational distinction of this pair was difficult. This accessibility could be explained by the feature hypothesis proposed by McAllister et al. (2002). Their claim is based on the belief that “an L2 category will be difficult to acquire if it is based on a phonetic feature not exploited in the L1 to signal phonological contrast” (p. 231) and they investigated how Estonian learners of Swedish, English learners of Swedish and Spanish learners of Swedish perceived and produced the durational contrast between long and short vowels. Estonian was categorized as a language that used the durational cues phonologically, whereas English and Spanish were categorized as the type of language that did not. They thus hypothesized that Estonian learners of Swedish performed better than the other groups of learners, which was supported. They also claimed that the prominence of a phonetic feature in L1 plays some role in acquiring the L2 under the feature hypothesis, and English and Spanish learners were selected to verify the hypothesis. English and Spanish differ in that the temporal cue plays a more important role in the former language than the latter, which was reflected in the results that the English learners of Swedish perceived and produced Swedish long and short vowels better using the temporal cue. Given that Japanese is comparable to Estonian, the feature hypotheses support the results that the JL subjects successfully employed the temporal cues in their L2 thanks to their phonological function in their L1.

Although this study did not focus on the absolute duration of each vowel, it was also a notable finding that the subjects in the JL clusters produced /æ/ longest of all short vowels on average. A low front vowel, /æ/, was defined as a new phone under the framework of the SLM in the analysis of vowel quality, and was found to be an easy item to learn to produce,



supporting the hypothesis. This suggests that, to the JL subjects, the newness of this vowel was as salient in durational features as it was in spectral features.

### **5.3. Plosives**

#### **5.3.1. Findings**

Fourteen BN/AN subjects were grouped into one cluster, which was considered to represent native speaker performances of VOT. When the absolute VOT duration that this cluster produced was compared with the VOT categories that Lisker and Abramson (1964) and Shimizu (1993) proposed, they fell within the long-lag VOT range under Lisker and Abramson's category but not under Shimizu's category. In a comparison of the absolute VOT values, /p, t/ of the BN/AN cluster were comparable to /p, t/ in Lisker and Abramson, but their /k/ was shorter. The VOT values reported in Shimizu generally showed a longer duration, especially in /k/. Taken together, the BN/AN subjects in this study were likely to produce a shorter VOT for /k/ in particular. However, their VOT can still be recognized as a long-lag VOT, considering that the difference in the materials used in the experiment could lead to these differences.

Three JL subjects out of 67 were classified into the BN/AN cluster. This implies that it might be possible for some JL subjects to learn to discriminate between the plosives with VOT, but it was not easy as a whole. The remaining JL subjects formed two separate clusters of 31 JL subjects and 33 JL subjects, respectively. The former cluster contained four BN/AN subjects.

One of the primary differences between the BN/AN cluster and the JL clusters was the absolute VOT duration of /t/ and the relative VOT difference in /t-st/. Both of the JL clusters produced a shorter VOT for /t/ and a smaller VOT difference in /t-st/. These two items were thus defined as difficult for the JL subjects to learn to produce with native-like VOT, based on the criterion of more than half the JL subjects. The latter variable was closely related to the former variable, and the absolute VOT duration in /t/ in /st/ was longer for the two JL clusters than the BN/AN cluster. It follows that the difference in the relative VOT difference in /t-st/ between the BN/AN cluster and the JL clusters lay in both a shorter VOT

for /t/ and a longer VOT in /t/ in /st/. When the absolute duration was compared, one of the JL clusters of 31 subjects, in particular, did not clearly differentiate the aspirated /t/ and the unaspirated /t/ in /st/ in their VOT.

Similarly, the two JL clusters were discriminated from the BN/AN cluster by the absolute duration of /k/. However, this item was not defined as a difficult item because the two JL clusters were differentiated from the BN/AN cluster for the opposite reason: one JL cluster of 31 subjects produced a shorter VOT and the other JL cluster of 33 subjects produced a longer VOT. As the BN/AN cluster in the present study tended to produce a shorter VOT for /k/, it would be reasonable to define the longer VOT for /k/ by the JL cluster of 33 subjects as a native-like performance. The absolute VOT duration of /k/ averaged across all subjects in this cluster also fell within the range between 60 ms and 100 ms, presented by Lisker and Abramson (1964) and between 70 ms and 100 ms, by Shimizu (1993). This meant that more than half the JL subjects were not discriminated from the subjects in the BN/AN cluster for the absolute VOT duration of /k/, when the three JL subjects clustered into the BN/AN cluster were added. Thus, /k/ was defined as a learnable item for the JL subjects.

The remaining items, the absolute VOT duration of /p/ and the relative VOT difference in /k-sk/ were found to discriminate the JL cluster of 31 JL subjects from the BN/AN cluster. However, the JL cluster of 33 JL subjects was not discriminated from the BN/AN cluster by these variables. This suggests that more than half the JL subjects were not discriminated from the BN/AN cluster, when 3 JL subjects in the BN/AN cluster were added to 33 JL subjects. These items were therefore defined as learnable items for the JL subjects.

Regarding the items defined as difficult, the absolute VOT duration of /t/ and the relative VOT difference in /t-st/, they were categorized into the second type of difficult item, D2. One of the JL clusters of 33 subjects and the three JL subjects in the BN/AN cluster performed better than the other JL cluster of 31 subjects with regard to these variables. This means that while all JL subjects but three in the BN/AN clusters differed from the subjects in the BN/AN cluster, more than half the JL subjects improved the performance of these two

items to the level of the BN/AN cluster. These difficult items were thus defined as D2, suggesting some potential for learning.

### 5.3.2. Hypotheses regarding plosives

It was hypothesized that it would be difficult for Japanese learners of English to produce aspirated voiceless plosives with a long VOT and to discriminate clearly between aspirated and unaspirated voiceless plosives with VOT. The results showed that the VOT of /t/ and the discrimination of aspirated /t/ and unaspirated /t/ with VOT were difficult items, while the VOT of /p/ and /k/ and the discrimination of aspirated /k/ and unaspirated /k/ with VOT were learnable items. The hypotheses were thus supported for the aspirated /t/ and the discrimination of /t/ between aspirated and unaspirated voiceless plosives, but the hypotheses were rejected for the VOTs of /p, k/ and the discrimination of /k/ between aspirated and unaspirated voiceless plosives.

For aspirated voiceless plosives, the hypotheses were not upheld for /p, k/, defined as learnable items, whereas the hypothesis was supported for /t/, solely defined as a difficult item. However, what /p, k/ have in common with /t/ was that some JL subjects still tended to produce a shorter VOT for /p, k/. Shorter voiceless plosives in Japanese learners of English have been frequently pointed out in the literature. Riney and Takagi (1999) argued that Japanese learners of English did not improve their shorter VOT of voiceless plosives even after 42 months had passed. The difficulty of VOT was also replicated in this study. Shimizu (2008) reported that Japanese /p, t, k/ had a VOT of 41 ms, 30 ms and 66 ms, respectively. In the present study, the absolute VOT durations of the subjects in one JL cluster were 31.09 ms for /p/, 36.83 ms for /t/, and 56.26 ms for /k/ on average. A comparison of Shimizu's VOT values against these VOT values suggests that the absolute VOT durations of /p, k/ were even shorter in the present study, while that of /t/ was slightly longer. Assuming that the VOT duration in Japanese can be defined as the starting point for learning English VOT, these values suggest that some JL subjects in this study may not have made much progress in learning.

The difference between /p, k/ and /t/ was in that the former plosives were learnable

items that more than half the JL subjects had learned, while the latter plosive was a difficult item that differentiated both of the JL clusters from the BN/AN cluster. Although Flege and Hiillenbrand (1984) and Flege (1987) claimed that experienced American learners of French achieved an intermediate duration of VOT in their French, the results here suggested that some JL subjects in this experiment succeeded in improving their VOTs for /p, k/ to the level of native speakers. This conformed to Birdsong (2007), although his study is not directly comparable to the present study, as noted in Section 2.2.2. In contrast, for /t/, while both of the JL clusters were discriminated from the BN/AN cluster, one of the JL clusters improved toward a native-like level, defined as D2. This agreed with Flege and Hiillenbrand, and Flege. The difficulty of learning the VOT of /t/ was also found in Joto et al. (2007). They argued that /t/ was most difficult for Japanese learners of English, which was replicated in this study. The shortest VOT of /t/ of the three voiceless plosives in Japanese, reported by Shimizu (2008), would be one reason that /t/ was a difficult item. A detailed acoustic study focusing on this plosive was carried out by Azukisawa, Maeno, Yamada, and Wakita (2001). They investigated the production of /t/, using the following measurements: amplitude, frequency range, span of frequency range, power and duration. They maintained that Japanese learners of English produced a lower amplitude, a narrower frequency range, a weaker power and a shorter duration. Taken together with the results of this study, Japanese learners of English tended to produce voiceless plosives that were acoustically weaker.

As regards unaspirated plosives, the hypothesis of the VOT difference in /t-st/ was supported, but that for /k-sk/ was rejected. Apparently, the two JL clusters were likely to produce a shorter VOT for /t/ and a longer VOT for /t/ in /st/ than the BN/AN cluster, which defined the relative VOT difference in this pair as a difficult item. At the same time, it was also found that one JL cluster produced a longer aspirated /t/ than their unaspirated /t/ in /st/, although the other JL cluster produced the aspirated /t/ and the unaspirated /t/ in /st/ with a similar duration. This suggests that lengthening the VOT duration for the aspirated /t/ could be learned to a certain degree, which defined the relative VOT difference in /t-st/ as D2. The difficulty of the discrimination of /t-st/ was thus closely associated with the result of /t/, as

argued above. The difficulty in shortening VOT for unaspirated /t/ in /st/ would also show that VOT of Japanese voiceless plosives fell between unaspirated voiceless plosives and aspirated voiceless plosives, as discussed in Vance (1987). This suggests that although the discrimination in VOT in the /k-sk/ pair was found to be a learnable item, this would not be attributed to shortening an unaspirated VOT for /k/ in /sk/. One JL cluster closely approximated the BN/AN cluster, and these clusters were not statistically differentiated from one another by the relative VOT difference in /k-sk/. However, this was due to a longer VOT of /k/, not due to the shortening of a VOT for unaspirated /k/. It was thus found that the difficulty in discriminating between aspirated and unaspirated voiceless plosive varied across pairs, but depended more on how much they lengthened the VOTs of aspirated voiceless plosives than how much they shortened the VOTs of unaspirated voiceless plosives. Thus, the key to improving discrimination between aspirated voiceless plosives and unaspirated voiceless plosives would also concern lengthening the absolute VOT duration of aspirated voiceless plosives. The VOTs of unaspirated voiceless plosives themselves might not be defined as learnable items.

## **5.4. Fricatives**

### **5.4.1. Findings**

A comparison of the mean values of /θ/ and /s/ of the BN group in this study with the findings obtained in Jongman et al. (2000) shows that the following patterns were replicated in the present study: higher COG values for /s/ than /θ/, lower SD values for /s/ than /θ/, less positive skewness values for /s/ than /θ/ and more positive kurtosis values for /s/ than /θ/. The BN subjects in this study could thus be considered to represent the production of the native-speaker population for these fricatives.

All BN subjects were grouped into one cluster, where three JL subjects were categorized. This suggests that this element of pronunciation was difficult overall for the JL subjects. Three JL clusters were generated based on the results of the cluster analysis. They consisted of 18 JL subjects 27 JL subject, and 24 JL subjects. The BN cluster discriminated between the two fricatives /θ/ and /s/ in all four items, COG, SD, skewness and kurtosis. On

the other hand, while one JL cluster of 27 subjects differentiated /θ/ and /s/ in COG, the remaining JL clusters did not distinguish between the two fricatives concerning any of these items. The JL clusters therefore differed significantly from the BN cluster in that they failed to differentiate between the two target fricatives in the items tested, except one JL cluster's COG. More than half the JL subjects failed to discriminate between them, which suggests that these two voiceless fricatives were difficult items for the JL subjects to learn to discriminate in production.

When focusing on the articulation, not the discrimination, of each fricative, three variables, kurtosis of /s/ and /θ/ and SD of /s/, did not work to discriminate the BN cluster from the three JL clusters. Native-like production concerning these items was thus defined as easy for the JL subjects to learn to achieve. On the other hand, the remaining variables contributed to discriminating the JL clusters from the BN cluster: COG of /θ/ and /s/, skewness of /s/ and /θ/ and SD of /θ/. The skewness values of /θ/ and /s/ and SD of /θ/ were higher for the BN cluster than for the JL clusters, whereas the COG values of /θ/ and /s/ were both lower for the BN cluster than for the JL cluster. These items were thus defined as difficult for the JL subjects to learn.

There were also differences between the JL clusters. To begin with, SD of /θ/ and skewness of /s/ served to discriminate the JL clusters from the BN cluster. They suggested some potential of the JL subjects learning /θ/ and /s/. Two JL clusters of 18 subjects and 24 subjects approximated the BN cluster in terms of a higher SD value of /θ/. In the skewness of /s/, one JL cluster of 18 subjects performed better than that of 24 subjects. At the same time, the former cluster was not discriminated from the other JL cluster of 27 subjects by this item. This suggests that the two JL clusters of 18 subjects and 27 subjects were closer to the BN cluster for the performance of this item. However, if interpreted with the result that none of the JL clusters discriminated between /θ/ and /s/ using SD and skewness, it would be hasty to conclude that the difference between the JL clusters suggests learning. For SD of /θ/, for instance, the JL cluster of 18 subjects had the higher value than the two other JL clusters, but was much closer to the BN cluster's /s/ rather than /θ/. Thus, while there were some

differences among the JL clusters, the SD of /θ/ and skewness of /s/ were defined as the third type of difficult item, D3. Even this sort of difference among the JL clusters was not found in the rest of the difficult items, such as the COG of /θ/ and /s/ and skewness of /θ/. These items were thus identified as D3, which implies that improvement of these items would be very demanding.

Not only the SD of /θ/ and the skewness of /s/, but the kurtosis of /θ/ and /s/ and the SD of /s/ were also found to contribute to discriminating JL clusters. As reported above, however, the results showed that none of the JL clusters were discriminated from the BN cluster by these variables. In other words, the differences in these items in the JL clusters do not signal the distance from the BN cluster observed in the process of learning, and these differences were thus considered to be due to individual differences within the clusters.

In summary, regarding /θ/, COG, SD and skewness were defined as difficult items, while kurtosis was as an easy item. These difficult items were, furthermore, regarded as D3. Concerning /s/, COG and skewness were defined as difficult items, whereas SD and kurtosis were seen as easy items. The difficult items of /s/, COG and skewness, were both defined as D3. The result that of the four tested items, two or more items were defined as difficult items in D3 for both fricatives suggests that the target fricatives were difficult items to learn to produce as a whole. This definition would be reasonable, given that all four items form the overall spectral shape of each fricative and that more than half the JL clusters failed to discriminate between /θ/ and /s/ in all four items. The two voiceless fricatives /θ/ and /s/ were therefore defined as difficult items in D3, although /θ/ consisting of more difficult items than /s/ implies that /θ/ was more difficult for the JL subjects to learn than /s/.

#### **5.4.2. Hypotheses regarding fricatives**

It was hypothesized that both voiceless fricatives would be difficult items for Japanese learners of English to learn to produce. The results showed that it was difficult for the JL subjects to produce native-like spectral shapes in three spectral moments for /θ/ and two spectral moments for /s/. Both /θ/ and /s/ were thus defined as difficult items, which supported the hypotheses.

The difficulty of these fricatives was predicted under the framework of the SLM and based on its newness, and therefore, the result that /θ/ was challenging to learn is not surprising. Of the four spectral moments analyzed, higher SD values were regarded as characterizing non-sibilant fricatives, attributed to the diffusion of energy (Jongman et al., 2000). However, the subjects in the JL clusters tended to produce a lower SD value, which means that their articulation of /θ/ was accompanied by more energy than required. Thus, /θ/ was categorized as a difficult item, defined as D3 for the JL subjects. This suggests that Japanese learners of English have difficulty in producing /θ/ as a non-sibilant. As Guion et al. (2000) showed, while there was no perceptual counterpart to /θ/, it was difficult for Japanese learners of English to discriminate this fricative from /s/. Substitutions of other consonants were reported by Cairns (1999) and Bada (2002). The difficulty with /θ/ that was found in the productive test in the present study was in accordance with their findings. The difficulty of this fricative, in spite of the phonetic and phonological newness to Japanese learners of English, might be related to the intrinsic difficulty of learning certain sounds. As described above, English dental fricatives are regarded as difficult for learners with various language backgrounds (Mousa, 2014), and this also supported by Wells (2000) and Walker (2010), who noted that they are difficult even for native speakers of English. This clearly suggests that the difficulty in learning this sound could depend on intrinsic factors. Data for the AN subjects were not included in the analysis because of the different recording conditions, but it was found that two AN subjects tended to substitute it for [t].

The alveolar fricative /s/ was also found to be difficult for Japanese learners of English to learn, as hypothesized. While this fricative is phonologically distinct in both languages, English /s/ and Japanese /s/ differ phonetically. This is also acoustically supported by the findings of Sakata et al. (1997). Jongman et al. (2000) suggests that /s/ had the highest COG value, a lower SD value and the highest kurtosis value in a comparison of the four places of articulation for English fricatives. The results of the present study showed that the subjects in all JL clusters produced an even higher value of COG, but it was too high, and reached a statistical difference from those in the BN cluster. A lower SD value was achieved



by all JL subjects as for the BN subjects. Similarly, none of the JL clusters were discriminated from the BN cluster by kurtosis. In contrast, while Forrest et al. did not directly note the skewness of /s/, the present study found that skewness was a difficult item in D3 for all JL clusters. Taken together, whereas the subjects in the JL clusters tended to perform better in /s/ than /θ/, the overall spectral shape of /s/ needs improving to attain a native-like quality in production. This reflects the difficulty of learning this fricative due to the phonetic difference between Japanese and English.

## **5.5. Approximants**

### **5.5.1. Findings**

The following threshold values of F3 were defined for /r/ and /l/, respectively, to judge the target tokens as English /r/ or /l/: 1665 mel Hz and 1671 mel Hz. These thresholds of /r/ and /l/ are equal to 2177 Hz and 2185 Hz, respectively, when mel was converted back to Hz. Comparing them with the values reported in other studies, the threshold of /r/ in this study was close to the F3 value of /r/ shown in the previous research, since Saito and Lyster (2011) reported a value between 2200 Hz and 2300 Hz. This value for /r/ would therefore be reasonable. On the other hand, the threshold of /l/ defined in this study was lower than the value Saito and Lyster reported as F3 value of /l/, 2800 Hz.

Iverson and Kuhl (1996) also gave some indication of whether the threshold of this study was reasonable or not. Within the framework of the NLM, they investigated the perceptual similarity underlying /r/, /l/ and /w/. They assessed the acoustic distance, phonetic identification and category goodness, and found that an F3 value of the best exemplar for /r/ was 1473 Hz, and that for /l/ was 3478 Hz for one group, and 3329 Hz for the other. Compared with these values, the threshold values in the present study, 2177 Hz for /r/ and 2185 mel Hz for /l/, seemed higher and lower, respectively. However, as Iverson and Kuhl noted, the best exemplars tended to be more extreme, and different from average productions. The stimuli were created by synthesizing female speech, and the stimuli were tokens that constituted a simple syllable structure, CV, in Iverson and Kuhl. In contrast, the F3 values of 2177 Hz and 2185 Hz that this study defined were not F3 values for good exemplars or

averages, but thresholds. Male speakers produced them, rather than female speakers, and a passage was used as materials, rather than mono-syllable tokens. The phonetic boundary of F3 between /r/ and /l/, which Iverson and Kuhl found, was also located somewhere between 2067 Hz and 2523 Hz, where the thresholds of this study were located. Taken together, the thresholds of F3 defined in this study would be acceptable.

The cluster analysis, which was conducted using the variables rated based on these yardsticks, generated one BN/AN cluster, comprised of 25 subjects. All BN/AN subjects were grouped into this cluster. The fact that only six JL subjects were classified into the same cluster as the BN/AN subjects suggests the overall difficulty of this element of pronunciation.

Three JL clusters were formed, consisting of 20 JL subjects, 19 JL subjects and 27 JL subjects. All these clusters were discriminated from the BN/AN cluster in both productions of /r/ and /l/. Because more than half the JL subjects were differentiated from the BN/AN subjects, the results suggest that /r/ and /l/ were difficult items for the JL subjects to learn to produce.

On the other hand, there was a difference in performances of both /r/ and /l/ among the JL subjects, which points to some potential for learning, although none of the JL clusters reached the level of BN/AN cluster. The major difference was that the JL cluster of 20 subjects performed better in producing /r/, that the JL cluster of 19 subjects performed better in /l/ and that the JL cluster of 27 subjects performed poorly on both /r/ and /l/. The pattern of the errors they made showed that the JL cluster of 20 subjects preferred pronouncing /r/ even for the /l/ tokens, which suggests that /r/ was an easier item than /l/ for them to learn. In contrast, the JL cluster of 19 subjects was more likely to substitute /l/ for the /r/ tokens, and /l/ was thus an easier item than /r/ for them to learn. The cluster of 27 JL subjects tended to produce a flap-like sound for both /r/ and /l/ tokens more frequently than the two other JL clusters. This suggests that they did not have any preference for learning /r/ or /l/, but rather had been learning neither /r/ nor /l/. These results statistically confirmed that some of the JL subjects had been learning /r/ and/or /l/, while the others had not. However, the number of the JL subjects who had been learning /r/ and/or /l/ did not reach the criterion of more than

half the JL subjects, even after adding the six JL subjects clustered into the BN/AN cluster. Both approximants were therefore defined as difficult items in D3 for Japanese learners of English to learn to produce, although it was found that there was an individual preference about which approximant was learned faster.

### **5.5.2. Hypotheses regarding approximants**

It was hypothesized that /r/ and /l/ would be learnable and difficult for Japanese learners of English, respectively. The results revealed that more than half the JL subjects were discriminated from the BN/AN subjects by both approximants and fewer than half the JL subjects showed improvement in articulating /r/ and /l/. Both approximants were thus identified as difficult items in D3 for Japanese learners of English to learn. This rejected the hypothesis of /r/ and upheld that of /l/.

The salience of /r/ quality and the articulatory dissimilarity of this sound were predicted to facilitate learning this new phone, /r/. That the prediction was not supported suggests that these factors did not adequately predict learning by less experienced learners. The difficulty for less experienced learners of learning to produce /r/ and /l/ to the level of native speakers has often been found in previous studies (Flege, Takagi, et al., 1995; Goto, 1971). Yamada (1993) pointed out that experience of living in the U.S. could affect the perception of these approximants. Saito and Lyster (2011) report that training could help Japanese learners of English to improve the production of /r/. These arguments suggest that some treatment is necessary for Japanese learners of English to learn to use /r/ and /l/. In other words, the subjects in this study, who had never lived in an English-speaking country, had difficulty producing these approximants. The group of 27 JL subjects, nearly half the JL subjects, thus performed poorly for both /r/ and /l/, and showed a higher frequency of substitution of a flap-like sound for /r/ and /l/. The need to use less prominent phonetic features in Japanese phonology would fail to promote learning of these approximants, as the feature hypothesis by McAllister et al. (2002) proposed.

The findings that there were /r/ preference and /l/ preference for learning were unexpected. While the number of the JL subjects was not sufficient to argue that these

approximants were difficult items in D2, some JL subjects showed that they were going through the learning process of /r/. Similarly, other JL subjects were learning /l/. This could be attributed to individual differences in the way of learning, and it should be emphasized that some JL subjects were learning /l/ faster than /r/, in particular. Guion et al. (2000) demonstrated that no Japanese sound was perceptually similar to /l/ in Japanese, although the difference between English /r/ and Japanese /r/ would be more salient than that between English /l/ and Japanese /r/. The hypothesis in the present study, following these claims, defined /l/ as a new phone with a lesser degree of newness to Japanese learners of English, which led to the prediction that /l/ is a difficult item. Some subjects were in the process of learning /l/, however, and this was being learned faster than /r/. While this result did not support the hypothesis, a better performance of /l/ was implied by the results of Aoyama et al. (2004). Although they concluded that the subjects improved /r/ more than /l/, their /l/ tokens were identified as intended more frequently than their /r/ tokens at the first session of their experiment. English /l/ thus possibly sounds more distinct from Japanese /r/ to some Japanese learners of English, whereas it is generally agreed that the difference between English /r/ and Japanese /r/ was more salient.

## **5.6. Rhythm**

### **5.6.1. Findings**

There were two main analyses of rhythm: the PVI values of successive stressed and unstressed vowels and the production of weak vowels in weak forms. As regards the former, it was found that the JL subjects tended to gain a lower PVI value as a whole, which suggests that their stressed vowels and unstressed vowels were not well discriminated from each other with duration. At the same time, the results showed a tendency for frequent pauses in the production of Japanese learners of English. Thus, variability could not be measured accurately. One of the reasons for this would be related to the difficulty or length of materials. Analysis considering the frequency of pauses is required in further studies. The main analyses here were therefore conducted only with the production of weak vowels in weak forms.

According to the results of the production of weak vowels in weak forms, the

BN/AN subjects were classified into one cluster, so that they could form one group. At the same time, the result that only one JL subjects was categorized into this BN/AN group suggests that the phonetic items involving English rhythm were rather demanding for the JL subjects to learn to realize.

Three JL clusters were generated, consisting of 27 JL subjects, 15 JL subjects and 29 JL subjects. All three clusters were discriminated from the BN/AN cluster, concerning two items: the vowel centralization of weak vowels and the durational difference between the stressed and weak vowels. The subjects in the BN/AN cluster tended to centralize the weak vowels more, which was reflected in their vowel distribution of the target weak vowels in the vowel space. They had only one vowel category for the target vowels in their phonological vowel space, one for /ə/. In contrast, the subjects in the JL clusters formed more than one separate category in their phonological vowel space. Each of these categories was easily linked with the categories of the vowels of their corresponding strong form. Weakening the vowel quality was thus defined as a difficult item for the JL subjects to achieve because fewer than half the JL subjects were able to do so. The BN/AN cluster distinguished between the durations of stressed vowel and those of weak vowels more sharply than the three JL clusters, and the JL clusters tended to produce the weak vowels in a longer duration than the BN/AN cluster. All three JL clusters were likely to produce the weak vowels longer than the stressed vowels. It follows that the durational item was also difficult for the JL subjects to learn to use in realizing the authentic English rhythm.

In the remaining two items, the pitch and intensity, two JL clusters of 27 subjects and 29 subjects were found to be discriminated from the BN/AN cluster. The subjects in the BN/AN cluster produced a greater difference in both the maximum pitch and maximum intensity between the stressed vowel and weak vowel than those in the two JL clusters. The former subjects were likely to place a higher pitch and stronger intensity on the stressed vowel than the latter subjects. The JL cluster of 27 subjects even used a higher pitch and stronger intensity on the weak vowel than on the stressed vowel. Under the criterion of requiring more than half the JL subjects to achieve them, the pitch and intensity were also

defined as difficult items for the JL subjects to learn to use.

Although all four items were defined as difficult for the JL subjects, it was found that some JL subjects improved some items toward the level of BN/AN cluster. One JL cluster of 15 subjects was not differentiated from the BN/AN cluster in the pitch and intensity. Another JL cluster of 29 subjects was discriminated from the other JL cluster of 27 subjects by performing better in intensity, although they were still discriminated from the BN/AN cluster. This suggests that some JL subjects, not reaching more than half the JL subjects, developed their realization of intensity to place a stronger intensity on the stressed to the level of the BN/AN cluster. Intensity was therefore defined as D1, although it was a difficult item. One cluster of 15 subjects also differed from the other two JL clusters in terms of pitch, performing at a level where they were not discriminated from the BN/AN cluster, as with intensity. Pitch was thus also defined as D1. For the duration and the vowel centralization, all three JL clusters were discriminated from the BN/AN clusters, and there was no strong evidence that some JL clusters performed better than the others. This led to defining these two difficult items as D3. Pitch and intensity could likely be learned to a native-like level by JL subjects, although all four items were difficult.

### **5.6.2. Hypotheses regarding rhythm**

It would not be appropriate to discuss whether the hypothesis for durational variability was supported or not. The results showed that JL subjects were more likely to pause than BN/AN subjects, meaning that some amount of data were excluded from the analysis. The valid data showed a tendency for JL subjects to lack durational variability in the production of successive stressed vowels and unstressed vowels. However, in order to arrive at some conclusion, an analysis with more data would be required.

As regards the production of weak vowels in weak forms, it was hypothesized that the intensity would be learnable, and the pitch, duration and vowel quality would be difficult for Japanese learners of English to learn to use. The results showed that, in the realization of English rhythm, all four items were difficult items for the JL subjects to learn. The hypotheses of pitch, duration and vowel centralization were therefore supported, but that of

intensity was rejected.

Pitch was found to be a difficult item for JL subjects, upholding the hypothesis. Fujisaki et al. (1986) maintained that it was the only phonetic property used to realize a Japanese pitch accent. However, although this property is phonetically familiar to native speakers of Japanese, pitch accent differs functionally from English rhythm, as described in Section 2.5.1. One possible explanation is that the JL subjects failed to learn to use this item in realizing the weak vowels due to phonetic similarity and the phonological difference. This is in accordance with the theoretical construct of the SLM, which claims that similarities and differences should be considered at a phonetic level.

