AN ACOUSTIC AND PERCEPTUAL INVESTIGATION OF VOWEL LENGTH IN
JAPANESE AND POHNPEIAN

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAI'I IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
LINGUISTICS

MAY 2005

By
Tomoko Kozasa

Dissertation Committee:
Kenneth L. Rehg, Chairperson
Victoria B. Anderson
Byron W. Bender
Patricia J. Donegan
Leon A. Serafin
Amy J. Schaefer
David Hanlon
© Copyright 2005

By

Tomoko Kozasa

iii
for those who believed in me
ACKNOWLEDGEMENTS

I finally have reached the stage where I can write my acknowledgments. This dissertation was possible because I had the support of so many people. First of all, I am grateful to my committee members for their constant support. My special thanks go, of course, to my Chair adviser Ken Rehg. I met him the very first day I came to the University of Hawai‘i at Mānoa. It was destiny. He inspired me with his gentle manner to study phonology, with a focus on Pohnpeian phonology. Without him, I would not be here today. I also am greatly indebted to Amy Schafer. She was not on my committee when I started writing this dissertation, yet from the beginning she spent a lot of time helping me to design experiments, to understand statistics, and to analyze my data. Whenever I needed help, she was there. Without her high academic standard, energy, encouragement, and her amiable smile, I would not have been able to finish my dissertation. I would like to thank Victoria Anderson. I was able to study phonetics, because she joined the University of Hawai‘i faculty. I admire her knowledge of phonetics and sincere concern for students and their work. I also thank Professor Byron Bender for a valuable critique of my work. I also express my appreciation to Patricia Donegan. This dissertation began as a term paper for her seminar in phonology. Patricia’s keenness and passionate advocacy of Natural Phonology made me interested in the phonetics/phonology interface. Leon Serafim is another person to whom I owe a lot. He ardently challenged me with thought-provoking questions on Japanese phonology. As his last name suggests, he was my angel who brought light to guide me through this journey. Finally, but not least, I am glad to have David Hanlon as my outside committee member.
I appreciate his encouragement to study Pohnpeian and his concern over the progress of my study. In addition to these committee members, I would like to thank Yuko Otsuka and William O'Grady for their moral support. I especially appreciate the time Yuko spent with me laughing and weeping while I was writing my dissertation. She was and is always a great encouragement to me. In addition, it was subtle, but I always felt William’s support with his warm smiles whenever I worked in our computer room at the night. Of course, I will not forget how supportive the faculty members were at California State University, Fresno. I am grateful that I met Chris Golston, Ritva Laury, Gerald McMenamin, Shigeko Okamoto, Graham Thurgood, and Raymond Weitzman. If I had not gotten to know them, and had not had their encouragement, I would never thought of pursuing a Ph.D.

Secondly, I was so fortunate to have so many participants in my experiments. I am grateful that they showed an interest in my experiments. Special thanks go to Mr. Simion Kihleng for introducing me to many Pohnpeian speakers in Honolulu. I also thank Dr. April Komenaka and Mr. James Mellon for making arrangements to conduct experiments on the University of Hawai‘i at Hilo campus.

Also, I want to thank the organizations which financially supported me all these years. The Bilingual Education Fellowship (Title VII), the Department of Linguistics at University of Hawai‘i, College of Languages, Linguistics, and Literature, and the Center for Japanese Studies. Without their support, I would not have been able to study in Hawai‘i.
I am also grateful to the many wonderful staff members on Mānoa campus. Our department former secretaries – Wendy Onishi and Carolyn Paet – and current secretaries – Jennifer Kanda and Tamara Zielinski – made sure my paperwork was done and solved my problems immediately. Daniel Tom and Nancy Arakawa at Language Learning Center helped me to set up data collection experiments. Kurt Brunner made sure our computers in research laboratories function properly. Without all these people’s help, it would have taken much longer for me to finish this dissertation.

I thank all my friends who have been supporting me in various ways. Lynn and Michael Yada and Dorothy Ikeda in Fresno always treat me just like a member of their family. Sam and Joyce Tanji in Honolulu let me stay at their house and treated me like their daughter for three years. Yoko and Rev. Ernie Kilbourne, Pastor Jim and Anna Stern, and members at West Oahu Christian Church and Kapahulu Bible Church – I appreciate their prayers and faith in me. I thank Woody Mott and Ryo Stanwood for helping me to polish my English. I have not named all of my other friends who have been supporting and encouraging me to finish my Ph.D. work, because I would need pages and pages to do so. I thank you the most, YOU, the reader, who are reading this right now.

Finally, it was providential for me to get to know the Late Professor Stanley Starosta, who believed in me and encouraged me. Stan, I finished!

お父さん、お母さん、お姉ちゃん、長らくお待たせしました。私、博士になりました。

February, 2005
ABSTRACT

This dissertation is about long vowels in Tokyo Japanese and Pohnpeian, a Micronesian language. Vowel length is contrastive in these two languages, and it is phonologically represented by mora counts. Long vowels carry two moras and short vowels one. The results from three production and three perception experiments show that the production and perception of vowel length is not identical between Japanese and Pohnpeian. The proportion of the duration of long to short vowels is much greater in Japanese. Furthermore, not only duration but also pitch fall is crucial for Japanese speakers in marking vowel length, and Pohnpeian speakers’ use of durational cues differs from that of Japanese speakers. The phonetic implementations of long vowels seem to play an important role in the listeners’ categorization of vowel length.

The duration of vowels and the amount of pitch fall were measured. The ratio of the mean duration of long vowels to short vowels was not 2:1 in either language. In Japanese accented vowels, it was 1.65:1 in fast speech and 1.77:1 in slow speech, and in Japanese unaccented vowels, it was 1.93:1 in fast speech and 2.42:1 in slow speech. In Pohnpeian, it was 1.30:1 in fast speech and 1.47:1 in slow speech.

The duration of vowels was vulnerable to speech rates in both languages, but pitch fall in Japanese accented vowels was not. The durational distinction between long and short vowels was increased in Japanese unaccented vowels compared to accented vowels, and it was even greater as the speech rate decreased. These results indicate that
both duration and pitch fall are important for Japanese speakers to mark phonological vowel length, while Pohnpeian speakers primarily used duration to mark vowel length.

Patterns of vowel length categorization by each language listeners agreed with the results from production experiments. Japanese listeners' categorization of vowel length was affected by pitch fall. Even when the duration of vowels was equivalent to long vowels, Japanese listeners were not able to categorize vowel length without an adequate amount of pitch fall. Furthermore, they required longer duration than Pohnpeian listeners to perceive a vowel as long.
# TABLE OF CONTENTS

Acknowledgments ......................................................................................... v
Abstract ........................................................................................................ viii
List of Tables ................................................................................................ xiii
List of Figures ................................................................................................. xv
List of Abbreviations and Symbols ................................................................ xvi

## CHAPTER 1: INTRODUCTION .................................................................... 1
  1.1 Introduction ........................................................................................... 1
  1.2 Japanese .................................................................................................. 7
    1.2.1 Japanese phonemes ......................................................................... 7
    1.2.2 Japanese vowel length ...................................................................... 10
    1.2.3 Japanese syllable structure ............................................................. 11
    1.2.4 Japanese pitch-accent ...................................................................... 12
  1.3 Pohnpeian ............................................................................................. 14
    1.3.1 Pohnpeian phonemes ..................................................................... 14
    1.3.2 Pohnpeian vowel allophony ............................................................ 16
    1.3.3 Pohnpeian syllable structure ......................................................... 17

## CHAPTER 2: BACKGROUND ..................................................................... 19
  2.1 Introduction ........................................................................................... 19
    2.1.1 Rhythmic unit typology .................................................................. 20
    2.1.2 Concept of the mora ...................................................................... 25
    2.1.3 Moraic theory ................................................................................. 27
    2.1.4 The relation between phonology and phonetics in prosody ........ 32
  2.2 Japanese mora ....................................................................................... 37
    2.2.1 The mora as a unit of speech segmentation ..................................... 38
    2.2.2 The mora as a timing unit and its isochrony .................................. 43
    2.2.3 A function of Japanese pitch-accent ............................................. 50
  2.3 Pohnpeian Mora .................................................................................... 50
    2.3.1 Pseudo-compensatory lengthening .............................................. 52
  2.4 Conclusion ............................................................................................. 55

## CHAPTER 3: JAPANESE LONG VOWELS ............................................. 57
  3.1 Introduction ........................................................................................... 57
  3.2 The effect of pitch-accent on vowel duration ....................................... 60
  3.3 Experiment 1: Production experiment with Japanese accented vowels ......................................................................................... 65
    3.3.1 Participants ..................................................................................... 66
    3.3.2 Materials ......................................................................................... 67
    3.3.3 Procedure ....................................................................................... 70
    3.3.4 Hypotheses ................................................................................... 72
    3.3.5 Data analyses ................................................................................ 73
    3.3.6 Results .......................................................................................... 74
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1:</td>
<td>Japanese phonemes</td>
<td>8</td>
</tr>
<tr>
<td>1.2:</td>
<td>Pohnpeian phonemes</td>
<td>15</td>
</tr>
<tr>
<td>1.3:</td>
<td>Pohnpeian consonants (Adopted from Rehg &amp; Sohl, 1981, p.44)</td>
<td>15</td>
</tr>
<tr>
<td>1.4:</td>
<td>Pohnpeian vowel allophony (Adopted from Rehg &amp; Sohl, 1981, pp.44-45)</td>
<td>16</td>
</tr>
<tr>
<td>3.1:</td>
<td>Examples of test words for Experiment 1</td>
<td>67</td>
</tr>
<tr>
<td>3.2:</td>
<td>Analysis by speakers: Mean durations of accented vowels and ratios in Experiment 1</td>
<td>74</td>
</tr>
<tr>
<td>3.3:</td>
<td>Analysis by vowel types: Mean durations of accented vowels and ratios in Experiment 1</td>
<td>74</td>
</tr>
<tr>
<td>3.4:</td>
<td>Analysis by speakers: Mean pitch falls of accented vowels and ratios in Experiment 1</td>
<td>76</td>
</tr>
<tr>
<td>3.5:</td>
<td>Analysis by vowel types: Mean pitch falls of accented vowels and ratios in Experiment 1</td>
<td>76</td>
</tr>
<tr>
<td>3.6:</td>
<td>Mean pitch fall by individual participants</td>
<td>78</td>
</tr>
<tr>
<td>3.7:</td>
<td>Analysis by speakers: Mean durations of unaccented vowels and ratios in Experiment 2</td>
<td>85</td>
</tr>
<tr>
<td>3.8:</td>
<td>Analysis by vowel types: Mean durations of unaccented vowels and ratios in Experiment 2</td>
<td>86</td>
</tr>
<tr>
<td>3.9:</td>
<td>Analysis by speakers: Comparison of mean durations of vowels in Experiment 1 and Experiment 2</td>
<td>88</td>
</tr>
<tr>
<td>3.10:</td>
<td>Comparison of mean durations of vowels produced by participant M-1</td>
<td>89</td>
</tr>
<tr>
<td>4.1:</td>
<td>A sample set of conditions: reji ‘cash register’ vs. reeji ‘midnight’</td>
<td>101</td>
</tr>
<tr>
<td>4.2:</td>
<td>Example of conditions of filler words</td>
<td>103</td>
</tr>
<tr>
<td>4.3:</td>
<td>Analysis by listeners: Mean choice of target vowels in Experiment 3</td>
<td>107</td>
</tr>
<tr>
<td>4.4:</td>
<td>Analysis by items: Mean choice of target vowels in Experiment 3</td>
<td>107</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure | Page |
--- | --- |
3.1: Measurement points for vowel duration | 70 |
3.2: Measurement points for pitch fall | 71 |
3.3: Analysis by speakers: Interaction line plot for duration in Experiment 1 | 75 |
3.4: Analysis by speakers: Interaction line plot for pitch fall in Experiment 1 | 77 |
3.5: Mean pitch fall by individual participants | 78 |
3.6: Analysis by speakers: Interaction line plot for duration in Experiment 2 | 87 |
4.1: Manipulation of the vowel duration | 102 |
4.2: Analysis by listeners: Interaction Bar Plot for RT in Experiment 3 | 108 |
4.3: Analysis by listeners: Interaction Bar Plot for RT in Experiment 4 | 119 |
5.1: Scatter plots of mean durations of vowels in Experiment 5 | 130 |
5.2: Mean duration of each phonemic vowel | 131 |
5.3: Analysis by speakers: Interaction line plot in Experiment 5 | 134 |
5.4: Extracted spectrogram and formants in pihl and pil | 140 |
5.5: Extracted spectrogram and formants in luhs and lus | 142 |
5.6: Sequence of an ABX trial | 151 |
5.7: Interaction line plot in Experiment 6 | 154 |
5.8: Japanese listeners’ responses to Japanese stimuli | 156 |
5.9: Japanese listeners’ responses to Pohnpeian stimuli | 156 |
5.10: Pohnpeian listeners’ responses to Japanese stimuli | 157 |
5.11: Pohnpeian listeners’ responses to Pohnpeian stimuli | 157 |
5.12: Responses to Japanese stimuli as long | 158 |
5.13: Responses to Pohnpeian stimuli as long | 159 |
# LIST OF ABBREVIATIONS AND SYMBOLS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANOVA</td>
<td>analysis of variance</td>
</tr>
<tr>
<td>ACC</td>
<td>accusative case</td>
</tr>
<tr>
<td>NOM</td>
<td>nominative case</td>
</tr>
<tr>
<td>F0</td>
<td>fundamental frequency</td>
</tr>
<tr>
<td>F2</td>
<td>second formant</td>
</tr>
<tr>
<td>F3</td>
<td>third formant</td>
</tr>
<tr>
<td>H</td>
<td>high pitch</td>
</tr>
<tr>
<td>H-</td>
<td>hypothesis</td>
</tr>
<tr>
<td>Hz</td>
<td>hertz</td>
</tr>
<tr>
<td>L</td>
<td>low pitch</td>
</tr>
<tr>
<td>ms</td>
<td>millisecond</td>
</tr>
<tr>
<td>JP</td>
<td>Japanese</td>
</tr>
<tr>
<td>PN</td>
<td>Pohnpeian</td>
</tr>
<tr>
<td>RT</td>
<td>reaction time</td>
</tr>
<tr>
<td>μ</td>
<td>mora</td>
</tr>
<tr>
<td>σ</td>
<td>syllable</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

1.1 Introduction

In languages that have a vowel length distinction, phonological vowel length is represented by mora count. Long vowels carry two moras and short vowels one. In both Japanese and Pohnpeian, vowel length is contrastive. The speakers of Japanese and Pohnpeian are able to correctly produce long vowels and short vowels and correctly perceive vowel length in their own language. However, the phonetic properties of long vowels in these two languages are not identical. In this dissertation, I will investigate phonetic properties of long vowels, focusing on duration and pitch fall in Tokyo Japanese and Pohnpeian. Tokyo Japanese is a dialect of Japanese, spoken in Tokyo and its environs. Pohnpeian is a Micronesian language, spoken on the island of Pohnpei in the Federated States of Micronesia.

The impetus behind this study was a single observation. Native speakers of Japanese, including myself, frequently misperceive the length of Pohnpeian vowels in isolation. This observation that Japanese listeners misperceive Pohnpeian vowel length, could be a piece of evidence that phonetics not being fully predictable from phonology. If long vowels in the two languages had identical phonetic properties, Japanese speakers should not be confused by Pohnpeian vowel length. Phonetic realization of the phonological feature – vowel length – must be different in these two languages. This

1 In this dissertation, I use the term “Japanese” or “Tokyo Japanese” to refer to the Standard dialect or to the Japanese language in general. Otherwise, I will specify the dialect.
dissertation investigates the way speakers of each language employ duration and pitch fall when they mark vowel length and how these acoustic signals influence the categorization of vowel length.

Phonetics is the study of speech sounds and is concerned with describing speech sounds in terms of our articulatory and auditory mechanisms and the physical properties of those speech sounds. Since our speech is continuous and dynamic, phonetic realizations of speech sounds are gradient and complex, yet phonetic representations are concrete. Phonology, on the other hand, is the study of sound systems, both of a language and across languages. Phonology is the abstract system that underlies our mental activities in processing physical signals as meaningful units and that organizes these units internally. The focus of theoretical phonology is to represent our mental activities in the most parsimonious yet realistic way. Phonological representations are thus categorical and simple, yet abstract.

Some who strictly follow the tradition of *Generative Phonology* may consider phonetics and phonology autonomous. In fact, there was a period when “phonetics was placed outside of linguistics proper and phonology was conceived of as an autonomous discipline” (Ohala, 1997, p.681). Ohala expresses his concerns about the neglect of phonetics in the study of linguistics (phonology) and argues that, “phonology should not be conducted as an autonomous discipline but rather should embrace any means that will help it to get the answers it seeks” (p.693). Discovering sound patterns and labels that represent the patterns must be merely the first step, not the goal of phonology (Ohala, 1974). To complete the study, we must seek an explanation for the patterns that we found
(Ohala 1974, 2003; Myers, 2000), as phonological categorization must be embedded in the phonetic signals. Sapir (1949) also argued for the importance of seeking physical properties which signal phonological entities such as phonemes. The goal of the study of phonology, like that of phonetics, must be to find the way physical signals map onto categorical representations; that is, to find the way physical properties and production mechanisms of speech sounds inform a speaker’s phonological system and the way a speaker’s phonological knowledge is manifested in the physical properties of speech sounds.

For decades, some phoneticians and phonologists have endorsed the idea that “phonetics is an indispensable tool for the phonologist” (Ohala, 1974, p.253) or that phonetics is the central part of phonology (Stampe, 1972; Donegan & Stampe, 1979; Pierrehumbert, 1980; Ohala, 1990; Pierrehumbert, et al., 2000; Keating, 1984, 1988, 1991, 1996; Browman & Goldstein, 1992; Hayes, 1996; Johnson & Hume, 2001). They have been investigating the mechanism and effects of coarticulation on the perception of speech segments, phonological variations observed in different geographical regions or in different speech styles to account for the directionality of sound change, and the process of children’s acquisition of adult-like pronunciation of their language to support their claims.

The discrepancies between physical signals and categorization of those signals have been attested in various phenomena and in various languages. It is undeniable that each language has allophonic variations in each phoneme. Phonetic properties in allophones are different from one another yet they are categorized as a single phoneme in
the language. While both English and Spanish have phonemes /p, t, k/, there are differences in the values of voice onset time (VOT) of these voiceless stops between these two languages (Keating, 1984; Flege & Eefting, 1987). The ratio of the duration of a geminate nasal /n:/ to a singleton /n/ is greater in Finnish than in Japanese (Aoyama, 2001). These studies illustrate that phonetics is not always fully predictable from phonology and suggest that phonology is created by being exposed to phonetic experience in an ambient language. We must continue investigating physical properties of phonological representations by conducting acoustic and perceptual investigations in order to enhance understanding of our phonological behavior. The aims of this dissertation are to investigate differences between Japanese and Pohnpeian in phonetic properties of long vowels and to illustrate how phonetic implementations in each language play a crucial role in the listeners’ categorization of vowel length.

This dissertation is organized as follows: In this chapter, I will introduce some features of Japanese phonology that are important for this dissertation. Japanese is one of the best-studied languages in the world; yet, scholars still argue over whether descriptions of the Japanese sound system should contain historical developments, morphophonological alternations, or results from psycholinguistic experiments. I will suggest a model of the Japanese sound system based on currently spoken Japanese. Then I will present an introduction to Pohnpeian. Although Pohnpeian is one of the best-documented Oceanic languages, it is still relatively unknown in the field of linguistics in comparison to Japanese.
In Chapter 2, I will lay out the background for this dissertation. I will start by reviewing the study of rhythmic unit typology to introduce the concept of the mora, since many studies have shown that the mora plays important roles in Japanese and Pohnpeian phonology. Next, I will review previous studies on the mora in these two languages.

In Chapter 3, I will discuss two production experiments involving Tokyo Japanese vowels. Words in Tokyo Japanese are divided into two types, accented and unaccented, depending on the presence of pitch-accent. Pitch-accent is marked by relative downward movement of pitch, or fundamental frequency (F0). In words containing an accented long vowel, this pitch fall takes place within the vowel. Thus, Japanese accented long vowels have two phonological features, vowel length and pitch-accent, which correspond to phonetic properties, vowel duration and pitch fall respectively. Japanese unaccented long vowels have only vowel length, which is marked by vowel duration phonetically. The purpose of the experiments in Chapter 3 is to investigate the phonetic properties (duration and pitch fall) of Japanese long vowels. Experiment 1 uses Japanese accented vowels and Experiment 2 Japanese unaccented vowels. Even within a single language, phonetic properties of the mora realized in long vowels differ according to the function that the mora bears. When Japanese speakers produce accented long vowels, they produce an adequate amount of both durational distinction and pitch fall across various speech rates. When they produce unaccented long vowels, where pitch fall cannot be used, they magnify the durational distinction. These results suggest that for Japanese vowels pitch fall as well as vowel duration is important in marking vowel length.
In Chapter 4, I will examine the effect of pitch cues on the Japanese speakers’ perception of vowel length. I will describe two perception experiments. In these two experiments, I created ambiguous vowels by manipulating the duration of accented Japanese short vowels while maintaining their pitch contour. The duration of the ambiguous vowels was equivalent to that of long vowels but the timing of pitch fall was same as in short vowels. The listeners took longer to categorize the length of ambiguous vowels; moreover, their categorization of the length of ambiguous vowels was at the chance level. The results indicate that in order for Japanese listeners to categorize vowel length in accented vowels, there must be an adequate amount of pitch fall as well as the durational distinction present in accented vowels.

In Chapter 5, I will discuss two experiments. I investigated the phonetic properties of Pohnpeian vowels and compared them with the phonetic properties of Japanese vowels. Then I conducted a cross-linguistic perception experiment. Pohnpeian speakers do not use pitch movement to mark vowel length; however, the durational distinction between long and short vowels is much smaller than that of Japanese vowels. Pohnpeian speakers do seem to also use spectral cues in marking vowel length; however, this dissertation will not discuss details of spectral cues in Pohnpeian vowels.

In Chapter 6, I will briefly summarize this study and discuss the remaining issues for future research. In Appendices A to E, I will provide complete lists of words or items that I used in Experiments 1 to 5.
1.2 Japanese

Japanese is the language spoken by virtually everyone in Japan. According to the *Ethnologue* (Grimes & Grimes, 2000), there are approximately 121 million speakers. Linguistic studies in Japanese have a long history, (from *Kokugaku* to the framework of generative linguistics, which is since the late 50's). Japanese is one of the best-documented languages outside of the Indo-European language family, which, of course, does not necessarily mean that all aspects of Japanese phonology are explicated. In this section, I will briefly describe the aspects of the Japanese sound system that are relevant to this dissertation.

1.2.1 Japanese phonemes

There are 25 consonants and 5 vowels in Japanese, as in Table 1.1. Shibatani (1990) lists only 16 consonants (p.159), omitting all palatalized phonemes but including the coda nasal and coda plosive as phonemes, which are conventionally coded as N and Q respectively. The capital letter N is often used to represent the coda nasal, as the place feature of the Japanese coda nasal is not specified. For example, if the following consonant is a bilabial plosive /b/ as in the word *konbu* 'seaweed,' the nasal is realized as a bilabial nasal [m], and if an alveolar follows the nasal as in the word *ondo* 'temperature,' the nasal is realized as an alveolar nasal [n]. Likewise, if a velar follows the nasal as in the word *tanka* 'stretch,' the nasal is realized as a velar nasal [ŋ]. The capital

---

2 Studies examining classical literature, such as *Man'yōshū* and *Kojiki*, as a scientific (linguistic) subject began in the early 18th century. Major scholars in the school were Kada no Azumamaro, Kamo no Mabuchi, and Motoori Norinaga.

3 Ito (1988) proposed that this coda nasal characteristic is a 'coda condition,' where moraic consonants with their own place feature are forbidden.
letter Q is used to indicate a coda plosive, which is always identical to the following consonant which is the onset of the following syllable, such as in the word *kitte* ‘postage stamps’ and *Nippon* ‘Japan.’ However, N and Q are not distinctive speech sounds in Japanese. The place of articulation of a coda nasal is specified by its context, which is a natural process and does not require establishing it as a distinctive Japanese phonemic consonant. Geminate consonants are not a sequence of two consonants but long consonants (true geminates) within a single word in modern Japanese. McCawley (1968) lists all 25 consonants in Table 1.1 plus /ʔ/. The glottal stop /ʔ/ is not a distinctive phonemic consonant in Japanese either. He uses the glottal stop at the beginning of words starting with a vowel; however, Japanese does not distinguish /ie/ vs. /ʔie/ ‘house.’ I will use either [n], [m], or [ŋ] for a coda nasal and the length marker [:] to indicate geminate consonants in this dissertation, and I will not accept /ʔ/ as a Japanese consonant.4

<table>
<thead>
<tr>
<th>Table 1.1: Japanese phonemes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consonants</td>
</tr>
<tr>
<td>p  p̟  t  t̟  k  k̟</td>
</tr>
<tr>
<td>b  b̟  d  d̟  g  g̟</td>
</tr>
<tr>
<td>s  s̟  ɕ̟  h</td>
</tr>
<tr>
<td>m  m̟  n  n̟  r  r̟</td>
</tr>
<tr>
<td>w  j</td>
</tr>
</tbody>
</table>

4 Vance (1987, pp.43-44) points out that /ʔ/ may be inserted utterance-finally when a speaker adds an emphasis on an expression such as [haiʔ] ‘yes’ or [aʔ] ‘oh;’ however, it does not change the meaning of the word. Thus, he suggests that it is not necessary to propose this /ʔ/ as a Japanese phoneme (p.33).
The palatalized phonemes, except for /tɿ, dɿ/, are usually considered a combination of a consonant and a glide /j/ (e.g. /Cy/) (Vance, 1987) for historical reasons. However, no consonant cluster is allowed within a syllable in Japanese, and it is easy to find minimal pairs such as /k'oku/ ‘melody’ vs. /koku/ ‘aroma,’ /r'aku/ ‘abbreviation’ vs. /raku/ ‘comfortable.’ I treat these palatalized segments as single phonemes.

The palatalized consonants, /pɿ, tɿ, kɿ, bɿ, dɿ, gɿ, sɿ, çɿ, mɿ, nɿ, and rɿ/, and the palatal glide /j/ do not occur before the front vowels /i/ and /e/. The bilabial glide /w/ occurs only before the low vowel /a/. The Japanese liquid is often transcribed as /r/, but it is more like the English flap [r]. I use the symbol /r/ for this phoneme. The Japanese alveolar plosives /t, d/ have three allophones each: [t], [c], and [ts] for /t/ and [d], [ʃ], [dʒ] for /d/. The palatal affricates [c, ʃ] occur before the high front vowel /i/ and the alveolar affricates [ts, dz] occur before the high back vowel /u/. The phoneme /h/ has two allophones besides [h] itself, namely, the voiceless bilabial fricative [ɸ] and the voiceless palatal fricative [ç]. The voiceless bilabial fricative [ɸ] occurs before the high back vowel /u/ and the voiceless palatal fricative [ç] occurs before the high front vowel /i/. The phoneme /ɡ/ sporadically appears as [ŋ] word medially especially in some dialects of Japanese. Several explanations have been offered to account for the alternation (c.f.,

---

5 The allophone does not occur in some loanwords such as /ti:/ ‘tea,’ /patti:/ ‘party,’ and /komedi:/ ‘comedy.’

6 The phone [ɸ] now occurs with all five Japanese vowels in loanwords such as [fan] ‘fan’ or ‘fun,’ [fenʃu] ‘fence,’ [figua] ‘figure,’ and [fonto] ‘font.’ In other words, this phone [ɸ] has become a distinctive consonant in current Japanese.
Vance, 1987, Chapter 9). Since, however, the quality of Japanese consonants is beyond the scope of this dissertation, I will not discuss this issue any further.