In contrast, the hypothesis of intensity was not supported. Against the prediction, intensity was a difficult item, not a learnable item. The result was inconsistent with the findings of Lee et al. (2006). They found that weaker intensity was successfully implemented by both experienced Japanese learners and inexperienced Japanese learners of English. One reason for this discrepancy in the findings could be the difference in the materials used. The subjects in the current study read a phonetically-balanced passage, from which the target stressed vowels and weak vowels were extracted. On the other hand, the experiment Lee et al. designed involved the subjects reading target words comprising one stressed syllable and one or two weak syllables in a carrier sentence. Reading a passage requires the subjects to understand the passage, to place a stress on the correct syllable and to articulate a stream of sound. This takes a higher load than reading words in a carrier sentence from the perspective of cognitive load. Thus, the JL subjects may not have been experienced enough to employ them to read the passage used in the experiment, although they might have been able to use the pitch and intensity cues.

Both intensity and pitch were defined as D1, however. This suggests that these cues were difficult items to learn to use, but some JL subjects actually achieved a native-speaker level in using them in the realization of English rhythm. All the results imply that some treatment focusing on these items could improve their production greatly despite the difficulty in learning these items.

The hypothesis about the duration was upheld, which was found to be a difficult item. The JL subjects in this study even produced the weak vowels longer than the stressed vowels. While this supported the hypothesis of the present study, the result did not corroborate Lee et al. (2006), who found that even less experienced learners shortened weak vowels. As noted above, however, the materials of the experiment differed between their study and this study, which would make the two studies less comparable. This would be also applied to Hatano and Kitamura (2014), who measured words.

In contrast, the present study agreed with Sudo and Kaneko (2006) and Sudo (2010a, 2010b) and Aoyama and Guion (2007). They reported that the durational cue was difficult for Japanese speakers, measuring the duration of function words. The weak forms are all categorized as function words, and thus, the findings in these studies were replicated in the present study. There is a general agreement about the difficulty of using the durational cue in the realization of rhythm. Arai and Greenberg (1997) claimed that Japanese morae varied in length, exemplifying the durations of devoiced /i/ and /u/. However, they would not do so unless accompanied by the devoicing of /i/ and /u/.

While the findings about duration in this study were generally in accordance with those in the preceding studies, the potential for this item being learned was still inconclusive. Sudo and Kaneko (2006) argued that pronunciation training in English class could enhance learning to shorten unstressed syllables. However, the duration was defined as a difficult item in D3. Not only did the JL subjects fail to differentiate between the stressed vowels and weak vowels with temporal cues, but they also tended to lengthen the weak vowels more than the stressed vowels. This implies that basic, short treatment would not work well for Japanese learners of English to learn to use the durational cue for the rhythm. As suggested by Sudo (2010a), long-term treatment such as living in an English-speaking country would be required. After all, the better the learners could control duration, the more proficient they would be, as noted by Sudo and Kiritani (1991).

The hypothesis for vowel centralization was also supported. It was found to be a difficult item for Japanese learners of English to learn. Lee et al. (2006) reported that both



experienced Japanese learners and less experienced learners failed to centralize the vowel quality. Hattori and Kitamura (2014) also pointed out the difficulty of learning to articulate the centralized vowel for schwa. Anderson-Hsieh et al. (1994) found that reducing vowels in unstressed syllables was difficult even for Japanese learners of English with high proficiency. The results in this study corroborated these findings. Lee et al. maintained that vowel centralization was difficult for even experienced learners to attain in reading a word in a carrier sentence, which made it reasonable in particular that the JL subjects in this study were not able to centralize vowels in their quality. Not only did their vowels disperse in the vowel space, but the effect of the orthography or strong form was also clearly found. This suggests that they did not simply fail to weaken vowels, but that they did not learn that the vowels could be weakened.

The result for this element of pronunciation agreed with the findings in the analysis of vowel quality. The long counterpart of schwa, /ɜ:/, was defined as a difficult item in D3, as with the centralization of weak vowels. Japanese learners of English are likely to articulate this vowel by replacing it with /a/ in Japanese. This implies that they tend to set their tongue lower than required to articulate /ɜ:/. This tendency was observed in the JL subjects in this study, and none of the JL clusters showed any improvement in F1. Similarly, the target weak vowels in the present study were generally articulated with the tongue more lowered by the subjects in the JL clusters than those in the BN/AN cluster. This suggests the difficulty of the JL subjects attaining this ambiguous vowel quality, whether it is long or short.

## **5.7. Intonation**

### **5.7.1. Findings**

The target utterances were divided into long/non-final utterances and final utterances, which served as the variables of the nucleus placement. They were also categorized into non-falling utterances and falling utterances, which served as the variables of the nuclear tones choice in the test syntactic and pragmatic contexts. The categorization was based on the typical realization of the intonation in the utterance produced by the BN/AN subjects, which needs discussing, first of all.

Two target utterances were long utterances: *There was once a young rat named Arthur* and *There was a kindly horse named Nelly*. These utterances are two of the longest utterances analyzed in this study, starting with *there was*, and therefore, it would be logical to expect the JL subjects to divide the utterances into more pieces of IPs. The non-final utterances consisted of three target utterances: *go out with them*, *said to him* and *and a garden with an elm tree*. The first two utterances ended with a pronoun, which is called a function word from its syntactic function. The basic nucleus placement is the last accented word (O'Connor & Arnold, 1973), and thus, it is common that these pronouns do not work as the nucleus unless they convey new or contrastive information. The third target utterance of the non-final utterances ended with *tree*, which constitutes the second element of the compound noun *elm tree*. In the compound noun, the first element of the word receives more prominence, where the nucleus can occur (Hahn, 1994). Therefore, it is reasonable that *elm* rather than *tree* bears the nucleus, as found in the production of the BN/AN subjects in the present study.

The following five target utterances were identified as said with non-falling tones, categorized as non-falling utterances: *go out with them* in the end of the subordinate clause preceding the main clause context, *he would only answer* and *said to him* in the reporting clause before direct speech context and *There was a kindly horse named Nelly, a cow and a calf* in the lists context. A fall-rise or a level were used for the end of the subordinate clause preceding the main clause context, a level or a low rise for the reporting clause before direct speech context and a fall-rise or a low rise for the lists context. Fall-rise and level tones, found in the end of the subordinate clause preceding the main clause context and in the reporting clause before direct speech context, could signal that more information will follow (O'Connor & Arnold, 1973). These target utterances were, in fact, followed by more information. In this sense, these contexts could also be regarded as what Wells (2006) calls a leading dependent element, where a non-fall is commonly used. Thus, the tones used by the BN/AN subjects for the target contexts in this study can be seen as a typical type of tone in the contexts. A low rise is also a non-fall tone, and therefore, use of this tone is also likely for

the reporting clause before direct speech. Nagamine (2002) found that a rise tone was commonly used by two native speakers of American English in the lists context. O'Connor and Arnold (1973) define a low rise as a common tone for lists, and Wells also notes that a non-fall including a fall-rise and rise occurs commonly on the items of lists followed by another item. Although how the role of a low rise tone should be interpreted is a little controversial (Levis, 2002), it would be fair to define fall-rise and low rise tones as typical tones for these utterances, both of which could be recognized as a rise tone.

A cluster analysis was conducted using the variables obtained based on the judgment above. However, not all variables were included in the analysis. The two variables for the phonetic items of intonation, span and level, were not found to contribute to discriminating the BN/AN subjects from the JL subjects when the analysis was carried out separately. This suggests that it would be easy for the JL subjects to use a similar overall pitch height and pitch range as the BN/AN subjects. Therefore, span and level were easy items for the JL subjects to attain at the level of the BN/AN subjects.

The cluster analysis was thus conducted using only the variables of the phonological items of intonation. The results showed that all BN/AN subjects were classified into one cluster. Two JL subjects were segmented with them, and the remaining JL subjects were grouped together into separate clusters. This suggests that the BN/AN subjects performed in a similar way, different from the JL subjects. Although the nuclear tone choice was a little varied, the nucleus placement was highly consistent across BN/AN subjects. It follows that intonation is grammatical in a given context (Wennerstrom, 1994). At the same time, the result that almost all JL subjects were clustered separately from the BN/AN subjects implies that there was an important difference between the BN/AN subjects and JL subjects with regard to the phonological implementation of intonation.

Three JL clusters were formed, each consisting of 32 subjects, 16 subjects and 18 subjects. The items where major differences were not detected between these JL clusters and the BN/AN cluster were the score for the nucleus in the final utterances, the score for the non-nuclear words in the final utterances and the score for the nuclear tone choice in the

falling utterances. The first two variables showed a floor effect. The score for the majority of the subjects, including both BN/AN subjects and the JL subjects, ranged within the three highest possible points, such as 16, 17 or 18 out of 18, and 9, 10 or 11 out of 11. This suggests that, in the utterances where it was typical that the final word took the nucleus, the JL subjects tended to place the nucleus on the final word of the target utterances without the nucleus falling on an extra or wrong word. The nucleus placement in the final utterances was thus considered to be an easy item for the JL subjects to learn. None of the three JL clusters was statistically discriminated from the BN/AN cluster for the score for the nuclear tone choice in the falling utterances. This score concerned whether a falling tone was used by the subjects for the utterances where a falling tone was defined as typical. The results suggest that a falling tone was an easy item for the JL subjects to learn to use for the utterances where the BN/AN subjects preferably used it as a typical tone.

There were differences between the BN/AN subjects and JL subjects in the score for the nucleus in the long/non-final utterances, the score for the non-nuclear words in the long/non-final utterances and the score for the nuclear tone choice in the non-falling utterances. All three JL clusters were discriminated from the BN/AN cluster for these variables. The production of the JL subjects revealed that the nucleus went on an extra word when the utterance was long. One of the common extra nuclei in the present study occurred on *there* or *was* in the target utterances beginning with *there was*. They also tended to locate the nucleus on the final word of the utterances even when it was not a typical placement. As a result, the non-final words failed to bear the nucleus. Considering that the JL subjects gained higher scores for the final utterances, placing the nucleus on the utterance-final word would be familiar to them, no matter the context. Similarly, the JL subjects did not perform well in the use of non-falling tones, which generated a floor effect. Many of the subjects in the JL clusters attained a much lower score than the BN/AN cluster. The results showed that 90.6%, 87.5% and 77.8%, of each JL cluster achieved a score lower than 1 point, which means that the nucleus did not fall on the typical placement even for one target context. These phonological items of intonation were therefore considered to be difficult items for the JL

subjects to learn to employ. Because none of the JL clusters was discriminated from another JL cluster, these difficult items were defined as D3.

### 5.7.2. Hypotheses regarding intonation

The phonological items were hypothesized as follows: it would be easy for Japanese learners of English to place the native-like nucleus in the final utterances and to use a falling tone where it is common; it would be difficult to locate the native-like nucleus in the long/non-final utterances and to use non-falling tones where they are common. It was hypothesized for the phonetic items that the span would be difficult and the level would be easy for Japanese learners of English to learn to achieve in a native-like way. The results supported all hypotheses, except that about span.

Firstly, the hypotheses on the nucleus placement were upheld. The effect of length in long utterances was revealed. In the long utterance in the present study, an extra nucleus occurred at the earlier part of the utterances, as in the two target utterances, *There was once a young rat named Arthur* and *There was a kindly horse named Nelly*. The occurrence of an extra nucleus is equivalent to more IPs, and thus, factors to break an utterance into more IPs need to be considered. One possible factor is the frequent use of pause (Nagamine, 2002). A series of IPs are not necessarily divided with pauses, but it is one of the signals. By inserting pauses, more IPs and more nuclei may have occurred (Todaka, 1994). More IPs could also be explained by their L1. An accentual phrase in Japanese is a prosodic unit in Japanese. This phonological unit consists of one content word and a function word that follows, which is the smallest unit of prosody. These units in Japanese could have affected their performances of tonality and tonicity. Another possible influential factor is thus their L1 influence, which could make them divide utterances into smaller units. The preceding content words of these units have an intrinsic pitch accent, which would trigger a high pitch at the earlier part of utterance. However, the influence was not strong in short utterances, such as *I don't know*, although the claim of Saito and Ueda (2011) suggests that a higher pitch could be placed on *I* in this utterance. The case was not observed in this utterance produced by the majority of the JL subjects, where the nucleus fell on *know*. It could be assumed that there was an effect from

the utterance length.

The results showed that the JL subjects successfully located the nucleus on the final word of the utterance, but not on the non-final word. The difficulty in locating the nucleus on appropriate syllables has been pointed out (Arimoto et al., 2008; Joto, 1983; Wennerstrom, 1994). The results of this study were in accordance with these previous studies in that the difficulty of the nucleus placement was identified in a certain type of utterances. One possible cause leading the nucleus to exclusively occur on the final word could be the influence from their L1 intonation, as predicted. As noted in Sections 2.6.1 and 2.6.4, the major pitch movement as an intonational phenomenon in Japanese occurs mainly at the phrase-initial and the end of the prosodic phrase. Because of the positive transfer, placing the nucleus on the final word would be easy for Japanese learners of English. In contrast, while focus is a factor in placing prominence on other morae in Japanese, the nucleus in English does not always move forward only with the effect of focus. The nucleus falling on the non-final word is thus a similar phonological item that needs to be recognized.

Secondly, the hypotheses about nuclear tone choice were also supported. The more frequent use of a falling tone by Japanese learners of English was implied by Wennerstrom (1994), who reported that Japanese speakers used low boundary tones more often than English speakers, 46% and 12% of the time, respectively. The JL subjects in the present study used a falling tone even in the utterances where a typical tone is not falling. Joto (1983) also pointed out the difficulty of using a rising tone in declarative sentences. Because the BN/AN subjects in the present study also used low rise and fall-rise tones for the utterances where non-falling tones happened more typically, the results in this study corroborated Joto's findings. The BN/AN subjects used these tones at the end of a subordinate clause preceding a main clause, for a reporting clause before direct speech and for lists, but the JL subjects failed to use them in these contexts. The BN/AN subjects also used a level tone, at the end of a subordinate clause preceding a main clause and a reporting clause before direct speech. This was not successfully implemented by the JL subjects, either. This is possibly due to differences between Japanese and English in the phonological dimension and the semantic

dimension. Japanese does not have a fall-rise and it does not use a low rise or level in these contexts. A fall-rise was defined as a new phone with a low degree of newness because there has been no claim that it is possible for Japanese learners of English to learn to use this tone, to the author's knowledge. This definition led to the prediction that a fall-rise would be difficult, which was confirmed for the JL subjects in the present study. In contrast, low rise and level tones were predicted to be difficult, being defined as similar. The result confirmed the difficulty of these tones. This suggests a relevance of the semantic dimension in learning these tones. Thus, a closer comparison of these tones in the semantic dimension will help clarify what Japanese learners of English need to focus on in learning them.

Finally, regarding the phonetic items of intonation, while the hypothesis about level was upheld, that about span was not. The JL subjects tended to use a similar height of pitch to the BN/AN subjects, as predicted, and they also used a similar amount of the span to the BN/AN subjects, against the prediction. Ohara (1992) and Tsuji (2004) claimed that the Japanese male speakers maintained their pitch height regardless of the language that they spoke. This implies that there is no characteristic feature in the level of Japanese male speakers, unlike that of Japanese female speakers. The present study also found that level was not a phonetic item that discriminated any JL cluster from the BN/AN cluster. On the other hand, the prediction about span was not confirmed, although the span of the JL subjects was predicted to be lower than that of the BN/AN subjects. The average value was slightly smaller for the JL subjects, but it did not reach a statistically significant level. This was not consistent with the findings in previous studies (Joto, 1983; Maeda, 2005; Narita & Tanaka, 2012; Sato, 1999; Todaka, 1994). One of the possible reasons for this is that the BN/AN subjects in this study produced a narrower span than expected. Mennen (2007) reports 7.16 ST as the mean value for the span of monolingual speakers of English, but the present study obtained 4.81 ST as the BN/AN's mean value. This narrower ST in the present study could mean that none of the JL clusters were differentiated from the BN/AN cluster by the span. One possible cause of the narrower span of the BN/AN subjects could be the method of study, where only the last three sentences of the whole passage were measured. This part contained

only narrative sentences, concluding the story, which might have affected the reading style of the subjects. While span was identified as an easy item for Japanese learners of English to learn, further research would not only verify the results of the present study but also the method concerning where to measure (Mennen, Schaeffler, & Dickie, 2014) and what scale to use (Toivanen, 2014).

## **5.8. Connected speech phenomena**

### **5.8.1. Findings**

According to the results of the cluster analysis, two BN subjects were not grouped with the other BN/AN subjects, but the rest of the BN/AN subjects were clustered together. None of the JL subjects were classified into this BN/AN cluster, which suggests a clear difference between them in the use of connected speech phenomena tested in the present study. It was also found that the subjects in the BN/AN cluster were highly likely to use all three connected speech phenomena, elision, CC linking and CV linking, except CC linking where a plosive was followed by an approximant. It follows that the BN/AN subjects used these connected speech phenomena more frequently than the JL subjects.

Four JL clusters were formed, consisting of 15 JL subjects, 28 JL subjects, 14 JL subjects and 15 JL subjects. The two BN subjects not classified into the BN/AN cluster were grouped with 15 JL subjects. Whereas the JL cluster of 15 JL subjects was not discriminated from the BN/AN cluster, the remaining three JL clusters were differentiated from the BN/AN cluster for all five variables tested. Elision, CC linking and CV linking were thus defined as difficult items for the JL subjects to learn to use, despite a slight difference depending on the phonetic context, as noted later in this section.

The results showed that at least one JL cluster consisting of 15 JL subjects was not differentiated from the BN/AN cluster, which defines the type of all these difficult items as D1. More detailed differences among the JL clusters did not affect the definition of the difficulty level, as follows. According to the results, differences were found among the JL subjects in all five items. The JL cluster of 15 JL subjects were differentiated from the other three JL clusters for elision and CC linking, but not discriminated from the BN/AN cluster.



The cluster of 28 JL subjects, additionally, approximated the level of BN/AN cluster more than the remaining two JL clusters. This means that the subjects in the JL clusters of 15 and 28 subjects used elision and CC linking more often than those in the other JL clusters. The differences in CV linking depended on the phonetic context. There was a difference in CV linking of a voiceless consonant between the cluster of 15 JL subjects and the other JL clusters, where the former more frequently used this connected speech phenomenon. No difference was found among the rest of the three JL clusters. On the other hand, both the cluster of 14 JL subjects as well as the cluster of 15 JL subjects showed a higher frequency in using CV linking of a voiced consonant. This suggests that the cluster of 14 JL subjects performed closer to the BN/AN cluster, following the cluster of 15 JL subjects.

The above points concern the overall performance of elision, CC linking and CV linking in a comparison against the BN/AN cluster. The use of CC linking and CV linking were found to vary in more detailed phonetic contexts, especially for CC linking. For CC linking, there are two notable findings concerning the effect of the phonetic context. One is that the JL clusters tended to use CC linking at the same place of articulation and in a different manner of articulation more frequently than CC linking at a different place of articulation or in a different manner of articulation, as a whole. There was a clear pattern that all JL clusters used CC linking where a consonant was followed by the same consonant, in particular. The other was that they showed a lower frequency in using CC linking at a different place of articulation or in a different manner of articulation especially in a sequence of two plosives at a different place of articulation and that of a plosive and an approximant. In the latter context, however, the BN/AN cluster also used CC linking by far the least often of all the target phonetic contexts. In contrast, the effect of the phonetic contexts on the use of CV linking seemed smaller than observed in CC linking. The predicted effect of the voicing of the preceding consonant seemed rather limited, and only a slightly more frequent use of CV linking was found for CV linking of a voiceless plosive than for CV linking in the other phonetic contexts.

### 5.8.2. Hypotheses regarding connected speech phenomena

It was hypothesized that it would be difficult for Japanese learners of English to learn to use elision, CC linking and CV linking. The results showed that these connected speech phenomena were difficult items for Japanese learners of English, which suggests that the hypotheses were all supported. Elision, CC linking at the same place of articulation and in the same manner of articulation, CC linking at a different place of articulation or in a different manner of articulation, CV linking of a voiceless consonant and CV linking of a voiced consonant were all found to be difficult to learn to use.

Elision, CC linking and CV linking are connected speech phenomena that are not usual in Japanese. As described in Section 2.7.1, Japanese has a phenomenon corresponding to elision and linking in English. However, the critical difference in these connected speech phenomena between Japanese and English is that they frequently occur at word boundaries in English. This is less common in Japanese, and not a noticeable phenomenon, if any. While these connected speech phenomena would be new to Japanese learners of English, it was predicted that the newness would be low from an articulatory perspective. They were thus still difficult for Japanese learners at the proficiency level that this study targeted, who had learned English under the common curriculum in Japan. These results conformed to those of Matsui (1998), Hieke (1984), Anderson-Hsieh et al. (1994) and Maxwell (1997), corroborating the difficulty for less experienced learners learning to use elision and linking.

Although the results for the research question on connected speech phenomena showed that there seemed to be some effects of phonetic context on the use of CC linking and CV linking, all four conditions were defined as D1. This implies that the tested phonetic context did not have a considerable difference in the level of difficulty. However, more detailed observation of the phonetic contexts revealed some differences.

The JL subjects generally tended to show a higher frequency use of CC linking at the same place of articulation and in the same manner of articulation than CC linking at a different place of articulation or in a different manner of articulation, although it did not surface statistically. Possibly, it would be phonetically more natural for speakers to articulate the sequence of the same place of articulation and the same manner of articulation by linking

sounds. The phonetic context where the same consonants were next to one another in particular was found easier. The majority of the JL subjects connected these sounds, rather than separately articulating them.

Of the three phonetic contexts for CC linking at a different place of articulation or in a different manner of articulation, the sequence of a plosive and a fricative was easier than that of a plosive and an approximant and that of a plosive and a plosive at a different place of articulation. The target items for the former sequence were *like this*, *about face* and *said the*. Of these, *said the* was the one that caused most JL subjects, 61 out of 72, to use CC linking. In contrast, 15 and 21 JL subjects used CC linking for the first two items. All these figures were still greater than those of the subjects who used CC linking in the two other phonetic contexts. However, this difference suggests a difference in the difficulty even in the tokens categorized as one phonetic context, which would furthermore contribute to explaining the effect of the phonetic context. Articulatory differences between *said the* and the other target items, *like this* and *about face*, are that the preceding plosive is voiced, and the last sound of the preceding word and the first sound of the following word are apical consonants, sharing the same passive articulator. These two conditions would have made it easier for the JL subjects to use CC linking. Using the same passive articulator to promote CC linking is in accordance with the results that CC linking at the same place of articulation and in the same manner of articulation was more frequently used than CC linking at a different place or articulation or in a different manner of articulation, as a whole. A voiced consonant also tends to have weaker energy, which could facilitate the use of CC linking. More detailed phonetic features such as the passive articulator and voicing would thus also be involved in the use of CC linking, not only the place of articulation and the manner of articulation.

The difficulty in CV linking seemed lowest for CV linking of a voiceless plosive. There is no convincing explanation why CV linking of a voiceless plosive was easier for the JL subjects than CV linking of a voiceless fricative, which also constituted the target items for CV linking of a voiceless consonant. The most frequent use of CV linking of a voiceless plosive was found for the target item *make up*, where 65 JL subjects produced it, linking /k/

and /ʌ/. This target is one of the loan words in Japanese, meaning *cosmetics*. Therefore, the familiarity with this token might have resulted in a greater use of CV linking in this target, although it meant something different in the passage. However, it is not common that the linking heard in English occurs to this token in Japanese. It is pronounced by inserting a vowel between /k/ and /ʌ/, as in [meikuappu], and thus, it would be more logical to look for a possible explanation based on the phonetic context rather than the influence of loan words. One possibility is that plosives create the phonetic context where connected speech phenomena are more likely to occur. CC linking in the sequence of a plosive and an approximant was less frequent than that in the sequence of a plosive and a fricative, as noted above. This was applied even to the BN/AN subjects. This might suggest that connected speech phenomena more frequently occur with consonants with a higher degree of stricture, or a lower sonority.

## **5.9. Relationships between the elements of pronunciation**

The second research question asked whether there is any supportive relationship between the elements of pronunciation in the learning process. In order to deal with this, the profile of the JL subjects was studied and the correlation analyses were conducted. The results showed that there were some relationships between vowel quality and approximants and between approximants and fricatives. There seemed to be a lesser degree of relationship between vowel quality and rhythm. Although it was not confirmed in the correlation analysis excluding the BN/AN subjects, a summary of the profile and the correlation analysis including the BN/AN subjects suggested some potential relationships between these two elements, too.

In terms of the presence of supportive relationships, it would be reasonable that vowel quality and approximants, and vowel quality and rhythm have such relationships. Vowel quality and approximants are both segments, and the former is categorized into vowels and the latter into consonants conventionally. Approximants are the closest to vowels of all consonants, however, which are true of /r/ and /l/. The degree of the stricture when approximants are articulated is much smaller than that required to articulate other consonants.

A similarity between the two elements of pronunciation is also reflected in the acoustic measurement. The vowel quality is measured with F1 and F2, and approximants, F3. Some vowels such as /u:/ and /ʊ/ are also known to be characterized as F3 values. The features of lip rounding and tongue curling appear in F3, and it is thus theoretically logical that vowel quality and approximants are related to each other. This supports the finding that there is a supportive relationship between the two elements.

Similarly, it is also reasonable that there may be a supportive relationship between vowel quality and rhythm. There are various acoustic measurements to identify the characteristics of English rhythm, and the present study employed the method of measuring weak vowels as one measurement. However, it should be noted that the two variables selected from the rhythm for the analysis were the pitch and intensity differences between stressed and weak vowels, which could offer more profound suggestions. That is, the relationship between the two elements implies that the learning of pitch and intensity for the weak vowels have some relationship with that of vowel quality of monophthongs. As de Bot and Larsen-Freeman (2011) noted, all components of the elements involve the development, which is complicatedly interconnected, leading to a nonlinear effect. Possibly, the vowel quality of monophthongal vowels is directly related to the vowel quality of weak vowels. At the same time, the vowel quality of weak vowels would be affected or interrelated with the other components of rhythm, such as pitch, intensity and duration. This complexity of the relationship within the system of pronunciation might have made the presence of the relationship between rhythm and the vowel quality inconclusive in the results of the analysis.

In contrast, it is slightly difficult to give a lucid explanation of the unexpected finding that a supportive relationship lay in the approximants and fricatives in the learning process. From the theoretical perspective, the articulation or acoustic features have nothing in common. Although both are categorized as consonants, the degree of the stricture is much larger for fricatives, which accompany air turbulence. It makes them sound different from approximants. One possible explanation for their relationship would concern the level of difficulty. It is likely that the two elements are equally difficult for Japanese learners of

English. This could cause Japanese learners to learn one element as they learn another, corresponding to the supportive relationship defined in the current study. Learning to produce one difficult item in one element might even cultivate the ability to articulate difficult items, indirectly promoting learning another difficult item in another element. The results suggest that whether Japanese learners have learned one element could at least be a signal of whether they have learned the other element, but further examination is necessary as to whether this relationship could lead approximants and fricatives to support one another in the learning process.

No evidence of relationships between the remaining elements was found in this study. The results suggested in particular that plosives were isolated from other elements in the learning process. This implies that the learning of VOT would be less likely to be enhanced by the learning of other elements, or to enhance that of other elements. Similarly, no relationship was identified between connected speech phenomena and rhythm, for instance, although they were predicted to be related to one another, from the articulatory point of view that connected speech phenomena take place to maintain rhythm in a sense. Against predictions, the case was also applicable to combinations between vowel quality and vowel duration and between vowel duration and rhythm. The findings that plosives lacked relationships with other elements and that vowel duration was not related to vowel quality are especially intriguing. Further studies are required to explain the absence of relationships between these combinations, but this finding implies that the temporal cue is distinct for Japanese learners of English. This is consistent with the results in the analysis for the first research question, which demonstrated that plosives and vowel duration were more learnable to the JL subjects as a whole.

## **5.10. General Discussion**

This section will first discuss the key findings of this study in relation to the hypotheses based on the SLM and the LILt. The two research questions in this study will then be resolved: (1) which phonetic and phonological items are easy, learnable or difficult for Japanese learners of English who only have learned English under the curriculum of Japanese

English education; and (2) whether there is any relationship between the elements of pronunciation in the process of learning.

### 5.10.1. Hypotheses and models

A summary of whether the study hypotheses were confirmed is presented in Tables 5.1 and 5.2. Table 5.1 shows the items for which the hypotheses were supported in terms of prediction of their difficulty for learning: an easy item, a learnable item or a difficult item. Similarly, the items for which hypotheses were rejected are shown in Table 5.2, where the level of difficulty defined based on the results is provided in square brackets. One item for rhythm, the PVI values of successive stressed vowels and unstressed vowels, is not included in either table because it could not be measured due to the high frequency of pauses.

Table 5.1

*Summary of Items on which Associated Hypotheses were Supported*

Element	Easy items	Learnable items	Difficult items
Vowel quality			/ɪ/ (/u:/) <sup>a</sup>
Vowel duration	/i:-ɪ/ /ɑ:-æ/		/ɑ:-ʌ/
Plosives			/t/ /t-st/
Fricatives			/f/ /θ/
Approximants			/l/
Rhythm			Pitch Duration of weak vowels Vowel centralization
Intonation	Final utterances <sup>b</sup> Falling utterances <sup>c</sup> Level		Long/non-final utterances <sup>b</sup> Non-falling utterances <sup>c</sup>
Connected speech phenomena			Elision CC linking CV linking

*Note.* The results of the PVI values of successive stressed vowels and unstressed vowels are not included in the table. <sup>a</sup>The vowel quality of /u:/ was not statistically tested, and therefore, it is within the round brackets. <sup>b</sup>Nucleus placement was examined with these utterances. Final, long and non-final refer to utterances with a typical pattern where the nucleus fell on the final word, utterances that were long, and utterances with a typical pattern where the nucleus fell on the non-final word. <sup>c</sup>Nuclear tone choice was examined with these utterances. Falling and non-falling describe utterances where a falling tone was typical, and those where a non-falling tone was typical.

Table 5.2

*Summary of Items on which Associated Hypotheses were Rejected*

Element	Easy items	Learnable items	Difficult items
Vowel quality		/æ/ /ɔ:/ [easy] /ɜ:/ [difficult]	/i:/ /e/ /ʌ/ /ɑ:/ /ʊ/ [easy]
Vowel duration	/u:-ʊ/ [learnable]		
Plosives			/p/ /k/ /k-sk/ [learnable]
Fricatives			
Approximants		/r/ [difficult]	
Rhythm		Intensity [difficult]	
Intonation			Span [easy]
Connected speech phenomena			

*Note.* The level of difficulty found is presented within the square brackets.

As presented in Tables 5.1 and 5.2, the hypotheses formulated within the framework of the SLM were more likely to be rejected: 8 items out of 10 for vowel quality, 1 item out of 4 for vowel duration, 3 items out of 5 for plosives, 1 item out of 2 items for approximants. This suggests the overall difficulty in applying this model to the prediction of learning by less experienced learners. The poor prediction was especially notable for vowel quality, where the hypotheses were not upheld for 8 vowels of 10.

As described in Section 1.4.3, the SLM does not target less experienced learners, but focuses on predicting the learning capacity of experienced or proficient language learners. It defines age of arrival (AOA) or age of learning (AOL), experience of learning, as important factors that affect learning (Flege, 1995; Flege, Munro, & MacKay, 1995; Jia, Strange, Wu, Collado, & Guan, 2005; Piske, MacKay, & Flege, 2001), as noted by H4 in Table 1.1. In contrast, the subjects in the present study had studied English solely under the guidelines of the course of study (MEXT, 1998, 2009) and had no experience of living in an English-speaking country. It was therefore not straightforward to predict how new phones would be learned by less experienced learners, although the SLM considers it possible for learners to learn new phones more easily. In order to compensate for the gap in the SLM prediction between experienced learners and less experienced learners, the present study



attempted to consider the degree of newness. This study thus hypothesized that three vowels, /æ, ɔ:, ɜ:/, would be learnable items and the other three vowels, /ɪ, e, ʌ, ʊ/, would be difficult items whereas it defined them as new phones. However, only the hypothesis of /ɪ/ was supported, the results revealing that /e, æ, ʌ, ʊ, ɔ:/ were easy items, and /ɜ:/ was a difficult item. Even if the hypotheses for these new phones had been built purely based on the SLM, two vowels, /ɪ, ɜ:/, would have been against the hypotheses. Therefore, the accuracy of the prediction was not fully verified for vowel quality, in particular.

One of the difficulties in applying these models to the subjects in this study was in defining identical, similar or new phones under the framework of the SLM and defining the prominent phonetic features. Wester et al. (2007) also pointed this out in their study of Dutch learners of English learning dental fricatives, claiming that it would be necessary to make the definition of new and similar phones clear in order to apply this model. As shown in Table 1.1, H2 and H3 of the SLM proposed that new phones were defined based on perceived dissimilarity at a phonetic level. However, the methods for measuring dissimilarities and the extent of dissimilarities benefitting the learning are not clearly specified. To find these dissimilarities, the previous studies have conducted experiments, using identification tests, discrimination tests and goodness rating tests. This study, based on these results and the contrastive phonetics and phonology of the two languages, determined whether a given L2 phone could be learned or not. Newness was also defined for the new phones, based on the articulatory and acoustic features and the previous findings regarding which phone changed in the phonological space of learners over time. What the above all suggests is that there is no standard criterion for classifying L2 phones into identical, new or similar.

The failure to apply the model could be attributed to two points. One is that the definition of the perceptual distances between two languages is not clear-cut and they are difficult to measure. Even if perceptual distances could be identified, they might change over time as learners become more experienced with the L2 sound system. If so, the prediction based on the previous study targeting experienced learners does not fit the present study. Perceptual and productive tests had to directly be given to the subjects targeted in the

experiment. The other point concerns the argument that it takes more time for less experienced learners to improve production, even if there are recognizable perceptual distances between two phones. Flege (1995) claimed that the amount of language experience that provides more input matters in the learning process. This is a reasonable claim, but how long it will take for learners to learn and master a target is still unclear. The results of this study showed that the difficulty of new phones varied across items, implying that the period of time necessary to learn new phones would differ across new phones. It could be related to the density of the vowel distribution in the phonological vowel space and/or the articulatory difficulty, for example. Aoyama et al. (2004) noted, in interpreting their results regarding /w/ production, that the number of phones in L2 corresponding to one phone in L1 could affect learning L2 phones. The difficulty of /ɜ:/, a new phone, over other new phones can possibly be explained by both density of the vowel distribution in the area where it is located and difficulty in articulating it by setting the tongue at a resting position. In contrast, the difference in the difficulty between /ɪ/ and /ʌ, ʊ/, which were also defined as new phones, cannot be attributed only to temporal cues. Simply defining L2 phones as new was thus not enough to predict the level of difficulty. It would be necessary to consider other factors for a more precise prediction about learning new phones. This ambiguity leads to a limitation in the applicability of this model to new conditions, such as less experienced learners, while the equivalent classification noted by H5 of the SLM in Table 1.1 can explain the failure of these learners in category formation. This would be a major reason why the SLM could not be extended to predict less experience learners of English in this study.

It should also be noted that even if new phones are accurately defined, the classifications of similar and identical phones are not so straightforward. No phones are exactly the same between Japanese and English. As it is more difficult to prove the presence of objects than to prove the absence of objects, it would be difficult to define two phones in a separate language as identical. Flege (1984, 1995) did not clearly account for how to define an identical phone.

The failure of prediction based on these models was marked in the analysis of vowel

quality, including both monophthongal vowels and weak vowels, and needs to be discussed. An approximant /r/ is also relevant here given its articulatory similarity to vowels. A possible cause could stem from the articulatory characteristics of vowels that differ from those of consonants. The articulation of vowels tended to be more subtle than that of consonants. This is clear in that consonants are categorized not only by the place of articulation but also by the manner of articulation. This difference is reflected in more types of acoustic measurements used for the analysis of consonants. In contrast, vowels are classified mainly by tongue height and tongue position by moving the tongue intricately. This does not have as big an impact as changing the manner of articulation. This is why the same measurement, such as F1 and F2, can be applied to all vowels. This holds true for vowels that differ in lip rounding. The differences among each vowel are more subtle, and this is not simply dependent on the number of vowels in the phonological inventory of a language. The NLM (Kuhl, 1991, 2000; Kuhl & Iverson, 1995) suggests that all vowels intricately produced do not necessarily spread in all areas of vowel space equally. It is claimed that, as one ages, the phonological space becomes distorted so as to adjust it to L1, where the categories are formed with good exemplars located in their center. If so, these distortions also need to be counted in the prediction of learning, which will make it more difficult to define each phone as identical, similar or new. These kinds of complexity concerning vowels in particular could increase the difficulty in considering how long it would take for new phones to be learned and mastered.

Mennen (2015) described the LILt as a working model to predict the learning of intonation, and the prediction based on it was better verified than that based on the SLM. The hypotheses on intonation that developed within this framework were generally upheld. One target item, the span, failed to produce the result hypothesized; however, it would be still reasonable to claim that the prediction was overall accurate, considering that the reason the hypothesis about span was rejected could be explained by methodological issues, as discussed in Section 5.7.2. The current study predicted the learning of the target items mainly in the phonological and phonetic dimensions, not in all four dimensions. It would therefore be rather hasty to draw conclusions. The realization of tone is clearly relevant to the semantic

dimension, for instance, as argued in Section 5.7.2. If the intonation systems of Japanese and English could be compared in all four dimensions, including the semantic and frequency dimensions, this model would work well for predicting the learning of other intonational items more broadly.

### **5.10.2. Research questions**

The first research question was which phonetic and phonological items were easy items, learnable items or difficult items for Japanese learners of English who only had experience of learning English under the English curriculum in Japan. In order to address this research question, each phonetic and phonological item was defined as easy, learnable or difficult. A summary of the findings is given in Table 5.3. Difficulty levels of the difficult items are presented within the square brackets.

The table clearly shows which items were easy, learnable or difficult, giving the answer to the first research question. In general, both segments and prosodic features were difficult items for Japanese learners of English. The difficult items were further categorized into three types, which specify the difficult items that are less likely to be learned in a naturalistic learning under the English curriculum in Japan. These items are those categorized as difficult items in D3 and they include the following: three vowels, /ɪ, ʌ, u:/; two voiceless fricatives, /θ, s/; two approximants, /r, l/; two features of weak vowels, shorter duration and centralized vowel quality; three types of utterances for intonation, utterances that are long, utterances where the nucleus preferably fell on the non-final word and utterances where tone a non-fall tone is typically used. The results suggest that these items require longer to improve to a native-speaker level or need treatment to learn.

These results do not suggest which items should be learned first, or which should be the focus of learning or teaching pronunciation. The answer to these questions depends on what pronunciation goals learners or teachers aim for. To put forward such suggestions, therefore, these results will be discussed further in Chapter 6, where they will be compared against the descriptions for potential pronunciation goals for Japanese learners of English.

Table 5.3

*Summary of the Findings*

Element	Easy items	Learnable items	Difficult items
Vowel quality	/i:/ /e/ /æ/ /ʌ/ /ɑ:/ /ɔ:/		/ɪ/ [D3]
	/ʊ/		/u:/ [D3]
			/ɜ:/ [D3]
Vowel duration	/i:-ɪ/ /ɑ:-æ/	/u:-ʊ/	/ɑ:-ʌ/ [D1]
Plosives		/p//k/ /k-sk/	/t/ [D2]
			/t-st/ [D2]
Fricatives			/θ/ [D3]
			/s/ [D3]
Approximant			/r/ [D3]
			/l/ [D3]
Rhythm			Pitch [D1]
			Intensity [D1]
			Duration of weak vowels [D3]
			Vowel centralization [D3]
Intonation	Final utterances <sup>a</sup>		Long/non-final utterances <sup>a</sup> [D3]
	Falling utterances <sup>b</sup>		Non-falling utterances <sup>b</sup> [D3]
	Level		
	Span		
Connected speech Phenomena			Elision [D1]
			CC linking [D1]
			CV linking [D1]

*Note.* Difficulty levels of the difficult items are presented within the square brackets. The results of the PVI values of successive stressed vowels and unstressed vowels are not included in the table. <sup>a</sup>Nucleus placement was examined with these utterances. Final, long and non-final refer to utterances with a typical pattern where the nucleus fell on the final word, utterances that were long, and utterances with a typical pattern where the nucleus fell on the non-final word. <sup>b</sup>Nuclear tone choice was examined with these utterances. Falling and non-falling describe utterances where a falling tone was typical, and those where a non-falling tone was typical.

The second research question concerns the relationships between the elements of pronunciation: which elements of pronunciation have a supportive relationship with one another. The results of the profiles and the correlation analyses demonstrated that vowel quality and approximants, and fricatives and approximants had a supportive relationship with each other. Similarly, the possibility of a supportive relationship was also suggested by the

combination of vowel quality and rhythm. These findings have some pedagogical implications for pronunciation learning and teaching. A supportive relationship implies the possibility that learning one item could enhance learning another within the system. Effective learning and teaching could be realized by building guidelines that refer to these relationships between the elements of pronunciation.

## **Chapter 6 Practical implications**

### **6.1. A learning goal of English pronunciation**

One last purpose of the present study is to provide practical implications for the field of English pronunciation learning and teaching in Japan. This is closely related to the first research question. The experiment described in Chapters 3, 4 and 5 revealed the phonetic and phonological items that are easy, learnable and difficult to learn for Japanese learners of English. These results are based on the performance of the JL subjects being compared against that of the BN/AN subjects. However, the potential goals of teachers and learners are not restricted to the level of native speakers. In this chapter, other potential learning goals for pronunciation will first be described against the background of a new emerging attitude toward the goal of pronunciation. What Japanese learners of English need to learn in order to attain these goals will then be suggested by comparing the findings in the present study against what is defined in these goals for the phonetic and phonological items that this study dealt with.

#### **6.1.1. Need to reconsider a pedagogical goal of pronunciation**

As a result of globalization, caused by political, economical, technological and social development around the world, English has become the language most frequently used in various contexts. As Crystal (2003) argued, it has spread rapidly around the world since the 1950s due to “the expansion of the British colonial power, which peaked towards the end of the nineteenth century, and the emergence of the United States as the leading economic power of the twentieth century” (p.59). There is no room for argument against the claim that English is a language used globally.

What gives English an even more special status is that it is now a lingua franca for non-native speakers of English as well as native speakers. English functions as an essential medium of communication among non-native speakers even if there are no native speakers present in a given situation. To show the spread of English from the perspectives of world Englishes, Kachru (1985) proposed the three-circle model, which groups countries into three

kinds, depending on the status of English: the inner circle where English is the dominant, first language, called norm-providing; the outer circle where English is used as an official or second language due to a historical background of colonization by the British Empire, called norm-developing; and the expanding circle where English is taught as a foreign language and acknowledged as a lingua franca, called norm-dependent. No matter what proficiency levels are set as criteria to qualify as non-native speakers of English, countless speakers use English in the expanding circle. It would be difficult to come to an agreement on the number of non-native speakers of English, but Crystal (2003), for example, estimated that there were 750 million speakers of English in the expanding circle and 750 million in the inner and outer circles combined. Jenkins (2009) also reported an increasing number of English speakers. According to her calculations, while there were 350 million speakers each in the inner circle and the outer circle, around one billion would fall into the category of non-native speakers, who would satisfy the criterion of whether their English is reasonably understandable.

Against such an unprecedented background, new varieties of English have come into being, especially in the expanding circle, and they have gradually been accepted as legitimate varieties of English as those in the inner circle and outer circle through debate over the ownership of English (Graddol, 2006; Jenkins, 2000; Widdowson, 1994). They differ from English as a second language (ESL) or English as a foreign language (EFL) in that these varieties have developed through the interaction of non-native speakers, while ESL and EFL are native speaker-oriented.

These varieties of English have several names, referred to as World Englishes (Kachru, 1985), English as a Global Language (Crystal, 2003), English as an International Language (EIL; Smith, 1976), or English as a Lingua Franca (ELF; Jenkins, 2007; Seildhofer, 2004). All these terms reflect the way that English functions currently in a variety of contexts, but there are some differences in their perspectives on native-speaker English and other varieties. This study prefers using ELF to the other terms when discussing the English used by non-native speakers in international contexts. This is because it focuses not only on the acceptance of fluid English varieties but also on the commonality of English used in



international contexts, apart from native-speaker varieties. In Japan, where EFL has been highlighted in learning and teaching pronunciation, English in the inner circle countries, such as General American (GA) and Received Pronunciation (RP), has been traditionally viewed as an appropriate norm. In contrast, ELF does not completely exclude interaction between non-native speakers and native speakers, but still emphasizes communication among speakers from different L1 backgrounds. In order to satisfy the purpose of this chapter, therefore, presenting a potential goal within this paradigm would be especially worthwhile, as it provides a different, nation-independent view concerning what sort of English is taken as a goal. This is why this study will use the term ELF when considering a new pedagogical goal of English learning and teaching that differs from ESL or EFL.

ELF involves English used in communication among people with different L1 and cultural backgrounds, as noted above. This implies that ELF constitutes varieties of English with different accents. It is a language used no matter whether native speakers are involved in these non-native speaker interactions. This is why the view that ELF should be accepted as a legitimate variety of English requires reconsidering the pedagogical goal of pronunciation in the field of English learning and teaching: which accent of English should be adopted as their norm or goal. Some learners who need to interact with native speakers of English, in EFL contexts, would still see RP or GA as the strong candidates for their goal. On the other hand, those who have no contact with native speakers of English, in ELF contexts, may find it pointless to aim for meeting native speakers norms. Even if native speakers are involved in such interaction, this perspective applies here because ELF contexts require native speakers to tune themselves to non-native speakers.

A shift in attitudes toward English varieties is reflected in the way that *An Introduction to the Pronunciation of English* and *Gimson's Pronunciation of English* have been revised, as Shimizu (2011) noted. *An Introduction to the Pronunciation of English*, currently known as *Gimson's Pronunciation of English*, edited by Cruttenden since the fifth edition, is one of the most commonly read pronunciation textbooks worldwide by learners, teachers and phonetics experts. Examining it from the first edition to the current edition thus

makes it possible for us to realize how attitudes toward a learner's goal regarding English pronunciation have changed. In the first edition, Gimson (1962) did not create a separate part to give guidelines targeted to English learners, but in the latest edition, Cruttenden (2014) describes three possible targets in the part entitled *Language Learning and Teaching*, where GB and regional GBs<sup>3</sup>, Amalgan English and International English are described in detail. These three kinds of English refer to standard English in Britain, a hybrid English containing some local features used by native speakers and English as a lingua franca used primarily in international contexts, respectively. GA or RP are thus not necessarily the only goal for all learners, and more appropriate goals vary from learner to learner.

To address the issue of what an appropriate goal is, however, is beyond the scope of this study because the answer to the question of which accent of English learners should define as their goal primarily depends on their purpose in learning English, and it varies across learners, after all (Nelson, 2011; Walker, 2010). Instead, the focus here is on discussing potential goals, so that learners can select the most appropriate one by themselves, and teachers can suggest potential goals. Teachers especially play a significant role in setting goals for learners, because some learners do not realize ELF contexts and are likely to believe that a native speaker accent is the only target for which they should aim (Shimizu, 2011; Walker, 2010). Although it is surely logical to suggest that learners should choose whatever accent they wish to acquire as their goal, which Jenkins (2000) considered an option that learners should be offered, they also have to be guided to select an appropriate, realistic goal, and to be more aware of ELF, the ownership of English and many more issues involving ELF.

## **6.2. Potential pronunciation goals**

What are appropriate and potential goals in EFL contexts and ELF contexts respectively, then? In an EFL context, where learners use English primarily to communicate with native speakers of English, one of the possible goals could be a native-speaker accent. In an ELF context, however, it is not straightforward to define a single, appropriate goal

---

<sup>3</sup> Cruttenden (2014) calls standard English in Britain GB and regional GBs rather than RP in the latest issue of the book, and this dissertation therefore uses this term when citing from Cruttenden hereafter.

common across all learners. Jenkins (2000), for example, logically argued that RP is not the best choice of the pronunciation goal for ELF users, because only a limited number of speakers use RP; RP is not an easy accent for learners to master; RP changes over time; and some teachers and learners do not like learning RP. Factors such as the materials available to learners and teachers can also affect their choice of a goal.

The different targets in pronunciation learning and teaching, including those in EFL contexts and those in ELF contexts, are well described and compared by Shimizu (2011). Differences in descriptions between her work and this study are in the selection of the goals. Firstly, while Amalgan English (Cruttenden, 2014) was described in Shimizu, it was removed from the following discussion. It is mainly presupposed to be used in the outer circle, and therefore, it would be an unlikely goal for Japanese learners of English. Secondly, the model that aimed to keep minimum intelligibility proposed by Gimson (1980) was not discussed in this study, either. This is an old model, and Cruttenden (2014) consolidated models when suggesting the most recent targets. There is thus little reason to describe it in this study when attempting to offer practical implications. Thirdly, Shimizu dealt with neither GA nor RP as possible goals for EFL-oriented learners in her article because she focused mainly on goals for ELF users. However, Japanese EFL-oriented learners are more likely to set them as targets. This chapter aimed to describe several potential goals fairly, rather than to recommend that learners and teachers select one particular goal. The descriptions of GA and RP were thus also added as potential goals here.

### **6.3. Pronunciation goals for EFL-oriented learners and ELF-oriented learners**

This study involves GB and regional GBs described by Cruttenden (2014) and GA by Prator and Robinett (1985), which are used as descriptions of pronunciation goals for Japanese EFL-oriented learners. International English proposed by Cruttenden, the Lingua Franca Core (LFC; Jenkins, 2000; Walker, 2010) and Shimizu's (2011) guidelines based on the LFC were selected for Japanese ELF-oriented learners. International English, the LFC and Shimizu's guidelines all propose minimum targets to maintain intelligibility in

international contexts. Cruttenden explains that the general aim of International English is to achieve “minimal intelligibility in the use of English in international lingua franca situations” (p.344). Jenkins (2000) states that the LFC proposes the pronunciation features that learners need to acquire at least to maintain mutual intelligibility of ELF. Shimizu describes her guidelines as complying with the LFC. The basic idea is thus that it is sufficient for ELF-oriented learners to attain a minimum level of pronunciation features as long as mutual intelligibility is retained. Walker (2010) even argued that an approach setting ELF as the only goal would satisfy the three basic criteria in defining a goal of pronunciation for ELF-oriented learners, mutual intelligibility, identity and teachability. This means that the other two approaches that he points out, using a standard native-speaker accent and using a single world standard for pronunciation, do not warrant them.

According to this principle, Cruttenden (2014) offers a list of priorities and tolerances with which to set these targets, and which learners should emphasize in learning pronunciation. Similarly, in the LFC, Jenkins (2000) lists the pronunciation items that are required, using the term core. As noted above, retaining mutual intelligibility is the key to these targets, but International English and the LFC differ in the standard in selecting the targets. Cruttenden based proprieties and tolerances in International English on how frequently items are used and how many minimal pairs they have, stating that each item involving pronunciation has a different functional load. Core features in the LFC come from Jenkins’ observation and examination of spoken interactions between non-native speakers. The items that can cause communication breakdown were defined as core. Teachability and learnability are also incorporated in the definition. One thing to be noted about the LFC is that it is not a goal that is invariable and fixed, but a goal that can be modified, as Jenkins, Walker (2010) and Shimizu implied. Therefore, it can be predicted that more and more goals will be proposed as the awareness of ELF develops worldwide.

Nakano (2008), who summarized the LFC in Asian English contexts, is also in line with the LFC and Shimizu’s guidelines. Morizumi (2009) is another potential reference, also suggesting a possible goal for Japanese ELF-oriented learners. He noted the better

enunciation, the accordance of sounds to the spelling, the substitution of similar sounds of Japanese and the opening of the choices of pronunciation goals. However, Morizumi was not included here because it is a rather extreme version, without sufficient empirical data.

In contrast, the goals for EFL-oriented learners require learners to learn all features of the models. They must be acquired to attain a native-speaker level. However, Cruttenden (2014) gave priority among these features according to functional loads or GB varieties, while Prator and Robinett (1985) did not specify a priority. The phonetic and phonological items that Cruttenden, Jenkins or Shimizu listed, regarded as important in pronunciation learning and teaching, will thus be preferentially selected and discussed below.

In the following part, the descriptions of the five potential pronunciation goals will be shown first according to the elements of pronunciation, including segmental features and prosodic features. The former involves vowels, plosives, fricatives, affricates, approximants, nasals, syllabic consonants and consonant clusters. The latter concerns rhythm, stress, intonation and connected speech phenomena. The extent to which the items of each element are necessary for learners will be described, using terms such as essential, necessary, recommended, tolerated or avoided. When it is not clearly expressed, the degree of necessity was read out from the descriptions, which is indicated by an asterisk mark, as in essential.\* When nothing was specified about the item concerned, no description is provided in the tables below. After these descriptions are given, the main findings of this study will be presented and compared against these descriptions to suggest which phonetic and phonological items need to be learned or how much they need to be learned to attain each goal.

### **6.3.1. Vowels**

Table 6.1 shows the descriptions of the vowels in the five potential goals. The major difference between the goals for EFL-oriented learners and those for ELF-oriented learners in learning English vowels concerns the quality. In general, it is essential for EFL-oriented learners to retain quality, whereas it is not for ELF-oriented learners because the distinction of the vowels using temporal cues is accepted.

Table 6.1

*Potential Targets for Vowels*

	EFL		ELF		
	GB and Regional GBs	GA	International English	The LFC	Shimizu's guidelines
Quality and quantity of monophthongs	Essential	Essential*	5 vowel system <sup>a</sup> + length tolerated	Quantity more important, consistent quality required and /ɜ:/ essential	5 vowel system <sup>a</sup> + length tolerated and /ɪ, æ, u:, ɜ:/ necessary
Quality of diphthongs <sup>b</sup>	Distinct quality of first element necessary		Only /aɪ, ɔɪ/ needed	Sufficient length needed	Realized with 5 vowels tolerated
Length: lenis-fortis <sup>c</sup>		Essential	Preferred	Essential	

<sup>a</sup>5 vowel system = /i, e, a, o, u/. <sup>b</sup>GB and regional GBs have five closing diphthongs /eɪ, aɪ, ɔɪ, aʊ, əʊ/ and two centering diphthongs /ɪə, ʊə/. Cruttenden (2014) considers a centering diphthong [eə] to have developed to be /ɛ:/ recently. In GA, the second element of centering diphthongs is realized as /ɪ/. <sup>c</sup>Consonants articulated with weaker energy are called lenis, those with stronger energy, fortis. Vowel duration is shortened before fortis consonants than before lenis consonants.

EFL-oriented learners of GB and regional GBs are required to differentiate between /i:/ and /ɪ/ and between /ʌ/ and /æ/,<sup>4</sup> which is especially emphasized. Mastering the word-final /i/ as in *happy* is also recommended because the substitution of /ɪ/ sounds old-fashioned. In contrast, it is tolerated that /ɑ:/ and /ɔ:/ are replaced by another vowel. They have variants such as [ɑ:, ä:] and [ɔ:, ɔ̃:], which are accepted, if they do not overlap with [æ] and [u:], respectively. However, a mid back vowel, [ɔ̃:], a higher quality of [ɔ:], can be used on the condition that /əʊ/ is not pronounced as [ɔ:]. Although the quality of the first element is considered to be more important than that of the second element for diphthongs, extensive places of articulation for the first element are accepted. Two variants, [oʊ] and [o:~], are accepted as a pronunciation of /əʊ/ while [əʊ] is most recommended. Regarding the second element, it is defined that /ɪə/ and /ʊə/ must not be more open than half-open. However, it is more and more accepted that they are pronounced with long monophthongs [ɪ:] and [ʊ:], as

<sup>4</sup> Cruttenden (2014) used a phonetic symbol /a/ for this vowel, but /æ/ is still prevalently used. Therefore, this dissertation applies the latter symbol.

in the development from [eə] to /ɛ:/, which has resulted in /eə/ losing its position as a diphthong in Cruttenden's (2014) description<sup>5</sup>. For EFL-oriented learners of GA, all different vowel qualities are similarly features to acquire because the substitution of a vowel for another in the stressed syllable is regarded as a serious error, which Prator and Robinett (1985) argued could be caused by L1 influence or orthography.

The goals for ELF-oriented learners accept a transfer from the five-vowel system, as in Japanese, as characterized in International English, the LFC and Shimizu's guideline. International English argued that the five vowels /i, e, a, o, u/, the most common among world languages, plus a distinction of length for each vowel is more important than learning the quality. For example, /i:/ and /ɪ/ are differentiated by length, not quality, represented as [i:] and [ɪ]. Similarly, International English defines /e/ and /eɪ/ as pronounced as [e] and [e:], /æ/ and /ɑ:/ as [a] and [ɑ:], /ɒ/ and /ɔ:/, əʊ/ as [o] and [o:], and /ʊ, ʌ/ and /u:/ as [u] and [u:]. As a result, only two diphthongs, /aɪ/ and /ɔɪ/, must be retained in International English. A rhotic accent, where an orthographic *r* is always pronounced, is recommended in International English, which leads to a loss of /ɪə, ʊə, eə(ɛ:), ɜ:/.

The LFC also recommends that ELF-oriented learners choose a rhotic accent, leading to a reduction in the number of diphthongs. At the same time, it emphasizes the need for learners to achieve a quality of /ɜ:/, considering intelligibility problems caused by the confusion with /ɑ:/. For the remaining vowels, the LFC suggests that the distinction of short and long vowels is important, as with International English, defining the vowel quality as non-core. However, unlike International English, the pairing of vowels is not clearly noted in the LFC, which raises a question as to how the English /æ, ʌ, ɑ:/ could be paired and discriminated from one another. The LFC also requires ELF-oriented learners to shorten vowels followed by fortis consonants in closed syllables, compared to those followed by lenis consonants in closed syllables and in open syllables, so that they could highlight the distinction of length. This is a little more challenging goal than International English suggests.