There are five phonemic vowels in Japanese, /a, e, i, o, u/. The high-back vowel /u/ is often represented as the unrounded vowel [u] (Homma, 1973, p.350; Keating & Huffman, 1984, p.193; Shibatani, 1990. p.161; Ladefoged, 2001, p.209). Han (1962) uses the symbol [u] to suggest that this high back vowel is centralized rather than unrounded (p.11). Vance (1987) states that there is obvious lip activity involved in the articulation of this vowel, which is lip compression rather than lip rounding (p.11). The formant plots shown in Vance (p.11) and Keating & Huffman (p.197) indicate that the Japanese high back vowel /u/ is indeed centralized. That is, the frequency of the first formant (F1) of the vowel, is between that of the high front vowel /i/ and the mid front vowel /e/, and the frequency of the second formant (F2) is higher than the mid back vowel /o/ and is similar to the low central vowel /a/. I will use the symbol /u/ for this Japanese vowel.

1.2.2 Japanese vowel length

A distinction in vowel length is one of the major features of Japanese phonology. Each phonemic vowel has a long and short manifestation and they are distinctive as shown in (1).

(1) Long vowels               Short vowels
    biiru  'beer'               vs.   biru  'building'
    tooshi 'fighting spirit'    vs.   toshi  'city'
Vowel length is associated with mora counts – long vowels carry two moras and short vowels one. Pickett (1999) explains “phonetically, short vowels contained relatively short targets (i.e., where formant frequencies are steady-state or formant movement is relatively slow) and longer offglides (i.e., formant movements from the target to consonant closure), while long vowels have relatively long targets and short offglides, and when a vowel is diphthongized, it has a single target and extensive offglides” (pp.69-70). Keating & Huffman (1984) claim that no Japanese vowels are diphthongized (p.193). In addition, their experiment shows that there is relatively little allophonic variation in Japanese vowels. In other words, phonemic vowel length distinctions in Japanese vowels mainly correlate to duration phonetically (Hirata, 2004, p.560). As far as I know, no study has found significant spectral differences associated with vowel length distinctions in Japanese.

1.2.3 Japanese syllable structure

Syllables may take the following forms in Japanese: (C)V, (C)VV, and (C)VC. The onset consonant is optional in Japanese. In fact numerous Japanese words start with a vowel, such as asa ‘morning,’ eki ‘train station,’ ike ‘pond,’ okite ‘statute,’ and ushiro ‘behind.’ The coda consonant must be either a nasal or the initial part of a geminate obstruent. Coda nasals can occur word finally, such as in soroban ‘abacus’ and iken ‘opinion,’ but obstruents can never occur phonologically word finally.
1.2.4 Japanese pitch-accent

Another phonological feature of Japanese is contrastive pitch-accent. That is, the position or existence of pitch-accent in a word may change its meaning. In Tokyo Japanese, there are two types of words, accented and unaccented (Pierrehumbert & Beckman, 1988; Haraguchi, 1991), and approximately 55 percent of nouns are unaccented (Haraguchi, 1999). When a word is accented, there must be relative pitch movement from high pitch (H) to low pitch (L) within the word, as shown in (2), and the last vowel bearing the H is accented. When a word is not accented, the word does not have the HL pitch contour, as shown in (3).

(2) Accented words

\[ \text{hashi-ga} \quad \text{‘chopsticks-NOM’} \quad \text{kaki-ga} \quad \text{‘oyster-NOM’} \]

H L L

H L L

(3) Unaccented words

\[ \text{hashi-ga} \quad \text{‘edge-NOM’} \quad \text{kaki-ga} \quad \text{‘persimmon-NOM’} \]

H H H

H H H

The pitch on a phrase-initial vowel is always realized with a L pitch, when it is not accented (Pierrehumbert & Beckman, 1988). Therefore, in dictionaries and in the literature, the pitch contour of unaccented disyllabic words such as in (3) may be indicated LHH. However, phrase-initial L is realized as H phrase-medially. To avoid confusion, I will not use L for phrase-initial pitch.

---

7 There are many other dialects in Japanese and each dialect has distinct pitch patterns. Some dialects of Japanese such as the dialect spoken in Sendai, a northern part of Japan, and Kumamoto, a southern part of Japan, do not have pitch-accent (Shibatani, 1990).
Since pitch-accent is distinctive in Japanese, if a speaker changes the pitch contour of a word, it may be recognized as a different word or be considered unacceptable. When some words bearing a phonemically accented mora word-finally are pronounced at the end of an utterance, they show superficial neutralization, as in (4).

\[
\begin{align*}
\text{hashi-} & \quad \text{‘bridge-NOM’} & \rightarrow & \quad \text{hashi} & \quad \text{‘bridge’} \\
H & H & L & H & H \\
\text{kaki-} & \quad \text{‘fence-NOM’} & \rightarrow & \quad \text{kaki} & \quad \text{‘fence’} \\
H & H & L & H & H
\end{align*}
\]

When a long vowel is accented, only the first mora in the vowel carries a H pitch. Consequently, the HL pitch contour must occur within the long vowel (McCawley 1968, pp.133-134) as illustrated in (5).

\[
\begin{align*}
\text{Rōshi} & \quad \text{‘lecturer’} & \rightarrow & \quad \text{Rōshi} & \quad \text{‘fighting spirit’} \\
H & L & L & H & L
\end{align*}
\]

Several psycholinguists have investigated the role of pitch-accent in speech recognition by Japanese speakers (Sekiguchi & Nakajima, 1999; Cutler & Otake, 1999; McQueen et al., 2001). Their results all seem to indicate that pitch-accent provides Japanese speakers a cue for lexical segmentation. That is, pitch-accent is a significant part of lexical information for Japanese speakers. I will explore this issue again in Section 2.2.3.
1.3 **Pohnpeian**

Pohnpeian is a nuclear Micronesian language spoken on the island of Pohnpei in the Federated States of Micronesia. It used to be spelled *Ponapean*, but it has been changed to *Pohnpeian* to reflect its pronunciation. There are approximately 25,000 speakers (Rehg, 1998). It belongs to the language family called Austronesian,\(^8\) one of the largest language families in the world. Among Micronesian languages, Pohnpeian is one of the best-documented (Fischer, 1955; Garvin, 1971; Rehg & Sohl, 1979, 1981; Rehg, 1986, 1993; Goodman, 1995; Kennedy, 2003). The Pohnpeian sound system in particular has been extensively studied by Rehg & Sohl (1981) and Rehg (1984a, 1984b, 1984c, 1986, 1991, 1993). I will extract the parts of their analyses that are crucial for this dissertation and present them in the following sections.

### 1.3.1 Pohnpeian phonemes

The Northern dialect of Pohnpeian has 12 phonemic consonants and 7 vowels\(^9\) as in Table 1.2.

The phonemes \(/p^w/\) and \(/m^w/\) are labialized and velarized consonants. The laminal plosive \(/\ell/\) was described as retroflex in Rehg (1981, p.34). However, Rehg recognized later that it is rather a laminal alveolar plosive (Rehg, K., personal communication, December, 2001). The Pohnpeian liquid \(/r/\) is trilled. All voiced consonants can become

---

\(^8\) The Austronesian family was formerly called Malayo-Polynesian. Now Malayo-Polynesian is considered to be a subgroup of the Austronesian language family along with Formosan (SIL International, 2005).

\(^9\) Kiti dialect, spoken in the Southwest region of Pohnpei, does not contrast the vowels \(/e/\) and \(/e/\). Some words containing the vowel \(/e/\) in the northern dialect are pronounced with \(/a/\) in the Kiti dialect. See Rehg & Sohl (1981) for further discussion.
geminate. Pohnpeian consonants are divided into two types, front and back, based on place of articulation as in Table 1.3. This categorization is important for vowel allophony. I will discuss vowel allophones in Section 1.3.4.

Table 1.2: Pohnpeian phonemes

<table>
<thead>
<tr>
<th>Consonants</th>
<th>Vowels</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>i</td>
</tr>
<tr>
<td>t</td>
<td>e</td>
</tr>
<tr>
<td>k</td>
<td>a</td>
</tr>
<tr>
<td>p(^w)</td>
<td>u</td>
</tr>
<tr>
<td>t</td>
<td>o</td>
</tr>
<tr>
<td>s</td>
<td></td>
</tr>
<tr>
<td>m</td>
<td>e</td>
</tr>
<tr>
<td>n</td>
<td>o</td>
</tr>
<tr>
<td>m(^w)</td>
<td></td>
</tr>
<tr>
<td>l</td>
<td>a</td>
</tr>
<tr>
<td>r</td>
<td></td>
</tr>
<tr>
<td>w</td>
<td></td>
</tr>
<tr>
<td>j</td>
<td></td>
</tr>
</tbody>
</table>

Table 1.3: Pohnpeian consonants (Adopted from Rehg & Sohl, 1981, p.44)\(^{10}\)

<table>
<thead>
<tr>
<th>Front</th>
<th>Back</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>p(^w)</td>
</tr>
<tr>
<td>m</td>
<td>m(^w)</td>
</tr>
<tr>
<td>t</td>
<td>t(^\ddagger)</td>
</tr>
<tr>
<td>l</td>
<td>r</td>
</tr>
<tr>
<td>n</td>
<td>η</td>
</tr>
<tr>
<td>s</td>
<td></td>
</tr>
<tr>
<td>—</td>
<td>k</td>
</tr>
<tr>
<td>w</td>
<td>j</td>
</tr>
</tbody>
</table>

\(^{10}\) The glides /w/ and /j/ are not included in the original table but added to this table (Table 1.3) according to the description provided in the text (i.e., Rehg & Sohl, 1981).
There are seven phonemic vowels in Pohnpeian, /a, e, ɐ, i, o, ɔ, u/. All these vowels are contrastive for length. The letter $h$ following a vowel is used to indicate length, as shown in (6).

(6) Long vowels        Short vowels
    $pah /pa:/$ ‘to fight’ vs. $pa /pa/$ ‘under’
    $dohl /to:1/$ ‘mountain’ vs. $dol /to:/$ ‘to mix’

Rehg & Sohl (1981) state that long vowels might also be interpreted as a sequence of two identical vowels, or as double vowels. Pohnpeian vowel length thus is associated with mora count: long vowels carry two moras and short vowels one.

1.3.2 Pohnpeian vowel allophony

One of the most striking phonological phenomena of Pohnpeian is the allophony of the vowels. The surface quality of Pohnpeian short vowels is determined by adjacent consonants. For example, when Pohnpeian short front vowels, /i/, /e/, and /e/ occur between back consonants, they are centralized. Likewise, when short back vowels /u/, /o/, and /ɔ/ occur between front consonants, they become centralized. Examples are shown in Table 1.4.
Table 1.4: Pohnpeian vowel allophony (Adopted from Rehg & Sohl, 1981, pp.44-45)

<table>
<thead>
<tr>
<th>Vowel</th>
<th>Between Front Cs</th>
<th>Between Back Cs</th>
</tr>
</thead>
<tbody>
<tr>
<td>/i/</td>
<td>[pi:l]</td>
<td>['also'</td>
</tr>
<tr>
<td>/e/</td>
<td>[lel]</td>
<td>‘to be wounded’</td>
</tr>
<tr>
<td>/e/</td>
<td>[mEm]</td>
<td>‘sweet’</td>
</tr>
<tr>
<td>/u/</td>
<td>[lul]</td>
<td>‘to flame’</td>
</tr>
<tr>
<td>/o/</td>
<td>[pos]</td>
<td>‘to explode’</td>
</tr>
<tr>
<td>/ɔ/</td>
<td>[lɔl]</td>
<td>‘deep’</td>
</tr>
<tr>
<td>/ɑ/</td>
<td>[pæp]</td>
<td>‘to swim’</td>
</tr>
</tbody>
</table>

In a mixed environment, between a front consonant and a back consonant or between a back consonant and a front consonant, the following consonant has a stronger effect on the quality of vowels. Rehg & Sohl state that the effect of adjacent consonants is, however, not as noticeable on long vowels; that is, the primary front, central, or back quality of long vowels is largely preserved regardless of adjacent consonants. There might be a perceptible transition at the beginning or at the end of a long vowel in an environment where vowel allophones are expected to occur in short vowels (pp.45-46).11

1.3.3 Pohnpeian syllable structure

The core syllable in the native vocabulary of Pohnpeian is (C)V(X)(<C>), where X represents a segment without an independent place of articulation. X can be the second part of a long vowel, a glide, the first element of a sonorant geminate, or a homorganic nasal-obstruent cluster. <C> is extrametrical that may occur word-finally. Therefore, the

---

11 Contrary to Rehg & Sohl, Goodman (1995) claimed that adjacent consonants do have an effect on long vowels and that they had in some cases stronger effects on long vowels than on short vowels. However, she did not pursue this topic any further and did not provide concrete evidence. She averaged out the formant values from long vowels and short vowels in her analysis. Thus, the average of formants of vowels might be misleading with the respect to quality of the vowel. Further acoustic analysis of Pohnpeian vowel allophony is worth pursuing.
only consonant clusters that occur word-internally are sonorant geminates such as in *mall* ‘natural clearing in the forest’ and homorganic nasal-obstruent clusters such as *mand* /mant/ ‘tame,’ except in loanwords such as *nappa* ‘Chinese cabbage’ and *kesso* ‘the final race and interjections such as *esse* ‘ouch’ and *akka* ‘oh, my! (expressing surprise).’ Across morpheme boundaries, however, all possible consonant cluster types may occur, except for the combination of C plus a /η/. 
CHAPTER 2

BACKGROUND

2.1 Introduction

This chapter will provide the background for this dissertation. I will first briefly review the study of prosody focusing on the study of the mora. The mora is a prosodic unit representing the length of segments and may also be a pitch-/tone-bearing unit. It is an important concept for understanding Japanese and Pohnpeian phonology. The study of prosody has a long history, and acoustic and perceptual investigations of prosodic phenomena have been making remarkable progress due to the improvement of acoustic instruments. Prosody, without doubt, has been becoming a central key to understand human language and communication. One of the difficulties of studying prosody is that the term itself is used to refer to various phenomena. Researchers in various fields use the term in a way appropriate to their own field of study, such as intonation of an utterance, adding emphasis to a certain word in an utterance by increasing intensity and/or duration, expressing emotions by shifting voice quality, and so forth. In this dissertation, I will focus only on the part of prosodic aspect related to any vowel length distinctions.

I will begin this chapter by reviewing the study of prosody focusing on rhythmic unit typology and the concept of the mora. Then I will review the study of the phonetics-phonology interface in prosody, which is the motivation behind the present study. In Section 2.2, I will review the research relating to the Japanese mora. Section 2.3 is a literature review on the role of the mora in Pohnpeian. I believe that investigating the
phonological system in Pohnpeian from a phonetic perspective will contribute to the development of phonology, since Pohnpeian prosody – in particular – has been an understudied area.

2.1.1 Rhythmic unit typology

Cutler (1991) states, “Speech is rhythmic” (p.157). Crystal (1987) says, “Pitch, loudness, and tempo together enter into a language’s expression of rhythm” (p.169). All languages have elements of rhythm. Spoken languages use pitch, loudness, and tempo. Although sign language does not have a form of speech, it also has prosodic properties and rhythm (Brentari, 1998). Rhythm must be a universal property of languages. Spoken languages are often classified into mora-timed, syllable-timed, or stress-timed languages based on their rhythmic unit. For example, Japanese is a mora-timed language, Romance languages such as French and Italian are syllable-timed, and Arabic and Germanic languages such as English and Dutch, are classified as stress-timed languages. Finding out the principal rhythmic unit of a language and how the unit is organized and manifested in speech have been concerns for both phonologists and psycholinguists.

In the Renaissance, poetics was already vital in the study of prosody, and rhythm was the primary focus of prosody (Allen, 1973). Furthermore, a group of linguists recognized poetics as an important part of linguistic study in the late 1950’s. Jakobson (1960) argues that poetics is an integral part of linguistics (p.350). Later Jakobson & Waugh (1975) press this point even further, emphasizing that all elements of verse demand “an exact linguistic analysis with respect to the sound system of the given
language” (p.216). Abercrombie (1967) also recognizes the importance of poetics to enhance the understanding of the rhythm of everyday speech (p.98).

Most versification or poetic meter in the world’s languages is based on a prosodic unit called the foot. The investigation of scansion, or the metrical foot, is driven by the assumption that the rhythm of the poetry matches to the minimal rhythmic unit of the spoken form of the language. For example, the metrical foot of English is based on stress, and English is a stress-timed language (Kiparsky, 1975, 1977; Hammond, 1991). Italian meter is based on the syllable (Helsloot, 1995), and it is a syllable-timed language (Nespor, 1990); Classical Greek meter is based on the mora (McCawley, 1968; Golston & Riad, 2000), and it is a mora-timed language (McCawley, 1968).

It has been argued that the foot must be binary under a syllabic or moraic analysis (Prince, 1980; Prince & Smolensky, 1993; Hayes, 1995) and that the foot must be linked to the stress system (Liberman, 1975; Liberman & Prince, 1977). In other words, the number of either moras or syllables in a foot must be exactly two, and the foot is the stress-bearing unit. In opposition to the claim that the metrical foot must be linked to the stress system, Hansen & Kiparsky (1996) propose FOOT-BASED meter. They suggest that the foot is the fundamental rhythmic constituent and that the important components of the foot are two metrical positions: the head position STRONG and the non-head position WEAK. In other words, Hansen & Kiparsky argue that the alternation of head and non-head, not binarity, is the critical aspect of EURHYTHMY. In fact, Hansen & Kiparsky report that in some languages, like English and Finnish, the foot does consist of a head and non-head, but neither the syllable nor mora count in a foot is always two; rather, the
syllable count varies. Blevins & Harrison (1999) argue that there is evidence for ternary metrical feet in Gilbertese (Kiribatese), a Micronesian language spoken in Kiribati.

Meter realized through an association with stress is what Hansen & Kiparsky refer to as STRESS-BASED meter. In FOOT-BASED meter, the prominence position (STRONG) is language specific, but it is not necessarily associated with stress. That is, while Hansen & Kiparsky do not specifically state it, but their proposal supports the possibility that a language could have a meter or a foot that is not associated with a stress system. A significant finding supporting FOOT-BASED meter is Japanese foot structure proposed by Poser (1990). Japanese generally is not considered to have a stress system. By investigating hypocoristic formations, kinship terms, truncation patterns, reduplication patterns, and compound formation processes, Poser concludes that the basic rhythmic unit in Japanese is a bimoraic foot. The versification of traditional Japanese poetry also supports Poser's proposal. Traditional Japanese poetry has a strict metric system, which is based on mora count. That is, each line must contain either five- or seven- moras accompanied by a pause that occupies the non-head position of the foot; occasionally, a line is allowed to have an extra mora called ji-amari (which literally means 'an excessive letter') filling the non-head position of the foot (Kozasa, 2000).

Meter in Pohnpeian oral literature is also based on mora count (Fischer, 1959), which supports the proposal that the minimal rhythmic unit in Pohnpeian is the mora (Rehg 1993). The patterns observed in the poetic meter of a language reflect its prosodic patterns, particularly the constituents of the foot, such as stress, syllable, or mora, which correspond to the rhythmic unit of a language.
While some phonologists have been investigating poetic meter to determine the rhythmic unit of a language, psycholinguists have found some evidence that the rhythmic unit of a language corresponds to a unit of segmentation when the listeners process speech sounds (Cutler, 1994). Segmentation means making a division at some point in the acoustic signal into identifiable units, such as phonemes, syllables, or feet (Cutler & Norris, 1988). Cutler & Butterfield (1992) and Cutler et al. (1986, 1992) investigated the segmentation patterns of native speakers of French and English. French is considered a syllable-timed language and English a stress-timed language. Participants in their experiments were asked to detect a CV fragment such as ba- or a CVC fragment such as bal- in a French word such as balance 'balance' or balcon 'balcony' and in an English word such as balance or balcony. The results show that French listeners consistently used syllabification both for familiar French words and for words that are rather difficult to syllabify. That is, French participants were able to detect a fragment such as ba- in the word balance more quickly than in the word balcon or balcony. Conversely, a fragment such as bal- was more quickly detected in the word balcon or balcony than in the word balance, as the target fragment matched the initial syllable of the testing words; that is, the syllable ba- in balance and bal- in balcon and balcony.1 English listeners, however, did not use the syllabification strategy to detect fragments.

English is a stress language, and it has two different categories of syllables: strong and weak. Strong syllables contain a full vowel, while weak syllables have a reduced vowel. Furthermore, English lexical words predominantly begin with a strong syllable,

1 The period '․' in the words indicates syllable boundary.
and occurrence of words beginning with a strong syllable is twice as frequent as words beginning with a weak syllable (Cutler & Carter, 1987). Cutler & Norris (1988) conducted word-spotting experiments, in which participants were asked to detect a real word such as mint in a nonsense disyllabic utterance such as mintáyve and mintesh, and then repeat the real word aloud. The results suggested that English listeners use stress-based segmentation. It was harder for English listeners to detect words such as mint in mintáyve than in mintesh, as táyve, the second syllable of mintáyve, is strong (i.e., containing a strong vowel /aj/). Thus the listeners might have segmented mintáyve into min and táyve, while the word mintesh might not have had such a segmentation unit within the word and processed as a single unit mintesh. Therefore, it was easier for English listeners to detect the word mint in mintesh than in mintáyve. Sebastián-Gallés et al. (1992) conducted experiments similar to Cutler et al. (1986) with Catalan and Spanish listeners. Catalan and Spanish are both syllable-timed languages. The results show that Catalan and Spanish listeners also use syllabification to detect the target fragment in a word. I will discuss the details of research on the relation between the rhythmic unit and segmentation unit of Japanese in section 2.2.1.

A classification of spoken languages according to their rhythmic unit was suggested by Pike (1945) and adhered to by Abercrombie (1967). Pike and Abercrombie did not even mention the notion of the mora or mora-timed language. It was later added, as Trubetzkoy (1969) introduced the notion mora-counting languages. This rhythmic classification developed an assumption, i.e., the isochrony of the rhythmic unit of a
language. In other words, in a mora-timed language, each mora takes almost the same amount of time in an utterance; in a syllable-timed language, each syllable takes almost the same amount of time in an utterance; and in a stress-timed language, the stress takes an interval after the same amount of time in an utterance. It was thought that the regularity of the unit would create the rhythm of a language. The isochrony of the rhythmic unit was soon challenged by a number of phoneticians, and some phoneticians found no significant evidence for isochrony. Consequently, several scholars have proposed that the impression of different types of rhythm is a result of phonological properties (Dasher & Bolinger, 1982; Dauer, 1983; Warner & Arai, 2001a). There may not be absolute physical evidence proving the exact isochrony or existence of rhythmic units; however, the unit that the speaker uses to compose a poem or that the listener uses to segment an utterance into a meaningful speech sound can be represented with some categorical notations. I will discuss the isochrony of the mora further in Section 2.2.2. In the following section, I will explore the concept of the mora, the rhythmic unit of Japanese and Pohnpeian.

2.1.2 Concept of the mora

Mora is a phonological concept used to account for phenomena at a prosodic level. Trubetzkoy (1969 [1939]) uses the terms mora-counting and syllable-counting language to classify languages according to their rhythmic structure. McCawley (1968, 1978) defines the term mora as “something of which a long syllable consists of two and a short syllable consists of one” (1968, p.58).
His definition is illustrated in (1). The syllable (1a) has a single mora and the syllables (1b-c) have two moras.

\[
\begin{align*}
(1) \quad & \text{(C)V} & \text{(C)VV} & \text{(C)VC} \\
& \sigma & \sigma & \sigma \\
& \mu = [\text{ta}] & \mu \mu = [\text{ta}] & \mu \mu = [\text{tan}] \\
& \mu = \text{mora} & \mu = \text{mora} & \mu = \text{mora} \\
& \text{t a} & \text{t a} & \text{t a n}
\end{align*}
\]

Why does McCawley call this definition “imprecise” (ibid.)? What is missing from this definition? McCawley additionally states that when we need to detect the stress pattern or to assign stress in a word, we need to use a prosodic unit to count “the distance.” This statement implies that the mora is a prosodic unit with at least two functions: a unit representing phonological distance and a possible landing site for stress/pitch. McCawley’s definition of the mora that I introduced in the beginning of this section lacks the latter function; that is, a mora as a landing site for stress/pitch. These two functions – as a unit of phonological length/weight and as a landing site for stress/pitch, motivate Moraic theory, as proposed by Hyman (1984, 1985) and later adopted by McCarthy and Prince (1986, 1993) and others. In the following section, I will sketch the basic premise of Moraic theory.

---

2 These schemata are different from the ones generally used in literature in which the onset consonant attaches to the syllable node rather than to the mora node, which were originally proposed by Hyman (1984, 1985). Hyman used the symbol \(x\) to represent the mora node instead of the symbol \(\mu\). I will propose that structures like those in (1) better represent the concept of mora. The difference between this structure and the generally used ones is the treatment of onset. The structure in (1) puts the onset consonant under the mora node, which others put it under the syllable node. I will discuss the details in a later section.
2.1.3 Moraic theory

*Moraic theory* was developed to represent tone systems (Hyman, 1984, 1985) and stress systems of languages (Hayes, 1995). Its representations are often compared to those of *CV phonology* (Clements & Keyser, 1983) and *X theory* (Levin, 1985), as all three of these theories are concerned with the prosodic tier. In CV theory, the prosodic structure is represented schematically as in (2). In X theory, the prosodic structure is represented as in (3).4

![Diagram of CV theory](image1)

![Diagram of X theory](image2)

Both CV theory and X theory can represent relationships among segments in a syllable and how each segment functions in the syllable. Clements & Keyser suggest the representations in (2), which represent light syllables by a double branching from the

---

3 This term is used in Hayes (1989).
4 There are, of course, other theories that are concerned with prosodic structure; however, here I introduce just these three.
syllable node and heavy syllables by a triple or more branching. There is no nucleus or rhyme node. In X theory, syllable weight is represented by branching nucleus or branching rhyme as in (3). Blevins (1995, pp.214-215) provides the definitions in (4) based on her typological investigation of syllable structure.

(4) Structural definitions of syllable weight

<table>
<thead>
<tr>
<th>Language Type</th>
<th>Syllable weight</th>
<th>Heaviest</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CV</td>
<td>CVX&lt;sup&gt;5&lt;/sup&gt;...</td>
</tr>
<tr>
<td>2</td>
<td>CVC</td>
<td>CVV...</td>
</tr>
<tr>
<td>3</td>
<td>CV</td>
<td>CVV...</td>
</tr>
<tr>
<td>3</td>
<td>CVC</td>
<td>CV{V, R&lt;sup&gt;6&lt;/sup&gt;}...</td>
</tr>
</tbody>
</table>

In Type 1 languages such as Sierra Miwok and Hausa, if there is a constituent following a nuclear vowel in a syllable, the syllable is heavy. In Type 2 languages such as Huasteco and Hawaiian<sup>7</sup>, coda consonants do not contribute to the syllable weight, only vowels. In Type 3 languages such as Klamath, Yupik, and Creek, there is a three-way weight distinction.

Because neither CV theory nor X theory is developed with the intention of capturing syllable weight, it is not easy to determine the syllable weight by looking at either representations (2) or (3). In the representations in (2), all syllables have complex branching, except those consisting of only a vowel, and in order to determine the syllable weight, we must examine their components. In the representations in (3), each segment is

<sup>5</sup>X can be either a consonant or a vowel.
<sup>6</sup>Blevins (1995) does not specify what this R stands for. I assume that it is a sonorant consonant.
<sup>7</sup>Hawaiian does not have closed syllables.
dominated by an X, and the syllable weight is represented by either the nucleus or rhyme node, which are nodes at different levels.