---

<sup>5</sup> This study dealt with /eə/ as a diphthong because it is still more common recognition.

Shimizu's guidelines also propose a slightly strict target in a different way: /ɪ/, /æ/, /u:/ and /ɜ:/ need to be pronounced carefully, not being replaced with Japanese pure vowels, while /i:/ can be pronounced as Japanese [i:], /e/ as Japanese [e], /ʌ/ and /ɑ:/ as Japanese [a] and [a:], /ɒ/ and /ɔ:/ as Japanese [o] and [o:], and /ʊ/ as Japanese [u]. Her guidelines accept that Japanese ELF-oriented learners use Japanese [i, e, a, o, u] to produce seven or eight diphthongs commonly recognized in GB and regional GBs. It does not especially focus on recommending a rhotic accent, unlike International English and the LFC.

Of the descriptions in Table 6.1, what was relevant to the findings in this study is the row of quality and quantity of monophthongs. In the present study, the vowel quality of /i:, e, æ, ʌ, ɑ:, ɔ:, ʊ/ and the durational distinctions of /i:-ɪ, ɑ:-æ/ were found to be easy for Japanese learners of English; the durational distinction of /u:-ʊ/ was learnable; the vowel quality of /ɪ, ɜ:, u:/ and the durational distinction of /ɑ:-ʌ/ were difficult.

EFL-oriented learners are required to articulate all vowels with distinct quality, whether the goal is GB and regional GBs or GA. This suggests that Japanese learners of English need to focus on improving /ɪ, ɜ:, u:/. Given that the durational difference as well as the vowel quality contributes to the distinctness of the vowel category, the durational distinction is not enough for /ɑ:-ʌ/ and /u:-ʊ/ when these goals are defined.

It is recommended that ELF-oriented learners highlight the durational distinction to differentiate vowels. This suggests that not all learnable and difficult items have to be improved. While the quality of /ɪ/ is difficult for Japanese learners of English, Japanese ELF-oriented learners do not need to learn to attain this quality in the goal of International English and the LFC. These two goals consider the durational distinction in the /i:-ɪ/ pair to compensate for the inability to attain the quality of /ɪ/. Only Shimizu's guidelines define a more challenging goal for this item, suggesting the need to learn to produce /ɪ/ and /u:/.

As regards the quality of /ɜ:/, the LFC and Shimizu's guidelines, but not International English, view it as necessary. The JL subjects failed to attain a native-like quality for this vowel, and were unlikely to differentiate it from others. One of the foci in learning English vowels should thus be on learning the quality of /ɜ:/ for EFL-oriented



learners who define their goal as the LFC or Shimizu's guidelines.

What all goals for ELF-oriented learners require Japanese learners of English to improve, in contrast, is the durational distinction of /u:-ʊ/ and /ɑ:-ʌ/. In the /u:-ʊ/ distinction, not only did the JL subjects tend to fail to learn to produce the quality of /u:/, but some also failed to discriminate /u:-ʊ/ in duration. As regards their vowel quality, /u:/ overlapped /ʊ/. Japanese ELF-oriented learners thus need to improve their durational distinction of /u:-ʊ/ first. Because this distinction was found to be a learnable item, they will possibly be able to learn it with relative ease. Similarly, it is necessary for Japanese learners of English to improve the durational distinction of /ɑ:-ʌ/. The results of the present study imply that the JL subjects attained a native-like quality for these two vowels, but they were not discriminated in terms of duration. Assuming that the goals for ELF-oriented learners emphasize the importance of length, the durational distinction in the /ɑ:-ʌ/ pair would be necessary from a conservative point of view.

### 6.3.2. Plosives

Table 6.2 shows what each potential goal defines for plosives. Among the several features of plosives, all goals refer to the aspiration of /p, t, k/. This suggests that it is an indispensable feature with which to characterize voiceless plosives. All potential goals but International English consider this target as essential or necessary mainly to discriminate between voiceless plosives and voiced plosives at the same place of articulation.

The devoicing of approximants /l, r, j, w/ following /p, t, k/ and no audible release of word-final /p, t, k/ are also especially important features for EFL-oriented learners of GB and regional GBs and GA and EFL-learners of GA, respectively, while they are not regarded as an important feature to master in the goals for ELF-oriented learners.

In contrast, two features should be avoided by ELF-oriented learners. One is a kind of undershoot, where voiced plosives /b, d, g/ are produced, replaced by fricatives. International English and Shimizu's guidelines suggest this. The need to avoid a weakening of the voiced plosives /b, d, g/ to fricatives /v, z, ɣ/ is particularly noted in Shimizu's guidelines, which explain that it occurs frequently in Japanese learners of English. The other

is the phenomenon known as tapping or flapping for /t, d/, which is typical in GA. This is essential for EFL-oriented learners of GA (Prator & Robinett, 1985). The LFC recommends that learners avoid using it and pronounce every /t/ as in RP (Jenkins, 2000), and Shimizu's guidelines and International English are in line with the LFC as to the use of this feature. The proposition to avoid these two features in the goals for ELF-oriented learners is because the sound changes relating to the manner of articulation are likely to increase the possibility of unintelligibility.

Table 6.2

*Potential Targets for Plosives*

	EFL		ELF		
	GB and regional GBs	GA	International English	The LFC	Shimizu's guidelines
Aspiration of /p, t, k/	Essential	Essential*	Preferred but lack tolerated	Essential	Necessary*
Devoicing of /l, r, j w/ after /p, t, k/	Essential		Preferred but lack tolerated		
No audible release in the final position	Not necessary	Essential*			
/b, d, g/ > Fricative			Avoided		Avoided for /b, g/
Intervocalic /t, d/ > Tap		Essential*	Not recommended*	Avoided	Avoided

*Note.* A > B = the substitution of B for A.

The row for aspiration of /p, t, k/ in Table 6.2 shows the relevant descriptions of VOT, which was examined in the experiment of the current study. It was found that the VOT durations of /p, k/ were learnable items, while that of /t/ was a difficult item. The findings of the present study that the distinction between /t/ and /t/ in /st/ and that between /k/ and /k/ in /sk/ were a difficult item and a learnable item, respectively, suggests that Japanese learners of English would generally fail to differentiate aspirated plosives from unaspirated plosives.

As shown in Table 6.2, the aspiration of VOT is regarded as an important phonetic property to learn under all goals but International English, as noted above. Both EFL-oriented

learners and ELF-oriented learners following the LFC or Shimizu’s guidelines therefore have to learn to produce enough VOTs for /p, t, k/. Shorter VOTs for the voiceless plosives could cause confusion with the voiced counterparts. While the learnable items /p, k/ may be learned with relative ease, the difficult item /t/ will require more time, effort and attention in order for Japanese learners of English to learn it. However, Japanese ELF-oriented learners who define International English as their goal do not need to spend much time on learning VOT. It considers longer VOTs for the voiceless plosives to be preferred, but lack of this feature is tolerated.

### 6.3.3. Fricatives

Table 6.3 lists the potential goals for fricatives. All goals refer to the place of articulation, which implies the need to maintain the different quality of fricatives articulated at the different place of articulation.

Table 6.3

*Potential Targets for Fricatives*

	EFL		ELF		
	GB and regional GBs	GA	International English	The LFC	Shimizu’s guidelines
Place of Articulation	More important than voicing	Essential*	Distinction important except for /θ, ð/	Important except for /θ, ð/ and use of /f, s, t/ and /v, z, d/ for /θ/ and /ð/ accepted	/f, v/ variations avoided and use of Japanese [s, z] for /θ, ð/ accepted
Voicing			/f-v/ /θ-ð/ /s-z/ /ʃ-ʒ/ expendable	Important	Important
/s, z/ variations in the place of articulation			Dental or retroflex tolerated	Acceptable, but [ç, ʒ] avoided	Japanese [s, z] accepted, but [ç, ʒ] avoided
/h/ > drop, velar or uvular			Acceptable	[ϕ, j] avoided	[ϕ, j] avoided

*Note.* A > B = the substitution of B for A; A-B = the distinction between A and B.

The place of articulation, compared to the distinction of voicing, is an important feature for learners to retain, both in the goals for EFL-oriented learners and those for ELF-oriented learners. However, International English, the LFC and Shimizu's guidelines allow dental fricatives /θ, ð/ to be replaced by other phones such as [t, d, t̚, d̚, s, z, f, v], noting that they are not needed for intelligibility in ELF contexts. The LFC especially emphasizes the omission of these phonemes. As in Jenkins (2000), however, it does not allow the replacement of /θ, ð/ by Japanese [ç] according to Jenkins' observation, although using /s/ is considered acceptable. Only the Japanese [s] can be substituted for /θ/. This point is also highlighted in Shimizu's guidelines as indicated in Table 6.3, which added that [z], but not [ʒ, dʒ] before /i:, ɪ, i/, can be substituted for /ð/.

For other features of fricatives, Cruttenden (2014) considers it tolerable that, in International English, alveolar fricatives /s, z/ and voiceless glottal fricatives /h/ can be replaced by variants produced at different places of articulation. In contrast, a slightly more demanding goal is set for Japanese ELF-oriented learners in Shimizu's guidelines. Taking the phonetic features of Japanese into consideration, Shimizu (2011) suggests that Japanese ELF-oriented learners should avoid using variations for /f, v/, nonexistent in Japanese, and substituting [ç, ʒ, φ] for /s, z, h/ in the /si:, sɪ, zi:, zɪ, hi:, hɪ, hu:, hʊ/ environment, respectively.

The present study examined the production of /s/ and /θ/ by Japanese learners of English, which concerns the descriptions of the place of articulation and /s, z/ variations in the place of articulation in Table 6.3. The results revealed that they did not clearly discriminate between /θ/ and /s/, which were both defined as difficult items.

Japanese EFL-oriented learners need to learn to articulate both items at the authentic place of articulation. In the three potential goals for ELF-oriented learners, however, as noted above, the substitution of non-confusing consonants for /θ/ is allowed. This implies that EFL-oriented learners do not need to discriminate between /s/ and /θ/ as long as the substitutions occur within these consonants. One thing that Japanese ELF-oriented learners have to bear in mind is that they have to avoid confusing the substitution of [s] for [θ] with

that of [s] for [ç] if they wish to attain the LFC or follow Shimizu’s guidelines. Considering that variations in the place of articulation for /s/ are accepted, thus, Japanese EFL-oriented learners do not need to improve their production of /θ/ and /s/.

#### 6.3.4. Affricates

Although the present study did not focus on an examination of affricates, five potential goals specify what is required for consonants in this manner of articulation. While there are not many descriptions of affricates, Cruttenden (2014) claims that EFL-oriented learners of GB and regional GBs have to distinguish alveolar affricates /tʃ, dʒ/ from other confusing consonant clusters. According to Cruttenden, the only difference between the goal for EFL-oriented learners, GB and regional GBs, and that for ELF-oriented learners, International English, is that the latter considers the substitution of [ʃ, ʒ] to be tolerable, whereas the former does not. The substitution of another consonant for /tʃ, dʒ/ by learners is also pointed out by Prator and Robinett (1985). In the goals for Japanese ELF-oriented learners, the substitution of the Japanese affricates /tç dʒ/ for /tʃ, dʒ/ seems to cause no confusion, as implied in International English and described in Shimizu’s guidelines.

Table 6.4

*Potential Targets for Affricates*

	EFL		ELF	
	GB and regional GBs	GA	International English	Shimizu’s guidelines
/tʃ/-/dʒ, tr, ʃ/ /dʒ/-/tʃ, dr, ʒ/	Essential			
/tʃ/-/dʒ, tr/ /dʒ/-/tʃ, dr/			Essential	
/tʃ/-/ʃ/ /dʒ/-/j/		Necessary		
/tʃ/-/dʒ/ variations in the place of articulation			Acceptable	/tç, dʒ/ acceptable*

*Note.* A-B = the distinction between A and B.

### 6.3.5. Approximants

Table 6.5 shows what each of the five potential goals defines regarding approximants. One of the major differences between the goals for EFL-oriented learners and those for ELF-oriented learners is how they views allophonic variations of /l/ and /r/.

Table 6.5  
*Potential Targets for Approximants*

	EFL		ELF		
	GB and regional GBs	GA	International English	The LFC	Shimizu's guidelines
/l/	[l, ɫ, ɫ̥] essential	[l ɫ] essential	All allophonic variations tolerated	[ɫ] not necessary and [ɫ̥] > [ʊ] acceptable	[ɫ] > [u] acceptable
Post-alveolar /r/	Essential	Essential	Flap preferable and differentiated from /l/	Other variations acceptable	Necessary and differentiated from /l/ and Japanese flaps
/w/			Essential and differentiated from /v/	Japanese [ɰ] avoided	Elision of /w/ avoided, esp. for /wu:/, wu/*

*Note.* A > B = the substitution of B for A.

Both goals for EFL-oriented learners require learners to differentiate allophones of /l/ and to produce post-alveolar /r/. On the other hand, all three ELF goals accept a lack of allophonic variations of /l/ or substitutions of other sounds for dark /l/. The LFC, above all, clearly states that [ɫ], dark /l/, is not necessary for intelligibility in ELF contexts, and it is absolutely acceptable to replace dark /l/ with [ʊ] because this phone is also used by various speakers in London and South East England. Shimizu's guidelines adhere to it. The LFC allows more variations of /r/, proposing that a trill [r], a flap [ɾ] and a uvular [ʀ], are accepted alternatives to /r/. Some variants of /l/ and /r/ not found in native speakers of English are accepted in this way. The LFC recommends that learners attain a rhotic accent, common in GA, because it would not cause any confusion about when to pronounce /r/. International English agrees with the LFC in this respect. In contrast, Shimizu's guidelines

emphasize that the distinction between /l/ and /r/ is essential, differentiated from Japanese flaps.

The other approximants /w, j/, glides, are not discussed as heatedly as /l/ and /r/, liquids. Cruttenden (2014) notes, concerning International English, that a voiced labio-velar approximant /w/ has to be differentiated from a voiced labiodental fricative /v/, which is less likely to be a problem for Japanese learners of English. By contrast, Jenkins (2000) claims that the approximation of /w/ as Japanese [ɰ] should be avoided because it could cause intelligibility problems. Shimizu (2011) also points out that instruction is needed not to elide /w/ before /u:, ʊ/. Her guidelines, in contrast, explain that a voiced palatal approximant /j/ in Japanese functions as English /j/ with no problem, not affecting intelligibility.

The present study examined the production of /r/ and clear /l/ by scoring whether Japanese learners of English were able to produce a native-like quality for these consonants. The results showed that both were difficult items for them to learn. Some of the JL subjects were in the process of learning to produce /r/ or /l/, but the substitution of a flap-like sound was frequent. According to the row of /l/ and post-alveolar /r/ in Table 6.5, it is clear that these approximants are regarded as essential for EFL-oriented learners. Japanese EFL-oriented learners thus need to train themselves to improve the production of these consonants, although it is difficult for them. Because this study targeted only clear /l/, no consideration can be given to the other allophones.

The same applies to the learning goal of clear /l/ under the three goals for ELF-oriented learners. They do not require discrimination of allophonic variations of /l/. However, this does not necessarily mean that they allow substitutions of other phonemes such as a flap-like sound for clear /l/, which was the most common error among the JL subjects. Therefore, the results that this study obtained suggest that Japanese ELF-oriented learners need to improve their production of /l/ so that they will not replace it with Japanese flaps.

In contrast, what the three goals require learners to achieve as regards /r/ differs; Shimizu's guidelines set an much higher goal than International English and the LFC. According to Shimizu (2011), /r/ should be differentiated not only from English /l/ but also

from Japanese flaps. International English and the LFC allow variations such as a flap [ɾ] for /r/. Thus, while Japanese ELF-oriented learners following Shimizu’s guidelines need to improve /r/ along with /l/, those who prefer the goals of International English or the LFC have only to improve /l/. It should also be noted, however, that Shimizu suggests using an allophone of Japanese flaps for /l/, such as a flap followed by [N], which has a similar quality to English /l/. This is partly why she suggests that flaps should not be used for /r/.

### 6.3.6. Nasals

Nasals were another consonantal category that the present study did not target. The five potential goals define the targets of nasals as shown in Table 6.6. According to the goals for EFL-oriented learners of both GB and regional GBs and GA, /n/, a voiced alveolar nasal, and /ŋ/, a voiced velar nasal, must be produced at the accurate place of articulation. In contrast, International English does not regard a less native-like articulation of these phones as a possible factor in deteriorating intelligibility. It accepts /n/ produced at another place of articulation and /ŋ/ produced as a sequence of /ŋ/ and /g/. Shimizu’s guidelines agree with International English on this point; however, the LFC and Shimizu’s guidelines propose that Japanese ELF-oriented should learn to produce the alveolar nasal at the correct place of articulation. Jenkins (2000) points out that they should avoid dropping or substitution of postvocalic /n/ and nasalization of the preceding vowel. Similarly, Shimizu’s guidelines require that Japanese EFL-oriented learners do not substitute /N/ for /n/ in the word-final position.

Table 6.6  
*Potential Targets for Nasals*

	EFL		ELF		
	GB and regional GBs	GA	International English	The LFC	Shimizu’s guidelines
/n/ place of articulation	Essential	Essential*	Other places acceptable	Postvocalic /n/ > /m, N/ avoided	/n/ > /N/ in the word-final avoided
/ŋ/ > /ng/	Avoided	Avoided*	Acceptable		Acceptable

*Note.* A > B = the substitution of B for A.



### 6.3.7. Syllabic consonants

Syllabic consonants, which are also outside the scope of this study, are described in Table 6.7. Native speakers of English commonly use syllabic consonants. Therefore, Prator and Robinett (1985) regard /ə/ inserted before /l/ and /n/ that are supposed to be pronounced with syllabic consonants as “definitely an element of ‘foreign accent’” (p. 118). On the other hand, EFL-oriented learners of GB and regional GBs are not expected to acquire syllabic consonants. This means that not using syllabic consonants is accepted. Although both of the goals for EFL-oriented learners note something about the function of syllabic consonants, the three goals for ELF-oriented learners do not. This possibly implies that pronunciation without syllabic consonants would not cause a serious impediment to intelligibility in ELF contexts.

Table 6.7

#### *Potential Targets for Syllabic Consonants*

	EFL		ELF		
	GB and regional GBs	GA	International English	The LFC	Shimizu’s guidelines
Syllabic [l, r, m, n] > [ə] + [l, r, m, n]	Acceptable	Avoided*			

*Note.* A > B = the substitution of B for A.

### 6.3.8. Consonant clusters

Consonant clusters were not examined in this study because the articulation of these clusters is strongly affected by the original articulation of consonants composing clusters. However, the failure of Japanese learners of English to articulate consonant clusters is often pointed out in pronunciation textbooks. Table 6.8 illustrates what is accepted and is to be avoided in articulating consonant clusters.

Epenthesis is a strategy used by English learners to articulate consonant clusters. However, the goal for EFL-oriented learners of GB and regional GBs claims that inserting vowels after, before or between consonants at initial clusters, such as /l, r, w, j/ followed by another consonant and /sp, st, sk/, should be avoided. Similarly, epenthesis for consonant clusters in any position is not accepted in GA, which Prator and Robinett (1985) note that can

make one's English less understandable.

Table 6.8

*Potential Targets for Consonant Clusters*

	EFL		ELF		
	GB and regional GBs	GA	International English	The LFC	Shimizu's guidelines
Epenthesis at initial clusters	Avoided		Tolerated	Acceptable	Avoided
Epenthesis in any position		Avoided*		Acceptable	Avoided
Elision of /t, d/ of word-medial and word-final C+ +t, d/+C clusters	Allowed*	Allowed*	Acceptable	Acceptable and recommended except for /t/ of /nt/ cluster*	

*Note.* C = consonant.

Of the three goals for ELF-oriented learners discussed here, only Shimizu's guidelines suggest that Japanese ELF-oriented learners try not to add vowels to consonant clusters in any position. International English and the LFC allow ELF-oriented learners to insert vowels when producing consonant clusters. In International English, it is claimed that a medial intrusive vowel is preferred to an initial intrusive vowel, which is, furthermore, preferred to a complete deletion of consonant. Both International English and the LFC believe that it is acceptable to elide /t, d/ of word-medial and word-final /t, d/ clusters where another consonant precedes and follows them. The elision of these phones is fairly common, as in GB and regional GBs and GA (Cruttenden, 2014; Takebayashi, 1996), and it is probably natural for any speaker to elide them. This is why these two goals for ELF-oriented learners accept elision of /t, d/ in this phonetic environment, which is a less common attitude in the goals for ELF-oriented learners because they tend to regard a deletion of phones as a probable factor causing miscommunication in ELF contexts. Shimizu's guidelines do not clearly refer to whether this elision should be accepted or avoided. However, Shimizu (2011) only describes the necessity of training to learn to perceive elided /t, d/, which implies that Shimizu's guidelines do not positively recommend that learners elide these phones in their

production.

### 6.3.9. Rhythm

Table 6.9 indicates what each potential goal defines about the realization of stress-timed rhythm in English. The two goals for EFL-oriented learners and the three goals for ELF-oriented learners show clearly different requirements for this feature.

Table 6.9

*Potential Targets for Rhythm*

	EFL		ELF		
	GB and regional GBs	GA	International English	The LFC	Shimizu's guidelines
Stress-timed rhythm	Necessary*	Essential*	Syllable-timed acceptable	Not necessary	

It is considered essential to maintain so-called stress-timed rhythm in the EFL goals, which is mainly achieved by using unstressed vowels /ə/ and weak forms. However, International English and the LFC define it as unnecessary to realize rhythmic features in ELF contexts, supposing that an unsuccessful achievement of English rhythm does not affect intelligibility. International English clearly notes that syllable-timed rhythm is acceptable in ELF contexts. The LFC does not require ELF-oriented learners to acquire this feature, either, stating that the difference between stress-timed rhythm and syllable-timed rhythm has not been physically proved. Shimizu's guidelines do not refer to anything regarding the necessity of realizing English stress-timed rhythm, but because the guidelines are based on the basic philosophy of the LFC, there is no strong reason to maintain that there is a need for Japanese ELF-oriented learners who wish to follow Shimizu's guidelines to learn to realize English rhythm.

As described in Section 2.5, the realization of rhythm is closely related to the vowel weakening. Table 6.10 presents what is required by the five potential goals for this.

Table 6.10

*Potential Targets for Vowel Weakening*

	EFL		ELF		
	GB and regional GBs	GA	International English	The LFC	Shimizu's guidelines
Unstressed vowels like /ə, ɪ/	Essential and /ə/ preferred to /ɪ/	Necessary for naturalness	No /ə/ accepted	Not necessary	Not necessary
Weak form	Essential except for uncommon weak forms	Essential*	Not necessary	Strong LFC: not necessary Less strong LFC: shortening unstressed vowels useful	Not necessary, but contractions can be learned

As shown in the descriptions of unstressed vowels and weak form in Table 6.10, vowel weakening is essential for EFL-oriented learners. In the goals for EFL-oriented learners of GB and regional GBs, unstressed vowels are commonly pronounced as /ə/ or /ɪ/, and /ə/ is overall preferred in any position. The use of common weak forms, such as /ən/ for *and*, /bət/ for *but*, /ət/ for *at*, /əv/ for *of*, /kən/ for *can* and /tə/ for *to*, is also recommended to maintain a native-like rhythm, but it is recommended that some weak forms such as /jə/ for *your* and /mə/ for *my* be avoided. Similarly, in GA, /ə/, /ɪ/ or /ʊ/ are commonly used in unaccented syllables, as in GB and regional GBs. Prator and Robinett (1985) point out that learners tend to replace unstressed vowels with stressed vowels, being influenced by their spelling. While they note that this replacement is less likely to cause misunderstanding, learners are also advised to obscure unstressed vowels by pronouncing /ə/ and /ɪ/ to make their English sound more natural. *A, an, and, of, or, the, to, are, can, had, has, have, that* and *was* are introduced as the words most frequently weakened, pronounced. Above all, the first seven words from *a* to *to* are almost always pronounced in their weak forms, according to Prator and Robinett.

In contrast, all three goals for ELF-oriented learners explain that weak forms are not necessary and that their replacement with strong forms is acceptable. The LFC categorizes all items relating to vowel reduction as non-core, stating that only *a* and *the* are used by many

fluent bilinguals. This naturally implies the allowance of disappearance of /ə/. However, a less strong version of the LFC suggests that it would be helpful for ELF-oriented learners to shorten vowels in unstressed syllables, retaining their quality. Similarly, Shimizu's guidelines propose that Japanese ELF-oriented learners do not need to avoid acquiring contractions. This means that, in goals such as the LFC and Shimizu's guidelines, not all strategies to weaken unstressed syllables are considered absolutely unnecessary, but some strategies relating duration or word connection can be recommended to maintain or increase intelligibility in ELF contexts.

The present study investigated how unstressed vowels and weak vowels in weak forms are articulated in terms of the four acoustic cues, pitch, intensity, duration and vowel centralization. The results showed that Japanese learners of English had difficulty in using a lower pitch, weaker intensity, shorter duration and centralized vowel quality for weak vowels in weak forms than for stressed vowels. This suggests that all four items were difficult for Japanese learners of English to learn to use. It was also found that they were likely to insert pauses within utterances, which could lead to a failure to realize English rhythm.

A comparison of these findings with the descriptions in Table 6.10 shows that EFL-oriented learners need to improve all these items to realize English stress-timed rhythm. While they were identified as difficult items, it was also found that some Japanese learners learned pitch and intensity to approximate the level of native speakers. It is thus more likely that they will learn to use these two items with more ease than duration and vowel quality. On the other hand, none of the three goals for ELF-oriented learners regards the realization of weak vowels as necessary, except that the weak version of the LFC views shortening vowels in unstressed syllables as useful. As long as it is not defined as their goal, ELF-oriented learners do not need to put extra effort or energy into learning to produce weak vowels.

### **6.3.10. Stress**

Lexical stress is strongly connected to rhythm and intonation although it was not investigated in this study. Table 6.11 lists what each potential goal requires learners to achieve concerning lexical stress.

Table 6.11

*Potential Targets for Lexical Stress*

	EFL		ELF		
	GB and regional GBs	GA	International English	The LFC	Shimizu's guidelines
Word stress	Essential	Essential	Necessary	Not necessary, but worth achieving	Necessary
Accent shift	Necessary				

The placement of stress is regarded as an essential feature in native-speaker English. Therefore, EFL-oriented learners need to learn to put lexical stress on the correct syllable, according to the two goals for EFL-oriented learners. Prator and Robinett (1985) state that words where stress is placed on the wrong syllable are probably impossible to understand. Cruttenden (2014) claims that the correct use of stress shift is needed to attain an EFL level of GB and regional GBs.

In contrast, the LFC proposes that word stress is not necessary for ELF-oriented learners because it rarely influences intelligibility in ELF contexts. However, it adds that “it may be worth paying some attention to word stress” (Walker, 2010, p.40). This is because the degree to which lexical stress affects intelligibility is still not clear, and the ability to place the nucleus correctly, which is core in the LFC, presumes the ability to place lexical stress correctly. By contrast, Shimizu (2011) notes in her guidelines that Japanese ELF-oriented learners should pay attention to the placement of the stress and try to lengthen vowels in stressed syllables, which suggests that word stress is considered necessary. International English also implies that only lexical stress is important among all prosodic features.

### 6.3.11. Intonation

Table 6.12 shows what is required for English intonation. Both Cruttenden (2014) and Prator and Robinett (1985) devote much space to intonation, which reflects how important intonation is for EFL-oriented learners. The three Ts, tonality, tonicity and tone, are described in detail for EFL-oriented learners of GB and regional GBs, and thus possible

targets regarding intonation in each goal will be discussed below according to these three domains of intonational features.

Table 6.12

*Potential Targets for Intonation*

	EFL		ELF		
	GB and regional GBs	GA	International English	The LFC	Shimizu's guidelines
Intonation phrase (IP)	Necessary* [syntactic subjects & adverbial phrases]	Necessary*	Not necessary*	Appropriate IP important	Appropriate IP important
Nucleus placement	Necessary* [last lexical word in IP & no accent on old information]	Essential [last lexical word in IP & depends on focus]	Not necessary*	Essential	Essential
Nuclear tone	Necessary* [high fall, low rise & fall-rise for attitudinal uses]	Necessary* [high fall <sup>a</sup> & rise at the end of a sentence]	Not necessary*	Not necessary*	Not necessary
Nuclear tone in non-final positions in sentences	Necessary* [fall-rise for syntactic function]	Necessary* [high fall <sup>a</sup> , fall-to-normal & rise]			Not necessary*
Pre-nuclear patterns	Necessary [different nuance between glides-down and glides-up] [low level avoided]	Necessary [high tone on the former idea in contrasts and comparisons]	Not necessary*		

*Note.* What should especially be noted is described within the square brackets. <sup>a</sup>Prator and Robinett (1985) use the term, rising falling intonation. However, the tone that he calls rising falling intonation is different from a rise-fall in Cruttenden's (2014) description, and it is more similar to a high fall. The term, a high fall, was therefore used to refer to the rising falling intonation in Prator and Robinett to avoid confusion.

Firstly, dividing speech into one or more intonation phrases (IPs), called tonality, is necessary in both goals for EFL-oriented learners. Cruttenden (2014) especially notes that in GB and regional GBs, one separate intonation phrase can also be given to syntactic subjects and adverbial phrases. However, there is no simple rule about how to divide speech in both

GB and regional GBs and GA whereas Prator and Robinett (1985) maintain that divisions may be made between any large syntactic divisions. The LFC and Shimizu's guidelines, goals for ELF-oriented learners, also consider it necessary to divide speech appropriately. On the other hand, International English does not regard it as necessary.

Secondly, the nucleus placement, tonicity, is also viewed as essential in the goals for EFL-oriented learners of both GB and regional GBs and GA. EFL-oriented learners of these accents need to bear in mind that although the basic rule is that the last content word in an IP receives the prominence, this prominence can come earlier (or later) when the last accented content word is old information. This is also discussed by Prator and Robinett (1985), where focus is described as a factor in deciding nucleus placement. Similarly, two of the goals for ELF-oriented learners, the LFC and Shimizu's guidelines, also require learners to attain the ability to put the nucleus on an appropriate syllable, depending on the context. In contrast, International English considers it unnecessary.

Thirdly, the proper use of nuclear tones is defined as necessary for EFL-oriented learners, but not for ELF-oriented learners. A fall, rise and fall-rise are especially highlighted in the goal for EFL-oriented learners of GB and regional GBs. According to Cruttenden (2014), a high fall and a low rise are more common than a low fall and a high rise, respectively, unlike many other languages. This suggests that a greater change in pitch movement is likely to be preferred in this accent. Both of these tones are equally used for wh-questions and yes-no questions, although they differ in attitudinal function from one another; the former sounds more business-like and the latter, more polite. A fall-rise is the most peculiar tone in English and is used to express syntactic structures in the non-final positions of sentences such as "dependent clauses, adverbials and subjects" (Cruttenden, 2014, p. 335) and to express attitudes of speakers such as "warnings, reservations and contradictions" (Cruttenden, 2014, p. 335). EFL-oriented learners of GB and regional GBs are required to acquire all of these tone features.

Prator and Robinett (1985) also describe tone use in GA, and clarify what EFL-oriented learners of GA should learn concerning nuclear tones. Prator and Robinett



maintain that the end of a sentence is commonly said with a high fall and a rise in GA. Each of them is the most basic tone used at the end of declarative sentences, commands and wh-questions, expressing completeness, and at the end of yes-no questions, expressing incompleteness. There are also three patterns of nuclear tones frequently used in the non-final positions in sentences, which occur when one sentence is divided into more than one IP: a high fall, fall-to-normal and rise. There is no rule for choosing one tone between them, because speakers use them depending on their attitude rather than the syntactic structure or the meaning of the sentence (Prator & Robinett, 1985). Nonetheless, in some context, it is clear which is more likely to be chosen. For instance, a rise is more appropriate on the nucleus before *and* and *or*. This tone is also regarded as the safest nuclear tone for learners when addressing names or titles when talking with the person directly. There are more varieties of intonation patterns in GA in particular, called lexical intonation to express emotion such as surprise, shock, anger, approval or many more. However, Prator and Robinett conclude that it is good enough for EFL-oriented learners to begin by mastering the basic patterns of intonation because the use or meaning of lexical intonation varies from speaker to speaker. EFL-oriented learners of GA thus first need to learn to use a high fall and rise on the final nucleus of the sentence, and then, a high fall, fall-to-normal and rise on the non-final nucleus of the sentence.

A native-like use of the tones is not considered to be necessary in any of the three goals for ELF-oriented learners. Walker (2010) argues that there does not seem to be agreement about the meaning of each tone, and arrives at the conclusion that it is impossible to teach tones.

Finally, Cruttenden (2014) and Prator and Robinett (1985) describe the pre-nuclear tones to some extent, although none of the goals for ELF-oriented learners suggests the necessity of learning them. Cruttenden states that the different pre-nuclear patterns like glides-down, glides-up and a low level convey different nuances in GB and regional GBs, and a low level for a long sequence of syllables should particularly be avoided. Prator and Robinett explain that when contrasts or comparisons are focused on in a sentence, the former

idea is pronounced with a high tone in GA. The pre-nuclear intonational patterns are thus also what EFL-oriented learners need to attend to.

To summarize, while intonation has often been regarded as one of the most significant prosodic features contributing to nativeness, the complicated elements of intonation, such as the choice of appropriate tones from all those possible, are considered unnecessary for ELF-oriented learners. The LFC and Shimizu's guidelines therefore only emphasize the importance of tonality and tonicity, supposing that these two features help ELF users to comprehend messages. International English even suggests that ELF-oriented learners do not need to attain any native-like intonational pattern.

The results of the present study are related to tonality, tonicity and tone, which are featured in the rows of IP, nucleus placement, nuclear tone and nuclear tone in non-final positions in sentences in Table 6.12. IP and nucleus placement are both relevant to nucleus placement that this study examined. Nuclear tone and nuclear tone in non-final positions are connected with the investigation of nuclear tone choice.

The results revealed that Japanese learners of English had difficulty in placing the nucleus on the correct syllable when it occurred on the non-final word of IPs. The long utterances that they produced also tended to have an extra nucleus, which suggests that Japanese learners of English were likely to divide the utterances into more IPs. According to Table 6.12, how to divide utterances into IPs and where to put the nucleus are both regarded as important in the goals for EFL-oriented learners and ELF-oriented learners, except in International English. Japanese learners of English thus need to improve these features if they define GB and regional GBs, GA, the LFC or Shimizu's guidelines as their goal. Above all, nucleus placement in long utterances and that in utterances where the nucleus falls on the non-final word need to be focused on in learning the tonality and tonicity of English intonation. In contrast, Japanese ELF-oriented learners who set International English as their goal do not need to attend to these intonational items.

The present study revealed that non-falling tones, such as a low rise, a fall-rise and a level, were difficult for Japanese learners of English to use in the syntactic and pragmatic

contexts tested, such as the end of the subordinate clause preceding the main clause, the reporting clause before direct speech and lists. On the other hand, a falling tone successfully occurred where this tone was preferred. The goals for EFL-oriented learners and those for ELF-oriented learners obviously present different views about learning nuclear tone choice. The former goals consider the basic use of native-like tones to be necessary, although the latter goals do not. Thus, Japanese EFL-oriented learners need to be aware that Japanese learners of English tend to use a falling tone more frequently than the other tones, and to learn to use different tones depending on contexts. In contrast, this is not required in the goals for ELF-oriented learners, suggesting that Japanese ELF-oriented learners do not need to worry about nuclear tone choice.

### **6.3.12. Connected speech phenomena**

Table 6.13 shows what English learners need to focus on in learning connected speech phenomena. Of the various phenomena observed in connected speech, such as elision, linking and assimilation, elision is the most frequently discussed in the goals for EFL-oriented learners and those for ELF-oriented learners. Note that the first row of the descriptions in Table 6.13 is exactly the same as the third row of the description for consonant clusters in Table 6.8.

Basically, sound changes that occur in rapid speech can greatly influence intelligibility, and therefore, connected speech phenomena are not items that learners are recommended to master. For example, Cruttenden (2014) clearly maintains that there is no need to learn native English patterns of elision and assimilation in International English. However, elision of medial /t/ and /d/ between consonants is rather exceptional, as noted in Section 6.3.8. Not only EFL-oriented learners but also ELF-oriented learners, according to International English and the LFC, are allowed to elide /t, d/ in this position, although elision in other positions is considered to be unnecessary. The LFC explains that EFL-oriented learners are recommended to use elision of medial /t/ and /d/ because this elision changes three-consonant clusters to two-consonant clusters and makes it easier for learners to produce the sequence. However, using this type of elision too much might not be advised. Prator and

Robinett (1985) note that while medial consonants such as /p, t, k, d, θ, n/ within the three-consonant cluster of a word are frequently elided in GA, it may make speakers sound uneducated if they elide these consonants too often, especially in formal situations. In contrast, Shimizu (2011) does not note anything regarding elision of the medial /t, d/. Considering that Shimizu’s guidelines only suggest the need of training for Japanese learners of English to learn to perceive elided /t, d/, it would not be recommended that they elide these phones in their production, although the guidelines basically adhere to the LFC.

Table 6.13

*Potential Targets for Connected Speech Phenomena*

	EFL		ELF		
	GB and regional GBs	GA	International English	The LFC	Shimizu’s guidelines
Elision of /t, d/ of word- medial and word-final C+ +/t, d/+C clusters	Allowed	Allowed*	Allowed	Acceptable and recommended except for /t/ of /nt/ cluster*	
Elision of /ə/ in post-nuclear positions	Optional		Not necessary	Not necessary*	
Elision of /ə/ in pre-nuclear positions	Avoided		Not necessary	Not necessary*	

*Note.* C = consonant.

The results of the present study on elision concern the description in elision of medial /t, d/ between consonants in Table 6.13. It was found that this elision was difficult for Japanese learners of English. Apart from Shimizu’s guidelines, which did not note this phenomenon, the other potential goals for EFL-oriented learners and ELF-oriented learners showed a positive attitude toward using this connected speech phenomenon. It is thus recommended that both Japanese EFL-oriented learners and ELF-oriented learners learn to use this item. While elision was found to be a difficult item, some JL learners approximated the level of native speakers. Some treatment, such as pronunciation practice, might thus be

able to facilitate their realizing this connected speech phenomenon.

This study revealed that consonant-to-consonant (CC) linking and consonant-to-vowel (CV) linking were also difficult for Japanese learners of English. However, none of the five goals above refers to these connected speech phenomena. No description of these items implies that these phenomena are not essential features for English learners to learn. At least ELF-oriented learners would not be required to learn to use them, considering their overall philosophy about the items necessary or recommended.

#### **6.4. Summary of the chapter**

As shown in the descriptions above, it is necessary for ELF-oriented learners to learn a smaller number of target items than EFL-oriented learners. Table 6.14 shows a summary of what Japanese EFL-oriented learners and ELF-oriented learners need to prioritize in improving English pronunciation to meet each goal, comparing the learnable and difficult items identified in this study with the targets described in each potential goal. The cells where no item is provided indicate that no item needs to be improved.

Two points need to be noted in reading Table 6.14. Firstly, the three potential goals for ELF-oriented learners allow the production of vowel quality with the combination of the durational difference and five vowels. However, this does not lead to successful discrimination among /æ/, /ʌ/, /ɑ:/ and /ɜ:/, the characteristics of Japanese learners' productions being taken into account. Based on the results of the present study, which showed that /æ/ was easy for Japanese learners of English, the quality of /ɜ:/ should be improved to differentiate these vowels, as proposed by the LFC and Shimizu's guideline. These two goals would thus be preferable to International English for Japanese EFL-oriented learners as far as the learning of vowels is concerned. Secondly, while the use of weak vowels is required to fulfill the two goals for EFL-oriented learners, Cruttenden (2014) and Prator and Robinett (1985) do not specify which phonetic items should be learned in particular, pitch, intensity, duration or vowel centralization. Considering the claim of Cruttenden (1997) and Fujisaki et al. (1986) that pitch was the most prominent cue to express English stress, a lower pitch would be one of the items prioritized in learning weak vowels.

Table 6.14

*Learnable and Difficult Items to be Learned to Attain Potential Goals*

	EFL		EFL		
	GB and regional GBs	GA	International English	The LFC	Shimizu's guidelines
Vowel quality	/ɪ/ /ɜ:/ /u:/	/ɪ/ /ɜ:/ /u:/		/ɜ:/	/ɪ/ /ɜ:/ /u:/
Vowel duration	/ɑ:-ʌ/ /u:-ʊ/	/ɑ:-ʌ/ /u:-ʊ/	/ɑ:-ʌ/ /u:-ʊ/	/ɑ:-ʌ/ /u:-ʊ/	/ɑ:-ʌ/ /u:-ʊ/
Plosives	/p/ /t/ /k/	/p/ /t/ /k/		/p/ /t/ /k/	/p/ /t/ /k/
Fricatives	/θ/ /s/	/θ/ /s/			
Approximants	/r/ /l/	/r/ /l/	/l/	/l/	/r/ /l/
Rhythm	weak vowels	weak vowels		shorten vowels <sup>a</sup>	
Intonation	Nucleus placement for long/non-final utterances & non-falling tones	Nucleus placement for long/non-final utterances & non-falling tones		Nucleus placement for long/non-final utterances	Nucleus placement for long/non-final utterances
Connected speech phenomena	Elision	Elision	Elision	Elision	

*Note.* <sup>a</sup>This target is only required by the less strong version of the LFC (Jenkins, 2000).

## Chapter 7 Conclusion

### 7.1. Conclusion

At the very beginning of this dissertation, the author noted that pronunciation is unique in that it is impossible to learn or teach it selectively. All kinds of phones in the phonetic and phonological inventory of a language occur in any context. Even a simply-structured sentence can show rhythmic and intonational features of the language. At the same time, it is absolutely possible for one to speak foreign languages, employing one's L1 phonetic and phonology. It makes it even harder to identify where a problem lies in one's pronunciation of the target language.

The present study focused on Japanese learners of English who had learned English under the guidelines of the course of study (MEXT, 1998, 2009). It has been reported that the experience of living in an English-speaking country benefits learning pronunciation. It sounds instinctively true, even without any empirical support. However, not all learners can enjoy opportunities of this sort. Most Japanese learners of English live in Japan, and this is where they are exposed to English. Targeting these learners in research is thus more necessary for practical purposes.

In order to address questions about the pronunciation learning by Japanese learners of English, the current study conducted an experiment to measure their productive aspects of pronunciation including the following elements: vowel quality, vowel duration, plosives, fricatives, approximants, rhythm, intonation and connected speech phenomena. It aimed to reveal the phonetic and phonological items of these elements of pronunciation that are easy, learnable or difficult for Japanese learners of English, and how they are related to one another in the learning process. Predictions were established for the first research question, based on the difficulty of the items and potential for learning them, under the framework of one of the most influential learning models of segments, the SLM, and the working model of learning L2 intonation, the LILt. Acoustic analyses were carried out using different measurements for each phonetic and phonological item. Then, based on the results obtained in the statistical analyses of the data, the phonetic and phonological items that are easy, learnable or difficult

were identified. A study for the second research question was conducted under the framework of DST, and the relationships between the elements of pronunciation were discussed. The primary findings were as follows.

As regards vowel quality, of the 10 tested monophthongs, /i:, e, æ, ʌ, ɑ:, ɔ:, ʊ/ are easy and /ɪ, ɜ:, u:/ are difficult. These three vowels were all identified as difficult items that almost none of the Japanese learners of English could learn. The analysis of vowel duration revealed that the distinctions of /i:-ɪ/ and /ɑ:-æ/ were easy, and that of /u:-ʊ/ was learnable. In contrast, the durational distinction of /ɑ:-ʌ/ was found to be difficult; however, some Japanese learners, although not the majority, achieved a native-like distinction in this vowel pair. It can therefore be learned if effective training is provided, while it is difficult.

Plosives, fricatives and approximants were targeted for consonants. VOT durations were measured for plosives, which showed that the VOTs of /p/ and /k/ were learnable, while that of /t/ was difficult. The aspirated and unaspirated distinction of /k/ is learnable, although that of /t/ was difficult. Some Japanese learners of English could approximate the level of native speakers in the production of both aspirated /t/ and unaspirated /t/, but it was difficult to achieve. The other consonants analyzed, voiceless fricatives /θ/ and /s/ and approximants /r/ and /l/, were also found to be difficult items. These consonants were defined as even more difficult than aspirated /t/ and unaspirated /t/ for Japanese learners to learn to produce.

For rhythm, the production of weak vowels in weak forms was mainly investigated. Although they were expected to be produced with a lower pitch, with weaker intensity, of shorter duration and with centralized vowels, they were all defined as difficult items for Japanese learners of English to realize. However, it was also revealed that some Japanese learners were more likely to learn to produce weak vowels with a lower pitch and weaker intensity than with a shorter duration and centralized vowel quality.

The realization of intonation was found to have both easy items and difficult items. It is easy for Japanese learners of English to place the nucleus on the final word of an utterance. In contrast, it is difficult for them to place the nucleus on an appropriate word when it occurs in the non-final word. The nucleus placement in long utterances, where extra



nuclei fall at the beginning of the utterances for instance, was also identified as a difficult item. Concerning nuclear tone choices in the syntactic and pragmatic context tested, it is easy for Japanese learners of English to use a falling tone for the utterances in which it is common, but it is difficult for them to use other types of tones, such as a low rise, fall-rise, and level, for the utterances where non-fall tones are more commonly used. Span and level were measured for phonetic items in realizing English intonation, and they were both defined as easy. This means that Japanese learners of English are able to speak in a native-like pitch height and with a native-like pitch range in a certain context.

Connected speech phenomena were examined for elision, CC linking and CV linking. Some Japanese learners of English successfully attained a native-like use of these items; however, the majority of the subjects did not. All these items are therefore difficult items.

These findings indicate which phonetic and phonological items require time, energy and effort to improve and master English pronunciation. If the target item is easy, learners will not need to worry about learning it. If it is learnable, they will not need to put much emphasis on learning it because it will be probably learned naturally to a certain extent. If it is learnable, they will attempt to improve it more because they may be able to learn and master it with a little effort. If it is difficult, they will need to devote intensive time to learning it. If it is difficult, they might give up learning it and spend more time on enhancing other linguistic skills, such as listening or writing. In the present study, the difficult items were categorized as D1, D2 and D3 representing different difficulty levels, which have further implications for learning these items. Based on this categorization, Japanese learners of English can select which difficult items to focus on in their learning and how to learn them.

The results for the second research question suggest that vowel quality and approximants had a supportive relationship in the learning process. There is probably also a weaker relationship between vowel quality and rhythm. These relationships were interpreted as being due to articulatory behavior that they have in common. A supportive relationship was also found between approximants and fricatives. Unlike the above relationships, this was assumed to be attributed to the similar difficulty level that these elements of pronunciation

would have when Japanese learners of English learn English pronunciation. The presence of these relationships will provide implications for effective pronunciation learning and teaching.

## **7.2. Limitations**

The present study contributes to describing the learning of pronunciation by Japanese learners of English. Extensive analyses were carried out, including eight elements of pronunciation, and they make it possible to observe the learning of L2 pronunciation from broader perspectives. However, there were several limitations, as follows. Future directions of research in this field are noted.

There were methodological limitations relating to acoustic measurements, subjects and materials. First of all, while the present study employed various acoustic measurements to examine the productive aspects of pronunciation by Japanese learners of English, there were more potential measurements that could have been used to characterize them. For example, Mennen, Schaeffler, et al. (2014) found that learners expanded their pitch range in the initial peak of IPs, while they compressed it in the later peaks of IPs. This suggests that there can be position-sensitive influences on the pitch range, rather than uniform influences on any position of IPs. They even claimed that it would be possible for learners to first learn to produce global pitch range differences and then to produce position-sensitive pitch range differences. This suggests that both local pitch range and global pitch range should be investigated, although this study highlighted only the latter feature. Although span was found to be an easy item, which did not support the hypothesis, further examination to compare the realization of both global pitch range and local pitch range is required to evaluate the results of the present study. There is also another option for the scale to be used in the experiment. Toivanen (2014) argued that the scale of equivalent rectangular bandwidth (ERB) was a better measurement than semitones (ST), while ST was also better than the linear scale. Toivanen, who conducted an experiment using the ERB scale, reported that Finnish learners of English produced less pitch variation than native speakers of English. The measurements of this study were reasonable because they were carefully selected based on the previous

research, but more measurements have been proposed. More detailed studies can thus be conducted from different angles if different measurements are employed. This will benefit both Japanese learners of English and their teachers.

The second methodological limitation involves the subjects. The sample size of the native speakers was small compared with that of the Japanese learners of English in this study. The subjects of the native speakers' group showed less consistent within-group pattern for the production of vowel duration and plosives than that of the other elements examined. More data would be required to verify whether or not this was due to individual differences. Another limitation involving the subjects was that, in order to examine relationships between the elements of pronunciation, Japanese learners of English with different proficiency levels will be needed. The JL subjects in the present study were generally homogeneous in that they were from the same population, although some differences among them were found. This might have disguised possible relationships between the elements in the learning process. This point will be noted later, as related to the issue of the application of the theory.

The final methodological limitation involves materials. This study used a phonetically-balanced passage to collect data, which has the advantage of investigating various aspects of pronunciation. However, it also has the drawback that this kind of passage sometimes provides unbalanced target words or utterances. For example, only one target word was analyzed for the VOTs of /p, t/. Using different types of materials, such as the combination of a word list and a passage, may make it possible to examine more detailed features in L2 pronunciation.

The other limitations concern the applicability of the theory, DST, to address the second research question, which aimed to identify a supportive relationship between the elements of pronunciation. It could have been approached in a different way. A correlation analysis was not necessarily suitable for detecting relationships between the variables within the entire pronunciation system. DST defines the learning stage where the same state remains for a relatively long time, called an *attractor state*, roughly equivalent to fossilization. This means that there could be a stable stage before a change to move to the next stage. DST also

recognizes a competitive relationship between the variables. However, a correlation analysis cannot unearth an attractor state and a competitive relationship. As noted in Section 1.4.4, the linearity of the effect in learning is not always presumed in DST. Nonetheless, a correlation analysis assumes a linear relationship between the two variables. The findings in the present study are still helpful because the profile of the subjects was also observed to interpret the results of the correlation analysis and this study focused on finding a supportive relationship. However, a method to assess more complicated relationships directly needs to be developed, which will also make it possible to detect a competitive or conditional relationship.

The present study did not consider possible differences among the subjects at an initial stage of learning, which is another limitation involving the applicability of the theory. DST highlights the impact of earlier learning on later learning. Although the subjects in this study was selected from a population with uniform English language backgrounds, another option could have been to divide the subjects considering more details in their initial English learning, and not to analyze them together as one group. This is noted by de Bot et al. (2007), who also emphasize the importance of looking at the learning of individual learners.

Longitudinal studies are also needed to find relationships between the elements of pronunciation. DST has been more commonly applied in longitudinal research than cross-sectional research. The former helps to observe learning of the target items more closely, whereas the latter is advantageous in that a general tendency could be inferred from a large sample. Longitudinal studies from the earliest stage of learning and cross-sectional studies including learners with more varieties of proficiency levels will both confirm the findings of the present study.

### **7.3. Further studies**

Although eight elements of pronunciation were investigated for various phonetic and phonological items in the present study, there are more to be explored in future studies. These include other elements of pronunciation, such as affricates, nasals, consonant clusters and lexical stress. They were not included in the present study because the difficulty of affricates and nasals are less frequently discussed than the consonants targeted, the production of

consonant clusters are immediately relevant to the articulation of a single consonant and lexical stress is rather word-dependent. However, the study of these elements should lead to more comprehensive research.

Similarly, some items not examined in the present study could be targeted in future studies: voiced plosives /b, d, g/, fricatives /f, v, ð, z, ʃ, ʒ, h/ and approximants /j, w/. Not only could these consonants be measured, but experiments beyond one category of consonants would also be worth carrying out. For instance, the substitution of /b/ for /v/ could be a target in one of those studies. The same applies to other phonetic and phonological items of intonation. While a falling tone was found to be easy for Japanese learners of English to use, this only holds true of the syntactic and pragmatic contexts that this study targeted, for which a falling tone is typically used. Another focus for further research could thus be on whether learners are able to use a falling tone for questions, for instance. Also, the findings of the present study that Japanese learners of English used a native-like falling tone simply mean that they can choose to use a falling tone when it is a typical tone for an utterance. This only concerns the phonological dimension of intonation, and does not ensure that a falling tone is realized authentically from a phonetic point of view. A study of this sort to examine pitch in the phonetic dimension is essential for learners who would like to attain a native-speaker level of pronunciation. Although the realization of pitch is likely to be affected by individual differences, standardization methods have been developed to analyze these phonetic features acoustically. It is of practical value to explore the more detailed realization of pitch, using these techniques.

While the current study aimed to conduct an extensive study of the productive aspects of pronunciation learning to offer practical implications for a major part of the English pronunciation system, there are many more features of pronunciation to be empirically examined. However, even the phonetic and phonological items dealt with in this study are not all required to be learned for every setting of communication. What learners need to learn and master all depends on the English that they would like to learn and master. This means that the findings here will be of even more value by defining a specific goal.

Teachers need to know what goals of pronunciation learners should pursue, and how to guide them, providing an effective pronunciation practice based on empirical evidence. Learners need to think about what English they would like to learn and to work steadily to achieve their goal. When learners define the goal that they will aim for, the findings of the present study will be more helpful guidelines. One of the definite purposes of this dissertation was to make a solid, worthwhile contribution to the field of pronunciation learning and teaching.

## References

- Abe, H. (2011). Effects of form-focused instructions on the acquisition of weak forms by Japanese EFL learners. *ICPhs*, 17, 184-187.
- Abercrombie, D. (1966). *Elements of general phonetics*. Edinburgh, UK: Edinburgh University Press.
- Adachi, K. (2006). *Tahenryou deeta kaisekihoh: Shinri kyouiku shakai kei no tame no nyuumon* [Multivariate data analysis: An introduction to studies in the field of psychology, education and society]. Kyoto, Japan: Nakanishiya-Shuppan.
- Adank, P., Smits, P., & van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *Journal of the Acoustical Society of America*, 116(5), 3099-3107.
- Al-Tamimi, J., & Khattab, G. (2015). Acoustic cues weighting in the singleton vs geminate contrast in Lebanese Arabic: The case of fricative consonants. *Journal of the Acoustical Society of America*, 138(1), 344-360.
- Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody and syllable structure. *Language Learning*, 42(2), 529-555.
- Anderson-Hsieh, J., Riney, T., & Koehler, K. (1994). Connected speech modifications in the English of Japanese ESL learners. *IDEAL*, 7, 31-52.
- Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics*, 32(2), 233-250.
- Aoyama, K., & Guion, S. G. (2007). Prosody in second language acquisition: Acoustic analyses of duration and F0 range. In O.-S. Bohn & M. J. Munro (Eds.), *Language experiences in second language speech learning: In honor of James Emil Flege* (pp. 281-297). Amsterdam, The Netherlands: John Benjamins.
- Arai, T., & Greenberg, S. (1997). The temporal items of spoken Japanese are similar to those of English. *Eurospeech-97*, 2, 1011-1014.

- Arimoto, J. (2002). Teaching materials in English pronunciation. *The Bulletin of Kansai University of International Studies*, 3, 1-13.
- Arimoto, J., Yamamoto, K., Yamamoto, T., Kochiyama, M., & Makino, M. (2008). Nihonjin no eigo intoneeshon to sono kyoyoudo: EIL no kanten ni motodoku shidou eno teigen [Acceptability on English intonation in the utterances by Japanese speakers]. *Research Institute for Communication, Kansai University of International Studies, Studies on Communication*, 6, 2-12.
- Ashby, M., & Maidment, J. (2005). *Introducing phonetic science*. Cambridge, UK: Cambridge University Press.
- Auzou, P., Özsancak, C., Morris, R. J., Jan, M., Eustache, F., & Hannequin, D. (2000). Voice onset time in aphasia, apraxia of speech and dysarthria: A review. *Clinical Linguistics and Phonetics*, 14(2), 131-150.
- Bada, E. (2001). Native language influence on the production of English sounds by Japanese learners. *The Reading Matrix*, 1(2), 1-15.
- Beckman, M. E. (1982). Segment duration and the 'mora' in Japanese. *Phonetica*, 39, 113-135.
- Beckman, M. E., & Elam, G. A. (1997). *Guideline for ToBI Labelling* (Version 3). Retrieved from [http://www.cs.columbia.edu/~agus/tobi/labelling\\_guide\\_v3.pdf](http://www.cs.columbia.edu/~agus/tobi/labelling_guide_v3.pdf)
- Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 9-54). Oxford, UK: Oxford University Press.
- Benesse Corporation (2015). GTEC for STUDENTS shaken gaiyou setsumeii [A briefing on GTEC for STUDENTS] Retrieved from [http://www.mext.go.jp/component/b\\_menu/shingi/giji/\\_icsFiles/afieldfile/2015/03/25/1356122\\_04.pdf](http://www.mext.go.jp/component/b_menu/shingi/giji/_icsFiles/afieldfile/2015/03/25/1356122_04.pdf)
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-204). Timonium, MD: York Press.



- Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, *109*(2), 775-794.
- Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13-34). Amsterdam, The Netherlands: John Benjamins.
- Birdsong, D. (2007). Nativelike pronunciation among late learners of French as a second language. In Bohn, O.-S. & Munro, M. J. (Eds.), *Language experiences in second language speech learning: In honor of James Emil Flege* (pp. 99-116). Amsterdam, the Netherlands: John Benjamins.
- Bloch, B. (1950). Studies in colloquial Japanese IV phonemics. *Language*, *26*(1), 86-125.
- Blumstein, S. E., & Stevens, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *Journal of the Acoustical Society of America*, *66*(4), 1001-1017.
- Boersma, P., & Weenink, D. (2010). Praat: doing phonetics by computer (Version 5.2) [Computer program]. Retrieved from <http://www.praat.org/>
- Boersma, P., & Weenink, D. (2015). Praat: doing phonetics by computer (Version 6.0.05) [Computer program]. Retrieved from <http://www.praat.org/>
- Bohn, O.-S., & Flege, J. E. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition*, *14*, 131-158.
- Bolinger, D. L. (1965). Pitch accent and sentence rhythm. In I. Abe & T. Kanekiyo (Eds.), *Form of English: Accent, morpheme, order* (pp. 139-180). Cambridge, UK: Cambridge University Press.
- Cabrera-Abreu, M., Vizcaíno-Ortega, F., & Hernández-Flores, C. N. (2013). Production errors in the learning process of falling and falling-rising tones. In J. Przedlacka, J. Maidment, & M. Ashby (Eds.), *Proceedings of the Phonetics Teaching and Learning Conference 2013* (p. 23-26). London, UK: University College London.