On the other hand, Moraic theory is concerned with only the weight of a syllable; it is not concerned with segmental relationships within a syllable or the number of elements in a syllable. Hyman (1984) proposes the traditional notion of mora that is used as a unit of weight. Consequently, Moraic theory can describe prosodic processes associated with syllable weight, such as compensatory lengthening, more accurately,\(^8\) as the mora serves only as a weight unit, rather than representing syllable internal-structure. For the reader’s convenience, I have reproduced below the prosodic representations of Moraic theory from section 2.1.3. When Hyman (1984) originally proposed the central idea of Moraic theory, having the mora refer to a unit of syllable weight, he connected the mora node to both onset and nucleus, as in (5). These structures were adopted by Itô (1991) and a few others. There is another version of prosodic representation in Moraic theory, as in (6), suggested by McCarthy (1986), McCarthy & Prince (1986, 1993) and Hayes (1989).

\[\begin{align*}
(5) & \quad \text{a. (C)V} & \quad \text{b. (C)VV} & \quad \text{c. (C)VC} \\
\sigma & \quad \sigma & \quad \sigma \\
\mu = [\text{ta}] & \mu \_ \mu = [\text{ta}] & \mu \_ \mu = [\text{ta}] \\
\_ & \_ & \_ \\
\_ \_ & \_ \_ & \_ \_ \_ \\
\_ t \_ a & \_ t \_ a & \_ t \_ a \_ n
\end{align*}\]

\(^8\) For a detailed discussion comparing X theory and Moraic theory see Hayes (1989).
The representations in (6) indicate that the onset consonant has nothing to do with syllable weight, and that only the vowel and coda consonant are the sources of syllable weight. Now recall the definition of the mora that McCawley (1968) proposes. The mora functions as a unit of length; therefore, the mora count must correspond to the length of a segment. Vowel length can be represented under both systems as in (5b) and (6b), as two branches coming out from a vowel, linked to two different mora nodes. Nevertheless, consonant length cannot be represented by the prosodic structures in (6). In a word-medial geminate, the latter part of the long consonant is the onset of the following syllable. However, if the branch from the long consonant is linked to the syllable node as in (7), the representation does not reflect the consonant length, because the C is not associated with two moras; the syllable is not a unit of the length of a segment.

If I use the representation in (5c), however, the length of geminate consonants can be represented as in (8). Since the duration of geminates is longer than singletons, the
length should be represented in some phonological form. In (8), the second and third mora are linked to the nasal /n/ to indicate the consonant length.

\[ \sigma \sigma \]
\[ \mu \mu \mu = [\text{tanda}] \]
\[ t a n a \]

\( \sigma = \text{Syllable} \)
\( \mu = \text{mora} \)

Hayes (1995) favors the representations in (5) and (8) over those in (6) and (7), as they represent weight-bearing segments clearly (p.53-54). Ito (1988, 1991) also adopts the representations in (5) and (8) to account for Japanese syllable structure as they can predict the occurrence of vowel epenthesis. Kubozono (1995, 1996) suggests that some pieces of psycholinguistic evidence support the representations in (5) and (8) to represent Japanese prosodic structure. I will discuss Kubozono's suggestions in section 2.3.2.

The schema in (8) is not, however, problem free. Hyman (1984, 1985) already recognizes the problem and Broselow (1995) discusses it as well. The problem arises when the schema is used to represent syllable structure in a word that has a word-initial geminate consonant. For example, the word /m:im/ ‘inside me’ in Gokana, a language spoken in Nigeria (Hyman 1985, p.42) has an initial geminate consonant. Its prosodic structure is represented as in (9).

\[ \mu \mu \mu \]
\[ [\text{m:im}] \]

\( \mu = \text{mora} \)
A representation like (9) correctly reflects the phonological length of the geminate consonant. The problem with this representation is that it suggests that the onset is associated with a syllable weight. Hyman (1985) thus argues that weight units (moras) can be dominated by higher-order units such as the foot with no phonological syllables in this language. Kennedy (2003), by investigating reduplication mechanisms in Micronesian languages, suggests two separate tiers for moras and syllables as in (10).

\[(10) \quad \text{Foot} \quad \text{Foot} \quad \text{(adapted from Kennedy 2003, p.187)}\]

\[
\begin{array}{c}
\text{a} \quad \text{l} \quad \text{a} \quad \text{l} \quad \text{u} \\
\sigma \quad \sigma \quad \sigma
\end{array}
\]

*allalu* ‘to walk’ in Mokilese

In this structure, moras and syllables belong to two independent tiers and thus moras are not sub-constituents of syllables but of feet. The foot is constructed from moras or syllables depending on the language. Consequently, when we consider the rhythmic unit of languages, the foot rather than the mora or the syllable should be treated as the rhythmic unit of languages. The structure in (10) is a simple and more general representation of prosodic structure than the structures in (5-9).

### 2.1.4 The relation between phonology and phonetics in prosody

In the study of prosody, many scholars have been giving their full attention to the integration of phonetics and phonology. The methodology of their investigation of the phonetic-phonology interface is increasingly improving as modern technology progresses. On the one hand, phonetics considers prosodic properties as physical signals manifested
by intensity, duration, and pitch movement (Lehiste, 1970; Ladd & Cutler 1983; Nooteboom, 1997; Peters 1997). On the other hand, phonology deals with how these physical signals are processed and organized in the mind and how to represent this internal organization. Thus one of the goals of phonology is to find the simplest yet most realistic way of representing prosodic phenomena.

Pierrehumbert & Beckman (1988) point out that the phonological representation that SPE [The sound pattern of English (Chomsky & Halle, 1968)] proposes could not adequately represent phonetic findings, and that the representation has limitations in representing in prosodic structure. The point Pierrehumbert & Beckman emphasize is that making phonological representations without modeling the phonetic realization of tone and intonation cannot be accurate, as those prosodic properties are crucial to understanding the speech-sound system. In the tradition of generative phonology, metrical phonology (Lieberman, 1975; Lieberman & Prince 1977; Hayes, 1995) is an attempt to treat prosodic events using abstract phonological representations; as such, it does not incorporate acoustic analyses. In that respect, Pierrehumbert (1980) is an important work for integration of phonetics and phonology in prosody. She proposes a phonological representation system for English intonation, which had been thought to be a continuum and thus impossible to describe categorically. A phonological approach to intonation had began to progress in the 1970’s (Bolinger, 1986, 1989; Hirst & di Cristo, 1998). Liberman (1975), Bruce (1977), and Pierrehumbert (1980), and their acoustic analyses were developed into a conventional labeling system called ToBI (Tone and Break Indexes) (Silverman et al., 1992; Beckman & Ayers, 1994; Beckman & Hirschberg,
1994; Pitrelli et al., 1994), with which one can transcribe prosodic features of English (Ladd, 1996). The ToBI system has been adapted to transcribe non-English languages such as Japanese (Campbell & Venditti, 1995; Venditti, 1995), Korean (Jun, 1998, 2000), and German (Ohio State University Department of Linguistics, 1999).

In phonology, the prosodic organization of speech can be divided into roughly two dimensions: the rhythmic or temporal aspect of speech and the melodic9 aspect of speech. When we consider the rhythmic/temporal characteristics of a language, we use the term mora-, syllable-, or stress-timed to classify languages, whereas, when we consider the melodic characteristics of a language, we classify languages into tone, pitch-accent, or intonation languages. Investigating the way rhythmic units are manifested as phonetic or acoustic signals has been a challenge for linguists, as rhythm of speech is essentially a matter of perception. Moreover, what makes discovering the phonetic reality of prosodic features controversial and thus problematic is that these prosodic characteristics are not mutually exclusive; yet, the prosodic characteristics are bound together in a complex way within a string of speech. For example, tone languages may also have intonation. Mandarin Chinese, a tone language, can use intonation to convey some pragmatic meanings. Greek has a stress system, yet its timing unit is the mora, not stress. Moreover, each of these phonological features does not correspond to a single phonetic property. For example, English stress is known to be an accumulation effect of pitch, duration, and intensity (Hyman, 1975, p.207; Hayes, 1995, pp.6-7). It is also a

---

9 Some phonologists use the term melody or melody-tier to refer to each segment or the segment level of utterances (e.g., Itô & Mester (1986)). I am using the term melody here with the meaning that is generally accepted and used in a phrase such as melody of the song.
well-known fact that stress indirectly involves segmental cues such as change of vowel quality and amount of aspiration in plosives (Lehiste, 1970, pp.139-142).

One of the phonological assertions challenged by several phoneticians was the isochrony of rhythmic units of speech. Since rhythm is created by recurrent regular patterns, it is only natural that the idea of isochrony has arisen. The idea is that the duration of the rhythmic unit in an utterance is evenly distributed. Several scholars have argued that the perception of different types of rhythm is the result of the phonological properties of a language or of speakers, such as syllable and prosodic phrase structure or the stress system of the language. They argue that there is no phonetic evidence for the rhythmic unit (Beckman, 1982a; Dauer, 1983; Nespor, 1990). Beckman conducted an experiment to investigate the phonetic reality of the mora in Japanese and concluded that Japanese orthography was responsible for the ability of native speakers to count moras.10 Dauer examined the duration of each rhythmic unit in English, Thai, Spanish, Greek, and Italian and concluded that the perception of isochronous distribution of these units was created by the syllable structure and “language specific segmental variation” such as vowel reduction rules (p.59). Nespor also argued that the realization of rhythm was due to “nonrhythmic phonological rules” (p.172) such as vowel reduction and stress shift. It is still a controversial issue whether physical evidence supports the existence of isochrony in utterances. Nevertheless, what we perceive and process mentally is real; we are able to count the number of syllables/moras in a word and to compose/recite a poem with a

10 The orthographic system is one artifact of the way speakers segment utterances into meaningful units. Ladefoged (2001) defines phonemes as “abstract units that form the basis for writing down a language systematically and unambiguously” (p.24). Thus, in my opinion, Japanese orthography provides historical evidence for the phonetic reality of the Japanese mora. Whether it reflects current phonological phenomena needs to be empirically tested.
correct rhythmic pattern. The perception of rhythm must be triggered by some acoustic signals. The physical properties of speech rhythm may not be represented straightforwardly; that is, multiple elements might be interacting in a complex way. As Cutler (1991) points out, “rhythm in language is more than just timing” (p.157). We recognize rhythm in language by its duration (long/short) and by its qualities (strong/weak). Greek can categorize time from two different perspectives, chronos ‘period of time’ and kairos ‘a significant moment of a period of time. The concept of time in Greek resembles two dimensions of prosody, rhythm/tempo and melody. In scientific investigations, data need to be analyzed as objectively as possible. Therefore, it is important to separate factors corresponding to each prosodic notation into temporal aspects and qualitative aspects, and then investigate the way they interact with one another at the phonetic level as well as at the phonological level.

One of the aims of this dissertation is to attempt to explore the phonetic component of a phonological unit, the mora, by measuring acoustic properties of long and short vowels in Japanese and Pohnpeian. Each language’s speakers are able to produce and perceive vowel length; this phonological ability must be established in the speakers’ mind by being exposed to their own language. Finding acoustic properties of prosodic features such as vowel length in these two languages follows a direction that phonological studies need to take in the future.
2.2 Japanese mora

The mora plays a significant role in Japanese phonology. The mora is an important prosodic unit in Japanese. Kubozono (1999, p.32) states that a mora may be realized in five different ways: (i) a single short vowel either by itself (V) or preceded by a consonant (CV), (ii) a moraic nasal (CVN) (a nasal in the coda position), (iii) a moraic obstruent (CVC) (the first element of a geminate consonant), (iv) the second half of a long vowel (CVV), and (v) the second half of diphthongal vowel sequences\(^\text{11}\) (CVG), where G stands for a glide. Examples of each mora type are in (11i-v), respectively.

(11) i. 
\begin{align*}
\text{i.su.zu} & \quad \text{‘Isuzu’} \\
\text{a.ke.bo.no} & \quad \text{‘Akebono’}
\end{align*}

ii. 
\begin{align*}
\text{ho.na} & \quad \text{‘Honda’} \\
\text{ma.na.ku} & \quad \text{‘middle’}
\end{align*}

iii. 
\begin{align*}
\text{te-ka} & \quad \text{‘tuna roll sushi’} \\
\text{ki.te} & \quad \text{‘postage stamp’}
\end{align*}

iv. 
\begin{align*}
\text{so.ni-i} & \quad \text{‘Sony’} \\
\text{su.mo-o} & \quad \text{‘Sumo wrestling’}
\end{align*}

v. 
\begin{align*}
\text{ta-lko} & \quad \text{‘taiko drums’} \\
\text{ka.re-i} & \quad \text{‘sole fish’}
\end{align*}

Kubozono (1996) suggests that the role of the Japanese mora falls into four types:

a) a unit by which to measure phonological weight or distance, b) a timing unit, or a unit of temporal regulation of natural connected speech, c) a unit by which to segment in speech production, and d) a unit by which to segment speech in speech perception (p.78).

\(^{11}\) To the best of my knowledge, there are no studies specifically suggesting that Japanese has diphthongs. Vance (1987) suggests that there is a distinction between phonetic syllables and phonological syllables, which is roughly determined by the way listeners perceive the rhythmic unit and by the speech tempo. For example, English speakers perceive two units (or syllables) in the Japanese word \textit{kekkon} ‘marriage’ = \textit{kek} and \textit{kon}; while Japanese speakers may count four units (or moras) – \textit{ke}, \textit{k}, \textit{ko}, and \textit{n}. Likewise Japanese speakers count six units (moras) in a place name \textit{shinoomiya} (\textit{shi.n.o.o.mi.ya}) not four units (\textit{shi.noo.mi.ya}) that English speakers may count. As a native speaker of Japanese, I do not believe that there is a phonological diphthong in Japanese, as any two vowels can be combined freely in Japanese, although I do not deny that nonidentical vowel sequences may be realized as diphthongs phonetically.
In this section, I will review the literature on the Japanese mora, focusing on its role as a unit of speech segmentation (Section 2.2.1) and a timing unit (Section 2.2.2), and the literature on a function of Japanese pitch-accent (Section 2.2.3).

2.2.1 The mora as a unit of speech segmentation

Despite the fact that the phonetic reality of the mora is still controversial, quite a few scholars are convinced of the psycholinguistic reality of the mora. The ability of native speakers to count the number of moras in traditional Japanese verses or the Japanese writing system is a strong piece of evidence for the psychological reality of the Japanese mora. Contemporary psycholinguistic research has uncovered another piece of evidence for the psychological reality of the mora and has shed some light on the mora as an important concept for Japanese phonology.

Natural speech is continuous, and segmenting speech into meaningful units, such as phrases and words, is a necessary task of each listener. Cutler et al. (1986) found that French speakers utilize syllables as a segmentation unit. Their findings suggest that the segmentation unit corresponds to the rhythmic unit of a language. Their hypothesis has motivated several experiments on the Japanese mora as a unit of speech segmentation.

Otake et al. (1993) conducted fragment-detecting experiments similar to Cutler et al. (1986) with Japanese, English, and French listeners but used Japanese words. Participants were asked to detect CV fragments such as *ta*- and CVC fragments such as *tan*- in words such as *ta.ni.shi* or *tan.shi*. The results show that Japanese listeners were able to detect CV targets such as *ta*- relatively quickly and accurately in CV.CV.CV words such as *ta.ni.shi* and in CVN.CV words such as *tan.shi*. On the other hand, French
listeners used the syllabification strategy even when they listened to Japanese words. They were able to detect CV targets easily in CV.CV.CV words and CVN targets in CVN.CV words, but they had difficulty in detecting CV targets in CVN.CV words. English speakers did not show segmentation patterns that were observed in Japanese and French speakers. The researchers concluded that Japanese speakers used the mora for segmentation and French speakers used the syllable. Otake et al. (1996b) examined Japanese listeners’ recognition of mora nasals in on-line listening tasks. The results show that Japanese listeners detect a nasal consonant in CVN.CV words such as tonbo ‘dragonfly’ significantly more rapidly than in CV.NV.CV words such as tenisu ‘tennis’. The results suggest that Japanese listeners are sensitive to mora structure. Cutler et al. (2003) tested speakers of Japanese and Telugu, a Dravidian language. The results support the findings of previous experiments; Japanese speakers used the mora and Telugu speakers used the syllable for the fragment-detecting task.

The results of these experiments show that the segmentation unit of a language matches the minimal rhythmic unit of the language. Furthermore, the segmentation strategy of each language’s listeners corresponds to the minimal rhythmic unit of their language whether they process their own language or foreign languages.

Kakehi et al. (1996) conducted another type of perception experiment in which a consonant identification task was used with native speakers of Dutch and Japanese. The participants were asked to identify a consonant in a VCV segment, where the Vs were Japanese vowels (/a, e, i, o, u/) and the Cs were Japanese /p/, /t/, or /k/. In order to control phonemic information, a cross-splicing technique was used. The types of stimuli Kakehi
et al. used are listed in (12). The stimulus (12b) VC$_1$-C$_2$V was created from the original VCV segments by blending the pre-closure\textsuperscript{12} part of a VC$_1$V and the post-closure part of a VC$_2$V, and (12c) was created by replacing the pre-closure part with a single vowel pronounced in isolation. The type (12c) was created by eliminating the pre-closure part of a VCV, and the type (12e) was created by eliminating the post-closure part of a VCV.

(12) a. Original VCV

\[ \text{SAMPLE: } /ape/ \]

b. VC$_1$-C$_2$V: \[ V C_1 V \xrightarrow{\text{VC} \_ + C_2 V} V C_1 + C_2 V \] 

pre-closure part of VC$_1$V /ate/ + post-closure part of VC$_2$V /ape/

c. V-CV: \[ \xrightarrow{V + C V} \text{SAMPLE: } /a-pe/ \]

single vowel /a/ + post-closure part of VCV

d. VC: \[ \text{VCV } \rightarrow \text{VCO } \text{SAMPLE: } /ap-/ \]

pre-closure part of VCV

e. CV: \[ \text{VCV } \rightarrow \text{ØCV } \text{SAMPLE: } /-pe/ \]

post-closure part of VCV

Furthermore, a partial noise replacement was used to control phonemic information; that is, the part from the beginning of the release burst for the stop of each stimulus was replaced by noise (V-noise-CV). The durations of noise were 0, 10, 30, 50 and 70 ms.

The results show that both Dutch and Japanese speakers identified the onset consonant in (12a, c, and e) types of stimuli 100% correctly. However, the accuracy of

\textsuperscript{12} That is, the closure for the utterance-medial stop. Thus, the pre-closure part must be the vowel preceding a stop. Likewise, the post-closure part must be from the released point of the stop to the end of the vowel following the stop. Neither pre-closure nor post-closure part contained supplemental phonemic information, such as aspiration.
Japanese speakers identifying the coda consonant – C₁ (12b) type stimuli and C in (12d) type stimuli – was significantly lower than those of Dutch speakers. Even without noise replacement, the accuracy rate of Japanese speakers was only 53% for VC₁ and 59% for VC, where Dutch speakers got 71% and 78% accuracy respectively. In other words, Japanese listeners had difficulties identifying coda consonants when they were heterorganic to the following consonant, while Dutch speakers did not seem to have problems identifying consonants regardless of their position. Kakehi et al. thus concluded that basic speech perception units in Japanese were larger than phonemes, i.e., the consonant must be followed by a vowel in order for Japanese listeners to identify it, and the geminate is an independent perception unit in Japanese (p.141). 13

The majority of studies investigating the relation between the speech segmentation unit and the rhythmic unit of a language have used a perception-oriented method such as a fragment-detecting task. Kubozono (1995), in contrast, used a word blending task, which is a production-oriented method, to investigate the speech segmentation unit in Japanese and English. Kubozono asked native speakers of English and Japanese to create new words by blending two existing words. Examples in (13) are adopted from Kubozono (1995, p.145, p.148).

13 See Otake et al. (1996b) for Japanese listeners' identification of moraic nasal. Japanese speakers' identification mechanism of moraic consonants (the first element of geminate consonants) needs to be further investigated.
Pairs of input words produced by a native speaker of English were given as audio stimuli. The participants were asked to repeat the input words aloud and then to say the blended word. If speakers used the onset from the first input word and the following part from the rhyme of the second input word (C/VC blend), the output words would fall under the first column of the output word list. If speakers used the onset and the vowel from the first input word and the coda from the second input word (CV/C blend), the output words would fall under the second column of the output word list. The results show that English speakers used dominantly the C/VC blending patterns (86% of the total responses), whereas, Japanese speakers used the CV/C blending patterns (79% and 80% of the total responses\textsuperscript{14}). From these results, Kubozono suggested that Japanese speakers segment words on the basis of moraic structure. Furthermore, because English speakers showed tendencies to segment a syllable into onset and rhyme, he suggests that the syllable structure commonly used in phonology such as in (14) is suitable for languages like English but does not work for languages like Japanese. Japanese prosodic structure is best represented in structure (15), as Japanese speakers segment a syllable into an onset-plus-vowel position and a coda.

\begin{table}[h]
\centering
\begin{tabular}{|l|l|l|}
\hline
Pattern of blend & C/VC & CV/C \\
\hline
a. cup / mitt \rightarrow & kit or cut & \\
b. pen / fat \rightarrow & pat or pet & \\
c. fan / put \rightarrow & foot or fat & \\
d. lot / take \rightarrow & lake or lock & \\
e. team / such \rightarrow & touch or teach & \\
f. five / sheet \rightarrow & feet or fight & \\
\hline
\end{tabular}
\end{table}

\textsuperscript{14} Kubozono conducted two experiments with Japanese speakers.
Kubozono (1996) studied cases of speech errors in Japanese. Speech error patterns in English showed that words were segmented before the nuclear vowel such as everybody/everyone → everybun and smart/clever → smever (p.78), while Japanese speakers’ speech error patterns showed that words were split after the nuclear vowel such as neko ‘cat’/nyanko ‘kitten’ → nenko and doomed ‘why’/nande ‘how come’ → donde (p.78). Along the lines of his previous experiment (Kubozono, 1995), Kubozono argues that the mora serves as “a segmentation unit in Japanese while in other languages it only serves as a weight unit, a unit whereby phonological timing is defined” (Kubozono, 1996, p.92).

2.2.2 The mora as a timing unit and its isochrony

When Pike (1945) introduced his idea of classifying languages based on their temporal qualities, American linguistics had not given much attention to the mora. He did

\[ \text{(14)} \]

\[
\begin{array}{c}
\sigma \\
o \text{onset} & \text{rhyme} \\
\text{peak} & \text{coda} \\
C & V & C
\end{array}
\]

\[ \text{σ = syllable} \]
\[ C = \text{consonant} \]
\[ V = \text{vowel} \]

\[ \text{(15)} \]

\[
\begin{array}{c}
\sigma \\
\mu & \mu \\
C & V & V/C
\end{array}
\]

\[ \text{σ = syllable} \]
\[ \mu = \text{mora} \]
\[ C = \text{consonant} \]
\[ V = \text{vowel} \]
not use the term *mora*, but instead categorized languages as syllable-timed or stress-timed. The mora is now a commonly accepted notion in phonology.

Japanese long vowels are often used to illustrate the mora as a unit of phonological length, as McCawley (1968, p.58) suggests. If a vowel is distributed over two moras, the vowel is phonologically long; if over one more, the vowel is phonologically short. It was generally assumed that moras have roughly equal duration phonetically as well.

Several phoneticians have commented on the duration of moras. Bloch (1950) uses the term *fraction* to refer to the mora. He states that all fractions are “about the same length” and that “two phrases containing the same number of fractions are heard as equal in duration” (p.91). He did not have any evidence from instrumental analysis to support his statement and admitted that it was his “auditory impression” that “staccato rhythm” as a characteristic feature of Japanese (p.90). Ladefoged (2001) also states, “Each mora takes about the same length of time to say” (p.233) but provides no phonetic evidence.

Han (1962) suggests that “The actual duration of each *onsetsu* [referring to the mora] is approximately the same” (p.81) based on her phonetic measurements. It is not clear the exact number of tokens and methods of data analysis she used, but some examples of test words were V.CV vs. VC.CV words such as *i.ki* ‘breath,’ *ik.ki* ‘one aircraft,’ *i.sho* ‘will,’ *is.sho* ‘together,’ *shi.mai* ‘sisters,’ *shim.mai* ‘untrained,’ and CV segments such as *pa, ta, ka, si, chi, ki, ra,* and *ri*. These words and segments were uttered in isolation. Han found that the ratio of the mean duration of V.CV words to VC.CV words was 2:3. Moreover, she found that the duration of consonant and vowel balanced
each other within a CV syllable so that the duration of two different utterances such as 
/ra/ and /pa/ were identical despite of the differences in intrinsic duration between /r/ and
/p/. Han further suggested, “an utterance of a certain number of onsetsu lasts essentially
the same length of time as another utterance with the same number of onsetsu” (p.82).

Homma (1981) measured closure duration, VOT, and vowel durations in two-
mora nonsense words CVCV and three-mora nonsense words CVC:V. The vowel used in
all stimuli was a low vowel /a/, and consonants were Japanese stops /p, t, k, b, d, g/ and
geminates of these stops. In each target word, the two stops shared the same place of
articulation, for example, /papa/, /paba/, /bapa/, /baba/, /tatta/, /tadda/, /data/, and /dada/,
and thus words such as /pata/ or /taga/ were not used. The pitch-accent was on the first
vowel. The results showed that all durations – closure, VOT, and vowel – varied
depending on 1) voicing feature, 2) point of articulation, and 3) adjacent segments.
However, the difference in word duration among the same mora count words was small,
and the ratio of the mean duration of 2 mora words to 3 mora words was roughly 2:3. She
concluded that the duration of each mora varied due to intrinsic segmental duration
differences; however, moras compensated the duration within a word to equalize it.

Port et al. (1987) conducted four production experiments using words with
various mora counts including different types of syllables (CV, CV:, and CVC, in which
the coda consonant was a first element of a geminate). The method used was similar to
Lehiste (1976), who used words like form, former, formerly, to investigate the isochrony
of the mora. Port et al. increased the number of moras in a word. Some examples of the
test words are in (16). All test words were nonsense words except ones used in Experiment 2.

<table>
<thead>
<tr>
<th>(16)</th>
<th>Experiment 1</th>
<th>Experiment 2</th>
<th>Experiment 3</th>
<th>Experiment 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>ra shi</td>
<td>ka</td>
<td>(all 2-mora)</td>
<td>buku</td>
<td></td>
</tr>
<tr>
<td>raku shita</td>
<td>kaku</td>
<td>kaka</td>
<td>buuku</td>
<td></td>
</tr>
<tr>
<td>rakuda shitaku</td>
<td>kakusi</td>
<td>kasa</td>
<td>buku</td>
<td></td>
</tr>
<tr>
<td>rakudaga shitakusu</td>
<td>kakushido</td>
<td></td>
<td>bukudo</td>
<td></td>
</tr>
</tbody>
</table>

The results showed that the word duration increased as the mora count increased “by roughly equal increments for each additional mora” (p.1578). Differences in intrinsic duration among segments were adjusted in a word, so that the duration of the word was close to the value that was predicted from the number of mora. Port et al. suggested that the duration of each mora is not equal but the mora exists as a timing unit only in longer strings of utterances like words (p.1584). What Homma, Port et al., and Bloch (1950), are describing is the overall tendencies of moraic behavior at the phrase level, which Warner & Arai (2001a) call phrasal compensation.