- Cairns, R. S. (1999). "I sink, therefore I am." Japanese learners and the English dental fricative. *JACET Bulletin*, 33, 1-8.
- Carey, M. D., Mannell, R. H., & Dunn, P. K. (2011). Does a rater's familiarity with a candidate's pronunciation affect the rating in oral proficiency interviews? *Language Testing*, 28(2), 201-219.
- Cauldwell, R. (2002). The functional irrhythmicality of spontaneous speech: A discourse view of speech rhythms. *Apples: Journal of Applied Language Studies*, 2(1), 1-24.
- Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (2010). *Teaching pronunciation* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Council of Europe (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge, UK: Cambridge University Press.
- Cruttenden, A. (1994). *Gimson's pronunciation of English* (5th ed.). London, UK: Edward Arnold.
- Cruttenden, A. (1997). *Intonation* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Cruttenden, A. (2014). *Gimson's pronunciation of English* (8th ed.). London, UK: Routledge.
- Crystal, D. (1985). *A dictionary of linguistics and phonetics* (2nd ed.). Oxford, UK: Basil Blackwell.
- Crystal, D. (2003). *English as a global language*. Cambridge, UK: Cambridge University Press.
- de Bot, K., & Larsen-Freeman, D. (2011). Researching second language development from a dynamic systems theory perspective. In M. H. Verspoor, K. de Bot, & W. Lowie (Eds.), *A dynamic approach to second language development* (pp. 5-23). Amsterdam, the Netherlands: John Benjamins.
- de Bot, K., Lowie, W., & Verspoor, M. (2007). A dynamic systems theory approach to second language acquisition. *Bilingualism: Language and Cognition*, 10(1), 7-21.
- de Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech

- rate automatically. *Behavior Research Methods*, 41(2), 385-390.
- de Manrique, A. N. B., & Massone, M. I. (1980). Acoustic analysis and perception of Spanish fricative consonants. *Journal of the Acoustical Society of America*, 69(4), 1145-1153.
- Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TOESOL Quarterly*, 39(3), 379-397.
- Derwing, T. M., & Rossiter, M. J. (2003). The effects of pronunciation instruction on the accuracy, fluency and complexity of L2 accented speech. *Applied Language Learning*, 13, 1-17.
- Deterding, D. (2001). The measurement of rhythm: A comparison of Singapore and British English. *Journal of Phonetics*, 29(2), 217-230.
- Espy-Wilson, C. Y. (1992). Acoustic measures for linguistic features distinguishing the semivowels /w j r l/ in American English. *Journal of the Acoustical Society of America*, 92(2), 736-757.
- Estebas, E. (2013). TL\_ToBI: A new system for teaching and learning intonation. In J. Przedlacka, J. Maidment, & M. Ashby (Eds.), *Proceedings of the Phonetics Teaching and Learning Conference 2013* (pp. 39-42). London, UK: University College London.
- Fant, G. (1968). Analysis and synthesis of speech processes. In B. Malmberg (Ed.), *Manual of phonetics* (pp. 173-177). Amsterdam, the Netherlands: North-Holland.
- Field, A. (2009). *Discovering statistics using IBM SPSS Statistics* (3rd ed.). London, UK: SAGE.
- Flege, J. E. (1987). The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalent classification. *Journal of Phonetics*, 15(1), 47-65.
- Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). Timonium, MD: York Press.
- Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In A. Meyer & N. Schiller (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 319-355). Berlin,

Germany: Mouton de Gruyter.

- Flege, J. E., & Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *Journal of the Acoustical Society of America*, 76(3), 708-721.
- Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America*, 106(5), 2973-2987.
- Flege, J. E., Munro, M. J., & MacKay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, 97(5), 3125-3134.
- Flege, J. E., Takagi, N., & Mann, V. (1995). Japanese adults can learn to produce English /ɹ/ and /l/ accurately. *Language and Speech*, 38(1), 25-55.
- Flege, J. E., Yeni-Komshian, G., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Learning*, 41, 479-491.
- Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: Preliminary data. *Journal of the Acoustical Society of America*, 84(1), 115-123.
- Fox, R. A., & Jacewicz, E. (2009). Cross-dialectal variation in formant dynamics of American English vowels. *Journal of the Acoustical Society of America*, 126(5), 2603-2618.
- Fujisaki, H., & Hirose, K. (1984). Analysis of voice fundamental frequency contours for declarative sentences of Japanese. *Journal of the Acoustical Society of Japan*, 5(4), 233-242.
- Fujisaki, H., Hirose, K., & Sugito, M. (1986). Comparison of acoustic features of word accent in English and Japanese. *Journal of the Acoustical Society of Japan*, 7(1), 57-63.
- Gimson, A. C. (1962). *An introduction to the pronunciation of English* (1st ed.). London, UK: Edward Arnold.
- Gimson, A. C. (1980). *An introduction to the pronunciation of English* (3rd ed.). London, UK: Edward Arnold.
- Gordon, M., Barthmaier, P., & Sands, K. (2002). A cross-linguistic acoustic study of voiceless

- fricatives. *Journal of the International Phonetic Association*, 32(2), 141-174.
- Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds “L” and “R.” *Neuropsychologia*, 9, 317-323.
- Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In N. Warner & C. Gussenhoven (Eds.), *Papers in laboratory phonology 7* (pp. 515-546). Berlin, Germany: Mouton de Gruyter.
- Graddol, D. (2006). *English next: Why global English may mean the end of ‘English as a foreign language.’* London: British Council. Retrieved from <http://englishagenda.britishcouncil.org/sites/ec/files/books-english-next.pdf>
- Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults’ perception of English consonants. *Journal of the Acoustical Society of America*, 107(5), 2711-2724.
- Hahn, L. D. (1994). The stress of compound nouns: Linguistic considerations and pedagogical implications. *IDEAL*, 7, 67-78.
- Hahn, L. D. (2004). Primary stress and intelligibility: Research to motivate the teaching suprasegmentals. *TESOL Quarterly*, 38(2), 201-233.
- Hallé, P. A., Best, C. T., & Levitt, A. (1999). Phonetic vs. phonological influences on French listeners’ perception of American English approximants. *Journal of Phonetics*, 27(3), 281-306.
- Halliday, M. A. K. (1967). *Intonation and Grammar in British English*. The Hague, The Netherlands: Mouton.
- Hammer, Ø., Harper, D. A. T., & Ryan, P. D. (2001a). PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*, 4(1), 9.
- Hammer, Ø., Harper, D. A. T., & Ryan, P. D. (2001b). PAST (Version 2.17) [Computer program]. Retrieved from <http://nhm2.uio.no/norlex/past/download.html>
- Hatano, H., & Kitamura, T. (2014). Acoustic and articulatory characteristics of English reduced vowels uttered by native speakers of English and Japanese: Analysis based on

- X-ray microbeam speech production database. *Journal of the Acoustical Society of Japan*, 70(3), 106-113.
- Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication*, 47, 360-378.
- Hermes, D. J., & van Gestel, J. C. (1991). The frequency scale of speech intonation. *Journal of the Acoustical Society of America*, 90(1), 97-102.
- Hieke, A. E. (1984). Linking as a marker of fluent speech. *Language and Speech*, 27(4), 343-354.
- Hillenbrand, J. M., Clark, M. J., & Nearey, T. M. (2001). Effects of consonant environment on vowel formant patterns. *Journal of the Acoustical Society of America*, 109(2), 748-763.
- Hirai, A. (2012). *Kyouiku shinri kei kenkyuu no tame no deeta bunseki nyuumon: Riron to jissen kara manabu SPSS katsuyouhou* [An introduction to data analysis for educational and psychological studies: Learning a method for utilizing SPSS from theories and practices]. Tokyo, Japan: Tokyotosho.
- Hirata, Y. (2004). Effects of speaking rate on the vowel length distinction in Japanese. *Journal of Phonetics*, 32(4), 565-589.
- Hisagi, M., Nishi, K., & Strange, W. (2008). Acoustic items of Japanese and English vowels: Effects of phonetic and prosodic context. In M. Endo-Hudson, P. Sells, & S.-A. Hum (Eds), *Japanese/Korean Linguistics 13* (pp. 223-224). Chicago, MI: University of Chicago Press.
- Hismanoglu, M., & Hismanoglu, S. (2010). Language teachers' preferences of pronunciation teaching techniques: Traditional or modern? *Procedia - Social and Behavioral Science*, 2, 983-989.
- Homma, Y. (1981). Durational relationship between Japanese stops and vowels. *Journal of Phonetics*, 9(3), 273-281.
- Huckvale, M. (2004). Speech Filing System (Version 4.5) [Computer program]. Retrieved from <http://www.phon.ucl.ac.uk/resource/sfs/download.php>
- Hughes, G., & Halle, M. (1956). Spectral properties of fricative consonants. *Journal of the Acoustical Society of America*, 28(2), 303-310.

- Igarashi, S. (1981). *Eibei hatsuon shinkou* (2nd ed.) [English: Its vocal expression]. Tokyo, Japan: Nanundou.
- Ingram, J. C. L., & Park, S.-G. (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics*, 25(3), 343-370.
- Ingram, J. C. L., & Park, S.-G. (1998). Language, context, and speaker effects in the identification and discrimination of English /r/ and /l/ by Japanese and Korean listeners. *Journal of the Acoustical Society of America*, 103, 1161–1174.
- International Phonetic Association (1999). *Handbook of the International Phonetic Association: A guide to the use of the international phonetic alphabet*. Cambridge, UK: Cambridge University Press.
- Iverson, P., & Kuhl, P. K. (1996). Influence of phonetic identification and category goodness on American listeners perception of /r/ and /l/. *Journal of the Acoustical Society of America*, 99(2), 1130-1140.
- Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2001). A perceptual interference account of acquisition difficulties for non-native phonemes. *Speech, Hearing and Language: Work in progress*, 13, 106-118.
- Jenkins, J. (2000). *The phonology of English as an international language*. Oxford, UK: Oxford University Press.
- Jenkins, J. (2007). *English as a Lingua Franca: Attitude and identity*. Oxford, UK: Oxford University Press.
- Jenkins, J. (2009). *World Englishes: A resource book for students* (2nd ed.). London, UK: Routledge.
- Jia, G., Strange, W., Wu, Y., Collado, J., & Guan, Q. (2005). Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure. *Journal of the Acoustical Society of America*, 119(2), 1118-1130.
- Johnson, K. (2003). *Acoustic and auditory phonetics* (2nd ed.). Oxford, UK: Blackwell.
- Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America*, 108(3), 1252-1263.

- Joto, A. (1983). Some intonational characteristic of Japanese learners of English. *Bulletin of Chugoku Junior College*, *14*, 136-143.
- Joto, A., Nagase, Y., & Funatsu, S. (2007). The effect of VOT on the intelligibility of English voiceless stops produced by native speakers of Japanese. *Online Proceedings of the Phonetics Teaching & Learning Conference 2007, University College London*.
- Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The English language in the outer circle. In R. Quirk & H. Widdowson (Eds.), *English in the world: Teaching and learning the language and literatures* (pp. 11-30). Cambridge, UK: Cambridge university Press.
- Kamura, M. (2011). Realization of English nuclear accent by non-native speakers (NNS): Relation of realized tonicity and intelligibility between NNS. *Journal of the Phonetic Society of Japan*, *15*(1), 73-86.
- Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgements of language learner proficiency in oral English. *The Modern Language Journal*, *94*, 554-566.
- Kashiwagi, A., Snyder, M., & Craig, J. (2006). Suprasegmentals vs. segmentals: NNS phonological errors leading to actual miscommunication. *JACET Bulletin*, *43*, 43-57.
- Kato, A., & Cox, F. (2006). Development of Japanese length contrast: A longitudinal study of L2 vowels produced by Australian learners of Japanese. *Proceedings of the 11th Australian International Conference on Speech Science & Technology*, 170-175.
- Keating, P. A., & Huffman, M. K. (1984). Vowel variation in Japanese. *Phonetica*, *41*, 191-207.
- Kent, R. D., & Read, C. (2002). *Acoustic analysis of speech* (2nd ed.). Clifton Park, NY: Delmar.
- Kewley-Port, D. (1983). Time-varying features as correlated of place of articulation in stop consonants. *Journal of the Acoustical Society of America*, *73*(1), 322-335.
- Kikuchi, H., Miyajima, T., & Shen, R. (2013). Clustering of boundary tones at the accentual phrase edge in the expressive speech corpus. *Proceedings of the 3rd Japanese Corpus*



*Linguistics Workshop*, 23-28.

- Kleber, F., Harrington, J., & Reubold, U. (2011). The relationship between the perception and production of coarticulation during a sound change in progress. *Language and Speech*, 55(3), 383-405.
- Kochiyama, M., Arimoto, J., & Nakanishi, N. (2013). English pronunciation teaching in teacher's license courses. *Language Education and Technology*, 50, 119-130.
- Kori, S. (2003). Intoneeshon [Intonation]. In Z. Ueno (Ed.), *Asakura nihongo kooza 3: Onsei to onin* [Asakura Japanese course 3: Phonetics and phonology] (pp. 109-131). Tokyo, Japan: Asakura.
- Kori, S. (2011a). Intoneeshon [Intonation]. In H. J. Jôo, T. Fukumori, & Y. Saito (Eds.), *Onseigaku kihon jiten* [Dictionary of basic phonetic terms] (pp. 338-348). Tokyo, Japan: Bensei.
- Kori, S. (2011b). Kyoocyo [Emphasis]. In H. J. Jôo, T. Fukumori, & Y. Saito (Eds.), *Onseigaku kihon jiten* [Dictionary of basic phonetic terms] (pp. 376-378). Tokyo, Japan: Bensei.
- Kubozono, H. (1999). *Nihongo no onsei* [Japanese phonetics]. Tokyo, Japan: Iwanami.
- Kuhl, P. K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*, 50, 93-107.
- Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences, USA*, 97, 11850-11857.
- Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493), 979-1000
- Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the “perceptual magnet effect.” In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121-154). Timonium, MD: York Press.

- Kusumoto, Y. (2012). Between perception and production: Is the ability to hear L1-L2 sound differences related to the ability to pronounce the same sounds accurately? *Polyglossia*, 22, 15-33.
- Ladd D. R. (1996). *Intonational Phonology*. Cambridge, UK: Cambridge University Press.
- Ladefoged, P. (2001). *Vowels and consonants: An introduction to the sounds of languages*. Oxford, UK: Blackwell.
- Ladefoged, P. (2003). *Phonetic data analysis: An introduction to field work and instrumental techniques*. Malden, MA: Blackwell.
- Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, 26, 227-247.
- Lee, B., Guion, S. G., & Harada, T. (2006). Acoustic analysis of the production of unstressed English vowels by early and late Korean and Japanese bilinguals. *Studies in Second Language Acquisition*, 28, 487-513.
- Levis, J. M. (2002). Reconsidering low-rising intonation in American English. *Applied Linguistics*, 23(1), 56-82.
- Li, F., Edwards, J., & Beckman, M. (2007). Spectral measures for sibilant fricatives of English, Japanese and Mandarin Chinese. *ICPhS16*, 917-920.
- Li, A., & Post, B. (2014). L2 acquisition of prosodic items of speech rhythm. *Studies in Second Language Acquisition*, 26, 223-255.
- Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America*, 35(11), 1773-1781.
- Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20(3), 384-422.
- Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *Journal of the Acoustical Society of America*, 49(2), 606-608.
- Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative characterisations of speech rhythm: 'Syllable-timing' in Singapore English. *Language and Speech*, 43, 377-401.

- Maeda, M. (2005). Intonation and duration in English questions of Japanese speakers. *Language*, 28, 59-88.
- Maekawa, K., Kikuchi, H., Igarashi, Y., & Venditti, J. (2002). X-JToBI: An extended J\_ToBI for spontaneous speech. *Proceedings of the 7th International Conference on Spoken Language*, 1545-1548.
- Maekawa, K. (1999). Prosody and communication. *Journal of the Acoustical Society of Japan*, 55(2), 119-125.
- Maekawa, K., Igarashi, Y., Kikuchi, H., & Yoneyama, S. (2004). *Nihongo hanashi kotoba coopasu no intoneeshon reberingu* (version 1.0) [Intonation labelling of the corpus of spontaneous Japanese (version 1.0)]. Retrieved from <http://www2.ninjal.ac.jp/kikuo/intonation.pdf>
- Major, R. C. (2007). Identifying a foreign accent in an unfamiliar language. *Studies in Second Language Acquisition*, 29, 539-556.
- Maniwa, K., Jongman, A., & Wade, T. (2008). Acoustic characteristics of clearly spoken English fricatives. *Journal of the Acoustical Society of America*, 125(6), 3962-3973.
- Markham, D., & Hazan, V. (2002). The UCL speaker database. *Speech, Hearing and Language: Work in Progress*, 14, 1-17.
- Matsui, J.-K. (1998). The production and perception of the 'deleted' [t] in English. *Annual Review of English Language Education in Japan*, 9, 97-106.
- Matsusaka, H. (1986). *Eigo onseigaku nyuumon* [An introduction to English phonetics]. Tokyo, Japan: Kenkyusha.
- Maxwell, C. (1997) Connected speech phenomena—assimilation, elision, linking, and weakening: A study of Japanese L2 learners. *CELE Journal*, 5, 66-77.
- Mayr, R., & Davies, H. (2011). A cross-dialectal acoustic study of the monophthongs and diphthongs of Welsh. *Journal of the International Phonetic Association*, 41(1), 1-25.
- McAllister, R., Flege, J. E., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics*, 30(2), 229-258.

- Mennen, I. (2007). Phonological and phonetic influences in non-native intonation. In J. Trouvain & U. Gut (Eds.), *Non-native prosody: Phonetic description and teaching practice* (pp. 53-76). Berlin, Germany: Mouton de Gruyter.
- Mennen, I. (2015). Beyond segments: Towards a L2 intonation learning theory. In E. Delais-Roussarie, M. Avanzi, & S. Herment (Eds.), *Prosody and language in contact: L2 acquisition, attrition and languages in multilingual situations* (pp. 171-188). Berlin, Germany: Springer.
- Mennen, I., & de Leeuw, E. (2014). Beyond Segments: Prosody in Second Language Acquisition. *Studies in Second Language Acquisition*, 36, 183-194.
- Mennen, I., Schaeffler, F., & Dickie, C. (2014). Second language acquisition of pitch range in German learners of English. *Studies in Second Language Acquisition*, 36, 303-329.
- Merfert, I. (1997). The alt.usage.english Audio Archive. Retrieved from <http://alt-usage-english.org/index.shtml>
- Minematsu, N. (2004). Automatic scoring of language learners' pronunciations based on the distortion of their universal structures. *Technical Report of IEICE, SP2003-180*, 31-36.
- Minematsu, N., Shiho, A., Murakami, T., Maruyama, K., & Hirose, K. (2005). Structural representation of speech and its distance measure. *Technical Report of IEICE, SP2005-13*, 9-12.
- MEXT (1998). *Koutougakkou gakushuu shidou youryou Gaikokugo* [The course of study for senior high school: Foreign languages]. Retrieved from [http://www.mext.go.jp/a\\_menu/shotou/cs/1320334.htm](http://www.mext.go.jp/a_menu/shotou/cs/1320334.htm)
- MEXT (2009). *Koutougakkou gakushuu shidou youryou: Gaikokugo* [The course of study for senior high school: Foreign languages]. Retrieved from [http://www.mext.go.jp/a\\_menu/shotou/new-cs/youryou/kou/kou.pdf](http://www.mext.go.jp/a_menu/shotou/new-cs/youryou/kou/kou.pdf)
- MEXT (2011). *Koutougakkou kyouiku no genjyou* [Upper secondary education in Japan]. Retrieved from

[http://www.mext.go.jp/component/a\\_menu/education/detail/\\_\\_\\_icsFiles/afieldfile/2011/09/27/1299178\\_01.pdf](http://www.mext.go.jp/component/a_menu/education/detail/___icsFiles/afieldfile/2011/09/27/1299178_01.pdf)

MEXT (2013). *Globalka ni taioushita eigokyouikukaikaku jisshikeikaku* [English education reform plan corresponding to globalization]. Retrieved from [http://www.mext.go.jp/a\\_menu/kokusai/gaikokugo/\\_\\_\\_icsFiles/afieldfile/2014/01/31/1343704\\_01.pdf](http://www.mext.go.jp/a_menu/kokusai/gaikokugo/___icsFiles/afieldfile/2014/01/31/1343704_01.pdf)

Morizumi, M. (2009). Japanese English for EIAL: What it should be like and how much has been introduced. In K. Murata & J. Jenkins (Eds), *Global Englishes in Asian Contexts: Current and future debates* (pp. 73-93). London, UK: Palgrave Macmillan.

Mousa, A. (2014). Acquisition of the inter-dental fricatives /θ/ and /ð/ in ESL/EFL and Jamaican Creole: A comparative study. *Open Journal of Modern Linguistics*, 4, 38-47.

Munro, M. J. (1993). Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech*, 36(1), 39-66.

Munro, M. J. (1995). Nonsegmental factors in foreign accent. *Studies in Second Language Acquisition*, 37, 17-34.

Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, 38, 289-306.

Munro, M. J., & Derwing, T. M. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 49(Supp. 1), 285-310.

Munro, M. J., Flege, J. E., & MacKay, I. R. A. (1996). The effects of age of second language learning on the production of English vowels. *Applied Psycholinguistics*, 17, 313-334.

Nagamine, T. (2002). An experimental study on the teachability and learnability of English intonational aspect: Acoustic analysis on F0 and native-speaker judgment. *Journal of Language and Linguistics*, 1(4), 362-399.

Nakamura, A., Suzuki, M., Minematsu, N., Hirose, K., & Makino, T. (2010). An experimental study on assessment of diphthongs of learners using pronunciation structure. *Proceedings of the 2010 Spring Meeting of the Acoustical Society of Japan*, 1-P-15, 439-442.

Nakano, M. (2008). Eigohatsuon no tokucyou to sekaikyoutsuugo toshiteno eigo

- [Characteristics of English pronunciation and English as a lingua franca]. In Y. Yano & M. Ikeda (Eds), *Eigo sekai no kotoba to bunka* [Language and culture in English world] (pp. 286-298). Tokyo, Japan: Seibundo.
- Narita, T., & Tanaka, K. (2012) A study on pitch patterns of Japanese speakers of English in comparison with native speakers of English. *Journal of the Acoustical Science and Technology*, 33(4), 247-254.
- Nation, P. (2005). Range program [Computer program]. Retrieve from <http://www.victoria.ac.nz/lals/about/staff/paul-nation>
- Nelson, C. L. (2011). *Intelligibility in world Englishes: Theory and application*. New York, NY: Routledge.
- Nishi, K., Strange, W., Akahane-Yamada, R., Kubo, R., & Trent-Brown, S. A. (2008). Acoustic and perceptual similarity of Japanese and American English vowels. *Journal of the Acoustical Society of America*, 124(1), 576-588.
- O'Connor, J. D., & Arnold, G. F. (1973). *Intonation of colloquial English* (2nd ed.). London, UK: Longman.
- Oh, G. E., Guion-Anderson, S., Aoyama, K., Flege, J. E., Akahane-Yamada, R., & Yamada, T. (2011). A one-year longitudinal study of English and Japanese vowel production by Japanese adults and children in an English-speaking setting. *Journal of Phonetics*, 39(2), 156-167.
- Ohara, Y. (1992). Gender dependent pitch levels: A comparative study in Japanese and English. In K. Hall, M. Bucholtz, & B. Moonwomon (Eds.), *Locating power: Proceedings of the Second Berkley Women and Language Conference*, 2, 468-477.
- Ohata, K. (2004). Phonological differences between Japanese and English: Several potentially problematic areas of pronunciation of Japanese ESL/EFL learners. *Asian EFL Journal*, 6(4), 1-19.
- Ortega-Llebaria, M., & Colantoni, L. (2014). The L2 acquisition of English intonation: Form-meaning associations and maintenance of auditory resolution to acoustic cues. *Studies in Second Language Acquisition*, 36(2), 331-353.

- Patterson, D. (2000). *A linguistic approach to pitch range modelling* (Doctoral dissertation, University of Edinburgh).
- Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds), *Intentions in communication* (pp. 271-311). Cambridge, MA: The MIT Press.
- Pike, K. (1945). *The intonation of American English*. Ann Arbor, MI: University of Michigan Press.
- Piske, T., MacKay, I. R., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29(2), 191-215.
- Port, R. F., Al-Ani, S., & Maeda, S. (1980). Temporal compensation and universal phonetics. *Phonetica*, 37, 235-252.
- Port, R. F., Dalby, J., & O'Dell, M. (1987). Evidence for mora timing in Japanese. *Journal of the Acoustical Society of America*, 81(5), 1574-1585.
- Prator, C. H., & Robinett, B. W. (1985). *Manual of American English pronunciation* (4th ed.). San Diego, CA: Harcourt College Publisher.
- Ramus, F., Nespors, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, 265-292.
- Rimac, R., & Smith, B. L. (1984). Acoustic characteristics of flap productions by American English-speaking children and adults: Implications concerning the development of speech motor control. *Journal of Phonetics*, 12(4), 387-396.
- Riney, T. J., & Flege, J. E. (1998). Changes over time in global foreign accent and liquid identifiability and accuracy. *Studies in Second Language Acquisition*, 20, 213-243.
- Riney, T. J., & Takagi, N. (1999). Global foreign accent and voice onset time among Japanese EFL speakers. *Language Learning*, 49(2), 275-302.
- Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In D. Crystal (Ed.), *Linguistic Controversies* (pp. 73-79). London, UK: Edward Arnold.
- Roach, P. (2002). *English phonetics and phonology* (4th ed). Cambridge, UK: Cambridge University Press.

- Robinson, B. F., & Mervis, C. B. (1998). Disentangling early language development: Modeling lexical and grammatical acquisition using an extension of case-study methodology. *Developmental Psychology, 34*(2), 363-375.
- Saito, K., & Lyster, R. (2011). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /ɹ/ by Japanese learners of English. *Language Learning, 62*(2), 595-633.
- Sakata, M., Azukisawa, A., Maeno, Y., Yamada, T., & Wakita, M. (2001). Comparison of the English sound “r” as pronounced by native speakers of Japanese and English, using the method of frequency analysis. *Memoirs of the Konan University, Science Series, 48*(2), 65-80.
- Sakata, M., Azukisawa, A., Shinoki, T., Yamada, T., & Wakita, M. (1997). Comparison of the English sound ‘s’ as pronounced by native speakers of Japanese and English, using the method of frequency analysis. *Memoirs of the Konan University, Science Series, 44*(1), 43-59.
- Sato, T. (1999). The phonetic differences between Japanese and English: The characteristics of English pronunciation by Japanese speakers. *Journal of the Phonetic Society of Japan, 3*(2), 40-50.
- Sato, H., & Ueda, I. (2011). Misplacement of nuclear stress by Japanese learners of English. *Journal of the Phonetic Society of Japan, 15*(1), 87-95.
- Satoi, H., Yoshimura, M., & Yabuuchi, S. (2005). The relationship between English speech rhythm and vowel reduction in production: Comparison between Japanese EFL learners and native English speakers. *Language, Education and Technology, 42*, 59-72.
- Schmid, M. S., & Hopp, H. (2014). Comparing foreign accent in L1 attrition and L2 acquisition: Range and rater effects. *Language Testing, 31*(3), 367-388.
- Seidlhofer, B. (2004). Research perspectives on teaching English as a lingua franca. *Annual Review of Applied Linguistics, 24*, 209-239.
- Setter, J., Stojanovik, V., & Martínez-Castilla, P. (2010). Evaluating the intonation of non-native speakers of English using a computerized test battery. *International Journal of*



- Applied Linguistics*, 20(3), 368-385.
- Shigemasa, T., Mori, Y., & Yanai, H. (2008). *Q & A de shiru toukei deeta kaiseki: DOs and DON'Ts* (2nd ed.) [Statistical data analysis learned by Q & A: DOs and DON'Ts]. Tokyo, Japan: Saiensu-sha.
- Shimada, T. (2005). On weak vowels in English: Towards an effective way to instruct Japanese learners of English. *Memoirs of the Muroran Institute of Technology*, 55, 1-7.
- Shimizu, K. (1993). A cross-language study of phonetic characteristics of stop consonants: With reference for voicing contrasts. *Journal of Asian and African Studies*, 45, 163-175.
- Shimizu, K. (1999). Eigoonseigakushuu ni okeru inyuu: Boin no hatsuon to sono onkyouteki tokucyou [Transfer in English pronunciation learning: Pronunciation of vowels and their acoustic characteristics]. *Nagoya Gakuin University Round Table on Languages, Linguistics and Literature*, 29, 1-9.
- Shimizu, K. (2008). L2 onseigakushuu to sono rirontekihaikei [Learning of L2 sounds and its theoretical backgrounds]. *Journal of Nagoya Gakuin University; Language and Culture*, 19(2), 81-87.
- Shimizu, A. (2011). English as a lingua franca and the teaching of pronunciation. *Journal of the Phonetic Society of Japan*, 15(1), 44-62.
- Slawinski, E. (1999). Acquisition of /r-l/ phonemic contrast by Japanese children and adults. *Psycholinguistics on the Threshold of the Year 2000*, 583-590.
- Smith, L. (1976). English as an international auxiliary language. *RELC Journal*, 7(2), 38-53.
- So, C. K., & Best, C. T. (2014). Phonetic influences on English and French listeners' assimilation of Mandarin tones to native prosodic categories. *Studies in Second Language Acquisition*, 36(2), 195-221.
- Stevens, J. P. (2007). *Intermediate statistics: A modern approach* (3rd ed.). New York, NY: Routledge.
- Stölten, K. (2006). Effects of age on VOT: Categorical perception of Swedish stops by near-native L2 speakers. *Lund University, Center for Languages & Literature, Department of Linguistics & Phonetics: Working Papers*, 52, 125-128.

- Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., Nishi, K., & Jenkins, J. J. (1998). Perceptual assimilation of American English vowels by Japanese listeners. *Journal of Phonetics*, 26(4), 311-344.
- Strange, W., Bohn, O.-S., Trent, S. A., & Nishi, K. (2004). Acoustic and perceptual similarity of North German and American English vowels. *Journal of the Acoustical Society of America*, 115(4), 1791-1807.
- Strange, W., Weber, A., Levy, E. S., Shafiro, V., Hisagi, M., & Nishi, K. (2007). Acoustic variability within and across German, French, and American English vowels: Phonetic context effects. *Journal of the Acoustical Society of America*, 122(2), 1111-1129.
- Sudo, M. M. (2010a). Nihonjincyuugakusei ni yoru eigo no rizumu pattern shuutoku [Acquisition of English rhythmic patterns by Japanese junior high school students]. In M. M. Sudo (Ed.), *Eigo no onseishuutoku ni okeru seisei to chikaku no mekanizumu: Nihonjineigogakushuusha no rizumu pattern shuutoku* [Mechanism of production and perception in the acquisition of English pronunciation: Acquisition of rhythmic patterns by Japanese learners of English] (pp. 39-57). Tokyo, Japan: Kazama.
- Sudo, M. M. (2010b). Nihonjindaigakusei ni yoru eigo no rizumu pattern shuutoku: Futatsu no kyoujyuhou no kouka [Acquisition of English rhythmic patterns by Japanese university students: Effects of two teaching methods]. In M. M. Sudo (Ed.), *Eigo no onseishuutoku ni okeru seisei to chikaku no mekanizumu: Nihonjineigogakushuusha no rizumu pattern shuutoku* [Mechanism of production and perception in the acquisition of English pronunciation: Acquisition of rhythmic patterns by Japanese learners of English] (pp. 59-75). Tokyo, Japan: Kazama.
- Sudo, M. M., & Kaneko, I. (2006). Effects of teaching methods on the acquisition of stress-related and focus-related durational control of English and Japanese junior high school students. *JACET Bulletin*, 42, 53-65.
- Sudo, M. M., & Kiritani, S. (1991). Production and perception of stress-related durational patterns in Japanese learners of English. *Journal of Phonetics*, 19(2), 231-248.
- Sussman, H. M., McCaffrey, H. A., & Matthews, S. A. (1991). An investigation of locus

- equations as a source of relational invariance for stop place categorization. *Journal of the Acoustical Society of America*, 90(3), 1309-1325.
- Suzuki, M., Minematsu, N., & Hirose, K. (2010). Non-native pronunciation assessment based on pronunciation structure and multilayer regression. *Transactions of Information Processing Society of Japan*, 52(5), 1899-1909.
- Suzuki, M., Qiao, Y., Minematsu, N., & Hirose, K. (2010). Integration of multilayer regression analysis with structure-based pronunciation assessment. *INTERSPEECH 2010*, 586-589.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Pearson Education.
- Takebayashi, S. (1996). *Eigo onseigaku* [English phonetics]. Tokyo, Japan: Kenkyusha.
- Takefuta, Y. (1982). *Nihonjin eigo no kagaku: Sonogenjoo to asu eno tembo* [Scientific analysis of the English of the Japanese people: Current issues and future prospects]. Tokyo, Japan: Kenkyusha.
- Thomson, R. I. (2013). ESL teachers' beliefs and practices in pronunciation teaching: Confidently right or confidently wrong? In J. Levis & K. LeVelle (Eds.), *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (pp. 224-233). Ames, IA: Iowa State University.
- Todaka, Y. (1994). Japanese students' English intonation. *Bulletin of Miyazaki Municipal University Faculty of Humanities*, 1(1), 23-47.
- Toivanen, J. (2005). ToBI or not ToBI? Testing two models in teaching English intonation to Finns. In J. Maidment (Ed.), *Proceedings of the Phonetics Teaching and Learning Conference 2005* (pp. 1-4). London, UK: University College London.
- Toivanen, J. (2014). Aspects of second language speech prosody: Data from research in progress. In M. Heldner (Ed.), *Proceedings from FONETIK 2014* (pp. 101-104). Stockholm, Sweden: Stockholm University.
- Torgersen, E. N., & Szakay, A. (2012). An investigation of speech rhythm in London English. *Lingua: New Horizons in Sociophonetic Variation and Change*, 122(7), 822-840.

- Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, 28(1), 251-276.
- Tsuji, A. (2004). The case study of high pitch register in English and in Japanese: Does high pitch register related to politeness? *Seijo English Monographs*, 37, 227-260.
- Ueda, H., & Otsuka, T. (2010). An analysis of pronunciation instruction in Japanese junior high school English textbooks: Indications from the early stage of input. *Journal of Osaka Jogakuin Univeristy*, 7, 15-32.
- van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech*, 38(3), 253-265.
- Vance, T. J. (1987). *An introduction to Japanese phonology*. Albany, NY: State University of New York Press.
- Venditti, J. J. (2005). The J\_ToBI model of Japanese intonation. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 172-200). Oxford, UK: Oxford University Press.
- Verspoor, M., & van Dijk, M. (2011). Visualizing interaction between variables. In M. H. Verspoor, K. de Bot, & W. Lowie (Eds.), *A dynamic approach to second language development* (pp. 85-98). Amsterdam, The Netherlands: John Benjamins.
- Walker, R. (2010). *Teaching the pronunciation of English as a lingua franca*. Oxford, UK: Oxford University Press.
- Watanabe, K. (1994). *Eigono intoneeshon ron* [English intonation]. Tokyo, Japan: Kenkyusha.
- Wells, J. C. (2000). Overcoming phonetic interference. *Journal of the English Phonetic Society of Japan*, 3, 9-21.
- Wells, J. C. (2006) *English intonation: An introduction*. Cambridge, UK: Cambridge University Press.
- Wennerstrom, A. (1994). Intonational meaning in English discourse: A study of non-native speakers. *Applied Linguistics*, 15(4), 399-420.

- Wester, F., Gilbers, D., & Lowie, W. (2007). Substitution of dental fricatives in English by Dutch L2 speakers. *Language Sciences*, 29, 477-491.
- Widdowson, H. G. (1994). The ownership of English. *TESOL Quarterly*, 28(2), 377-389.
- Winke, P., Gass, S., & Myford, C. (2012). Raters' L2 background as a potential source of bias in rating oral performance. *Language Testing*, 30(2), 231-252.
- Yamada, R. (1995). Age and acquisition of second language speech sounds perception of American English /ɹ/ and /l/ by native speakers of Japanese. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 305-320). Timonium, MD: York Press.
- Yamada, T. (2013). *Kokugo kyoushi ga shitteokitai nihongo onsei onseigenngo* [Japanese phonetics and spoken languages Japanese teachers have to know] (Rev. ed.). Tokyo, Japan: Kuroshio.
- Yamazawa, H., & Hollien, H. (1992). Speaking fundamental frequency patterns of Japanese women. *Phonetica*, 49, 128-140.
- Zhang, Y., & Elder, C. (2011). Judgments of oral proficiency by non-native and native English speaking teacher raters: Comparing or complementary construct? *Language Testing*, 28(1), 31-50.

## **Appendix A: Materials**

### *The Story of Arthur the rat* (for BN and JL subjects)

There was once a young rat named Arthur who would never take the trouble to make up his mind. Whenever his friends asked him if he would like to go out with them, he would only answer, “I don’t know.” He wouldn’t say “Yes” and he wouldn’t say “No” either. He could never learn to make a choice.

His Aunt Helen said to him “No-one will ever care for you if you carry on like this. You have no more mind than a blade of grass.” Arthur looked wise but said nothing.

One rainy day the rats heard a great noise in the loft where they lived. The pine rafters were all rotten, and at last one of the joists had given way and fallen to the ground. The walls shook and the rats’ hair stood on end with fear and horror. “This won’t do,” said the old rat who was chief, “I’ll send out scouts to search for a new home.”

Three hours later the seven scouts came back and said, “We’ve found a stone house which is just what we wanted. There’s room and good food for us all. There’s a kindly horse named Nelly, a cow, a calf and a garden with an elm tree.” Just then the old rat caught sight of young Arthur. “Are you coming with us?” he asked. “I don’t know,” Arthur sighed, “The roof may not come down just yet.” “Well,” said the old rat angrily, “We can’t wait all day for you to make up your mind. Right about face! March!” And they went off.

Arthur stood and watched the other rats hurry away. The idea of an immediate decision was too much for him. “I’ll go back to my hole for a bit,” he said to himself, “just to make up my mind.”

That night there was a great crash that shook the earth, and down came the whole roof. Next day some men rode up and looked at the ruins. One of them moved a board, and under it they saw a young rat lying on his side, quite dead, half in and half out of his hole.

### *Arthur the Rat* (for AN subjects)

Once there was a young rat named Arthur, who could never make up his mind. Whenever his friends asked him if he would like to go out with them, he would only answer,

“I don't know.” He wouldn't say “yes” or “no” either. He would always shirk making a choice.

His aunt Helen said to him, “Now look here. No one is going to care for you if you carry on like this. You have no more mind than a blade of grass.”

One rainy day, the rats heard a great noise in the loft. The pine rafters were all rotten, so that the barn was rather unsafe. At last the joists gave way and fell to the ground. The walls shook and all the rats' hair stood on end with fear and horror. “This won't do,” said the captain. “I'll send out scouts to search for a new home.”

Within five hours the ten scouts came back and said, “We found a stone house where there is room and board for us all. There is a kindly horse named Nelly, a cow, a calf, and a garden with an elm tree.” The rats crawled out of their little houses and stood on the floor in a long line. Just then the old one saw Arthur. “Stop,” he ordered coarsely. “You are coming, of course?” “I'm not certain,” said Arthur, undaunted. “The roof may not come down yet.” “Well,” said the angry old rat, “we can't wait for you to join us. Right about face. March!”

Arthur stood and watched them hurry away. “I think I'll go tomorrow,” he calmly said to himself, but then again “I don't know; it's so nice and snug here.”

That night there was a big crash. In the foggy morning some men—with some boys and girls—rode up and looked at the barn. One of them moved a board and he saw a young rat, quite dead, half in and half out of his hole. Thus the shirker got his due.

## Appendix B: Target tokens for AN subjects

### Monophthongal vowels (vowel quality and vowel duration)

1. /i:/: *either* and *tree*
  2. /ɪ/: *this* [2], *in*, *think* and *big*
  3. /e/: *never*, *whenever*, *friends*, *yes*, *Helen*, *said* [6], *end*, *send*, *ten*, *Nelly*, *yet*, *men*, *dead*, *again*, *fell* and *then* [2]
  4. /æ/: *rat* [6], *carry*, *back*, *that*, *crash*, *asked*, *answer*, *aunt*, *grass*, *last*, *can't*, *half* [2], *rather*, *captain* and *angry*
  5. /ʌ/: *young*, *up* [2], *coming*, *come* and *sung*
  6. /ɑ:/: *Arthur* [4], *garden*, *march*, *barn* [2], *stop*, *not* [2], *watched* and *calmly*
  7. /ɔ:/: *more*, *horse*, *coarsely*, *board* [2], *floor*, *order*, *course*, *move* and *mourning*
  8. /u:/: *do*, *room*, *roof* and *moved*
  9. /ʊ/: *wouldn't* [2], *looked* [2], *shook* [2] and *stood* [3]
  10. /ɜ:/: *heard*, *searched*, *shirk*, *certain*, *hurry* and *girls*
- 6 or 7. /ɑ: / or /ɔ:/: *long*, *horror*, *all* [3], *always*, *saw*, *undaunted*, *crawled* and *saw* [2]

### Plosives

1. /p/: *pine*
2. /t/: *ten*
3. /k/: *care*, *carry*, *kindly*, *cow*, *can't*, *calf* and *calmly*
4. /st/: *stood* [3], *stone* and *stop*
5. /sk/: *scouts* [2]

### Fricatives

1. /θ/: *Arthur* [4] and *think*
2. /s/: *yes*, *choice*, *said* [5], *grass*, *unsafe*, *send*, *search*, *house*, *horse*, *saw* [2], *course*, *certain*, *face* and *nice*



## Approximants

1. /r/: *rat, rainy, room, roof, right, rode, carry* and *hurry*
2. /l/: *like, look, loft, little, long, line, Helen* and *Nelly*

## Durational variability of successive stressed and unstressed vowels

1. *aunt Helen said to him*
2. *carry on like this*
3. *no more mind than a blade of grass*
4. *rats heard a great noise in the loft*
5. *send out scouts to search for a new home*
6. *garden with an elm tree*
7. *roof may not come down just yet*
8. *half in and half out of his hole*

## Weak vowels in weak forms

1. Pitch and intensity: *a* [4], *and* [3], *at* [1], *could* [1], *he* [2], *his* [2], *of* [1], *some* [1], *them* [1] and *to* [2],
2. Duration and vowel quality: *a* [10], *an* [1], *and* [10], *the* [9], *to* [5], *them* [3], *than*, *of* [5], *some* [2], *just* [1], *but*, *that* and *at* [2]

## Stressed vowels to compare against weak vowels

1. Against *a*: *great, stone* or *house, cow* and *garden*
2. Against *and*: *garden, watched* and *looked*
3. Against *at*: *last*
4. Against *could*: *never*
5. Against *he*: *only* and *wouldn't*
6. Against *his*: *friends* and *aunt*
7. Against *of*: *moved*

8. Against *some*: *men*
9. Against *them*: *moved*
10. Against *to*: *go* or *out* and *search*

#### Span and level

*That night there was a big crash. In the foggy morning some men—with some boys and girls—rode up and looked at the barn. One of them moved a board and he saw a young rat, quite dead, half in and half out of his hole. Thus the shirker got his due.*

#### The nucleus placement and the nuclear tone choice

1. Antecedent modified by the relative clause (ANT): *Once there was a young rat named Arthur (,who would never ...)*
2. End of the subordinate clause preceding the main clause (SD end): *go out with them*
3. Reporting clause before direct speech (bf DS): *he would only answer and said to him*
4. Short dialogue (DIA): *I don't know* and *This won't do*
5. Topic (TOPIC): *His aunt Helen*
6. Adverbial phrase (AdP): *one rainy day, at last, just then* and *that night*
7. Lists (LIST): *There was a kindly horse named Nelly, a cow and a calf*
8. Last component of closed lists (lastLIST): *and a garden with an elm tree*
9. Exclamation (EXCL): *Well*
10. Command (COM): *Right about face* and *March*

#### Elision

*won't do, don't know [2], named Nelly, can't wait, end with, watched the and mind than*

## CC linking

1. The same place of articulation and the same manner of articulation:

PPS: *said to* [2] and *quite dead*

PN: *rat named, great noise* and *that night*

CC: *with them* and *some men*

2. A different place of articulation or a different manner of articulation:

PPD: *like to* and *not come*

PA: *would like*

PF: *like this, about face* and *said the* [2]

## CV linking

1. A voiceless consonant:

PV: *make up* [3] *shook and, back and, right about, out of* and *looked at*

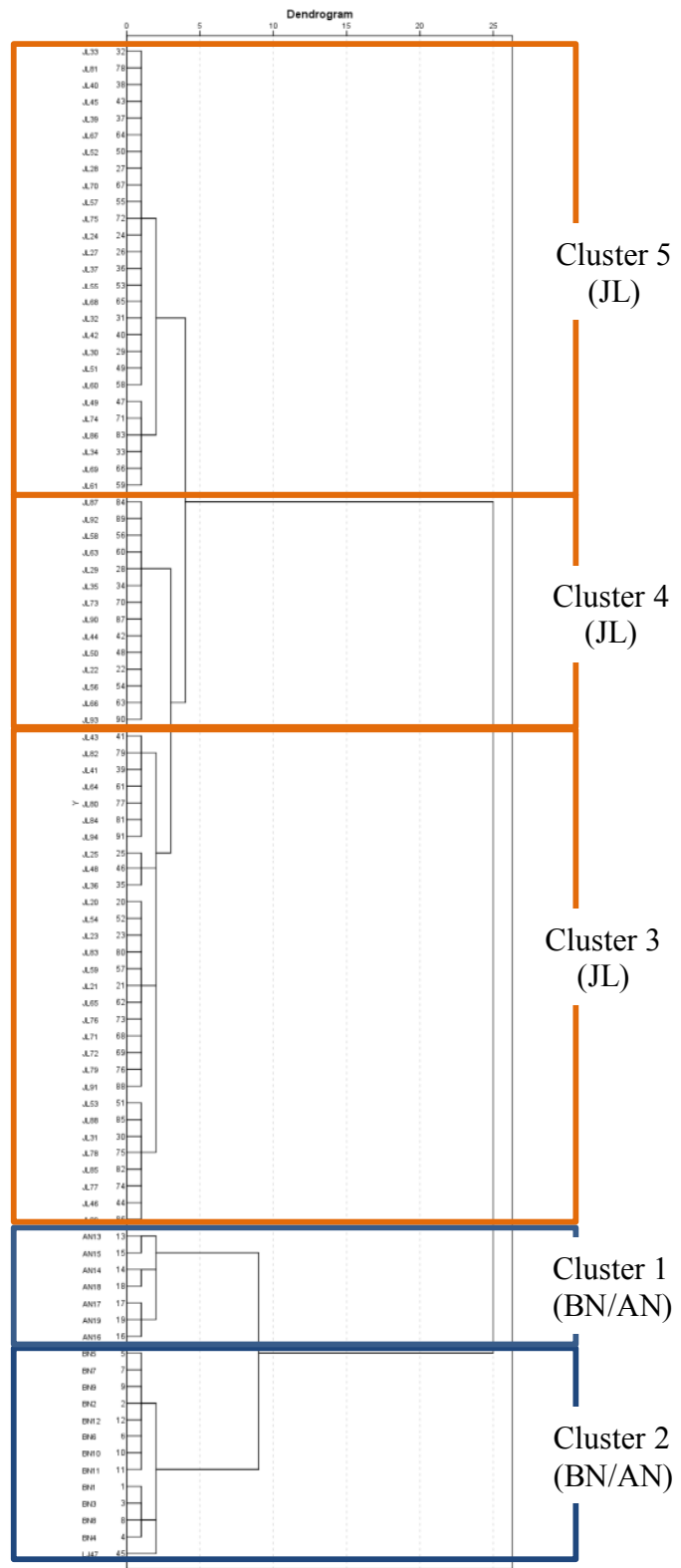
FlessV: *noise in, us all, half in* and *half out*

2. A voiced consonant

DV: *heard a, found a, moved a, named Arthur, would only, stood on, stood and, send out* and *rode up*

FedV: *friends asked, there's a, was a* and *with an*

# Appendix C: Dendrogram for vowel quality



## Appendix D: Correlations between the F1 variables

*Correlations between the standardized F1 mel values*

Variable	1	2	3	4	5	6	7	8	9
1. /ʌ/	—								
2. /æ/	-.32**	—							
3. /ɑ:/	.07	-.01	—						
4. /e/	-.35**	.39**	-.46**	—					
5. /ʊ/	-.22*	.44**	.13	.22*	—				
6. /ɜ:/	.01	-.65**	-.39**	-.36**	-.58**	—			
7. /ɪ/	-.32**	.65**	.01	.56**	.52**	-.62**	—		
8. /ɔ:/	.19	-.40**	.31**	-.48**	-.17	-.02	-.55**	—	
9. /u:/	.31**	-.42**	-.24*	-.36**	-.31**	.46**	-.63**	.19	—
10. /i:/	.14	-.18	.27**	-.36**	-.59**	.34**	-.43**	-.03	-.07

\*  $p < .05$ . \*\*  $p < .01$ .

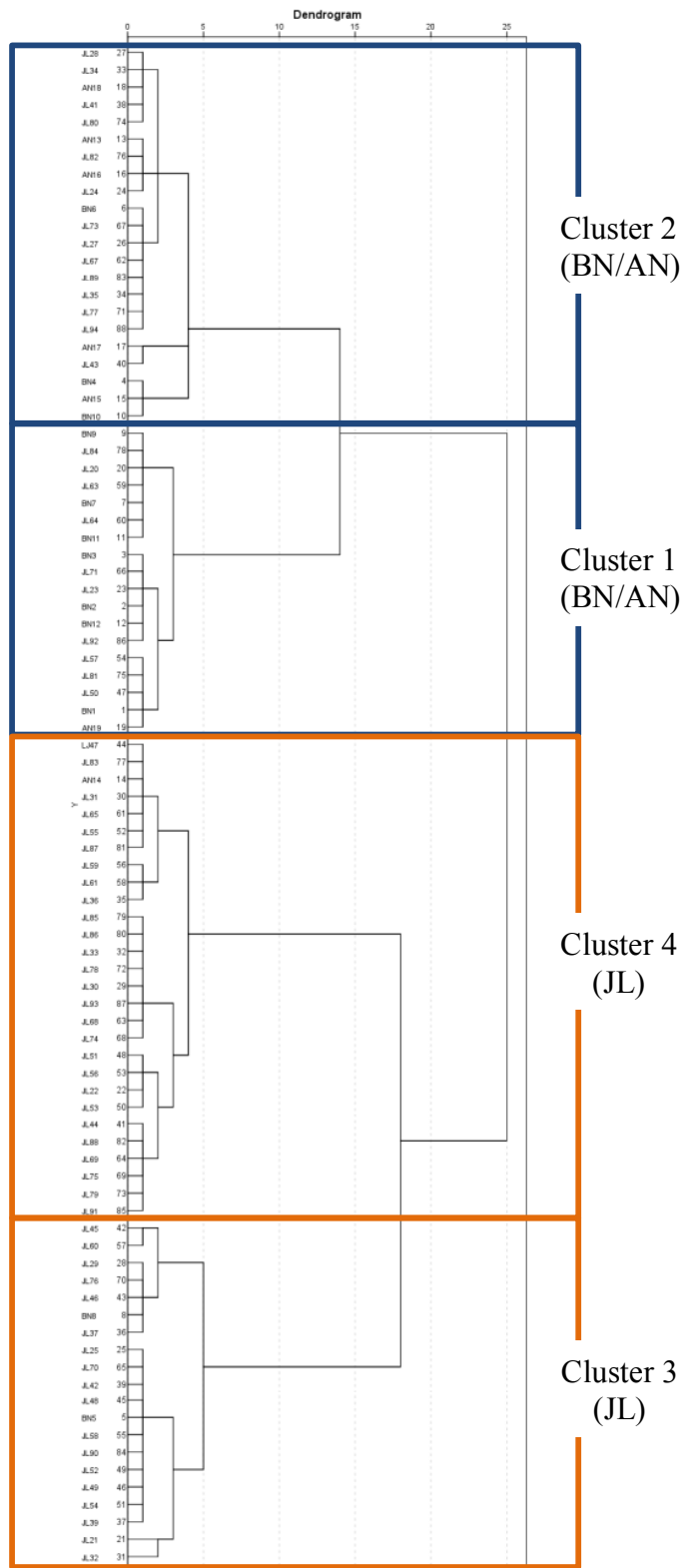
## Appendix E: Correlations between the F2 variables

*Correlations between the standardized F2 mel values*

Variable	1	2	3	4	5	6	7	8	9
1. /ʌ/	—								
2. /æ/	.51**	—							
3. /ɑ:/	.01	-.25*	—						
4. /e/	-.29**	-.25*	.30**	—					
5. /ʊ/	-.49**	-.60**	-.20	.10	—				
6. /ɜ:/	.51**	.42**	-.10	-.60**	-.45**	—			
7. /ɪ/	-.45**	-.49**	.52**	.49**	.07	-.47**	—		
8. /ɔ:/	-.46**	-.17	.01	.39**	-.09	-.57**	.59**	—	
9. /u:/	-.16	-.28**	-.46**	-.33**	.42**	-.02	-.51**	-.46**	—
10. /i:/	.19	.33**	-.13	-.35**	-.46**	.23*	-.27**	.28**	-.28**

\*  $p < .05$ . \*\*  $p < .01$ .

# Appendix F: Dendrogram for vowel duration



## Appendix G: Correlations between the variables for vowel duration

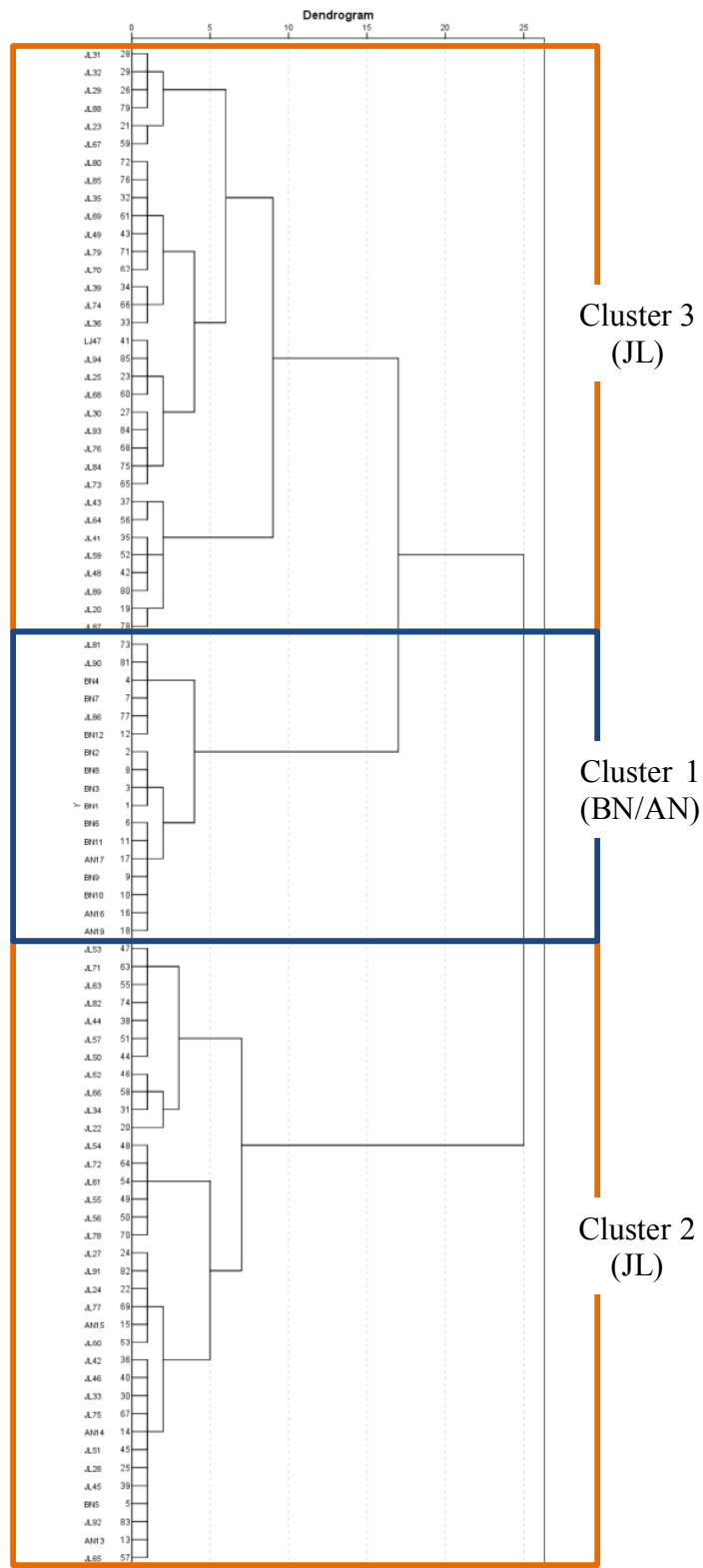
*Correlations between the Variables for Vowel Duration*

Variable	1	2	3
1. /i:-ɪ/	—		
2. /u:-ʊ/	.10	—	
3. /ɑ:-æ/	-.05	.16	—
4. /ɑ:-ʌ/	.20	.15	.37**

\*\*  $p < .01$ .