Contrary to Han (1962), Homma (1981), and Port et al. (1987), Beckman (1982a) is the first instrumental study that concluded that the mora was just a perceptual timing unit with no phonetic reality. Beckman investigated the duration of moras in 75 real Japanese words, 2 to 4 moras in length, with various combinations of consonants and vowels, such as siku ‘spread,’ gakuhu ‘musical score,’ kato ‘transition,’ katto ‘haircut,’ kamu ‘gnaw,’ genkin ‘ban,’ and so forth. The results did not show even a tendency toward isochrony; rather what she found were “universal physiological constraints” (p.133). Beckman focused on measuring the duration of single moras, mainly CV moras.
and moraic consonants, to investigate if the Japanese moras have constant durations. She argued that no data support “a strict claim of an exact isochronous mora” (p.114), but they are only showing tendencies toward isochrony. She compared the mean duration of CV (CV syllables with a devoiced vowel) with the mean duration of C (prevocalic consonants in CV syllables with a voiced vowel). If the intra-mora isochrony is physically real, the duration of CV will be much longer than C, to fulfill the durational slot of the devoiced vowel. However, Beckman did not find a significant difference between the durations of CV and C. She also compared the mean durations of word medial CCV segments and CV segments to investigate the duration of coda consonants. If the mora were a roughly constant unit of timing, the duration of CCV should be twice as long as the duration of CV. The results showed that the mean duration of CCV segments were merely 66% longer than the mean duration of CV segments. Beckman also measured the proportions of the durations of C and V within CV syllables. She did not find systematic negative correlation between the duration of the consonant and the duration of the vowel. In other words, the duration of CV syllables varied environment to environment. What Beckman pointed out are that affricates and voiceless fricatives are intrinsically longer than stops, nasals and flaps, and that voiceless stops are longer than voiced stops in many languages in which vowel deletion does not occur. Affricates, voiceless fricatives, and voiceless stops are “precisely the types of consonants most favoring” devoicing the following [u] in Japanese (p.129). Beckman argued that the durational compensation between consonant and vowel within a CV syllable cannot be
attributable to the isochrony of the mora but to universal physiological constraints. Consequently, Beckman concluded that Japanese speakers’ ability to count the number of moras in utterances was due to their knowledge of how the utterance could be written in *kana* letters. Furthermore, she suggested that the ‘staccato rhythm’ was due to the larger number of syllable counts compared to English utterances and the lack of radical variations in the length of Japanese syllables because Japanese syllables did not have the lexical stress that English syllables have.

Another investigation that challenged the isochrony of the mora is Warner & Arai (2001a). They examined data from spontaneous speech of native speakers of Japanese and found no evidence for the isochrony of the mora. They measured not only the durations of entire words but also the durations of partially truncated words that were words without their final two moras (e.g., the duration of *kagosi* in the word *kagoshimashi* ‘Kagoshima city’). They extracted segments from sentence initial, medial, and final positions and truncated words, and examined the correlation between duration and the number of moras. Warner & Arai used both lexical words and *prosodic words*, which usually include function words – typically case markers in Japanese – for their analysis. They found a very strong effect of final lengthening; that is, the duration of two-mora segments located in the final position of words, phrases, and sentences was significantly longer than the duration of two-mora segments in initial or medial position. If final lengthening is due to phrasal compensation, the duration of phrase-final two-mora segments can be predictable from the duration and the mora count of the rest of the phrase. However, the degree of phrase final lengthening was not predictable. In other
words, final lengthening was not due to phrasal compensation. In addition, their results showed that the duration of truncated words was better predicted by the mora count than by the duration of whole prosodic words. That is, as the mora count in truncated words increased, the duration of whole words became less predictable, which does not support the ‘durational compensation’ hypothesis proposed by Homma (1981) and Port et al. (1987). If the duration of moras compensated within a whole prosodic word, the duration of the truncated word should be predictable by the duration of whole prosodic word and as the mora count in truncated words increased the duration of the truncated words should have become more predictable by the duration of whole prosodic words. These results led Warner and Arai to the conclusion that Japanese mora-timing was not due to the isochrony or durational compensation among moras but due to the structure of the language. Warner & Arai do not explain explicitly what the ‘structure of the language’ is. It could be the syllable structure, as Dauer (1983) suggests, the phrasal structure, how prosodic words are constructed, the intonation pattern of whole utterances, the entire syntactic/semantic structure, and/or a combination of these linguistic properties.

In their review of the studies on the Japanese mora, Warner & Arai (2001b) emphasize the importance of investigations on the relationship between phonetic and psycholinguistic effects for understanding the timing of Japanese and its role in the rhythm of Japanese. Perceptual investigations are crucial to understand the Japanese mora as a timing unit.
2.2.3 A function of Japanese pitch-accent

Pitch-accent is another significant phonological feature in Japanese. It has not only a morphological function, (that is, the position and presence of pitch-accent may change the meaning of a word) but it also functions as a cue for segmentation. Cutler & Otake (1999) investigated how pitch-accent accelerates Japanese listeners' recognition of an extracted syllable. For example, the participants were asked to determine the source of an extracted syllable from two words differing in accentual pattern such as a syllable ka from baka HL ‘idiot’ or gaka HH ‘painter.’ Japanese listeners more correctly responded when the target syllable was accented. Otake & Higuchi (2004) conducted similar experiments to Cutler & Otake (1999) with native speakers of the Fukushima dialect in which pitch-accent is not a distinctive feature. The results show that accurate perception of accented word by Tokyo Japanese speakers was significantly higher than that by Fukushima dialect speakers. These results support the claim of another study on the effect of pitch-accent on word recognition with the speakers of the Kagoshima dialect, another accentless dialect of Japanese (Otake & Cutler, 1999). Otake & Cutler suggest that pitch-accent constraints the activation and selection of candidates for spoken-word recognition for Tokyo Japanese speakers.

2.3 Pohnpeian Mora

Several studies have attempted to account for the prosodic organization of Micronesian languages including Pohnpeian (for Proto-Micronesian, Rehg, 1993; for Gilbertese, Blevins & Harrison, 1999; for Marshallese, Abo et al., 1976; Bender, 1968,
1969, forthcoming). Among them, a study on meter of Micronesian languages by Fischer (1959) is noteworthy. Fischer described the meter of oral text or chant in Pohnpeian based on data that he collected in Ponape (Pohnpei), Truk (Chuuk), Ngatik, Mokil, and Pingelap. Fischer used the term *mora* as a unit of phonological length; that is, a short vowel has one and a long vowel has two. Considering the time at which he wrote the paper (i.e., 1959), his linguistic knowledge was remarkably sophisticated. He suggested that the meter of Pohnpeian was based on quantity; that is, its meter is based on number of moras. Each line of Pohnpeian chant consists of either five or seven moras. Although this versification is identical to versification of classical Japanese poetry, he denied any language relationship between Japanese and Pohnpeian. Fisher recognized that the prosodic organization of languages of Micronesia was distinct from that of European languages. He argued that “regular meter has been overlooked in a number of languages because the investigators were looking for a type of meter like their own or like that in some other European language” (p.50). He added, “Perhaps further examination will reveal more instances and place the Eastern Carolinian examples in a wider historical and typological perspective” (p.51).

In addition to Fisher’s proposal, there is some evidence for the mora as the minimum prosodic unit in Micronesian languages. Lynch (2000) suggests that the stress assignment site in Proto-Oceanic languages was not on the syllable but on the *mora*. Blevins & Harrison (1999) also use the mora as a unit of phonological length and propose that the foot of Gilbertese (now Kiribati), another nuclear Micronesian language is

---

16 The large number of Japanese loan words in Pohnpeian was due to relatively recent language contact; Pohnpei was under Japanese administration between World War I and World War II (1917-1945).
trimoraic. Their claim is based on the Proto-Micronesian accent unit, which was reconstructed by Rehg (1984a, 1993), and this trimoraic unit plays an important role in accounting for historical sound changes and morphophonological rules in Pohnpeian. Rehg (1984a) uses the mora to account for a sound change in Micronesian languages and suggests that there was a constraint that noun phrases be minimally trimoraic in Proto-Micronesian.

Very few acoustic investigations have been done on Pohnpeian prosody. Goodman (1995) merely claims that the average duration of long vowels is almost twice that of short vowels. However, she does not provide data to support her claim. Moreover, the focus of her study is not Pohnpeian prosodic structure, and she barely mentions the role of the mora in Pohnpeian phonology. In the following sections, I will discuss the way in which the mora characterizes the sound system in Pohnpeian.

2.3.1 Pseudo-compensatory lengthening

Hayes (1989) defines real compensatory lengthening as “the lengthening of a segment triggered by the deletion or shortening of a nearby segment” (p.260). When the segment /s/ following a vowel before anterior sonorants was deleted in Latin, the vowel became long. The example in (17) is adopted from Ingria (1980 [cited in Hayes, 1989, p.260]).

\[
\begin{align*}
(17) & \quad s \rightarrow \_ \_ \_ [+ \text{son, +ant}] \\
& \quad *\text{kasnus} \quad \rightarrow \quad \text{kamus} \quad \text{‘gray’} \\
& \quad *\text{kosmis} \quad \rightarrow \quad \text{komis} \quad \text{‘courteous’} \\
& \quad *\text{fideslia} \quad \rightarrow \quad \text{fidelia} \quad \text{‘pot’}
\end{align*}
\]
In Pohnpeian, there was a sound change that appeared to be compensatory lengthening. Rehg (1984a), however, points out that this is not compensatory lengthening; rather, this sound change took place to satisfy the constraint that noun phrases had to be minimally trimoric (p.57). Below in (18) are examples adopted from Rehg (1984a, p.53). He evaluates previous analyses and states that the phenomenon was “mistakenly called ‘compensatory lengthening’” (ibid.).

(18) Reconstructed Form | Free Form | Gloss
--- | --- | ---
*kili | ki:l | skin
*seki | se:k | sea
*mwarE | mWar | title
*əsə | əs | thatch
*roŋə | roŋ | news
*u:m\textsuperscript{w}i | u:m\textsuperscript{w} | earth oven

If we look only at these forms, the sound change does look like an instance of compensatory lengthening: the deletion of the final vowel triggered the medial vowel lengthening. However, this sound change can be observed only in nouns whose form was *(C)V(C)V. Thus, the nouns in (19), whose original forms were not *(C)V(C)V, do not exhibit the sound change.

(19) Reconstructed Form | Free Form | Gloss
--- | --- | ---
*nsara | nsar | snore
*ɛmpi | emp | coconut crab
*aramasa | aramas | person
Another piece of evidence comes from Woleian, another nuclear Micronesian language closely related to Pohnpeian. In Woleian, the final vowels are retained as voiceless vowels; nevertheless, vowel lengthening can be observed as in (20).

\[(20) \begin{array}{l}
*\text{laŋo} \quad \text{laŋŋo} \quad \text{‘fly’} \\
*\text{yaŋi} \quad \text{yaŋŋi} \quad \text{‘fire’} \\
*\text{ita} \quad \text{itŋ} \quad \text{‘name’} \\
*\text{mata} \quad \text{matŋ} \quad \text{‘eyes’}
\end{array}\]

Furthermore, Kiribati, another Micronesian language, provides crucial evidence that this diachronic lengthening was a phrase governed phenomenon (Rehg, 1984, p.57), which illustrates the important role of the mora in these Micronesian languages. In Kiribati, vowel lengthening took place not only in noun phrases but also in verb phrases as in (21).

\[(21) \begin{array}{ll}
\text{e piri} & \text{‘He ran’} \\
\text{e nako} & \text{‘He went’} \\
\text{e siku} & \text{‘He stayed’} \\
\text{e kipa} & \text{‘He jumped’}
\end{array} \quad \begin{array}{l}
\text{piri} & \text{‘Run!’} \\
\text{nako} & \text{‘Go!’} \\
\text{siku} & \text{‘Stay!’} \\
\text{kipa} & \text{‘Jump!’}
\end{array}\]

These examples in (21) illustrate the minimal phrasal constraint on verb phrase. In other words, when a verb occurs alone in a phrase, the initial vowel must be long, which makes the verb phrase three moras in length. Therefore, it is plausible to conclude that the vowel lengthening that occurred in Micronesian languages was not real compensatory lengthening, but occurred to satisfy the minimal phrasal constraint. Rehg states, “a constraint existed to the effect that all phrases had to be minimally trimoric. Since all
lexical items belonging to major word classes were apparently minimally disyllabic in
this language (as in Proto-Oceanic), noun phrases were normally at least trimoric as a
consequence of the presence of an article” (p.57). Examples of reconstructed forms of
Proto-Pohnpeic noun phrases are in (22).

(22) *te imw’a  ‘the house’  *i:mw’a  ‘houses’ (generic)
     *te wara ‘the canoe’  *wa:ra  ‘canoes’ (generic)

The lengthening of the phrase initial vowel, Rehg further suggests, might be
triggered by stress assignment in Proto-Micronesian. There are other ways to satisfy the
trimoric phrase constraint: by making the medial consonant geminate, by adding a
prothetic vowel, by lengthening the final vowel, etc. Since it is common that stress is
associated with increasing the duration of the stressed vowel, it is reasonable to assume
that the stress pattern was responsible for the vowel lengthening in Proto-Micronesian.

The sound change I have discussed in this section illustrates the significant role of
the mora as a unit of weight/length in Proto-Micronesian.

2.4 Conclusion

I have provided the background of my study in this chapter. The study of prosody
has become a central key to deepen our understanding of human language. Phonetic
properties of prosodic representations and the way these properties map onto our
phonological knowledge have been waiting for a contemporary investigation.
Phonological categorizations usually discount acoustic details. However, some phonetic
implementations of vowel length must exist. Japanese is a well-studied language;
however, some contradictory claims regarding the functions of the mora are unsolved. Phonological vowel length is manifested by mora count, yet no convincing evidence to support a stable phonetic implementation of the mora has been discovered. I will investigate how Japanese speakers integrate these two phonetic signals, duration and pitch, when they produce long vowels and how these two cues influence Japanese listeners’ perception of vowel length. To my knowledge, only a few studies have investigated the relationship of phonetic implementations with the perception of Japanese vowel length.

Pohnpeian phonetics is almost an unstudied area. I will investigate duration of long and short vowels, and compare the durations with Japanese vowel durations and examine if the difference in phonetic properties makes a difference in vowel length perceptions of Japanese and Pohnpeian speakers.
CHAPTER 3

JAPANESE LONG VOWELS

3.1 Introduction

This chapter will investigate the interaction between duration and pitch in the production of Japanese vowels. Japanese vowels can have two prosodic features – length and pitch-accent. Phonological vowel length is associated with mora count – long vowels carry two moras and short vowels one. As I discussed in section 2.2.2, the isochrony of the mora is still a controversial issue. Nonetheless, the duration of long vowels must be phonetically longer than the duration of short vowels. Hirata (2004) states that the main acoustic correlate of phonological vowel length is duration (p.566).

Pitch-accent is another prosodic feature that vowels can have in Tokyo Japanese. Words in Tokyo Japanese are either accented or unaccented. An accented word contains a high and low pitch sequence (HL) as shown in (1), while an unaccented word does not have the HL pitch sequence as in (2). The vowel carrying the high pitch (H) is called the accented vowel. The primary acoustic realization of pitch-accent is the abrupt downward movement of fundamental frequency (F0).

In order for phonological pitch-accent to be phonetically realized as an abrupt downward movement of F0 in Tokyo Japanese, there must be at least two moras. The
mora thus has another role in Japanese phonology, i.e., a pitch-bearing unit\(^1\) in Japanese. In accented words that are monomoraic or that contain the accented vowel on the last mora, pitch fall is realized only when the words are accompanied by a case marker or other particle as in (1b, 1c, and 1d).

(1) **Accented words**

a. \(\text{hana}-\text{ga}\) \(\text{‘Hana-NOM’}\)  
   \(\text{H L L}\)  
   (girl’s name)

b. \(\text{hanak}-\text{ga}\) \(\text{‘flower-NOM’}\)  
   \(\text{H H L}\)

c. \(\text{ki}-\text{ga}\) \(\text{‘tree-NOM’}\)  
   \(\text{H L}\)

d. \(\text{takar}-\text{ga}\) \(\text{‘treasure-NOM’}\)  
   \(\text{H H H L}\)

(2) **Unaccented words**

a. \(\text{hana}-\text{ga}\) \(\text{‘nose-NOM’}\)  
   \(\text{H H H}\)

b. \(\text{ki}-\text{ga}\) \(\text{‘interest-NOM’}\)  
   \(\text{H H}\)

c. \(\text{miyako}-\text{ga}\) \(\text{‘capital-NOM’}\)  
   \(\text{H H H H}\)

When a long vowel is accented, a H pitch cannot be assigned to both moras in the vowel but only to the first mora, and a L pitch must be assigned to the second mora (McCawley, 1968, p.134). In other words, when a long vowel is accented, the HL pitch sequence must take place within the vowel as in (3a). In a long vowel without pitch-accent, on the other hand, the pitch remains high on both moras of the vowel. In addition,

\(^1\) In the literature, the term tone-bearing unit is generally used as a function of the mora. Since Japanese is a pitch-accent language, not a tone language, the term tone-bearing unit is misleading. Thus, I will use the term pitch-bearing unit in this dissertation rather than the commonly used term tone-bearing unit. Theoretically, the first half of a geminate consonant can bear a pitch. However, to my knowledge, no acoustic investigation has been done on the pitch movement in geminate consonants due to the difficulty in the measurement. It might be a further study to see the interaction between duration and pitch in geminate consonants, in words such as \(\text{kata}\) (HL) ‘shoulder,’ \(\text{kata}\) (HH) ‘style,’ \(\text{katta}\) (HLL) ‘win-PAST,’ and \(\text{katta}\) (HHH) ‘buy-PAST.’
the second mora cannot be followed by a L pitch, as the L pitch makes the second mora of the long vowel accented, and that is ungrammatical in Tokyo Japanese as in (3c). Therefore, accented long vowels in Tokyo Japanese have two phonetic properties – longer duration than short vowels and F0 fall within the long vowels. Unaccented long vowels have only longer duration.

(3) a. pitch contour in an accented long vowel
\[\text{kôôshi} \quad \text{‘lecturer’}\]
\[\text{HH H}\]
\[\text{HL L}\]

b. pitch contour in an unaccented long vowel
\[\text{koo.shi} \quad \text{‘Confucius’}\]
\[\text{HH H}\]
\[\text{HL L}\]

c. ungrammatical pitch contour in an accented long vowel
\[\text{kôô.shi}\]
\[\text{HH L}\]

In this chapter, I will discuss two production experiments that I conducted in order to investigate how Tokyo Japanese speakers use these two phonetic signals when they produce long vowels. First, I investigated differences in duration and pitch-fall between accented long vowels and accented short vowels (Experiment 1). Then, I investigated durational differences between unaccented long vowels and unaccented short
vowels (Experiment 2). Finally I compared the results from the two experiments to examine the effect of pitch-accent on the duration of vowels.

3.2 The effect of pitch-accent on vowel duration

It is a well-accepted fact that stress affects duration, intensity, and vowel quality in English (Fry, 1958). However, there are two contradicting views on Japanese pitch-accent. One group of studies suggests that pitch-accent does not affect vowel duration and the other suggests it does. These two opposing claims are due to the differences in experimental method. The first group includes McCawley (1968), Homma (1973), Beckman (1982a, 1982b), and Larish (1989), while the second includes Han (1962), Hoequist (1983a, 1983b), Kuriyagawa & Sawashima (1987, 1989).

McCawley (1968) states, “The accented mora is characterized solely by its high pitch relative to the following mora; it does not differ in length or intensity from the other moras” (p.135). He does not provide any acoustic evidence for his statement. Presumably, his statement was based on impressionistic observations. Homma (1973) argues that pitch-accent has no influence on vowel duration. She used data collected from a single participant, a native speaker of the Kyoto dialect. Durations of accented and unaccented vowels were measured from vowels within a bimoraic word. In other words, the mean duration of \( V_1 \) and \( V_2 \) in disyllabic words of the shape \( CV_1CV_2 \) were compared. Some example words that Homma used were \( hana \) (HL\(^2\)) ‘flower,’ \( hana \) (HH) ‘nose,’ \( kaki \) (HL)

---

\(^2\) The target words in Homma’s experiment were uttered with a Kyoto accent in which pitch patterns differ from the Tokyo dialect.
‘fence,’ *kaki* (HH) ‘persimmon,’ *kaki* (LH) ‘vase,’ *kaki* (L.HL)³ ‘oyster.’ She found that
the duration of $V_2$ was longer than the duration of $V_1$ for all accent types. Based on this
finding, she concluded that pitch-accent does not affect the duration of vowels. One
phenomenon Homma does not mention is that of phrase final lengthening. Warner & Arai
(2001a) argue that the effect of phrase final lengthening is strong in Japanese. Homma
does no describe how the target words were uttered; however, they were apparently
uttered in isolation. Thus, phrase final lengthening might have affected the duration of $V_2$.
Moreover, adjacent consonants were different between $V_1$ and $V_2$. In fact, Homma also
found an effect of adjacent consonants on vowel duration in her data (pp.360-361). Comparing the duration of $V_1$ and $V_2$ might not have accurately represented the effect of
pitch-accent on the duration of vowels.

Unlike Homma, Beckman (1982a) did not depend on data from a single
participant. Beckman used several native speakers of Tokyo Japanese, and the target
words were uttered in a carrier sentence *Sosite, to iimasu* ‘Then, I'll say ____.’
However, she conducted measurements similar to Homma's; that is, comparing the mean
duration of $CV_1$ with $CV_2$ in disyllabic words of the shape $CV_1CV_2$. This method can be
used to compare the durations of moras, but is not suitable to examine the effect of pitch-
accent on vowel duration. Since the positions of the target vowels $V_1$ and $V_2$ in a word
are different, we cannot assess the effect of pitch-accent on the duration of vowels by
comparing the duration of $CV_1$ and $CV_2$.

³The sequence of HL realized in the final vowel of this word *kaki* ‘oyster’ is considered a contour pitch.
How many moras the second vowel bears is still an unsolved issue.
The method of the experiments that Larish (1989) used was different. However, the material words used to examine the influence of pitch-accent on duration were only two minimal pairs, contrasting in pitch-accent. These were *koko* (HL) ‘houses’ vs. *koko* (HH)4 ‘here’ and *kookoo* (HL) ‘filial piety’ vs. *kookoo* (HH) ‘highschool’ in (4a). These words were read in a carrier sentence *Sono ___ desu ‘It’s that ___ ’* at a natural speaking rate by six native speakers of Tokyo Japanese (three female and three male). Larish compared various types of ratios of the duration of accented syllables to unaccented syllables as in (4b) and found that the effect of pitch-accent on duration was negligible.

(4) a. target words

<table>
<thead>
<tr>
<th>accented words</th>
<th>unaccented words</th>
</tr>
</thead>
<tbody>
<tr>
<td>short vowel</td>
<td></td>
</tr>
<tr>
<td><em>koko</em> (H) ‘houses’</td>
<td><em>koko</em> (L5) ‘here’</td>
</tr>
<tr>
<td>long vowel</td>
<td></td>
</tr>
<tr>
<td><em>kookoo</em> ‘filial piety’</td>
<td><em>kookoo</em> ‘highschool’</td>
</tr>
</tbody>
</table>

b. comparisons

1. \[ko(H)/ko(L)\] ↔ \[koo(HH)/koo(HL)\]
2. \[koo(HL)/ko(L)\] ↔ \[koo(HH)/ko(L)\]
3. \[koo(HL)/ko(H)\] ↔ \[koo(HH)/ko(H)\]
4. \[koo(HL)/ko(L)\] ↔ \[koo(HH)/ko(L)\]
5. \[koo(HL)/ko(H)\] ↔ \[koo(HH)/ko(L)\]

The ratios of the comparison in (4b-i), the mean duration of accented short vowels and unaccented short vowels and the mean duration of accented long vowels and unaccented long vowels, were 1.01 : 1.00 and 1.05 : 1.00 respectively. Although Larish

---

4 In Larish (1989) the pitch pattern of this word is coded as LH and the initial long vowel in the word *kookoo* ‘highschool’ is LH, as the L occurs at the word initial.
5 I use “L” to mark the unaccented vowel in the first syllable in this unaccented word for clarification.
found durational increases in accented syllables, he argues that the influence of pitch accent on duration is “linguistically insignificant” (p.114).

Another group of studies, Han (1962), Hoequist (1983a, 1983b), Kuriyagawa & Sawashima (1987, 1989), suggests an effect of pitch-accent on vowel duration. The methods that this group of studies used are better for investigating the effect of pitch-accent on vowel duration than the methods used in the former group of studies.

In her study, Han (1962) compared the duration of accented vowels with the duration of unaccented vowels from sets of minimal pair words uttered “in various ways and by a number of native speakers” (p.104). For example, the duration of the vowel /a/ in hashi (HL) ‘chopsticks’ was compared with the duration of the vowel /a/ in hashi (HH) ‘bridge.’ She found that presence of pitch-accent slightly increased intensity and duration of vowels.

Hoequist (1983a, 1983b) found that pitch-accent has a significant effect on syllable duration, although he expresses some doubts about the perception of the durational differences. The words used in his study were of various lengths, and the words included all five Japanese phonemic vowels. The participants read the target words in a carrier sentence kinoo ___ ga kita ‘Yesterday ___ arrived’ at a comfortable speech rate. There were five participants who were Tokyo Japanese speakers. The mean duration of accented syllables was only 1.02 times longer than that of unaccented syllables (1983b, p.210). Hoequist found that pitch-accent had a significant effect on syllable duration [F(1,4) = 9.47, p < .05], when he combined the duration of light syllables and heavy syllables (1983a, p.26) and compared the mean duration of accented
syllables and unaccented syllables. However, he suggests that durational increase in accented vowels does not play a role in the perception of Japanese pitch-accent.

Kuriyagawa & Sawashima (1987) propose that vowel duration increases significantly when vowels carry pitch-accent. They conducted a production experiment, specifically designed to investigate the effect of pitch-accent on the duration of vowels. They used four sets of disyllabic minimal pair words contrasting different pitch patterns, pitch-accent on the first mora (\(u_1\)) vs. pitch-accent on the second mora (\(u_2\)). They used only a high-back short vowel /u/ as the target vowel as in (5).

(5) Accent on the first mora  |  Accent on the second mora  
---|---
\(Cu_1Cu_2\)  |  \(Cu_1Cu_2\)  
\(humu\) (interjection)  |  \(humu\) ‘to step’  
\(sumu\) ‘to live’  |  \(sumu\) nonsense word  
\(tsumu\) ‘checkmate’  |  \(tsumu\) ‘to pick’  
\(kumu\) ‘to fold’  |  \(kumu\) ‘to ladle out’

A single male participant read material words in two different carrier sentences, \(tsugi-wa \__ \) \(daroo\) ‘the next will be \__ ’ and \(tsugi-wa \__ \) \(to-iu\) ‘the next is said \__ ’, at two different speech rates, fast and slow. They compared the mean durations of accented vowels in the first syllable (accented \(u_1\)) and unaccented vowels in the first syllable (unaccented \(u_1\)), the mean durations of accented first syllables (accented \(Cu_1\)) and unaccented first syllables (unaccented \(Cu_1\)), the mean durations of unaccented vowels in the second syllable (unaccented \(u_2\)) and accented vowels in the second syllable (accented \(u_2\)), and the mean durations of unaccented second syllables (unaccented \(Cu_2\)) and accented second syllables (\(Cu_2\)). They found that regardless of the type of carrier
sentence, the mean durations of accented u₁ and accented Cu₁ were significantly longer than the mean durations of unaccented u₁ and unaccented Cu₁ respectively. A similar pattern held when pitch-accent was on the second vowel u₂. That is, the mean durations of accented u₂ and accented Cu₂ were longer than the mean durations of unaccented u₂ and unaccented Cu₂ respectively. Whether durational increase is perceptible remains a question; nonetheless, pitch-accent seems to have an effect on vowel duration in Japanese.

3.3 Experiment 1: Production experiment with Japanese accented vowels

The first experiment of this study is a production experiment using Japanese accented vowels. Two phonetic properties, the duration that characterizes vowel length and the pitch fall that characterizes pitch-accent, were measured from Japanese accented vowels uttered at different speech rates.

It is a widely known fact that speech rate affects production and perception of speech sounds in various ways at the segmental level. Articulatory gestures tend to compress or overlap more in fast speech, resulting in the reduction of certain segments. For example, in English, the phrase-medial coronal [t] may be deleted or not be heard in the phrase *perfect memory* (Browman & Goldstein, 1992); the phrase *six sheep* may be heard without phrase-medial /s/ in the word *six /siks* as [sǐkʃip] (Donegan, 2001); and the sentence *Did you eat?* may be heard as *Jew eat?* [dʒu iːt]. Vowel devoicing in Japanese is more likely to occur in fast speech (Han, 1962, pp.23-25; Vance, 1987, p.49, p.55). At the segmental level, Browman & Goldstein (1992) report that articulatory gestures are present even if the coronal [t] is not heard in the phrase *perfect memory*. In
other words, the speakers are not always deleting entire phonetic signals of unheard segments even in fast speech.