# Appendix H: Dendrogram for plosives



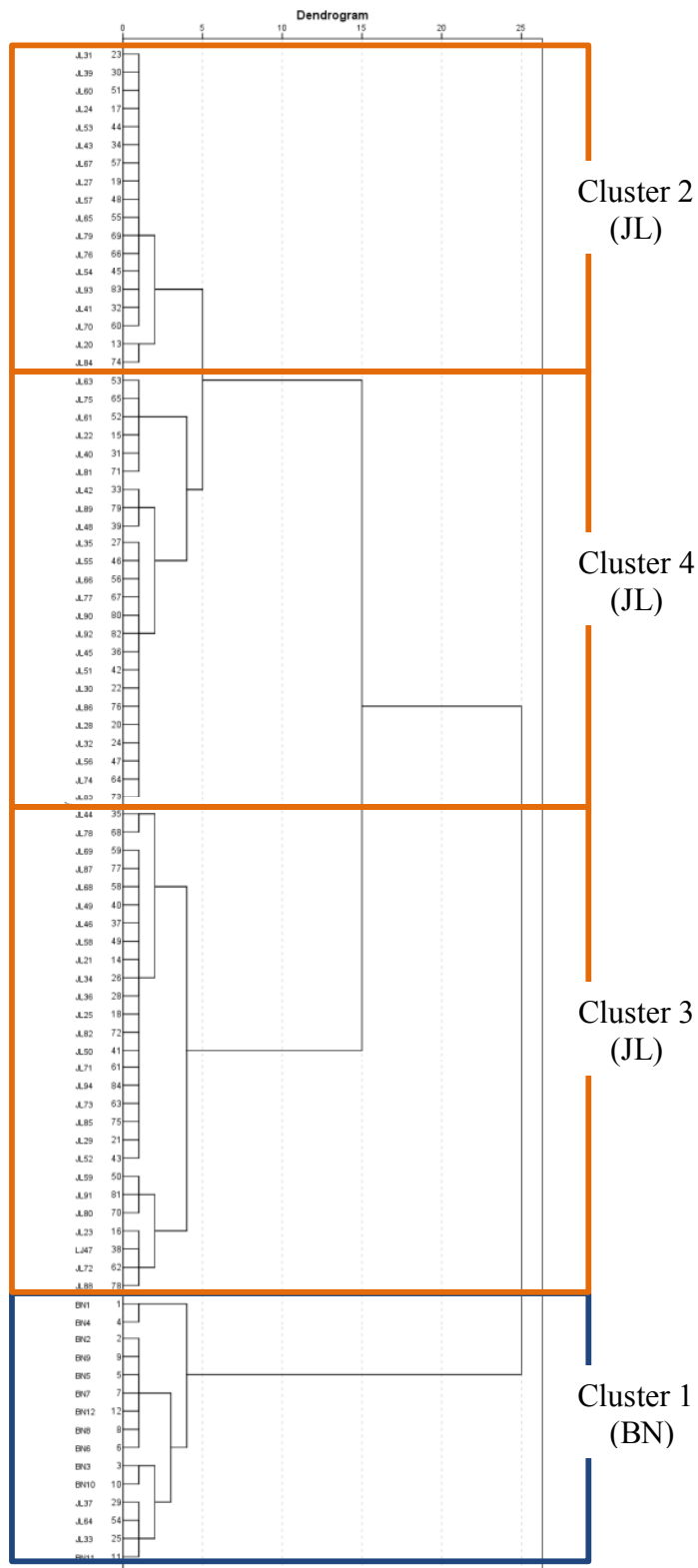
## Appendix I: Correlations between the variables for plosives

### *Correlations between the Variables for Plosives*

Variable	1	2	3	4
1. /p/	—			
2. /t/	.42**	—		
3. /k/	.36**	.33**	—	
4. /t-st/	-.29**	-.72**	-.13	—
5. /k-sk/	-.16	-.42**	-.37**	.46**

\*\*  $p < .01$ .

# Appendix J: Dendrogram for fricatives



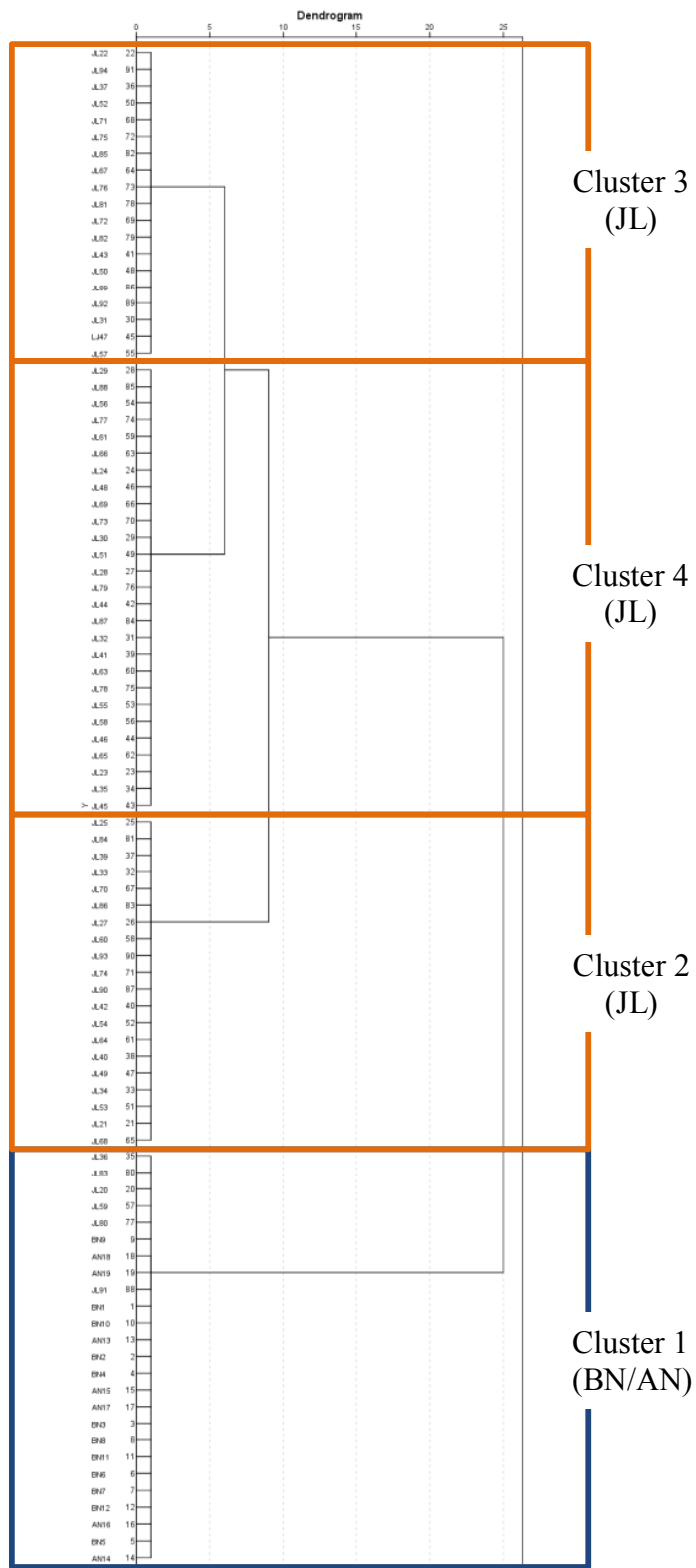
## Appendix K: Correlations between the variables for fricatives

*Correlations between the Variables for Fricatives*

Variable	1	2	3	4	5	6	7
1. /θ/ COG	—						
2. /θ/ SD	-.54**	—					
3. /θ/ skewness	-.90**	.35**	—				
4. /θ / kurtosis	-.22*	-.60**	.40**	—			
5. /s / COG	.90**	-.50**	-.81**	-.18	—		
6. /s/ SD	-.19	.51**	.09	-.44**	-.15	—	
7. /s/ skewness	-.71**	.36**	.81**	.25*	-.82**	-.05	—
8. /s/ kurtosis	-.42**	-.02	.47**	.52**	-.47**	-.68**	.61**

\*  $p < .05$ . \*\*  $p < .01$ .

# Appendix L: Dendrogram for approximants



## Appendix M: Correlation between the variables for approximants

### *Correlation between the Variables for Approximants*

Variable	1
1. /r/	—
2. /l/	.48**

\*\*  $p < .01$ .

## Appendix N: Correlations between the variables for rhythm

### *Correlations between the Variables for Rhythm*

Variable	1	2	3
1. Pitch	—		
2. Intensity	.53**	—	
3. Duration	.58**	.53**	—
4. Vowel centralization	-.35**	-.39**	-.78**

\*\*  $p < .01$ .

## Appendix O: Results of the pitch patterns for the target utterances

Only the tone of the target nucleus is presented for nuclear tone choice in the tables as in Table 4.27, whether the subjects placed extra nuclei on other words or not.

*There was once a young rat named Arthur ...*

	Nucleus placement		Nuclear tone choice			
	BN/AN ( <i>n</i> = 19)	JL ( <i>n</i> = 72)			BN/AN ( <i>n</i> = 19)	JL ( <i>n</i> = 72)
rat & Arthur	13 <sup>a</sup>	11	named/Arthur	Fall	17	46
young & rat	6	2		Level	2	20
Arthur		3		Fall-rise		3
was & Arthur		1		High rise		1
was, rat & Arthur		5		Low rise		2
was, a, young & Arthur		1				
there & Arthur		5				
there, named & Arthur		3				
there, rat & Arthur		38				
there, a & Arthur		1				
there, once & Arthur		1				
there, once, young & Arthur		1				

<sup>a</sup>The target token for the AN group was *Once there was a young rat named Arthur*. All AN subjects placed a nucleus on *once*. However, the nucleus on *once* in the tokens produced by the AN subjects was not regarded as an extra placement of nucleus because it was estimated to be due to a different syntactic structure and all the AN subjects showed the uniform pattern.

*... go out with them,*

	Nucleus placement		Nuclear tone choice			
	BN/AN ( <i>n</i> = 19)	JL ( <i>n</i> = 72)			BN/AN ( <i>n</i> = 19)	JL ( <i>n</i> = 72)
out	13	2	go/out/with/them	Fall-rise	6	3
them	6	66		Level	6	3
out & them		4		Fall	5	62
				High rise	2	4



*... he would only answer, ...*

	Nucleus placement		Nuclear tone choice			
	BN/AN	JL	only/answer	Level	BN/AN	JL
	(n = 19)	(n = 72)			(n = 19)	(n = 72)
answer	15	69			16	18
only	3	1		Fall	1	40
only & answer		1		Fall-rise	1	7
Error	1	1		High rise		6
			Error		1	1

*I don't know.*

	Nucleus placement		Nuclear tone choice			
	BN/AN	JL	don't/know	Fall	BN/AN	JL
	(n = 19)	(n = 72)			(n = 19)	(n = 72)
know	19	70			14	64
don't		2		Low rise	5	1
				High rise		1
				Level		6

*... said to him,*

	Nucleus placement		Nuclear tone choice			
	BN/AN	JL	said	Level	BN/AN	JL
	(n = 19)	(n = 72)			(n = 19)	(n = 72)
said	18				9	4
him	1	72		Low rise	6	3
				High rise		3
				Fall	4	62

*His aunt Helen ...*

	Nucleus placement		Nuclear tone choice			
	BN/AN	JL	aunt/Helen	Fall	BN/AN	JL
	(n = 19)	(n = 72)			(n = 19)	(n = 72)
Helen	18	49			18	24
aunt & Helen	1	21		Fall-rise	1	15
aunt		1		Level		31
His & Helen		1		High rise		2

*One rainy day, ...*

	Nucleus placement		Nuclear tone choice			
	BN/AN	JL			BN/AN	JL
	(n = 19)	(n = 72)			(n = 19)	(n = 72)
day	19	64	day/rainy	Low rise	8	5
rainy		5		Fall	4	35
rainy & day		2		Level	4	25
Error		1		Fall-rise	3	3
				High rise		3
			Error			1

*This won't do.*

	Nucleus placement		Nuclear tone choice			
	BN/AN	JL			BN/AN	JL
	(n = 19)	(n = 72)			(n = 19)	(n = 72)
do	18	67	this/won't/do	Fall	13	67
this	1			Level	4	3
won't		4		Low rise	2	
Error		1		Fall-rise		1
			Error			1

*At last, ...*

	Nucleus placement		Nuclear tone choice			
	BN/AN	JL			BN/AN	JL
	(n = 19)	(n = 72)			(n = 19)	(n = 72)
last	19	63	last	Fall	9	17
No nucleus		7		High rise	8	18
Error		2		Level	1	25
				Fall rise	1	2
				Low rise		1
			No nuclear tone			7
			Error			2

*Just then, ...*

Nucleus placement			Nuclear tone choice			
	BN/AN	JL			BN/AN	JL
	(n = 19)	(n = 72)			(n = 19)	(n = 72)
then	19	69	just/then	Fall	11	21
No nucleus		1		Level	6	44
Error		2		Fall-rise	1	2
				Low rise	1	1
				High rise		1
			No nuclear tone			1
			Error			2

*That night,*

Nucleus placement			Nuclear tone choice			
	BN/AN	JL			BN/AN	JL
	(n = 19)	(n = 72)			(n = 19)	(n = 72)
night	18	70	night	Level	7	42
Error	1	2		Fall	6	20
				Low rise	4	1
				High rise	1	6
				Fall-rise		1
			Error		1	2

*There was a kindly horse named Nelly, ...*

	Nucleus placement		Nuclear tone choice			
	BN/AN	JL			BN/AN	JL
	(n = 19)	(n = 72)			(n = 19)	(n = 72)
horse & Nelly	16	9	named/Nelly	Fall-rise	8	4
Nelly	3	6		Low rise	5	1
named & Nelly		1		High rise	1	3
kindly & Nelly		1		Fall	3	59
kindly, horse & Nelly		1		Level	2	5
a, horse & Nelly		1				
a, kindly, horse, Nelly		1				
was, kindly, horse & Nelly		1				
there & Nelly		12				
there, named & Nelly		1				
there, horse & Nelly		34				
there, horse, named & Nelly		1				
there, kindly & Nelly		2				
there, kindly, named & Nelly		1				

*... a cow, ...*

	Nucleus placement		Nuclear tone choice			
	BN/AN	JL			BN/AN	JL
	(n = 19)	(n = 72)			(n = 19)	(n = 72)
cow	19	72	cow	Low rise	15	5
				High rise	2	8
				Fall-rise	1	4
				Fall	1	37
				Level		18

*... a calf, ...*

	Nucleus placement		Nuclear tone choice			
	BN/AN	JL			BN/AN	JL
	( <i>n</i> = 19)	( <i>n</i> = 72)			( <i>n</i> = 19)	( <i>n</i> = 72)
calf	19	71	cow	Low rise	8	1
Error		1		High rise	2	7
				Fall-rise	6	4
				Fall	1	25
				Level	2	34
			Error			1

*... and a garden with an elm tree.*

	Nucleus placement		Nuclear tone choice			
	BN/AN	JL			BN/AN	JL
	( <i>n</i> = 19)	( <i>n</i> = 72)			( <i>n</i> = 19)	( <i>n</i> = 72)
garden & elm	17	4	with/an/elm/tree	Fall	18	70
elm	2			Low rise	1	
an & tree		6		Level		1
an, elm & tree		1		High rise		1
garden & tree		4				
a, & tree		1				
a, an & tree		11				
a, an & elm		1				
a, garden & tree		3				
a, garden, an & tree		1				
and & tree		2				
and, an & tree		19				
and, an & elm		1				
and, garden & tree		16				
and, garden, an & tree		2				

*Well, ...*

Nucleus placement			Nuclear tone choice			
	BN/AN	JL			BN/AN	JL
	(n = 19)	(n = 72)			(n = 19)	(n = 72)
well	19	71	well	Fall	15	35
Error		1		Level	3	28
				Low rise	1	1
				High rise		6
				Fall-rise		1
			Error			1

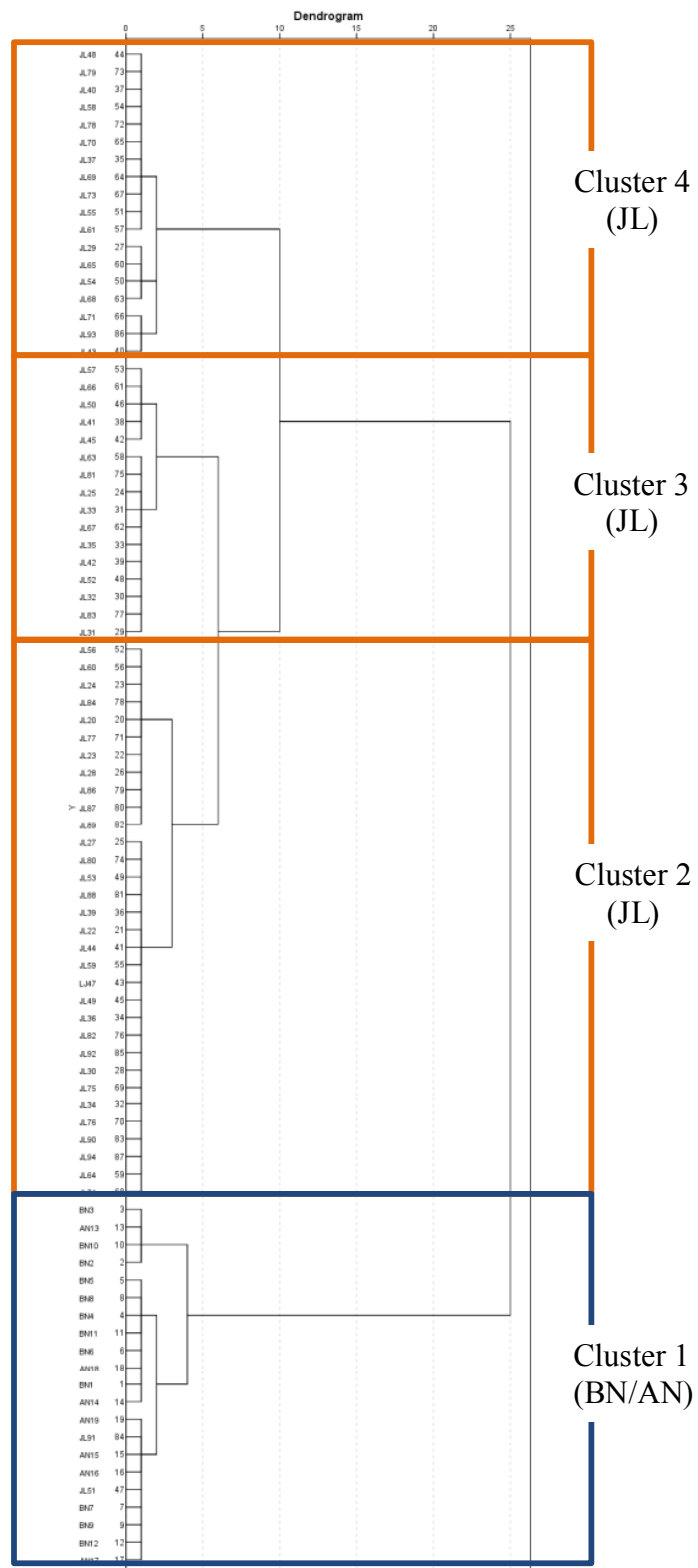
*Right about face.*

Nucleus placement			Nuclear tone choice			
	BN/AN	JL			BN/AN	JL
	(n = 19)	(n = 72)			(n = 19)	(n = 72)
face	17	67	right/face	Fall	10	66
right & face	2	4		Low rise	5	2
Error		1		Level	4	2
				Fall-rise		1
			Error			1

*March.*

Nucleus placement			Nuclear tone choice			
	BN/AN	JL			BN/AN	JL
	(n = 19)	(n = 72)			(n = 19)	(n = 72)
march	19	72	right/about/face	Fall	19	67
				Level		4
				High rise		1

# Appendix P: Dendrogram for intonation



## Appendix Q: Correlations between the variables for intonation

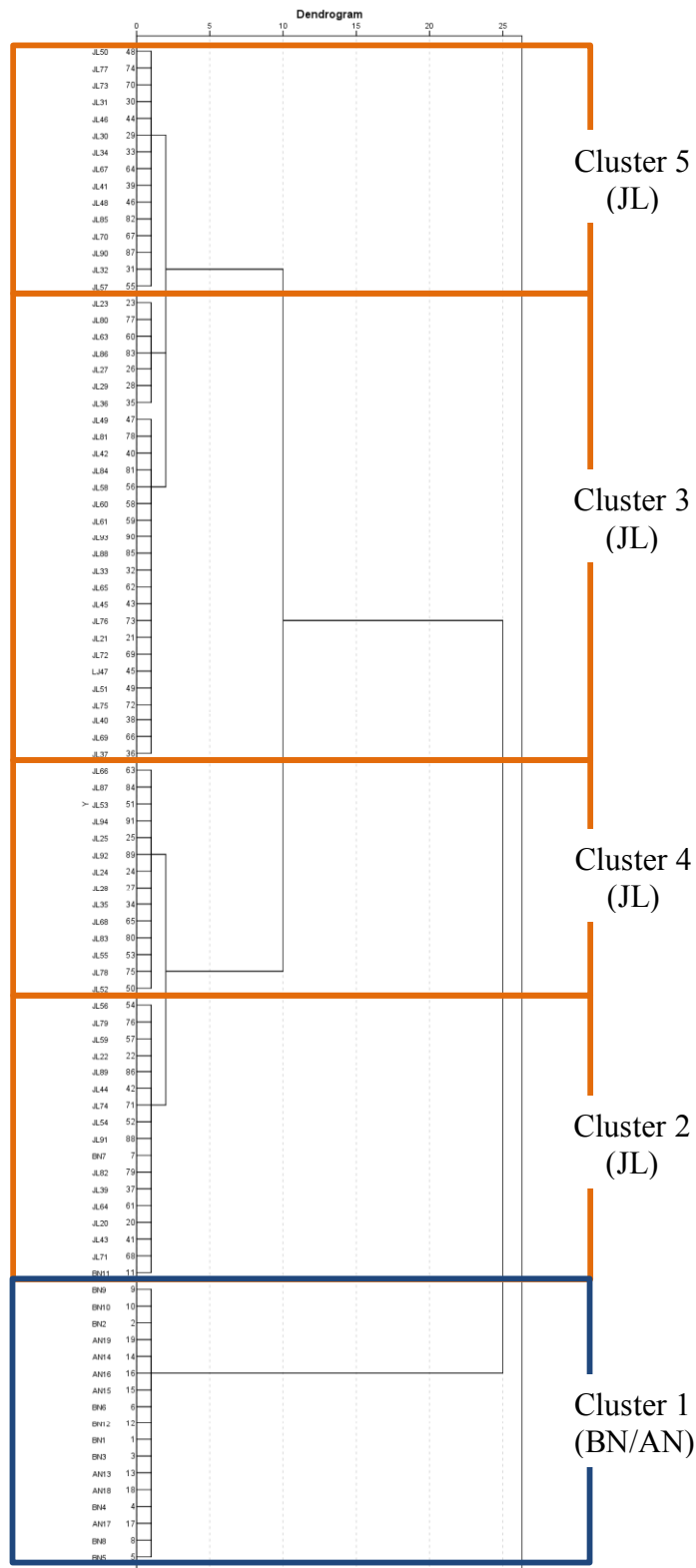
### *Correlations between the Variables for Intonation*

Variable	1	2
1. Non-nuclear words in the long/non-final utterances	—	
2. Nucleus in the long/non-final utterances	.85**	—
3. Falling utterances	.30**	.35**

\*  $p < .05$ . \*\*  $p < .01$ .



# Appendix R: Dendrogram for connected speech phenomena



## Appendix S: Correlations between the variables for connected speech phenomena

### *Correlations between the Variables for Connected Speech Phenomena*

Variable	1	2	3	4
1. Elision	—			
2. CC linking same	.66**	—		
3. CC linking different	.70**	.71**	—	
4. CV linking voiceless	.61**	.48**	.58**	—
5. CV linking voiced	.65**	.54**	.67**	.81**

*Note.* CC linking same = linking between two consonants at the same place of articulation and in the same manner of articulation; CC linking different = linking between two consonants at a different place of articulation or in a different manner of articulation; CV linking voiceless = linking between a voiceless consonant and a vowel; CV linking voiced = linking between a voiced consonant and a vowel.

\*\*  $p < .01$ .

## Appendix T: Results of the rate of level agreement between the two elements of pronunciation for the JL subjects

*Rate of Level Agreement between the Two Elements of Pronunciation for the JL subjects*

	Profile	VQ-VD	VQ-PL	VQ-FR	VQ-AP	VQ-R	VQ-INT	VQ-CO	VD-PL	VD-FR	VD-AP	VD-R	VD-INT	VD-CO
Levels agreed	11	13.85	8.57	15.94	18.92	15.79	10.61	12.12	10.77	13.04	18.92	11.76	10.94	10.80
	22	9.23	12.86	8.70	13.51	17.11	12.12	7.68	7.69	13.04	13.51	11.76	10.94	13.85
	33	13.85	10.00	11.59	13.51	13.16	10.61	12.12	9.23	15.94	13.51	11.76	12.50	13.83
1-level gap	12	7.69	12.86	7.25	8.11	10.53	10.61	7.58	12.31	10.14	8.11	10.29	10.94	9.23
	21	13.85	12.86	11.59	9.46	7.89	9.09	15.15	12.31	11.59	9.46	11.76	14.06	12.31
	23	12.31	10.00	15.94	10.81	7.89	12.12	15.15	10.77	8.70	10.81	10.29	9.38	9.23
	32	15.38	7.14	14.49	13.51	9.21	9.09	10.61	10.77	7.25	13.51	10.29	9.38	6.15
2-levels gap	13	7.69	12.86	7.25	8.11	10.53	10.61	7.58	12.31	10.14	8.11	10.29	10.94	9.23
	31	6.15	12.86	7.25	4.05	7.89	15.15	12.12	13.85	10.14	4.05	11.76	10.94	15.39
Total	112233	36.92	31.43	36.23	45.95	46.05	33.33	31.82	27.69	42.03	45.95	35.29	34.38	38.46
	1221	21.54	25.71	18.84	17.57	18.42	19.70	22.73	24.62	21.74	17.57	22.06	25.00	21.54
	1331	13.85	25.71	14.49	12.16	18.42	25.76	19.70	26.15	20.29	12.16	22.06	21.88	24.62
	3223	26.15	20.00	27.54	24.32	21.05	22.73	27.27	20.00	24.64	24.32	22.06	21.88	23.08

*Note.* The values are expressed in percentages. VQ = vowel quality; VD = vowel duration; PL = plosives; F = fricatives; AP = approximants; R = rhythm; INT = intonation; CO = connected speech phenomena.

(continued)

*Rate of Level Agreement between the Two Elements of Pronunciation for the JL subjects*

	Profile	PL-FR	PL-AP	PL-R	PL-INT	PL-CO	FR-AP	FR-R	FR-INT	FR-CO	AP-R	AP-INT	AP-CO	R-INT	R-CO	INT-CO
Levels agreed	11	11.27	8.45	12.31	12.70	14.49	18.84	9.33	10.45	9.59	13.89	10.14	6.49	14.08	13.04	13.12
	22	9.86	9.86	10.77	12.70	14.49	11.59	14.67	11.94	10.96	13.89	13.04	7.79	16.9	5.80	8.20
	33	12.68	9.86	10.77	12.70	8.70	15.94	10.67	11.94	9.59	9.72	14.49	3.90	9.86	10.15	11.48
1-level gap	12	14.08	15.49	10.77	11.11	11.59	5.80	13.33	11.94	12.33	9.72	11.59	15.58	9.86	8.70	6.56
	21	12.68	11.27	7.69	11.11	4.35	10.14	9.33	11.94	12.33	5.56	14.49	12.99	5.63	14.49	14.75
	23	5.63	7.04	12.31	7.94	10.14	13.04	8.00	8.96	9.59	13.89	5.80	10.39	11.27	14.49	14.75
	32	11.27	11.27	9.23	9.52	8.70	13.04	8.00	8.96	9.59	9.72	10.14	12.99	9.86	14.49	11.48
2-levels gap	13	14.08	15.49	10.77	11.11	11.59	5.80	13.33	11.94	12.33	9.72	11.59	15.58	9.86	8.70	6.56
	31	8.45	11.27	15.38	11.11	15.94	5.80	13.33	11.94	13.7	13.89	8.70	14.29	12.68	10.15	13.12
Total	112233	33.80	28.17	33.85	38.10	37.68	46.38	34.67	34.33	30.14	37.50	37.68	18.18	40.85	28.99	32.79
	1221	26.76	26.76	18.46	22.22	15.94	15.94	22.67	23.88	24.66	15.28	26.09	28.57	15.49	23.19	21.31
	1331	22.54	26.76	26.15	22.22	27.54	11.59	26.67	23.88	26.03	23.61	20.29	29.87	22.54	18.84	19.67
	3223	18.31	16.9	23.08	20.63	18.84	28.99	18.67	20.90	19.18	23.61	20.29	14.29	21.13	24.64	26.23

*Note.* The values are expressed in percentages. P L = plosives; FR = fricatives; AP = approximants; R = rhythm; INT = intonation; CO = connected speech phenomena.