What is happening at the prosodic level? Does reduction of prosodic features also occur, when the speech rate increases? In the case of Japanese accented long vowels, the speaker must maintain the relative durational distinction between a long vowel and a short vowel in order to pronounce the intended word correctly at various speech rates. Likewise, the speaker needs to signal pitch-accent even in fast speech; otherwise the utterance may become ambiguous. In other words, these two prosodic features – vowel length and pitch-accent – are lexical; if the speaker does not maintain these contrasts, lexical information may be lost. Thus, these prosodic features in Japanese accented long vowels must be retained. The question is how speakers maintain these prosodic properties in fast speech. The purpose of Experiment 1 is to investigate the way native speakers of Tokyo Japanese maintain the phonological distinctions of vowel length and pitch-accent at various speech rates. The results of the experiment indicate that speakers use not only duration but also F0 information to produce accented long vowels.

3.3.1 Participants

Two female and two male native speakers of Tokyo Japanese participated in Experiment 1. Their parents were also either native speakers of this dialect or lived in a Tokyo Japanese-speaking community most of their lives. The participants’ ages ranged from the late-20s to the mid-30s. One female participant was a professor at the University of Hawai‘i at Mānoa, and the other female participant was a graduate student in the Department of Linguistics at the University of Hawai‘i at Mānoa. One male participant
was a Japanese instructor at the University of Hawai‘i at Mānoa, and the other male participant was a student at Kapi‘olani Community College in Honolulu, Hawai‘i. Three of the four participants have experience teaching Japanese as a foreign language to adults. They had all lived in an English-speaking community for several years; however, the author did not observe any indication of English influence in their Japanese pronunciation. The participants had several opportunities to use Japanese on a daily basis. They all participated in the experiment voluntarily.

3.3.2 Materials

The test words used for the data collection were 25 minimal pairs contrasting in vowel length: 5 words for each of the 5 Japanese vowels. Examples of test words are shown in Table 3.1 (see Appendix A for the entire word list).

<table>
<thead>
<tr>
<th>Target vowel</th>
<th>Long vowel</th>
<th>gloss</th>
<th>Short vowel</th>
<th>gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/</td>
<td>kaado</td>
<td>‘card’</td>
<td>kado</td>
<td>‘corner’</td>
</tr>
<tr>
<td>/e/</td>
<td>beeru</td>
<td>‘veil’</td>
<td>beru</td>
<td>‘bell’</td>
</tr>
<tr>
<td>/i/</td>
<td>biiru</td>
<td>‘beer’</td>
<td>biru</td>
<td>‘building’</td>
</tr>
<tr>
<td>/o/</td>
<td>rooba</td>
<td>‘elderly woman’</td>
<td>roba</td>
<td>‘donkey’</td>
</tr>
<tr>
<td>/u/</td>
<td>kuuru</td>
<td>‘cool’</td>
<td>kuru</td>
<td>‘to come’</td>
</tr>
</tbody>
</table>

The majority of phonetic studies investigating the duration of the mora have used a single frame sentence for all target words, which might induce the participant to read the sentences mechanically or to subconsciously put an emphasis on the target words. In

---

The items used in an experiment to collect data or test the participants’ production/perception processes for phonetic/psycholinguistic analysis are called “materials.”

67
order to minimize participants’ attention to target words, a different frame sentence was composed for each minimal pair. In other words, each pair was actually a minimal pair sentence, such as  *kawaii beeru dane* (with a long vowel) ‘What a pretty veil!’ vs.  *kawaii beru dane* (with a short vowel) ‘What a cute bell!’, and  *kono biru ga takai* ‘This beer is expensive’ vs.  *kono biru ga takai* ‘This building is tall.’ In each pair, only the length of the target vowel contrasted, yet the meaning of the sentences differed. All sentences were both syntactically and semantically well-formed. In addition, there was no ambiguity in material sentences.\(^7\)

Since retention of pitch-accent in fast speech was one of the foci of this study, accented words were chosen for the test words. Most material words were disyllabic, and the target vowels were always in the first syllable. Thus, the first mora of the material words was accented. For example, in a material word containing a short vowel, a high pitch (H) occurred on the target vowel and a low pitch (L) on the vowel in the following syllable such as  *be.ru* (HL). In a material word containing a long vowel, the HL pitch sequence was within the target vowel such as  *bee.ru* (HL.L).

Measurements were done on 22 of 25 words with a long vowel and 17 of 25 words with a short vowel. Some materials had to be excluded, since some target vowels were devoiced or read with different pitch patterns from the expected HL pitch contour. For example, the target vowel in the word  *kuki* ‘stem’ was devoiced. In the minimal pair  *ojiisan* ‘grandfather’ vs.  *ojisan* ‘uncle’, pitch-accent is on the target long vowel /i:/ in the word  *ojiisan*, but it is on the last syllable rather than on the target vowel /i/ in the word

\(^7\) I am grateful to Michiko Nakamura and Naoko Takahashi for their help in checking grammaticality of the material sentences.
The materials that were not used for analyses are indicated by an asterisk ‘*’ in the list in Appendix A.

The material sentences were written in Standard Japanese orthography, which is a mixture of kanji and kana; roman letters were used where they were appropriate, such as in the sentence sakki aaru to iimashita ‘I said R just now.’ In order to make the distinction easier, each sentence was read at three different speech rates – fast, normal, and slow, rather than in a two-way contrast – fast vs. slow. The order of material sentences was controlled, so that the participants did not read the same sentence repeatedly at different speech rates. For example, participants did not read a sentence at a fast speech rate immediately followed by the same sentence at another speech rate. The desired reading rate was written in the upper left corner of each sheet in a smaller font size than the one used for the material sentences, and described as dekirudake hayakuchi-de ‘as quick as you can read’ for a fast speech rate, futsuno hayasa-de ‘at normal speed’ for a normal speech rate, and otosiyori-ni hanasuyoo-ni ‘as if you are speaking to an elderly person’ for a slow speech rate. Since a pause influences the pitch pattern of a sentence (Pierrehumbert, 1980; Pierrehumbert & Beckman, 1988), the participants were instructed to read the materials without inserting any pauses into the sentences, particularly in slow speech. Before the participants started reading the sentences, they were asked to pay particular attention to the reading rate, which additionally diverted their attention from the target words. When the participant or the investigator was not satisfied with the way a material sentence was read, he/she repeated it before proceeding to the next sentence.
3.3.3 Procedure

The participants read the material sentences in a sound-attenuated recording studio on the University of Hawai‘i, Mānoa campus. The utterances were recorded on a TANDBERG cassette-recorder through a 3M tabletop microphone and digitized using Pitchworks at a sampling rate of 11,025 Hz. PC quirer was also used to detect formant movement when necessary.

All measurements were taken from the digitized files. The duration of each sentence, word, and target vowel was measured to ensure the material sentences were read at the desired speech rates. All durations of a sentence, word, and vowel in slow speech were longer than the durations of the same sentence, word, and vowel read as normal speech or fast speech. Likewise, all durations of a sentence, word, and vowel in normal speech were longer than the durations of the same sentence, word, and vowel in fast speech. The target measurements were duration of vowels and pitch fall. The beginning and ending points of vowels were determined by presence of the second formant (F2) and higher formants (Figure 3.1).

\[ /a:/ \]

\[ \text{sono} \quad \text{kaado} \quad \text{fuite.} \quad \text{‘Please clean the card.’} \]

**Figure 3.1:** Measurement points for vowel duration
Although pitch fall takes place within accented long vowels, pitch fall in accented short vowels occurs in the interval between the accented vowel and the following syllable. For consistent measurements, I measured the difference between the maximum F0 value in a target vowel and the minimum F0 value in the following syllable as the pitch fall in both short and long vowels\(^8\) (Figure 3.2).

\[\text{Figure 3.2: Measurement points for pitch fall}\]

Although the measurements were taken from all of the utterances at the three different speech rates, in order to see the effect of speech rate clearly, I used data only from two extremes – fast and slow speech.

\(^8\) However, since adjacent consonants influence the value of F0, the F0 maximum was determined by searching for the highest F0 point that was closest to the center of the target vowel and the F0 minimum was the lowest point that was closest to the center of the vowel in the following syllable.
3.3.4 Hypotheses

I constructed the following hypotheses. Hypotheses (H-1) and (H-2) concern vowel length and (H-3) and (H-4) concern pitch-accent.

(H-1) Because vowel length is contrastive in Japanese, the mean duration of long vowels will be significantly longer than that of short vowels, regardless of speech rate.

(H-2) Speech rate will not have any effect on the distinction between long and short vowels. In other words, the ratio of the mean duration of long vowels to short vowels will not change significantly, regardless of speech rate. For example, if the ratio of the mean duration of long vowels to short vowels is approximately 1.8:1 in fast speech, the ratio in slow speech will also be approximately 1.8:1.

(H-3) Since the test words are accented, the amount of pitch fall in long vowels will be identical to the amount of pitch fall in short vowels, regardless of speech rate. I assume that F0 value must fall by a certain degree to mark pitch-accent. Once this degree of F0 fall occurs, it is not necessary to fall any further whether the vowel is long or short. Therefore, contrary to duration, the amounts of pitch fall in long vowels and short vowels will not be significantly different.

(H-4) Speech rate will not have a significant effect on the amount of pitch fall. Since pitch-accent is a distinctive feature in Japanese, in order for the speakers to produce pitch-accent correctly, the speaker must maintain the
amount of pitch fall. Thus, the amount of pitch fall will be stable, regardless of speech rate in both long and short vowels.

3.3.5 Data analyses

In this experiment, there were two sets of independent variables\(^9\): speech rate (Fast vs. Slow) and vowel length (Long vs. Short), creating a “2x2 within-subject design.” I took multiple measurements from four different speakers and from five different vowels; therefore, a repeated measures two-way analysis of variance (2-way ANOVA) is the most suitable statistical analysis for experiments like this one. The dependent variables\(^10\) in Experiment 1 were vowel duration and amount of pitch fall. I performed a repeated measures 2-way ANOVA for the two independent variables, speech rate and vowel length. I analyzed the data treating speakers (participants) as a sample of the population. In addition, since I used five different vowels, I performed another repeated measures 2-way ANOVA treating vowel type (the five Japanese vowels /a, e, i, o, u/) as a sample of the population to see if there would be any differences in duration and pitch fall dependent on vowel type.\(^11\)

\(^9\) Independent variable = “the variable that the experimenter manipulates in order to explain the differences in the dependent variable or to cause changes in the dependent variable” (Runyon R.P. et al., 2000, p.4).

\(^10\) Dependent variable = “an outcome of interest that is observed and measured by the researcher in order to assess the effects of the independent variable” (ibid., p.4).

\(^11\) Since the vowel duration would be influenced by adjacent consonants, the data have to be sorted out by the type of adjacent consonants for more precise analyses. However, since I used minimal pair words as the materials of the experiment, presumably that influence would not be significant for the present analyses. Future research should investigate the influence of the consonants on the duration of vowels. I am grateful to Ian Maddieson at the 30th conference of the Berkeley Linguistics Society in February, 2004 and to Yoshinori Sagisaka at the 2nd International Conference: Speech Prosody 2004, in March 2004 for their helpful comments and suggestions on this issue.
3.3.6 Results

Mean durations of vowels and ratios of the mean duration of long vowels to short vowels are shown in Tables 3.2 and 3.3.

Table 3.2: Analysis by speakers: Mean durations of accented vowels and ratios in Experiment 1

<table>
<thead>
<tr>
<th></th>
<th>Long vowel (ms)</th>
<th>Short vowel (ms)</th>
<th>Ratio (Long : Short)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast Speech</td>
<td>119.194</td>
<td>72.104</td>
<td>1.65 : 1.00</td>
</tr>
<tr>
<td>Slow Speech</td>
<td>176.988</td>
<td>100.167</td>
<td>1.77 : 1.00</td>
</tr>
<tr>
<td>Ratio (Fast : Slow)</td>
<td>1:00 : 1.48</td>
<td>1.00 : 1.39</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.3: Analysis by vowel types: Mean durations of accented vowels and ratios in Experiment 1

<table>
<thead>
<tr>
<th></th>
<th>Long vowel (ms)</th>
<th>Short vowel (ms)</th>
<th>Ratio (Long : Short)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast Speech</td>
<td>118.790</td>
<td>71.650</td>
<td>1.65 : 1.00</td>
</tr>
<tr>
<td>Slow Speech</td>
<td>176.541</td>
<td>98.750</td>
<td>1.79 : 1.00</td>
</tr>
<tr>
<td>Ratio (Fast : Slow)</td>
<td>1.00 : 1.49</td>
<td>1.00 : 1.38</td>
<td></td>
</tr>
</tbody>
</table>

The results of repeated measures 2-way ANOVAs showed that the mean duration of long vowels was significantly longer than the mean duration of short vowels ([F(1,3) = 1377.831, p < .0001] by speakers and [F(1,4) = 442.154, p < .0001] by vowel type).

---

12 I set a significance level at .05; that is, if \( p \) is smaller than .05, we can reject the null hypothesis. A null hypothesis is a statement that specifies hypothesized values for one or more of the population parameters. In the case of this experiment, there were two null hypotheses, one for each factor: 1) speech rate does not affect the mean duration of vowels and 2) vowel length does not affect the mean duration of vowels.

13 \( F = F\text{-Ratio} \), indicates if the difference among the means is greater than chance. If the F-Ratio is not significantly greater than 1.0 we must assume that the treatment effect had no statistically significant effect on the date. By contrast, if the F-Ratio is greater than 1.0 we assume that part of the variance in the data is due to random effects, but that a greater part is due to the treatment effect (Runyon, R. P. et al. 2000, p.413). The numbers in parentheses are the degrees of freedom for the two parts of the ratio. The degrees of freedom are always 1 less than the number of observations contributing the variance estimate (ibid., p.293). In the case of this analysis, the first value is 2 minus 1 since there are two factors (e.g., fast vs. slow), and the second 4 minus 1 since there are four speakers.
types). However, although Japanese long vowels phonologically carry two moras, phonetically the ratio of the mean duration of long vowels to short vowels was not 2 : 1.

Speech rate had a significant main effect on the duration of vowels ([F(1,3) = 66.729, \( p = .0038 \)] by speakers and [F(1,4) = 125.560, \( p = .0004 \)] by vowel types).

![Interaction Line Plot for Duration](image)

**Figure 3.3:** Analysis by speakers: Interaction line plot for duration in Experiment 1

In addition, there was a significant interaction\(^{14}\) between two independent variables, speech rate and vowel length ([F(1,3) = 12.463, \( p = .0386 \)] by speakers and [F(1,4) = 94.394, \( p = .0006 \)] by vowel types). Figure 3.3 indicates that the effect of speech rate on the duration of short vowels and long vowels was not the same. That is, the duration of both types of vowels was lengthened in slow speech; however, long vowels and short vowels were not lengthened by the same amount. The duration of long vowels in slow speech was about 1.50 times longer than that in fast speech, while the

\(^{14}\) An interaction indicates that “the effects of the independent variables are not consistent across all treatment conditions” (ibid., p.406).
duration of short vowels in slow speech was only about 1.40 times longer than that in fast speech. The duration of long vowels was always significantly longer than that of short vowels, and the duration of both lengths of vowels was lengthened in slow speech. However, the degree of lengthening of long vowels in slow speech was greater than the degree of lengthening of short vowels in slow speech. That is, the ratio between fast speech and slow speech showed that long vowels were more sensitive to the speech rate. The speech rate had a stronger effect on the duration of long vowels than on the duration of short vowels.

Let me turn to pitch fall. I also performed a repeated measures 2-way ANOVA to evaluate the effect of speech rate on pitch fall. Tables 3.3 and 3.4 show the mean values of pitch fall and ratios of the mean values of pitch fall in long vowels versus short vowels.

**Table 3.4:** Analysis by speakers: Mean pitch falls of accented vowels and ratios in Experiment 1

<table>
<thead>
<tr>
<th></th>
<th>Long vowel (Hz)</th>
<th>Short vowel (Hz)</th>
<th>Ratio (Long : Short)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast Speech</td>
<td>82.433</td>
<td>61.573</td>
<td>1.34 : 1.00</td>
</tr>
<tr>
<td>Slow Speech</td>
<td>96.332</td>
<td>68.329</td>
<td>1.41 : 1.00</td>
</tr>
<tr>
<td>Ratio (Fast : Slow)</td>
<td>1.00 : 1.17</td>
<td>1.00 : 1.11</td>
<td></td>
</tr>
</tbody>
</table>

**Table 3.5:** Analysis by vowel types: Mean pitch falls of accented vowels and ratios in Experiment 1

<table>
<thead>
<tr>
<th></th>
<th>Long vowel (Hz)</th>
<th>Short vowel (Hz)</th>
<th>Ratio (Long : Short)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast Speech</td>
<td>83.868</td>
<td>63.164</td>
<td>1.33 : 1.00</td>
</tr>
<tr>
<td>Slow Speech</td>
<td>97.861</td>
<td>70.722</td>
<td>1.38 : 1.00</td>
</tr>
<tr>
<td>Ratio (Fast : Slow)</td>
<td>1.00 : 1.17</td>
<td>1.00 : 1.12</td>
<td></td>
</tr>
</tbody>
</table>
Vowel length had a significant main effect on the pitch fall ([F(1,3) = 22.058, 
p = .0183] by speakers and [F(1,4) = 17.177, p = .0143] by vowel types). Speech rate also
had a significant main effect on the pitch fall ([F(1,3) = 16.824, p = .0262] by speakers
and [F(1,4) = 14.899, p = .0181] by vowel types). These results indicate that the mean
value of pitch fall in long vowels was significantly greater than that in short vowels, and
the mean value pitch fall in slow speech was significantly greater than that in fast speech.

However, there was no significant interaction between the two independent
variables, speech rate and vowel length ([F(1,3) = 8.093, p = .0654] by speakers and
[F(1,4) = .715, p = .4455] by vowel types) (Figure 3.4). That is, the amount of pitch fall
in long vowels and short vowels was maintained in both fast speech and slow speech.
Unlike the results seen for vowel duration, the way speech rate affected the pitch fall was
similar for long vowels and short vowels.

Figure 3.4: Analysis by speakers: Interaction line plot for pitch fall in Experiment 1
An interesting result found in the data was that amount of pitch fall was much greater in female utterances than male utterances. Moreover, female participants produced greater pitch fall in fast speech than slow speech, particularly the speaker F-1 produced much greater pitch fall in short vowels.

Table 3.6: Mean pitch fall by individual participants

<table>
<thead>
<tr>
<th>Participant</th>
<th>Long vowel (Hz)</th>
<th>Short vowel (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fast</td>
<td>Slow</td>
</tr>
<tr>
<td>F-1</td>
<td>136.9</td>
<td>134.6</td>
</tr>
<tr>
<td>F-2</td>
<td>95.5</td>
<td>106.3</td>
</tr>
<tr>
<td>M-1</td>
<td>56.1</td>
<td>68.3</td>
</tr>
<tr>
<td>M-2</td>
<td>52.8</td>
<td>68.3</td>
</tr>
</tbody>
</table>

Figure 3.5: Mean pitch fall by individual participants
3.3.7 Discussion

As I discussed in Chapter 2, the isochrony of the Japanese mora has been an interest of phoneticians since Bloch (1950). Bloch and Han (1962) suggest that Japanese moras are of about equal duration. However, acoustic investigations reveal that long vowels are not twice as long as short vowels (Campbell & Sagisaka, 1991 cited in Warner & Arai 2001b). That is, Japanese long vowels are not simply a combination of two identical short vowels. The present experiment, Experiment 1, supported this claim. The mean duration of long vowels was 1.65 times longer than the mean duration of short vowels in fast speech and about 1.77 times longer in slow speech. In both speech rates, the mean duration of long vowels was not twice as long as short vowels.

However, the mean duration of long vowels was always longer than that of short vowels, regardless of speech rate, as I predicted in (H-1) in section 3.2.4. Furthermore, the duration of vowels was longer in slow speech and shorter in fast speech as I expected in (H-2). Speech rate had an effect on both long vowels and short vowels. Contrary to (H-2), however, the ratio of the duration of long vowels to short vowels varied depending on the speech rate. That is, the ratio of the mean duration of long vowels to short vowels is greater in slow speech than in fast speech. The duration of long vowels is more susceptible to changes in speech rate than the duration of short vowels. This may explain the different ratio of the duration of long vowels to short vowels suggested by different studies (Warner & Arai, 2001b). For example, Campbell & Sagisaka (1991) use a corpus of reading data, Hoequist (1983a) reports a ratio of 1.71:1 for the duration of CVV to CV words, but it was close to 2:1 in subsequent work with reiterated speech (Hoequist, 79
Each study was based on a different set of data, and so each set of data was collected in a different way. In most cases, speech rate was determined by each individual who participated in an experiment. Therefore, it is not surprising to see different speakers produced vowels with various durations. In fact, results of Experiment 1 show that the mean duration of long vowels varied from 107.430 ms to 130.667 ms in fast speech and from 161.322 ms to 195.150 ms in slow speech; the mean duration of short vowels varied from 63.178 ms to 80.639 ms in fast speech and from 86.276 ms to 117.724 ms in slow speech depending on the speaker. Furthermore, the duration of long vowels is more sensitive to speech rate than the duration of short vowels; thus, it is not surprising to see differences in the ratio, calculated based on data elicited from presumably different speech rates. Minagawa et al. (2003) examined a spontaneous speech corpus and reported that the ratio of the mean duration of long vowels to short vowels increases in slower speech. Hirata (2004) specifically investigated the effect of speech rate on the vowel length distinction in Japanese. She found that the effect of speech rate on the duration of long vowels was greater than that of short vowels. Results of Experiments 1 support these previous findings.

In phonological terms, there are two moras in long vowels and one in short vowels. Nevertheless, as far as actual the duration of vowels is concerned, the phonological function of the mora as a unit of length is not supported phonetically, as the mean duration of long vowels was not twice as that of short vowels. Moreover, the ratio of the duration of long vowels to short vowels is not stable across speech rates.
In contrast, the speakers maintained a certain amount of pitch fall when they produced accented vowels regardless of speech rates, as I expected in (H-4). Vowel length made a difference in the amount of pitch fall, contrary to what I expected in (H-3). Speakers produced greater pitch fall in long vowels than short vowels. However, there was no significant interaction between the two independent variables, speech rate and vowel length. The effect of speech rate on the pitch fall of long vowels and short vowels was constant.

The results of Experiment 1 show the way in which Tokyo Japanese speakers use duration and pitch fall to distinguish vowel length in accented vowels. The speakers correctly produce not only a durational distinction but also a certain amount of pitch fall in accented vowels, regardless of speech rate. The difference in the behavior of these two prosodic characteristics in accented vowels is that the relative duration of vowels is sensitive to speech rate but pitch fall is not. Duration of the vowels fluctuates widely, but pitch fall does not seem to be as sensitive as duration to speech rate. Pitch fall is also greater in long vowels than short vowels, but the speakers seem to maintain a certain amount of pitch fall across various speech rates. The way female speakers in particular produced pitch fall indicates that speakers rely more on pitch fall to mark vowel length than on duration, as duration is much more vulnerable to speech rate. These results raise the following questions:

First, how do Japanese speakers distinguish vowel length in unaccented vowels across various speech rates? Japanese speakers must pay attention both to appropriate duration and to appropriate pitch fall for accented vowels. However, they should not
produce pitch fall for unaccented vowels. They must only use duration to distinguish long and short vowels. Is the duration of unaccented vowels as sensitive to speech rate as the duration of accented vowels? If so, is there any difference in the way speakers distinguish vowel length between accented vowels and unaccented vowels? Experiment 2 will investigate how the speakers distinguish vowel length in unaccented vowels by using duration at various speech rates.

3.4 Experiment 2: Production experiment with Japanese unaccented vowels

Results of Experiment 1 showed that when Japanese speakers produced accented long vowels, they produced both longer duration and greater F0 fall than accented short vowels. In addition, the duration of vowels was sensitive to speech rate changes. The ratio of the mean duration of long vowels to short vowels varied across various speech rates. If the speakers indeed use both duration and pitch fall to distinguish vowel length of accented vowels, the speakers will be more careful producing durational distinction between long and short in unaccented vowels, as they must rely only on duration. In other words, there are two phonetic properties to denote vowel length in accented vowels – duration and pitch fall, while only one in unaccented vowels – duration. The duration of unaccented vowels should be more stable across various speech rates than that of accented vowels. This section will test this hypothesis.

3.4.1 Participants

There were four participants in this experiment (two female and two male). They were different participants from Experiment 1, except for one male participant. They
were all native speakers of Tokyo Japanese and their parents were either native speakers of Tokyo Japanese or had lived in this dialect-speaking community most of their lifetime. Therefore, I assume that there was no influence from other dialects of Japanese. Both male participants were Japanese instructors at the University of Hawai‘i, Mānoa campus. The female participants were both Ph.D. students at the University of Hawai‘i at Mānoa. The age range of the participants was from the late 20s to the early 40s. They voluntarily participated in the experiment, although they were offered $5.00 to compensate them for their time.

3.4.2 Materials

The materials used in this experiment were words containing unaccented vowels in Tokyo Japanese. I used 18 minimal or near minimal pairs contrasting in vowel length; at least two minimal pairs for each of the 5 Japanese vowels. Examples are se.koo (H.HH) ‘construction’ vs. see.koo (HH.HH)⁴⁵ ‘success’ and sha.ku (HH) ‘the Japanese foot (a unit of length)’ vs. jaa.ku (HH.H) ‘evil’. Since it is difficult to find unaccented minimal pairs, I also used 9 words containing unaccented long vowels such as su.ldi.joo (H.HH.HH) ‘a ski ground’ and yuu.kai (HH.HH) ‘kidnap.’ I used some long vowels that

⁴⁵The accent patterns for these lexical items may differ from dictionary entries, because when a syllable with a long vowel occurs initially in an accentual phrase (AP), the first mora is always realized with a low phrasal accent. Consequently, it creates a LH pitch-contour within the long vowel, which is the accent pattern shown in the dictionary. However, since these words were read in a carrier sentence, syllables with a long vowel did not occur AP initially; thus, the pitch-contour stayed HH, which was supported by Fujimura (personal communication, March 24, 2004). Moreover, McCawley (1968) states that in a widespread variety of Tokyo Japanese, “an unaccented initial syllable of the form CVV or CVN is pronounced entirely on a high pitch” (p.133). According to the measurements in Weitzman (1969) the difference in the mean F0 value between initial L and following H in unaccented words (LHH) is 8.9 Hz (18 tokens), whereas the mean value of pitch fall, the difference between H and the final L, in accented words (LHL) is 65.7 Hz (18 tokens). It would be interesting to investigate how Tokyo Japanese speakers perceive this LH pitch contour (only 8.9 Hz difference).
cross over morpheme boundaries,\(^{16}\) such as *aka.ri* (H.HH.H), ‘red ant,’ *aka* ‘red’ + *ari* ‘an ant,’ and *oki.shi* (H.HH.H) ‘milestone,’ *oki* ‘to place’ + *ishi* ‘stone.’ The accent patterns were checked against the *Nihongo Accent Jiten* ‘Dictionary of Japanese Accent’ (Kindaichi, 1966) (see Appendix B for the entire word list). Because of the small number of minimal pairs with unaccented vowels, I was not able to compose enough minimal pair sentences for this experiment. Consequently, all material words were read in the carrier sentence *Ima ____ to iimashita* ‘I said ____ now.’ The sentences were written in standard Japanese orthography.

### 3.4.3 Procedure

In order to ensure consistent results, the procedure for the present experiment was identical to Experiment 1. Participants read the material sentences at three different speech rates in a randomized order and their utterances were recorded in a sound attenuated recording studio with a TANDBERG TCR522 cassette-recorder through a 3M tabletop microphone and digitized using Pitchworks at a sampling rate of 11,025 Hz.

### 3.4.4 Hypotheses

I constructed the following hypotheses:

\((H-5)\) The mean duration of long vowels will always be significantly longer than the mean duration of short vowels, regardless of speech rate.

---

\(^{16}\) Although these words were made up of multiple morphemes, most of them are totally lexicalized as a single word; for example, *keshiin* ‘postmark’ and *okiishi* ‘milestone.’ Vance (1987) reports that sometimes insertion of a glottal stop might occur between vowels belonging to different morphemes (p.14). I checked for spectral cues at morpheme boundaries and confirmed that there were no glottal stops in my data.
(H-6) The mean duration of unaccented vowels in fast speech will be shorter than that in slow speech.

(H-7) If speakers use duration and pitch fall to mark vowel length in accented vowels, they will be more cautious with producing durational distinctions when they produce unaccented vowels, as they cannot use pitch fall. They must maintain a certain amount of durational distinction between long vowels and short vowels, regardless of speech rate. Therefore, the interaction between the effect of speech rate and vowel length will not be significant for unaccented vowels.

### 3.4.5 Results

Results of statistical analyses of Experiment 2 were not largely different from the results of Experiment 1. The mean durations of unaccented vowels and their ratios are in Tables 3.6 and 3.7 below.

**Table 3.7: Analysis by speakers: Mean durations of unaccented vowels and ratios in Experiment 2**

<table>
<thead>
<tr>
<th></th>
<th>Long vowel (ms)</th>
<th>Short vowel (ms)</th>
<th>Ratio (Long : Short)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast Speech</td>
<td>117.285</td>
<td>60.820</td>
<td>1.93 : 1.00</td>
</tr>
<tr>
<td>Slow Speech</td>
<td>201.718</td>
<td>83.205</td>
<td>2.42 : 1.00</td>
</tr>
<tr>
<td>Ratio (Fast : Slow)</td>
<td>1.00 : 1.72</td>
<td>1.00 : 1.37</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.8: Analysis by vowel types: Mean durations of unaccented vowels and ratios in Experiment 2

<table>
<thead>
<tr>
<th></th>
<th>Long vowel (ms)</th>
<th>Short vowel (ms)</th>
<th>Ratio (Long : Short)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast Speech</td>
<td>116.550</td>
<td>61.778</td>
<td>1.89 : 1.00</td>
</tr>
<tr>
<td>Slow Speech</td>
<td>202.604</td>
<td>86.216</td>
<td>2.35 : 1.00</td>
</tr>
<tr>
<td>Ratio (Fast : Slow)</td>
<td>1.00 : 1.74</td>
<td>1.00 : 1.39</td>
<td></td>
</tr>
</tbody>
</table>

I performed a repeated measures 2-way ANOVA with speech rate (fast vs. slow) and vowel length (long vs. short) as independent variables. As I expected in H-5 and H-6, both speech rate and vowel length had a significant main effect on the mean duration of vowels (for speech rate \([F(1,3) = 33.005, \ p = .0105]\) by speakers, \([F(1,4) = 270.319, \ p < .0001]\) by vowel types; for vowel length \([F(1,3) = 928.971, \ p < .0001]\) by speakers, \([F(1,4) = 175.271, \ p = .0002]\) by vowel types). In other words, durations of both short vowels and long vowels were significantly shorter in fast speech than in slow speech. Moreover, mean durations of long vowels were significantly longer than those of short vowels in both fast and slow speech. However, unexpectedly (cf. H-7), the interaction of the two factors (speech rate and vowel length) was also significant (\([F(1,3) = 18.090, \ p = .0238]\) by speakers, \([F(1,4) = 37.295, \ p = .0036]\) by vowel types), Figure 3 shows this interaction.
These results showed that durations of unaccented vowels were as vulnerable to speech rate as accented vowels. The patterns of the results from statistical analyses of Experiment 2 appear to be quite similar to the results from Experiment 1. Both speech rate and vowel length had main significant effects on the mean duration of vowels, and the interaction between these two factors was significant.

However, the way in which speakers made durational distinctions between long and short in unaccented vowels was quite different from that in accented vowels. The ratio of the duration of long vowels to short vowels for unaccented vowels was much greater than that for accented vowels, which I did not expect. I combined the results from Experiment 1 (Table 3.2) and Experiment 2 (Table 3.7) here in Table 3.9 for comparison.
Table 3.9: Analysis by speakers: Comparison of mean durations of vowels in Experiment 1 and Experiment 2

<table>
<thead>
<tr>
<th></th>
<th>Long vowel (ms)</th>
<th>Short vowel (ms)</th>
<th>Ratio (Long : Short)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fast Speech</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>accented</td>
<td>119.194</td>
<td>72.104</td>
<td>1.65 : 1.00</td>
</tr>
<tr>
<td>unaccented</td>
<td>117.285</td>
<td>60.820</td>
<td>1.93 : 1.00</td>
</tr>
<tr>
<td><strong>Slow Speech</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>accented</td>
<td>176.988</td>
<td>100.167</td>
<td>1.77 : 1.00</td>
</tr>
<tr>
<td>unaccented</td>
<td>201.718</td>
<td>83.205</td>
<td>2.42 : 1.00</td>
</tr>
</tbody>
</table>

In fast speech, the mean duration of unaccented long vowels was almost twice (1.93 times) that of unaccented short vowels, while in slow speech, the mean duration of unaccented long vowels was 2.42 times that of unaccented short vowels. I performed a repeated measures ANOVA with Experiment 1 and Experiment 2 as a between-subjects factor. There was a significant interaction between vowel length and Experiment ([F(1,6) = 59.127, p = .0003] by speakers), which indicates that the effect of vowel length (long vs. short) on the mean duration of vowels was significantly different between Experiments 1 and 2. Note that the factor that varied between Experiments 1 and 2 was existence of pitch-accent. In other words, the effect of vowel length (long vs. short) on the duration of accented vowels was significantly different from the effect of vowel length on the duration of unaccented vowels. As the ratios in Table 3.8 show, the duration of long vowels was greatly lengthened in unaccented vowels as compared with accented vowels. Furthermore, speakers magnified the duration of long vowels in slow speech. The ratio of the duration of unaccented short vowels in fast speech to slow speech was about 1 : 1.4 (see Tables 3.7 and 3.9), which was quite close to the ratio for accented
short vowels; that is, $1 : 1.4$ (see Tables 3.2 and 3.3); whereas, the duration of unaccented long vowels in slow speech was about 1.7 times longer than that in fast speech.

There is a possibility that these durational differences could be due to the different set of participants and different materials. Thus, I examined the data from the male participant (M-1), who participated in both Experiments 1 and 2. If the data from M-1 show similar results to the results shown in Table 3.9, it would be possible to conclude that pitch-accent does affect the duration of Japanese vowels. As Table 3.10 shows that results of M-1 exhibit similar patterns to overall results of Experiments 1 and 2.

| Table 3.10: Comparison of mean durations of vowels produced by participant M-1 |
|---------------------------------|-----------------|-----------------|-----------------|
|                                 | Long vowel (ms) | Short vowel (ms) | Ratio (Long : Short) |
| Fast Speech                     |                 |                 |                   |
| accented                        | 107.430         | 69.817          | 1.54 : 1.00       |
| unaccented                      | 94.018          | 52.753          | 1.78 : 1.00       |
| Slow Speech                     |                 |                 |                   |
| accented                        | 187.196         | 98.847          | 1.89 : 1.00       |
| unaccented                      | 193.975         | 74.788          | 2.59 : 1.00       |
| Ratio (Fast : Slow)             |                 |                 |                   |
| accented                        | 1.00 : 1.74     | 1.00 : 1.42     |                   |
| unaccented                      | 1.00 : 2.06     | 1.00 : 1.42     |                   |

The proportion of long vowels was greater in unaccented vowels than that of accented vowels, regardless of speech rate. The mean durations of both accented and unaccented short vowels were 1.42 times longer in slow speech than in fast speech; however, the mean duration of unaccented short vowels was 1.74 times longer in slow speech than fast speech and the mean duration of unaccented long vowels was 2.06 times longer in slow speech than fast speech. Speech rate affected the mean duration of vowels differently between long vowels and short vowels, and the effect of speech rate was
particularly strong on the mean duration of unaccented long vowels. Ratios of the mean duration of accented long vowels to accented short vowels were $1.54 : 1.00$ in fast speech and $1.89:1.00$ in slow speech, while ratios of the mean duration of unaccented long vowels to unaccented short vowels were $1.78 : 1.00$ in fast speech and $2.59 : 1.00$ in slow speech.

Results of a repeated measures of ANOVA show that both independent variables, speech rate and vowel length, had a significant main effect on the mean duration of vowels ($[F(1,1) = 235.361, p < .0001$ for speech rate, $F(1,1) = 364.032, p < .0001$ for vowel length]).

3.4.6 Discussion

The results from both Experiments 1 and 2 are evidence against absolute phonetic isochrony of the Japanese mora. The duration of vowels varies according to speech rate; the duration of long vowels especially becomes much longer in slow speech. Furthermore, comparison of the results of Experiments 1 and 2 shows the way in which native speakers of Tokyo Japanese control both duration and pitch fall to mark vowel length. When they produce accented long vowels in which two distinctive features (length and pitch-accent) need to be manifested, the speakers make the duration significantly longer and lower F0 by the appropriate amount at the accurate position. The effect of speech rate on the duration of vowels is significantly different between long and short vowels; however, the speakers are able to maintain a certain amount of pitch fall across various speech rates when manifesting pitch-accent. Unlike accented long vowels, when they produce unaccented long vowels in which F0 should not be lowered, they intensify the durational
distinction to show the length contrast. That is, the speakers make the durational difference between long and short greater in unaccented vowels than that in accented vowels.

There was another result observed in the actual mean durations of accented and unaccented vowels. Comparisons of Experiments 1 and 2 show that except for the mean duration of long vowels in slow speech, the mean duration of accented vowels was longer than the mean duration of unaccented vowels (Table 3.9 and 3.10). The effect of pitch-accent on duration of Japanese vowels is another unsolved issue. Some studies compared the mean duration of accented syllables with unaccented syllables following the accented vowels (e.g., Homma, 1981; Beckman, 1982a). Since the position of the two types of syllables is different, the effect of pitch-accent may not be accurately reflected in the measurements.

The present experiments, Experiments 1 and 2, cover all five Japanese phonemic vowels, unlike Larish (1984) and Kuriyagawa & Sawashima in which a single Japanese vowel was used, and thus compare the mean duration of accented vowels with that of unaccented vowels rather than syllables. The duration of long vowels and short vowels was separately compared. The data was elicited from four native speakers of the Tokyo dialect, unlike Homma (1981) and Kuriyagawa & Sawashima, where there was only one participant. The results clearly show that pitch-accent influences the duration of vowels as in Table 3.9 and 3.10.

I performed a repeated measures 2-way ANOVA to examine the effect of pitch-accent on the duration of vowels. There was a significant interaction between the factor
vowel length (long vs. short) and experiment (Experiment 1 vs. Experiment 2) ([F(1,6) = 59.127, p = .0003]). The conditions in the factor ‘experiment’ were ‘accented vowel’ and ‘unaccented vowel.’ That is, the presence of pitch-accent has a different effect depending on the vowel length. As the measurements show, the existence of pitch-accent made the duration of short vowels longer, but the absence of pitch-accent made the duration of long vowels especially long and the duration of short vowels short in slow speech. Results of participant M-1, who participated in both Experiments 1 and 2, show that pitch-accent had also a significant main effect on vowel duration ([F(1,1) = 10.120, p = .0018]). It is fair to conclude that pitch-accent affects the duration of vowels.

The question about the perceptual saliency of the durational change in vowels depending on the presence of pitch-accent remains unanswered. As far as the duration of short vowels is concerned, the mean duration of accented short vowels was longer than unaccented short vowels in both fast speech and slow speech. However, because of the effect of speech rate, the mean duration of unaccented long vowels was much longer than the mean duration of accented long vowels in slow speech. In addition, the experiments were not designed for investigating whether pitch-accent lengthens vowel duration, so it was not possible to assess the effect of accent correctly. It would be interesting to pursue the investigation in the future.

3.5 Conclusion

In this chapter, I have discussed two production experiments that investigate the way native speakers of Tokyo Japanese produce Japanese vowels that are phonologically
long or short, and phonologically accented or unaccented. I have investigated how speakers use acoustic properties of duration and pitch fall to manifest prosodic features of vowel length and pitch-accent in Japanese vowels. The results indicate that the speakers use longer durations to distinguish long vowels from short vowels in both accented and unaccented vowels. Furthermore, the speakers increase the durational distinction between the long and short categories – making the duration of long vowels longer and short vowels shorter – when they produce unaccented long vowels. Speakers use not only duration but also pitch fall to mark phonological vowel length when they produce accented vowels. In other words, there are two phonetic signals (duration and pitch fall) in accented vowels that distinguish vowel length; however, there is only one (duration) in unaccented vowels. Because the speakers cannot use pitch fall to mark the phonological vowel length in unaccented vowels, they increase the durational distinction between long and short vowels. Duration and pitch-accent are two distinct prosodic properties, yet they seem to be cooperating with each other in an effective way to produce Japanese vowel length, as if they were compensating for each other. The next question is how these results relate to the perception of vowel length.

Even though the duration of vowels varies, listeners perceive the number of moras in a vowel as one or two, not one and a half or three. How do listeners determine mora count? Are they depending solely on duration? In other words, are durational cues alone enough for the perception of vowel length? When speakers produce an accented vowel, they seem to use both duration and pitch information. If both of these phonetic signals are important for producing accented vowels, both of them should also be important cues for
the perception of vowel length. Do pitch cues influence the perception of vowel length?

These questions will be the topic of the following chapter.
CHAPTER 4

THE INTERACTION OF DURATION CUES AND F0 CUES IN PERCEPTION

4.1 Introduction

This chapter focuses on the perception of Japanese vowel length by Tokyo Japanese speakers. Japanese has five distinctive vowels and they all can be either long or short. Furthermore, they can be accented or unaccented. Results of Experiments 1 and 2 show that the way speakers distinguish long vowels from short vowels in production depends on whether the vowel bears accent or not. Accented long vowels have two acoustic characteristics: long duration and downward F0 movement. Unaccented long vowels in Tokyo Japanese have long duration but no pitch fall within the vowels. Comparing the results of Experiments 1 and 2, I found that the duration of unaccented short vowels was slightly shorter than that of accented short vowels. In addition, Japanese speakers magnified the durational distinction between long and short when they produced unaccented vowels. Pitch-accent seems to have an effect on the way speakers produce Japanese vowels. In other words, speakers seem to use pitch fall as well as duration to discriminate vowel length. If that is so, does pitch-accent have any effect on the perception of vowel length? This is the central topic of this chapter.

Han (1962) compared the F0, intensity, and vowel duration of minimal pairs contrastinng in pitch-accent in Japanese, such as kaki (HL) ‘oyster’ vs. kaki (HH) ‘persimmon,’ hashi (HL) ‘chopsticks’ vs. hashi (HH) ‘bridge,’ and ishi (HL) ‘medical doctor’ vs. ishi (HH) ‘stone.’ She proposes with respect to the relationship among these
phonetic properties in Japanese vowels that Japanese accented vowels are characterized primarily by relative high pitch, but the accented vowels also have longer duration and greater intensity than unaccented counterparts as supplemental cues for the accent. Hoequist (1983a) also suggests that the duration of accented vowels is longer than the duration of unaccented vowels in Japanese. He, however, suspects that this difference in duration may not play a role in the perception of Japanese pitch-accent. Han, Hoequist, and Kuriyagawa & Sawashima (1987) found that pitch-accent had an effect on the production of vowel duration as I discussed in Section 3.2, but none of them investigated the effect of pitch-accent on the perception of vowel length.

In order for Japanese speakers to produce phonologically accented long vowels, they must significantly increase vowel duration and substantially lower F0 by certain amount within the vowels. Results from Experiments 1 and 2 showed that durational information was not as steady as F0 information across various speech rates. Furthermore, speakers exaggerated the durational distinction between long and short vowels when they produce unaccented vowels. It seems that speakers are more cautious about the durational distinction between long vowels and short vowels when they produce unaccented long vowels than when they produce accented long vowels. In other words, speakers use both durational information and F0 information to distinguish vowel length in accented vowels, but in unaccented vowels, speakers need to rely solely on durational information to distinguish vowel length. Therefore, they produce especially long durations for unaccented long vowels and especially short duration for unaccented short vowels, to ensure that vowel length will be perceived correctly. Pitch change (or the lack of it) may
be as crucial as vowel duration, for speakers to produce the Japanese vowel length correctly. Production and perception of speech sounds are two sides of the same coin. If pitch fall is equally important as duration for the production of accented vowels in Japanese, it might have some effect on the perception of vowel length. Acoustic signals are filtered through listeners' phonological knowledge, and phonological knowledge determines the way speakers produce and perceive speech sounds. Investigating how pitch-accent influences the perception of phonological vowel length should give us insight into the role of pitch-accent in the rhythmic organization of Japanese; that is, what kinds of acoustic cues trigger Japanese speakers' perception of vowel length.

4.2 Nagano-Madsen (1990, 1992)

Based on her perception experiment, Nagano-Madsen (1990, 1992) suggests that F0 cues strongly affect the perception of mora count. She used a single set of minimal pair words contrasting vowel length as test material. They were accented words beru (HL) 'bell' vs. bee.ru (HL.L) 'veil.' She shifted the timing of the pitch drop every 32 ms from the original point in the vowel while maintaining the duration of target vowels. In the case of the short vowel /e/, it was shifted from the original point to three different leftward points – 32 ms earlier, 64 ms earlier, and 96 earlier than the original point. In the long vowel /e:/, the pitch drop was shifted from its original point to four different rightward points – 32 ms delayed, 64 ms delayed, 96 ms delayed, and 128 ms delayed from its original point. Thus, there were 4 stimuli used (1 original + 3 derived from the original) as short vowel stimuli and 5 (1 original + 4 derived from the original) for long

97
vowel stimuli. Audio stimuli were separated into three Groups (a, b, and c) and repeated, as in (1).

(1) a. stimuli from beru repeated 6 times each \( (4 \times 6 = 24 \text{ tokens}) \)
b. stimuli from beeru repeated 6 times each \( (5 \times 6 = 30 \text{ tokens}) \)
c. stimuli mixed from beru and beeru repeated 5 times each \( (9 \times 5 = 45 \text{ tokens}) \)

All tokens were randomly ordered and presented in word isolation form. Participants, 21 native speakers of Japanese, were asked to listen to each stimulus once and write down the word either beru or beeru in either hiragana or katakana. When F0 drop was shifted to the beginning of the short vowel /e/ in the word beru, the participants (Tokyo and Osaka\(^1\) dialect speakers) perceived the stimulus 81% as beeru in non-mixed context (i.e., in Group a) and 71% in mixed context (i.e., in Group c). Whereas, when F0 drop was shifted to the end of the long vowel /e:/ in the word beeru, 70% of the tokens was perceived as beru instead of beeru in non-mixed context (i.e., in Group b) and 86% in mixed context (Group c). In other words, if the pitch drop occurred at the beginning of a vowel, listeners perceived the vowel as long, even though the duration of the vowel was equivalent to that of a short vowel. However, if the pitch drop occurred at the end of a vowel, listeners perceived the vowel length as short, even though the duration of the vowel was equivalent to a long vowel.

Lehiste (1976) suggests that listeners tend to perceive a vowel to be longer if there is a pitch contour within the vowel. This may explain why a short vowel with an F0 drop

\(^1\) It is another dialect of Japanese. Pitch-accent is contrastive in the Osaka dialect, although its patterns are different from the Tokyo dialect.
at the beginning of the vowel was perceived as long, but it does not explain why a long vowel with delayed FO drop was perceived as short. One possibility is that listeners expect to hear pitch drop in the middle of an accented long vowel. If they do not hear the pitch drop in the middle of a long vowel or hear delayed FO drop, they will perceive the vowel as short. Nagano-Madsen proposed that Japanese listeners rely more on FO cues than durational cues in determining the mora count in an accented vowel.

Nagano-Madsen manipulated the position of FO drop to investigate its effect on the perception of vowel length. In this chapter, I will discuss Experiments 3 and 4, in which I manipulated the duration of vowels to investigate the effect of FO-fall on the perception of vowel length. The results from Experiments 3 and 4 indicate that duration alone is not sufficient for Tokyo Japanese listeners to correctly identify vowel length in accented vowels, and that pitch fall influences Tokyo Japanese listeners’ perception of phonological vowel length in accented vowels.

4.3 Experiment 3: Perception experiment involving real words

There are three differences in materials between Nagano-Madsen’s (1990, 1992) experiment and Experiment 3. First, Nagano-Madsen manipulated the position of pitch drop within an accented vowel, while I manipulated the duration of a vowel while maintaining the pitch fall. Second, Nagano-Madsen used only a single minimal pair, but Experiment 3 used all five Japanese phonemic vowels, since vowel quality affects intrinsic duration of vowels. Finally, she presented each stimulus in isolation, but I used sentences to present a material word. Since a word is presented in a carrier sentence in
Experiments 3 and 4, participants received additional cues to determine the vowel length of target vowels.

Nagano-Madsen found a strong effect of F0 drop on the perception of vowel length. If F0 cues do affect the perception of vowel length, vowels with longer duration but no pitch fall will be perceived as short.

4.3.1 Participants

The participants in this experiment were 24 native speakers of Tokyo Japanese. They were recruited on the campus of the University of Hawai‘i at Mānoa. Most participants were students at the University of Hawai‘i at Mānoa. Some were students at a language school in Honolulu. Their ages ranged from early 20s to mid 40s. They voluntarily participated in the experiment, but they were offered $ 5.00 to compensate them for their time.

4.3.2 Materials

I selected 18 minimal pairs of sentences contrasting in vowel length (see Appendix C for a complete list). The material sentences were selected from the data collected during Experiment 1, the production experiment involving Tokyo Japanese accented vowels, discussed in Chapter 3.3. The sentences were uttered at normal speech rate by a female native speaker of Tokyo Japanese. In order to test the effect of pitch-accent on the perception of vowel length, I set three conditions in target vowels – short, long, and ambiguous. Ambiguous vowels had equivalent vowel duration to the long vowels but the same timing of pitch fall as the original short vowels. I created them from
short vowels by lengthening the duration of the vowels to be equivalent to long vowels. That is, the duration of an ambiguous vowel had the same proportion of duration of vowel to word as its long vowel counterpart. For example, if the duration of the word reeji ‘midnight’ is 80 ms and the duration of the long vowel /e:/ in the word is 50 ms, the ratio of the duration of word to the long vowel is 1.6 : 1. So I increased the duration of short vowel /e/ in the word reji ‘cash register’ to make the ratio of word duration to ambiguous vowel duration 1.6 : 1. Ambiguous vowels, therefore, had ungrammatical pitch contour as accented long vowels as in Table 4.1; that is, the pitch fall took place after the target vowels in stead of within the vowels. I used SoundEdit to lengthen accented short vowels, creating ambiguous vowels. SoundEdit has a function that automatically lengthens a speech segment to a desired proportion of target duration to the original duration without changing the position of pitch fall.

Table 4.1: A sample set of conditions: reji ‘cash register’ vs. reeji ‘midnight’

<table>
<thead>
<tr>
<th>Condition</th>
<th>Short vowel</th>
<th>Long vowel</th>
<th>Ambiguous vowel</th>
</tr>
</thead>
<tbody>
<tr>
<td>target vowel</td>
<td>/e/</td>
<td>/e:/</td>
<td>/e:/</td>
</tr>
<tr>
<td>word</td>
<td>reji</td>
<td>reeji</td>
<td>reeji</td>
</tr>
<tr>
<td>pitch contour</td>
<td>H.L</td>
<td>HL.L</td>
<td>HH.L</td>
</tr>
<tr>
<td>grammaticality</td>
<td>OK</td>
<td>OK</td>
<td>NO</td>
</tr>
</tbody>
</table>

An example of set of conditions appears in Figure 4.1. The top figure is the pitch movement of the sentence watashi-ga reji-ni iku ‘I will go to a cash register’ with a short vowel /e/, the middle is the pitch movement of the sentence watashi-ga reeji-ni iku ‘I will go at midnight’ with a long vowel /e:/, and the bottom figure show the pitch movement of
sentences of each condition. Dividing vertical lines and an arrow in between the lines indicate the duration of the target vowel in each sentence.

Figure 4.1: Manipulation of the vowel duration
The material sentences were divided into three Lists (List A, B, and C) by using the Latin square method. Each List contained the same number of sentences with short, long, and ambiguous vowels, and contained only one condition for each word set. That is, if there was a sentence containing the word *reji* ‘cash register’ in List A, List A would not have a sentence with the word *reeji* ‘midnight’ or a sentence with the ambiguous vowel. Therefore, each participant heard the target vowels in only one condition.

**Table 4.2: Example of conditions of filler words**

<table>
<thead>
<tr>
<th>Type</th>
<th>Condition</th>
<th>Filler word</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>consonant length</td>
<td>来て ‘come-IMPERATIVE’ vs. 切って ‘cut-IMPERATIVE’</td>
</tr>
<tr>
<td></td>
<td><em>kite</em> (HL)</td>
<td>金メダル vs. 銀メダル</td>
</tr>
<tr>
<td></td>
<td>‘consonant length’</td>
<td><em>kitte</em> (HLL)</td>
</tr>
<tr>
<td>2</td>
<td>voicing</td>
<td><em>kin</em> (HL)-<em>medaru</em>  vs. <em>gin</em> (HL)-<em>medaru</em></td>
</tr>
<tr>
<td></td>
<td>‘voicing’</td>
<td>‘gold-medal’</td>
</tr>
<tr>
<td></td>
<td><em>kin</em> (RL)-<em>medaru</em>              vs. <em>gin</em> (RL)-<em>medaru</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td>‘silver-medal’</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>pitch-accent</td>
<td><em>hashi</em> (HL)         vs. <em>hashi</em> (HH)</td>
</tr>
<tr>
<td></td>
<td>‘pitch-accent’</td>
<td>‘chopstick’ vs. ‘bridge’</td>
</tr>
<tr>
<td>4</td>
<td>contrasting syllable</td>
<td><em>hadashi</em> (HHH)      vs. <em>hadaka</em> (HHH)</td>
</tr>
<tr>
<td></td>
<td>‘contrasting syllable’</td>
<td>‘barefoot’ vs. ‘naked’</td>
</tr>
</tbody>
</table>

Filler sentences were used to mask the presence of target minimal pairs that were based on vowel length. The filler sentences were four different types of minimal pairs as in Table 4.2. Type 1 contrasted consonant length (singleton vs. geminate). In this type, I chose pairs having the same pitch contour. Type 2 contrasted voicing. Here too, the pitch contour in the minimal pairs was identical. Type 3 was a set of minimal pairs contrasting pitch-accent and Type 4 was a set of words with contrasting syllables (see Appendix E.
for a complete list). Providing visual stimuli of the Type 3, minimal pairs contrasting pitch-accent, is possible in Japanese as each lexical item in Japanese has its own kanji character or written form, and kanji characters or written form for the two choices were always distinct, shown in Table 4.2.

Each participant listened to a total of 36 sentences (18 target sentences and 18 filler sentences) once. The filler sentences were mixed with the target sentences and the order of all the stimuli was controlled, so that the listeners did not listen to the same type of target stimuli (long, short, and ambiguous) or the same type of filler sentences consecutively. Visual stimuli for the choice were distributed in such a way that correct answers would display an equal number of times on the right and left side of the screen. Similarly visual stimuli for ambiguous vowels were distributed so that words containing a long and short vowel would be displayed an equal number of times on the right and left sides of the screen.

4.3.3 Procedure

The participants sat in front of a computer screen and listened to the material sentences one at a time through headphones. A button-box was set between the participant and the computer screen. The task for the participants in this experiment was to identify a word in the audio stimulus by pressing a button. They had to choose one of two visual stimuli displayed on the computer screen. The visual stimuli were displayed immediately after the audio stimulus finished. All visual stimuli were written in standard Japanese orthography. For example, レジ represents the word reji ‘cash register’ and the word 零時 is reeji ‘midnight’ in Japanese.
The participants pressed the left button if they thought the audio stimulus contained the word shown on the left side of the screen or pressed the right button if they thought the audio stimulus contained the word shown on the right side of the screen. The reaction time (RT) was recorded for each trial. It was the interval between the end of the audio stimulus and the moment the button was pressed. The participants advanced to the next material sentence by pressing a button located between the two response buttons, so that each participant was able to set their own pace to advance the experiment. Before they started the experiment, both verbal and written instructions in Japanese were given to the participants. There were four practice trials before the actual experiment trials. After the practice trials, each participant was asked if he/she understood the task of the experiment. The experiment took approximately 5 minutes per participant.

4.3.4 Hypotheses

(H-8) If longer duration is the only cue for phonologically long vowels, ambiguous vowels will be perceived as long.

(H-9) The results of Experiments 1 and 2 show that the Japanese speakers use pitch fall as well as longer duration to manifest phonologically long accented vowels. If pitch fall serves as a cue for vowel length in accented vowels, ambiguous vowels will be perceived as short, because there is no pitch fall within these vowels.

(H-10) If duration or the position of pitch fall alone is not an efficient way for participants to determine vowel length, even though they were able to categorize the length of ambiguous vowels according to either duration or
pitch fall, they would take significantly longer RT for ambiguous vowels to categorize their vowel length.

4.3.5 Results

Two of 24 listeners responded incorrectly to a filler sentence and chose an incorrect response to a short vowel stimulus. Two other listeners responded incorrectly to a filler sentence, but perceived the length of short vowels and long vowels 100% correctly. Other participants correctly responded to unambiguous stimuli and filler sentences. The accuracy of each participant's responses was better than 96.7% (more than 29 correct responses to 30 unambiguous stimuli).

I performed a repeated measures ANOVA with Lists A, B, and C as a between-subjects factor to determine the effect of the differences in List A, B, and C on listeners' choices and on RT. There was no significant difference in choices ([F(2,34) = .218, p = .8055]) and in RT ([F(2,34) = .061, p = .9411]) for target vowels among the three different Lists. That is, listeners' performances were not significantly different among the three Lists. Mean choices of ambiguous vowels as either long or short are shown in Tables 4.3 and 4.4. Table 4.3 was the result of a repeated measures of ANOVA, treating listeners as a sample population and Table 4.4 is treating items (target vowels) as a sample population. I performed a chi-square ($\chi^2$) test\(^2\) to see how categorizations of ambiguous vowels were made. The results show that the categorizations of ambiguous vowels were performed at the chance level ([$\chi^2$ (1, N = 23) = 6.333, p = 1.000]).

\(^2\) A $\chi^2$ test can evaluate how the data deviates from an expected value. I set an expected value as 3, because each listener listened to 6 different ambiguous vowels. If they had categorized vowel length randomly, their choices of vowel length would have been 3 shorts and 3 longs. The result of this $\chi^2$ test shows that p value was more than a significance level,.05, which suggests that the categorization was randomly made.
Table 4.3: Analysis by listeners: Mean choice of target vowels in Experiment 3

<table>
<thead>
<tr>
<th>Condition</th>
<th>Choice (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>long</td>
</tr>
<tr>
<td>Ambiguous vowels</td>
<td>46.5</td>
</tr>
<tr>
<td>Long vowels</td>
<td>100.0</td>
</tr>
<tr>
<td>Short vowels</td>
<td>1.4</td>
</tr>
</tbody>
</table>

Table 4.4: Analysis by items: Mean choice of target vowels in Experiment 3

<table>
<thead>
<tr>
<th>Condition</th>
<th>Choice (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>long</td>
</tr>
<tr>
<td>Ambiguous vowels</td>
<td>46.1</td>
</tr>
<tr>
<td>Long vowels</td>
<td>100.0</td>
</tr>
<tr>
<td>Short vowels</td>
<td>1.2</td>
</tr>
</tbody>
</table>

From these results, it is difficult to determine which cues, duration or pitch fall, were more important for listeners when categorizing vowel length. This was unexpected (cf. H-8 & H-9). Listeners’ RTs in identifying ambiguous vowels were longer than for both long vowels and short vowels (Tables 4.5 and 4.6), as I expected as in H-10. I performed a paired t-test to see if the differences in RTs among the three conditions – ambiguous, long, and short – were significant. Although listeners took more than 250 ms longer RTs for ambiguous vowels than for long vowels, the difference in RTs between long vowels and ambiguous vowels was not significant ([t(23) = 1.577, p = .1285]). The difference in RTs for short vowels and for ambiguous vowels was nearly 400 ms and it was significant ([t(23) = 2.118, p = .0452]). The difference in RTs between long vowels and short vowels was significant ([t(23) = 2.228, p = .0359]) (Figure 4.2).
Table 4.5: Analysis by listeners: Mean reaction time (RT) to target vowels

<table>
<thead>
<tr>
<th>Condition</th>
<th>RT (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambiguous vowels</td>
<td>1,441.840</td>
</tr>
<tr>
<td>Long vowels</td>
<td>1,183.771</td>
</tr>
<tr>
<td>Short vowels</td>
<td>1,051.174</td>
</tr>
</tbody>
</table>

Table 4.6: Analysis by items: Mean reaction time (RT) to target vowels

<table>
<thead>
<tr>
<th>Condition</th>
<th>RT (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambiguous vowels</td>
<td>1,457.215</td>
</tr>
<tr>
<td>Long vowels</td>
<td>1,175.823</td>
</tr>
<tr>
<td>Short vowels</td>
<td>1,046.807</td>
</tr>
</tbody>
</table>

Figure 4.2: Analysis by listeners: Interaction Bar Plot for RT in Experiment 3

Results from Experiment 3 show that listeners were able to perceive the vowel length of long vowels and short vowels correctly, and to determine their vowel length in a relatively short period. Moreover, the listeners did not seem to have any problem
categorizing long vowels and short vowels. The non-significant difference in RTs between ambiguous vowels and long vowels may indicate that the listeners were responding to the duration of ambiguous vowels. However, they were not able to categorize the length of ambiguous vowels any better than at a chance level. The duration of ambiguous vowels was equivalent to that of long vowels. The difference between ambiguous vowels and long vowels was location of pitch fall. Therefore, it seems that the ungrammatical placement of pitch fall in ambiguous vowels slowed down the listeners' categorization of vowel length. Moreover, the ungrammatical pitch contour made the listeners unable to reliably categorize vowel length.

4.3.6 Discussion

If vowel duration alone is a sufficient cue for the discrimination of vowel lengths, all ambiguous vowels should have been categorized as long, as predicted in hypothesis H-8. However, contrary to H-8, listeners did not categorize ambiguous vowels simply by duration. Results indicate that the listeners perceived 46.5% of ambiguous vowels as long and 53.5% of them as short. Moreover, contrary to the prediction in H-9, the listeners’ judgment of the length of ambiguous vowels was also at a chance level, although it slightly favored short vowels. Under the same test conditions, the listeners were able to perceive natural vowel lengths accurately. They responded to the length of long vowels and short vowels almost 100% correctly. Since ambiguous vowels were created by manipulating only the duration of short vowels while timing of pitch fall was maintained, it is plausible to assume that not just duration but appropriate F0 cues play an important role in perceiving vowel length. Results showed that listeners took the longest
RT for ambiguous vowels, among the three conditions; even though statistically it was
not significantly different from that for long vowels.

One potential problem with this perception experiment is that the stimuli used
were all existing words in minimal pair sentences. The pair of visual stimuli displayed on
the computer screen was occasionally written in different types of characters; that is, in
some cases, one was in kanji characters and another in katakana, as they are written in
standard Japanese orthography. For instance, the word rooba ‘elderly woman’ is written
in kanji as 老婆, while the minimal pair word roba ‘donkey’ is written in katakana ロバ.
These words look totally different from each other and the kanji characters in the word
‘elderly woman’ might not be as familiar to the listeners as the katakana characters in the
word ‘donkey.’ Although the listeners did not have problems choosing a correct word
from the visual stimuli for long vowels and short vowels, since they were confused by the
duration and pitch movement of ambiguous vowels, it was possible that listeners chose a
familiar word from the visual stimuli for the word containing an ambiguous vowel. In
addition, the lexical frequency and/or the semantics of the material sentences might have
had an influence on the listeners’ judgments. For example, the phrase dooki-no sakura is
an idiomatic expression for ‘classmates,’ while doki-no sakura ‘earthenware cherry
blossoms’ might be unusual items for the listeners. Presumably, the frequency of the
phrase dooki-no sakura is much higher than the phrase doki-no sakura; consequently, the
word with an ambiguous vowel was semantically biased towards dooki rather than doki.
In fact, all seven participants who listened to this ambiguous vowel categorized the vowel
as long. There was a possibility that vowel length of ambiguous vowels was not
determined simply by acoustic cues, but by the semantics of the material sentences or the
frequency of the material words. Thus, RTs as well as choices might have been affected
by the frequency or familiarity of the target words and their written forms. Furthermore,
the RT might have been influenced by the duration of material sentences, as the position
of each target word varied in each trial. Consequently, the analysis based on the results of
Experiment 3 might not be as convincing as I expected. Taking into account all these
factors that might influence the perception of vowel length, I conducted another
perception experiment using nonsense words to assess the effect of pitch-accent on the
perception of vowel length. This experiment is described in the following section.

4.4 Experiment 4: Perception experiment involving nonsense words

Many factors affect language processing, such as the frequency of the
words/sentences and other lexical information. In order to eliminate these possible
influences on the perception of vowel length, Experiment 4 used nonsense disyllabic
words as stimuli. Consequently, there is no lexical context in the words other than
phonological information.

In addition, I used a single carrier sentence; thus, the period from the offset of a
target word to the end of the sentence should be similar in each trial, which makes the
measurement of RT more accurate. Consequently, this perception experiment
(Experiment 4) should more accurately indicate whether F0 cues contribute to the
perception of vowel length than Experiment 3.
4.4.1 Participants

There were 27 participants. Most were different from the participants in Experiment 3; 4 out of 27, however, had participated in Experiment 3. Because there was almost a 6-month gap between the two experiments, I would assume that their responses were not influenced by the earlier experiment. They were all native speakers of Tokyo Japanese and students at the Mānoa or Hilo campuses of the University of Hawai‘i. They all had normal hearing and normal or corrected eyesight. They voluntarily participated in the experiment, but some received $5.00 as compensation for their time. Two of 27 participants were students in an introductory linguistics course, who chose to participate in the experiment as partial fulfillment of a course requirement. The participants’ ages ranged from the early 20s to the early 40s.

4.4.2 Materials

The test words used for Experiment 4 were nonsense disyllables in Japanese (see Appendix D for the complete list). There were three conditions for the target vowels — short, long, and ambiguous. All material words were read in a carrier sentence, *ima _____ to iimashita* ‘I said _____ now,’ which was a rather neutral sentence providing no semantic bias to the target words. The material sentences were read by a female native speaker of Tokyo Japanese who was a Ph.D. student in Second Language Studies at University of Hawai‘i at Mānoa. The materials consisted of 30 sets of minimal pair words (6 sets for each of the 5 Japanese vowels) such as *gapi* vs. *gaapi*. Considering the effect of adjacent consonants on the duration of vowel, I used the same Japanese consonant phonemes with five different Japanese vowels as much as possible. To facilitate
measurement of vowel duration, I omitted approximants and palatalized consonants in onsets. The consonants used were /b, d, g, z, m, n/ for word initial positions and /p, t, k, s, d, g, n, m/ for word medial positions. In order to avoid vowel devoicing, I used only voiced consonants word-initially. The consonant [j] is used, as the combination of /d/ plus /i/ becomes [ji], such as in items #15 jiimo and #16 jiipe in Appendix D, in Japanese. The material words were read with pitch-accent on the target vowel. That is, the speaker read gapi with a HL pitch contour (i.e., a high pitch on ga and a low pitch on the following pi) and gaapi with a HLL pitch contour. She read each sentence at a comfortable speech rate according to her judgment. All utterances were recorded on a TANDBERG TCR522 cassette-recorder through a 3M tabletop microphone in a sound attenuated recording studio. All sentences were digitized with Pitchworks at a sampling rate of 11,025 Hz.

The method of manipulating the vowel duration was the same as the one I used in Experiment 3. That is, I used SoundEdit for the durational manipulation. Each duration was determined by making the proportion of an ambiguous vowel to the word containing the vowel same as the proportion of a long vowel to the word containing the vowel. For example, if the long vowel /aː/ was 60% of the word gaapi, then the short vowel /a/ was lengthened to make it 60% of the ambiguous word gaapi. An example set of material words is shown in Table 4.7.
Table 4.7: A set of material words used for Experiment 4

<table>
<thead>
<tr>
<th>Condition</th>
<th>short vowel</th>
<th>long vowel</th>
<th>ambiguous vowel</th>
</tr>
</thead>
<tbody>
<tr>
<td>target vowel</td>
<td>/a/</td>
<td>/a:/</td>
<td>/a:/</td>
</tr>
<tr>
<td>word</td>
<td>ガピ</td>
<td>ガーピ</td>
<td>ガーピ</td>
</tr>
<tr>
<td>pitch contour</td>
<td>HL</td>
<td>HLL</td>
<td>HHL</td>
</tr>
<tr>
<td>grammaticality</td>
<td>OK</td>
<td>OK</td>
<td>NO</td>
</tr>
</tbody>
</table>

The target sentences were divided into three presentation lists (List A, B, and C) by using the Latin square method. Each participant listened to only one condition of the target vowels. Each list contained sentences with the same number of short vowels, long vowels, and ambiguous vowels. Along with 30 target sentences, 30 filler sentences were used; thus, each participant listened to 60 sentences, 30 material sentences and 30 filler sentences. Filler sentences were minimal pairs of nonsense words contrasting voicing, e.g. tapa vs. dapa, and consonant length, e.g. zaki vs. zakki. In Experiment 3, I was able to use minimal pairs contrasting pitch-accent, as the writing form is able to indicate the difference in the meaning of the pair. However, since I used nonsense words in Experiment 4, minimal pairs contrasting pitch-accent could not be used as a condition for filler sentences. Visual stimuli were written in katakana. In katakana, a long vowel is marked by a straight horizontal line such as ガーピ gaapi /gapi/, which is different from the word with a short vowel as ガピ gapi /gapi/. All material sentences were controlled so that the participants did not listen to sentences in the same condition, neither test sentences nor filler sentences, consecutively.
The visual stimuli were arranged so that the number of correct answers were displayed equally on the right and left hand sides of the screen. In addition, the visual stimuli for ambiguous vowels were arranged to display the words containing a short vowel and a long vowel an equal number of times on the right and left sides of the screen.

4.4.3 Procedure

The procedure followed in this experiment was identical to that of Experiment 3. The participants listened to each material sentence over headphones. They determined the word in the audio stimuli and chose a word from two visual stimuli by pressing the appropriate button. Visual stimuli were displayed on the computer screen immediately after the audio stimulus finished. If the selected word was on the left-hand side of the screen, the participants pressed the left button; if it was on the right-hand side of the screen, they pressed the right button. Written as well as verbal instructions were given in Japanese before the experiment. Since I used nonsense words in Experiment 4, the participants were told to listen to sentences containing a word like a character's name from computer games. The visual stimuli were written in katakana, which was perfectly suitable for character's names in computer games, as katakana is usually used to write loanwords or foreigners' names. There were four practice sentences for each participant to familiarize him/herself with the procedure of the experiment. The target words for the practice session were also nonsense words and visual stimuli were written in katakana. To go to the next material sentence, the participants needed to press a button which was located between the ones to respond to each trial. Therefore, the participants were able to
set their own pace to progress through the experiment. Each participant took approximately 10 minutes to finish the experiment.

### 4.4.4 Hypothesis

I should be able to assess the effect of pitch on the perception of vowel length more accurately in this experiment, since I eliminated the possible lexical and semantic influences on the perception of vowel length of ambiguous vowels. The hypotheses for this experiment are the same as for Experiment 3. I repeat the hypotheses constructed for Experiment 3 here.

(H-8) If longer duration is the only cue for phonological long vowels, ambiguous vowels will be perceived as long.

(H-9) The results of Experiments 1 and 2 show that the Japanese speakers use greater pitch fall as well as longer duration to manifest phonologically long vowels. If amount of pitch fall serves as a cue for vowel length, ambiguous vowels will be perceived as short, because there is no pitch fall within these vowels.

(H-10) If duration or the position of pitch fall alone is not an efficient way for participants to determine vowel length, even though they are able to categorize the length of ambiguous vowels according to either duration or pitch fall, they will take significantly longer in ambiguous vowels to categorize vowel length than in unambiguous vowels.
4.4.5 Results

One listener chose an incorrect response to the long vowel condition once and to fillers three times, in which there was no ambiguity in the stimuli. The accuracy of this listener's responses was 92.0% (46 correct responses to 50 unambiguous stimuli). In order to increase the accuracy of the analysis, the entire data from this particular listener was excluded from analysis. Consequently, 26 out of 27 listeners' responses were used. Among the 26 listeners, 8 chose an incorrect response to the filler once, but all responses to short and long vowels were 100% correct. Thus, their accuracy was better than 98.0% (more than 49 correct responses to 50 unambiguous stimuli). I conducted a repeated measures ANOVA to determine the effect of List on listeners' choice and RT. There were no significant differences in choice ([F(2,58) = 2.216, \( p = .1182 \)]) and in RT ([F(2,58) = 2.964, \( p = .0595 \)]) among the three Lists A, B and C. In other words, participants’ performances were not significantly different across the three Lists. Mean choices of ambiguous vowels as either long or short are shown in Tables 4.8 and 4.9. I performed a chi-square (\( \chi^2 \)) test to see how the categorizations of ambiguous vowels were made. The results show that categorizations of ambiguous vowels were performed at the chance level ([\( \chi^2 (1, N = 25 = 10.000), p = .991 \)]).

Table 4.8: Analysis by listeners: Mean choice of target vowels in Experiment 4

<table>
<thead>
<tr>
<th>Condition</th>
<th>Choice (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>long</td>
</tr>
<tr>
<td>Ambiguous vowels</td>
<td>64.6</td>
</tr>
<tr>
<td>Long vowels</td>
<td>100.0</td>
</tr>
<tr>
<td>Short vowels</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Table 4.9: Analysis by items: Mean choice of target vowels in Experiment 4

<table>
<thead>
<tr>
<th>Condition</th>
<th>Choice (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>long</td>
</tr>
<tr>
<td>Ambiguous vowels</td>
<td>65.2</td>
</tr>
<tr>
<td>Long vowels</td>
<td>100.0</td>
</tr>
<tr>
<td>Short vowels</td>
<td>0.0</td>
</tr>
</tbody>
</table>

These results show that although the categorization of ambiguous vowels was randomly made, the listeners favored the long category, as I predicted in H-8. This result supports Hirata’s (2004) claim that primary acoustic correlate of vowel length distinction is duration. The mean RT for ambiguous vowels was significantly longer than that of either short vowels or long vowels (Tables 4.10 and 4.11).

Table 4.10: Analysis by listeners: Reaction time (RT) in Experiment 4

<table>
<thead>
<tr>
<th>Condition</th>
<th>RT (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambiguous vowels</td>
<td>1,248.377</td>
</tr>
<tr>
<td>Long vowels</td>
<td>918.558</td>
</tr>
<tr>
<td>Short vowels</td>
<td>945.846</td>
</tr>
</tbody>
</table>

Table 4.11: Analysis by items: Reaction time (RT) in Experiment 4

<table>
<thead>
<tr>
<th>Condition</th>
<th>RT (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambiguous vowels</td>
<td>1,196.504</td>
</tr>
<tr>
<td>Long vowels</td>
<td>916.807</td>
</tr>
<tr>
<td>Short vowels</td>
<td>939.210</td>
</tr>
</tbody>
</table>

The mean RT for ambiguous vowels was 1,248.377 ms, 918.558 ms for long vowels, and 945.846 ms for short vowels. There was no significant difference in RT
between long vowels and short vowels \(t(25) = -1.313, p = .2012\). However, the difference in RTs between ambiguous vowels and long vowels and between ambiguous vowels and short vowels were both significant \((t(25) = 3.401, p = .0023)\) and \((t(25) = 2.516, p = .0017)\) respectively(Figure 4.3).

![Interaction Bar Plot for RT](image)

**Figure 4.3:** Analysis by listeners: Interaction Bar Plot for RT in Experiment 4

These results show that the listeners were able to categorize the length of long and short vowels correctly and significantly faster than ambiguous vowels. They were inclined to perceive ambiguous vowels as long; however, they took longer to determine the length of the vowels than long vowels.

Although identification was still at a chance level, almost two third of ambiguous vowels were perceived as long. If the listeners relied solely on the duration of the vowels, all ambiguous vowels should have been perceived as long vowels, as the duration of ambiguous vowels was equivalent to long vowels. The only difference between
ambiguous vowels and long vowels was placement of pitch fall. Therefore, it was the ungrammatical timing of pitch fall in ambiguous vowels that caused the delay in the listeners’ responses.

4.4.6 Discussion

Unlike Experiment 3, all material words were nonsense words, the meaning of the frame sentence was neutral, and the visual stimuli were all written in katakana in Experiment 4. The perception of vowel length in this experiment should not be affected by any kind of lexical content; e.g., semantics, usage, and frequency, but only by phonological knowledge. I used multiple consonant phonemes to balance out the effects of adjacent segments on vowel duration and used all five Japanese phonemic vowels to increase the assessment accuracy of the effect of pitch cues on the perception of vowel length. The results of Experiment 4 show how duration and F0 cues affect Tokyo listeners’ perception of vowel length.

If the duration of vowels alone determines the perception of vowel length, all ambiguous vowels should have been categorized as long, as in (H-8). More than one third of ambiguous vowels were perceived as short; however, this is not strong enough to support (H-9), either. The choices of vowel length for ambiguous vowels were made at a chance level. Durational cues alone did not dictate how listeners perceive vowel length. The listeners were unable to determine the vowel length if the pitch contour was not grammatical. Consequently, the results from Experiment 4 indicate that F0 cues do affect the perception of vowel length by Japanese speakers.
4.5 General discussion

I conducted two perception experiments involving accented Japanese vowels. I found results from Experiments 1 and 2 that Tokyo Japanese accented vowels have two acoustic cues for vowel length: durational cues and F0 fall, no matter how fast speakers utter accented vowels. In Experiments 3 and 4, I manipulated the duration of short vowels to create ambiguous vowels whose duration was that of long vowels but whose pitch pattern was that of short vowels, and examined the effect of F0 cues on the perception of vowel length. These two perception experiments confirm that for the listeners to discriminate vowel length correctly, duration cues must be accompanied by appropriate F0 cues.

In contrast with Nagano-Madsen's (1990, 1992) experiment, which used only a single vowel /e/, the stimuli for Experiments 3 and 4 included all five Japanese phonemic vowels. The listeners hesitated when determining the vowel length of ambiguous vowels. The results suggest that the ungrammatical pitch contour not only slows listeners' perception of vowel length, but can also cause ambiguous vowels to be perceived as short. In order for Japanese listeners to perceive the length of accented vowels correctly, long vowels need to have both correct duration cues and correct (i.e., grammatical) pitch contours. F0 cues do affect the perception of vowel length, in agreement with Nagano-Madsen's findings.
4.6 Conclusion

Distinctive vowel length and pitch-accent are important characteristics of Japanese phonology. Only a few investigations have been done on the interaction of durational cues and F0 cues in Japanese. Results from the two production experiments (Experiments 1 and 2) in this dissertation show that native speakers of Tokyo Japanese employ both duration and pitch fall in producing Japanese accented vowels correctly. When they produce accented long vowels, both duration and F0 adhere to a certain range of values. Speakers also aligned the position of F0 drop at the right place. On the other hand, when they produce unaccented long vowels, they magnify the durational contrast between long and short vowels, as they cannot use alignment of pitch contour to mark vowel length. Accented long vowels are not just merely long in duration, but appropriate F0 cues are also involved in their production.

The purpose of Experiments 3 and 4 was to investigate how the behavior of these acoustic properties of Japanese accented vowels affects the perception of vowel length by Japanese speakers. Experiments 3 and 4 used accented long, short, and ambiguous vowels. The latter combined the pitch contour appropriate for short vowels with the duration of long vowels, to examine the effect of pitch fall on the perception of vowel length. The listeners were not able to determine the vowel length without a grammatical pitch contour as quickly as they did with unambiguous vowels, and they sometimes perceived the ambiguous vowels as short. Duration and pitch fall are both important acoustic properties for production and perception of accented long vowels. Native
speakers of Tokyo Japanese employ these acoustic signals to produce and perceive the
Japanese vowel length correctly.

Investigating only the acoustic characteristics of phonological phenomena is not
enough to provide us with a full understanding of our phonological knowledge. Perception experiments provide evidence on how we process speech sounds, and they can help uncover our unconscious phonological knowledge. Production experiments (Experiments 1 and 2) indicate that Japanese speakers use both duration and pitch fall to mark vowel length. Perception experiments (Experiments 3 and 4) indicate that Japanese listeners require both longer duration and appropriately timed pitch contour to perceive vowel length correctly. The results of the production experiments and perception experiments corroborate each other; both duration and proper alignment of pitch fall must be considered important properties for Japanese vowels.

Then the next question is how long the duration must be long for Japanese
listeners to perceive a vowel as long? If there are no F0 cues in vowels, will Japanese
listeners tend to perceive them as short? Is the durational distinction between long and short similar in other languages that have contrastive vowel length? These questions are the topic of the next chapter.
CHAPTER 5

POHNPEIAN LONG VOWELS AND PERCEPTION OF VOWEL LENGTH

5.1 Introduction

This chapter will investigate the duration of Pohnpeian long and short vowels and effect of different phonetic implementation on vowel length categorizations by Japanese and Pohnpeian listeners. Pohnpeian vowels also have a two-way length distinction – long and short. It is relatively common for English speakers to have problems with perceiving or correctly producing vowel length distinctions, since vowel lengths are not contrastive in English. However, it seems surprising that Japanese speakers often misperceive Pohnpeian vowel length.

Processing speech sounds, of course, involves not only phonetic cues but also lexical information. However, it is worth noting that Pohnpeian long vowels often do not sound long to Japanese listeners. If the phonetic properties of Japanese long vowels and Pohnpeian long vowels were similar, native speakers of Japanese, would not have problems discriminating vowel length in Pohnpeian. There must be crucial differences in phonetic implementation of vowel length between Japanese and Pohnpeian.

Pohnpeian is another mora-timed language, but to the best of my knowledge, it is not a pitch-accent language. That is, there is no lexical pitch movement in vowels. Since this is so, do the speakers have a comparable durational distinction between long and short vowels as Japanese speakers do when they produce unaccented vowels? Are there

1 I have consulted two native speakers and have not been able to find any words contrasting in pitch-accent.
any other acoustic properties marking vowel length in Pohnpeian? These are the central questions of Section 5.2.

Results from Experiment 5 below show that Pohnpeian speakers do not make the durational distinction between long and short vowels as great as Japanese speakers do. There must be some differences in the perception of vowel length between Japanese and Pohnpeian listeners. In Section 5.3, I will investigate this question.

5.2 Experiment 5: Production experiment with Pohnpeian vowels

The purpose of this experiment was to investigate the durational distinction between long vowels and short vowels in Pohnpeian. I conducted a production experiment using Pohnpeian vowels similar to Experiments 1 and 2.

5.2.1 Participants

There were four participants (two female and two male) in this experiment. One of the female and one of the male participants were in their early 20s. The others were both in their mid 40s. Their first language was Pohnpeian, but they were all fluent bilinguals in English and Pohnpeian. They were living in Honolulu, Hawai‘i. The younger participants were students at the University of Hawai‘i at Mānoa. They were native speakers of Pohnpeian, but because they had been exposed to a bilingual environment since they were children, they did not speak English with a Pohnpeian accent. The older participants, on the other hand, had distinct Pohnpeian accents in their English, although they had been educated in English when they were in Pohnpei. The

---

2 Kenneth Rehg, who is a native speaker of English, concurred with my observations.
older male participant was well educated, working in the public sector. The older female participant was working at a company providing housekeeping services. All participants had been using the Pohnpeian language daily at home or with their friends and had been actively interacting with Pohnpeian-speaking communities in Honolulu, Hawai‘i. They participated in the experiment voluntarily.

5.2.2 Materials

The materials used in this experiment consisted of 20 sets of minimal and near minimal pairs such as kang /kaŋ/ ‘to eat’ vs. kahng /kaŋ/ ‘to refuse’ and lulu /lulu/ ‘to flame’ vs. ululu /ulu:l/ ‘pillow’ (see Appendix E for the complete list). The target words were read in the carrier sentence Ia wehwehn _____ ni lokaiahn Pohnpei /ja we:we:n _____ ni lokaj:m po:npej/ ‘What is the meaning of _____ in Pohnpeian?’

5.2.3 Procedure

The procedure followed in this experiment was the same as Experiments 1 and 2 using Japanese vowels, which I described in Chapter 3. Each sentence was written on an index card in standard Pohnpeian orthography. The desired speech rate was indicated in the upper left corner of each sheet in a smaller font than the one used for the material sentences. The speech rate was described to the participants as ‘As fast as you can’ for fast speech, ‘At the normal speed’ for normal speech, and ‘As if you are speaking to an elderly person’ for slow speech. The material sentences were ordered in a way that the participants would not repeat the same sentence consecutively at a different speech rate.
The participants were asked to read without pausing within a sentence, as a pause would change intonation patterns that might affect the pitch movement and word duration.

The utterances were tape-recorded on a TANDBERG TCR522 cassette-recorder through a 3M tabletop microphone in a sound-attenuated studio and digitized with Pitchworks at a sampling rate of 11,025 Hz. For formant readings, I used Praat (Boersma & Weenink, 2002). The durations of sentences, words, and target vowels were measured to ensure the sentences were read at desired speech rates.

5.2.4 Hypothesis

Since Pohnpeian is not a pitch-accent language, I expected that speakers would not use a pitch fall to mark vowel length. Presumably, the behavior of vowel duration should show similar patterns to Japanese unaccented vowels. Based on the findings from Experiment 2, I constructed the following hypotheses:

(H-11) The mean duration of long vowels will be significantly longer than that of short vowels, as vowel length is distinctive in Pohnpeian.

(H-12) The duration of Pohnpeian vowels will be as sensitive as Japanese vowels to speech rate. Speech rate will have a significant effect on the mean duration of vowels.

(H-13) If speakers use only a durational distinction to manifest vowel length, the distinction between long and short will be large just as the durational distinction in Japanese unaccented vowels.
5.2.5 Results

In some sentences the Pohnpeian speakers could not make distinction between fast, normal, and slow speech rates. For instance, the duration of a sentence in normal speech was identical to or longer than that in slow speech, or the sentence duration in fast speech was identical to or longer than in normal speech. The purpose of asking the speakers to read material sentences at the three different speech rates was to help the speakers make speed distinctions more easily, especially between fast and slow speech. I had expected to see a larger distinction in speed between fast and slow speech if they needed to make three levels of distinctions. I used only the data from fast speech and slow speech for the analysis.

In addition, I noticed that none of the speakers produced all target words as they were written. For example, the vowels in the minimal pair pehi ‘alter’ vs. pei ‘to float’ were both pronounced [peːj] with a long vowel, and the vowels in the minimal pair neh ‘leg’ vs. ne ‘to be distributed’ were both pronounced [neː] also with a long vowel. Therefore, I first closely examined each speaker’s data separately. Table 5.1 below shows the mean durations of each of seven Pohnpeian phonemic vowels produced by each participant. The numbers in boldface were the pairs in which the mean duration of long vowels was shorter than that of short vowels or the difference in the mean duration between longs and short vowels was less than 10 ms.
Table 5.1: Mean durations of vowels by each speaker in Experiment 5

<table>
<thead>
<tr>
<th>Speaker</th>
<th>F-1 Short (ms)</th>
<th>F-1 Long (ms)</th>
<th>F-2 Short (ms)</th>
<th>F-2 Long (ms)</th>
<th>M-1 Short (ms)</th>
<th>M-1 Long (ms)</th>
<th>M-2 Short (ms)</th>
<th>M-2 Long (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/a/</td>
<td>137.33</td>
<td>160.05</td>
<td>112.28</td>
<td>170.85</td>
<td>85.68</td>
<td>136.35</td>
<td>102.05</td>
<td>192.05</td>
</tr>
<tr>
<td>/e/</td>
<td>188.60</td>
<td>279.40</td>
<td>215.00</td>
<td>170.60</td>
<td>79.60</td>
<td>122.10</td>
<td>196.20</td>
<td>207.50</td>
</tr>
<tr>
<td>/i/</td>
<td>140.35</td>
<td>131.00</td>
<td>96.00</td>
<td>122.56</td>
<td>63.93</td>
<td>112.84</td>
<td>123.05</td>
<td>153.24</td>
</tr>
<tr>
<td>/o/</td>
<td>178.70</td>
<td>158.00</td>
<td>128.40</td>
<td>162.95</td>
<td>77.20</td>
<td>136.68</td>
<td>71.47</td>
<td>180.17</td>
</tr>
<tr>
<td>/u/</td>
<td>144.77</td>
<td>125.57</td>
<td>96.58</td>
<td>110.08</td>
<td>84.97</td>
<td>104.30</td>
<td>79.07</td>
<td>156.70</td>
</tr>
<tr>
<td>Slow</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/a/</td>
<td>187.03</td>
<td>251.38</td>
<td>149.85</td>
<td>229.88</td>
<td>112.63</td>
<td>191.60</td>
<td>102.80</td>
<td>262.73</td>
</tr>
<tr>
<td>/e/</td>
<td>221.90</td>
<td>296.40</td>
<td>236.30</td>
<td>237.60</td>
<td>98.80</td>
<td>180.70</td>
<td>220.50</td>
<td>295.10</td>
</tr>
<tr>
<td>/e/</td>
<td>157.40</td>
<td>128.90</td>
<td>136.20</td>
<td>122.10</td>
<td>128.40</td>
<td>156.50</td>
<td>156.00</td>
<td>171.70</td>
</tr>
<tr>
<td>/i/</td>
<td>232.03</td>
<td>252.83</td>
<td>127.73</td>
<td>156.42</td>
<td>82.08</td>
<td>157.52</td>
<td>121.63</td>
<td>246.82</td>
</tr>
<tr>
<td>/o/</td>
<td>179.37</td>
<td>267.70</td>
<td>145.13</td>
<td>199.98</td>
<td>88.23</td>
<td>198.58</td>
<td>91.77</td>
<td>211.23</td>
</tr>
<tr>
<td>/u/</td>
<td>140.50</td>
<td>214.57</td>
<td>134.10</td>
<td>175.08</td>
<td>86.57</td>
<td>155.60</td>
<td>95.87</td>
<td>206.98</td>
</tr>
</tbody>
</table>

Note: F-1 = 1st female speaker, F-2 = 2nd female speaker, M-1 = 1st male speaker, M-2 = 2nd male speaker

Figure 5.1 shows how the mean durations of each vowel deviate from each other within each individual speaker. The mean duration of each phonemic vowel by each individual speaker is shown in Figure 5.2. Except for the speaker M-1, the mean duration of each vowel was widely scattered. Moreover, both female speakers’ mean duration of short vowels in slow speech was longer than that of long vowels in fast speech.
Note: F-1 = 1st female speaker, F-2 = 2nd female speaker, M-1 = 1st male speaker, M-2 = 2nd male speaker, Vowel: E = /e/, oa = /ə/

**Figure 5.1:** Scatter plots of mean durations of vowels in Experiment 5
Figure 5.2: Mean duration of each phonemic vowel

One possibility for the misproduction of vowel length is that the speakers were not paying careful attention to the written form of the target words. Standardization of the Pohnpeian writing system has been a relatively recent movement. The initial meeting of the Ponapean Orthography Workshop was held in January 1972 (Rehg & Sohl, 1981, p.377). It is not uncommon to see misspellings for long vowels in Pohnpei; i.e., a vowel without the following letter h, although native speakers of Pohnpeian do not misproduce long vowels (Rehg, Personal communication in July 2004). In addition, the speakers might have felt awkward participating in an experiment producing Pohnpeian words in a semantically neutral carrier sentence for an acoustic analysis.

It could also be partly due to the fact that the form of the target words that was rather unfamiliar to the speakers. That is, most Pohnpeian verbs occur with suffixes in a natural phrase or sentence. Although all target words were grammatically well-formed, the form of the target words are not commonly used. For example, the word *pei* ‘to float’
occurs with a directional suffix -do as in peido ‘to float here’ and the word lus ‘to jump’
occurs with a directional suffix -di or -iei as in lusidi ‘to jump down’ and lusiei or lusehi
‘to jump out.’ Therefore, it is possible that the speakers produced the noun form pihl for
both words pihl ‘water’ and pil ‘also’ and neh for both neh ‘leg’ and ne ‘to be
distributed,’ because noun forms were more natural in a carrier sentence like the one I
used. Likewise, they chose the more familiar form lhls for both lhls ‘to lose’ and lus ‘to
jump.’ The variability was more obvious in the female speakers than in male speakers.
The Pohnpeian males are more careful about their speech, since they have more
opportunities to speak formally in public than the Pohnpeian females.

Since the durations of the vowel /e/ deviated the most throughout the speakers,
extcept for speaker M-1, I excluded all speakers’ measurements of this vowel from
analysis. I performed repeated measures 2-way ANOVAs. Table 5.2 is the results with
speech rate (fast vs. slow) and vowel length (long vs. short) as independent variables, and
treating speakers as a sample of the population. Table 5.3 is the results of a repeated
measures 2-way ANOVA treating vowel types (the six Pohnpeian vowels) as a sample of
the population. The results show that the mean duration of long vowels was significantly
longer than short vowels ([F(1,3) = 13.444, p = .0351] by speakers and [F(1,5) = 27.440,
$p = .0034$] by vowel types), as I expected in (H-11). In addition, speech rate had a
significant effect on the duration of vowels ([F(1,3) = 26.192, p = .0144] by speakers and
[F(1,5) = 126.617, $p < .0001$] by vowel types), as in (H-12).
Table 5.2: Analysis by speakers: Mean durations of vowels and ratio in Experiment 5

<table>
<thead>
<tr>
<th></th>
<th>Long vowel (ms)</th>
<th>Short vowel (ms)</th>
<th>Ratio (Long : Short)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast speech</td>
<td>150.129</td>
<td>115.210</td>
<td>1.30 : 1.00</td>
</tr>
<tr>
<td>Slow speech</td>
<td>213.149</td>
<td>145.469</td>
<td>1.47 : 1.00</td>
</tr>
<tr>
<td>Ratio (Fast : Slow)</td>
<td>1 : 1.41</td>
<td>1 : 1.27</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.3: Analysis by vowel types: Mean duration of vowels and ratio in Experiment 5

<table>
<thead>
<tr>
<th></th>
<th>Long vowel (ms)</th>
<th>Short vowel (ms)</th>
<th>Ratio (Long : Short)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast speech</td>
<td>142.599</td>
<td>108.033</td>
<td>1.32 : 1.00</td>
</tr>
<tr>
<td>Slow speech</td>
<td>199.224</td>
<td>137.771</td>
<td>1.45 : 1.00</td>
</tr>
<tr>
<td>Ratio (Fast : Slow)</td>
<td>1 : 1.40</td>
<td>1 : 1.28</td>
<td></td>
</tr>
</tbody>
</table>

The patterns of these statistical results are the same as for Japanese vowels. However, the interaction between the two independent variables, speech rate and vowel length, was not significant by speakers ([F(1,3) = 7.681, $p = .0695$], Figure 5.2), which was not what I expected. The interaction was significant by vowel types ([F(1,5) = 8.660, $p = .0322$], which might have been an influence of the differences in intrinsic duration of each vowel. The significant interaction between the two independent variables suggests that Pohnpeian speakers maintained similar proportional distinctions between the mean duration of long vowels and short vowels, regardless of speech rate.
There was no indication that the Pohnpeian speakers used F0 cues. This was not surprising, as Pohnpeian is not a pitch-accent language. The table below shows pitch fall in Pohnpeian. There was no significant effect of speech rate ([F(1,3) = .022, \( p = .8820 \)] by speakers and [F(1,5) = .488, \( p = .4854 \)] by vowel types) nor vowel length ([F(1,3) = .003, \( p = .9564 \)] by speakers and [F(1,5) = .616, \( p = .4332 \)] by vowel types) on the amount of pitch fall.

Table 5.4: Analysis by speakers: Mean pitch falls and ratios in Experiment 5

<table>
<thead>
<tr>
<th></th>
<th>Long vowel (Hz)</th>
<th>Short vowel (Hz)</th>
<th>Ratio (Long : Short)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast Speech</td>
<td>23.611</td>
<td>25.120</td>
<td>1.00 : 1.06</td>
</tr>
<tr>
<td>Slow Speech</td>
<td>23.840</td>
<td>24.216</td>
<td>1.00 : 1.02</td>
</tr>
<tr>
<td>Ratio (Fast : Slow)</td>
<td>1.00 : 1.01</td>
<td>1.04 : 1.00</td>
<td></td>
</tr>
</tbody>
</table>
A noticeable difference between Japanese vowels and Pohnpeian vowels was the difference in the mean duration of Pohnpeian vowels and the mean duration of Japanese vowels as in Table 5.5 below.

**Table 5.5:** Mean duration of vowels in Pohnpeian and Japanese

<table>
<thead>
<tr>
<th></th>
<th>Japanese vowels</th>
<th>Pohnpeian vowels</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accented vowels</td>
<td>Unaccented vowels</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Long (ms)</td>
<td>Short (ms)</td>
<td>Long (ms)</td>
</tr>
<tr>
<td>Fast speech</td>
<td>119.194</td>
<td>72.104</td>
<td>117.285</td>
</tr>
<tr>
<td>Slow speech</td>
<td>176.988</td>
<td>100.167</td>
<td>201.718</td>
</tr>
</tbody>
</table>

Results of Experiment 5 show that Pohnpeian speakers were able to produce durational distinctions between long and short vowels, regardless of speech rate. However, unlike Japanese speakers, Pohnpeian speakers maintained proportional distinctions between long vowels and short vowels at the various speech rates. I examined the mean durations and ratios of each speaker as in Table 5.6. The speaker F-2 showed that the ratio of the mean duration of long vowels to short vowels was 1.27 : 1.00 in both fast speech and slow speech. The speaker M-1's ratio of the mean duration of long vowels to short vowels in fast speech was not so much different from that in slow speech. (1.59 : 1.00 in fast speech and 1.55 : 1.00 in slow speech). The patterns of the ratios of the speaker M-2 were similar to those for Japanese unaccented vowels. That is, the ratio of the mean duration of long vowels to short vowels was greater in slow speech than that in fast speech. However, ratios were still small compared to the ratios of Japanese vowels.
Table 5.6: Mean duration of vowels and ratio in individual speakers

<table>
<thead>
<tr>
<th>Speakers</th>
<th>Long vowel (ms)</th>
<th>Short vowel (ms)</th>
<th>Ratio (Long : Short)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F-1</td>
<td>152.290</td>
<td>151.253</td>
<td>1.01 : 1.00</td>
</tr>
<tr>
<td>Slow speech</td>
<td>251.558</td>
<td>201.100</td>
<td>1.25 : 1.00</td>
</tr>
<tr>
<td>Ratio</td>
<td>1 : 1.65</td>
<td>1 : 1.33</td>
<td>1 : 1.65</td>
</tr>
<tr>
<td>F-2</td>
<td>141.404</td>
<td>111.220</td>
<td>1.27 : 1.00</td>
</tr>
<tr>
<td>Slow speech</td>
<td>194.188</td>
<td>153.155</td>
<td>1.27 : 1.00</td>
</tr>
<tr>
<td>Ratio</td>
<td>1 : 1.37</td>
<td>1 : 1.38</td>
<td>1 : 1.37</td>
</tr>
<tr>
<td>M-1</td>
<td>130.581</td>
<td>82.348</td>
<td>1.59 : 1.00</td>
</tr>
<tr>
<td>Slow speech</td>
<td>169.624</td>
<td>109.481</td>
<td>1.55 : 1.00</td>
</tr>
<tr>
<td>Ratio</td>
<td>1 : 1.30</td>
<td>1 : 1.33</td>
<td>1 : 1.30</td>
</tr>
<tr>
<td>M-2</td>
<td>166.700</td>
<td>101.853</td>
<td>1.64 : 1.00</td>
</tr>
<tr>
<td>Slow speech</td>
<td>224.776</td>
<td>107.074</td>
<td>2.10 : 1.00</td>
</tr>
<tr>
<td>Ratio</td>
<td>1 : 1.35</td>
<td>1 : 1.05</td>
<td>1 : 1.35</td>
</tr>
</tbody>
</table>

The ratio of the duration of long vowels to the short vowels, however, was smaller than that of unaccented Japanese vowels, contrary to (H-13). To facilitate the comparison, I created the table below. I included the ratios of Japanese accented long vowels to short vowels as well.

Table 5.7: Ratio of vowel duration (Long : Short)

<table>
<thead>
<tr>
<th></th>
<th>Japanese vowels</th>
<th>Pohnpeian vowels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accented</td>
<td>Unaccented</td>
</tr>
<tr>
<td>Fast speech</td>
<td>1.65 : 1.00</td>
<td>1.93 : 1.00</td>
</tr>
<tr>
<td>Slow speech</td>
<td>1.77 : 1.00</td>
<td>2.42 : 1.00</td>
</tr>
</tbody>
</table>

3 Goodman (1996) reports that the duration of long vowels is twice as long as short vowels in Pohnpeian (p.279). She does not explain how the measurements were taken or provide us with any data that she obtained.
The mean duration of Pohnpeian long vowels was less than 1.5 times longer than the mean duration of short vowels at both speech rates, while the duration of Japanese unaccented long vowels was 1.9 times longer than that of short vowels in fast speech and more than 2.4 times longer in slow speech. Pohnpeian speakers made the durational distinction between long and short much smaller than Japanese speakers did for their vowels.

5.2.6 Discussion

The results of Experiment 5 show the differences between Japanese speakers and Pohnpeian speakers in the way they produce vowels. As I expected in (H-11), the mean duration of long vowels was significantly longer than the mean duration of short vowels regardless of speech rate. However, there was not a significant interaction between the two independent variables, speech rate and vowel length, in the duration of Pohnpeian vowels. Japanese speakers did not maintain proportional distinctions between the mean duration of long vowels and short vowels, while Pohnpeian speakers seemed to maintain them at different speech rates.

Since Pohnpeian is not a pitch-accent language, I expected in (H-13) to see Pohnpeian speakers produce their long vowels in a way similar to the way Japanese speakers produce unaccented long vowels; that is, by maintaining a relatively large durational distinction between long and short vowels. Although the mean duration of long vowels was significantly longer than that of short vowels in Pohnpeian, the ratio of the duration of long vowels to short vowels was much smaller than that of Japanese unaccented vowels. It seems that even less than 1.5 times longer duration is enough for
Pohnpeian vowels to be long. Even the mean duration of Japanese accented long vowels, in which amount of pitch fall additionally indicates the vowel length, was more than 1.6 times as long as that of accented short vowels. Do Pohnpeian speakers use other cues to mark vowel length?

It is natural to think that Pohnpeian speakers use spectral cues for vowel length. In the case of English, although vowel length is not distinctive, durational differences in vowels are noticeable. For example, the duration of the higher front vowel /i/ in the word *heat* is longer than a lower high front vowel /ʌ/ in the word *hit*.

I need to consider vowel allophony here. It is a feature of Pohnpeian phonology. In Pohnpeian, the quality of short vowels varies depending on the quality of adjacent consonants. When front short vowels are surrounded by back consonants, the vowels become centralized and when back short vowels are surrounded by front consonants, the vowels become centralized. However, the quality of long vowels does not vary depending on adjacent consonants as much as short vowels. There is a possibility that Pohnpeian speakers use spectral cues to indicate vowel length.

I examined vowel quality for all four participants. I was not able to observe quality differences between long and short vowels clearly, except in the data from one male Pohnpeian participant. This particular speaker produced target vowels correctly except for one minimal pair out of 20 pairs, despite the fact that other speakers misproduced the length of several target vowels. Since Experiment 5 was not designed to investigate spectral cues to vowel length, I do not have adequate evidence to draw a strong conclusion; however, it seems that this Pohnpeian speaker used spectral cues as
well as duration cues to distinguish vowel length. Following are some observations that I made in this male Pohnpeian speaker’s speech samples.

Figure 5.3 below illustrates the comparison of the formant movement in the word *pihl* and *pil*. In order to make the comparison easy, I extracted the words from the sentences and put them side-by-side, and then extracted visible formants of the words by using *Praat* (Boersma & Weenink, 2002). The maximum formants extracted were set at 5,000 Hz, and extracted formants were marked every 5 ms by dots on the plot. Since both /p/ and /l/ are front consonants, the front vowels /i:1/ and /i:/ should not be affected by the adjacent consonants. Notice that the second formant (F2) of the long vowel /i:1/ in the word *pihl ‘water’* is much higher than that of the short vowel /i:/ in the word *pil ‘also’.* The mean F2 in the long vowel /i:1/ is 2,230 Hz and in the short vowel /i:/ is 1,801 Hz. That is, the tongue position when the speaker produced the long vowel /i:1/ in the word *pihl* was more advanced than the tongue position for the short vowel /i:/ in the word *pil*. In fact, the word *pil* sounds like [pil] rather than [pil]. This quality difference may be perceived as the tense /i:/ vs. lax /i:/ distinction for native speakers of English.
Similarly, a higher F2 in a long vowel than in a short vowel was observed in the minimal pairs *kihd* ‘garbage’ vs. *kid* ‘thousand’ and *ihd* ‘plant (sp.)’ vs. *id* ‘to make a fire.’ When a vowel is in a mixed environment (i.e., the preceding consonant is front and the following consonant is back or the preceding consonant is back and the following consonant is non-back), the following consonant has a stronger influence on the vowel.

The environment of the minimal pairs *kihd* /kiːt/ vs. *kid* /kit/ is an example of the mixed environment; that is, the consonant /k/ that is preceding the vowels is a back consonant and the consonant /t/ that is following the vowels is a front consonant. Thus, the influence of the neighboring consonants on the vowels in this pair is the front-consonant environment. Another pair *ihd* ‘plant (sp.)’ vs. *id* ‘to make a fire’ is also presumably in
the front-consonant environment. That is, these three minimal pairs *pihl* vs. *pil*, *kihd* vs. *kid*, and *ihd* vs. *id* are all in the front-consonant environment, and so the front vowels should not be affected by the adjacent consonants. The quality differences observed between /i:/ and /i/ in these three pairs are not due to the neighboring consonants, but they could be indicating vowel length distinctions.

Figure 5.4 below shows the difference in F2 between /u:/ in the word *luhs* ‘to lose’ and /u/ in the word *lus* ‘to jump.’ The F2 of /u:/ in *luhs* is much lower than that in *lus*. The mean F2 in the long vowel /u:/ is 1,154 Hz and in short vowel /u/ is 1,423 Hz. That is, the short vowel /u/ in the word *lus* was produced with more advanced tongue position or with less rounded lips than the long vowel /u:/ in the word *luhs*. The short back vowel /u/ was centralized in the word *lus*. In fact, it does sound like [u]. The minimal pair *luhs* ‘to lose’ vs. *lus* ‘to jump’ could be an example of the quality difference caused by the neighboring consonants, because prevocalic and postvocalic consonants, /l/ and /s/, are both front consonants.
Another spectral difference that I observed was that the third formant (F3) of the long vowel /ɔː/ in the word *poahd* /pɔːt/ 'individual planting' was lower than that of the short vowel in the word *poad* /pɔːt/ 'to be planed.' In some minimal pairs, such as *kahng* 'to refuse' vs. *kang* 'to eat' and *pehi* 'alter' vs. *pei* 'to float,' the differences in spectral cues were not obvious in the spectrograms; however, the quality differences were audible. In some minimal pairs, however, such as *neh* 'leg' vs. *ne* 'to be distributed,' no quality difference was detected visibly or audibly. These observations, needless to say, were identified in the data from a single Pohnpeian speaker. Since other Pohnpeian speakers produced many target minimal pairs without length and/or quality contrast, the amount of
valid data collected from Experiment 5 was not enough to conduct a statistical analysis of the effect of spectral cues on vowel length distinctions. There was a detailed phonetic investigation on vowel allophony in Marshallese, another Micronesian language (Choi, 1992), supporting Bender’s (1968) impressionistic analysis that Marshallese has similar allophonic vowel alternations to Pohnpeian. The effect of vowel quality on the vowel length distinction in Pohnpeian is worth exploring in future research.

The results from Experiment 5 show that phonetic characteristics of the Pohnpeian vowels differ from those in Japanese. The mean duration of Pohnpeian vowels appeared to be longer than that of Japanese vowels, yet the ratio of the mean duration of long vowels to short vowels in Pohnpeian is much smaller than that in Japanese. Both Japanese and Pohnpeian are mora-timed languages in which vowel length is distinctive; nonetheless, their phonetic implementations of long vowels differ. Although the mora is a unit of phonological length in both languages, what makes listeners recognize a mora as a mora is not the same.

How long must the duration of a vowel be in order for listeners to perceive it as long? Do Japanese listeners perceive vowel length differently from Pohnpeian listeners? Does the longer mean duration of vowels in Pohnpeian affect the Japanese listeners’ perception of vowel length? I will explore these questions in the following section.

5.3 Experiment 6: Cross-linguistic perception experiment

The results of the production experiments involving Japanese and Pohnpeian vowels show that the durations of vowels are affected by speech rate and that the duration
of long vowels is always longer than that of short vowels in both languages. However, the ratio of the duration of long vowels to short vowels is much smaller in Pohnpeian than in Japanese. All other things being equal, Japanese speakers produce long vowels with a longer relative duration to short vowels than Pohnpeian speakers do. How long must a vowel be for Japanese listeners and Pohnpeian listeners to categorize vowel length as long? The results from the production experiments showed that Japanese speakers produce a mean duration of unaccented long vowels 1.93 times longer than that of short vowels in fast speech and 2.42 times longer in slow speech. Pohnpeian speakers produce long vowels 1.30 times longer than short vowels in fast speech and 1.47 times longer in slow speech. If the perception of vowel length corresponds to these ratios, I can expect that long vowels must be at least 1.93 times longer than short vowels for Japanese listeners to perceive a vowel as long, but for Pohnpeian speakers 1.30 times longer duration than short vowels would be enough to perceive a vowel as long. In addition, since the mean duration of Pohnpeian vowels was longer than the mean duration of Japanese vowels, the perception of vowel length should reveal interesting differences between Japanese listeners and Pohnpeian listeners. Experiment 6 was designed to test these predictions.

5.3.1 Participants

There was a total of 51 participants, 26 native speakers of the Tokyo dialect of Japanese and 25 native speakers of Pohnpeian. They volunteered to participate in the experiments but were offered $5.00 or kava as compensation for their time. All of them were native speakers of either Tokyo Japanese or Pohnpeian but understood and spoke
English. They were living in Honolulu or Hilo, Hawai‘i. All Japanese participants were students at the University of Hawai‘i at Mānoa or Hilo campus. Their ages ranged from the early 20s to the early 40s. All of the Japanese participants had opportunity to use Japanese in their daily life and/or watch Japanese TV programs. Seven of 25 Pohnpeian participants were students at the University of Hawai‘i at Mānoa or Hilo campus. They were fluent bilinguals in English and Pohnpeian, and were still involved in a Pohnpeian-speaking community on a regular basis, such as going to church and having a casual meeting weekly. The other Pohnpeian participants were residents of Honolulu, Hawai‘i. They were living in the same neighborhoods with either their relatives or friends. They had a relatively less-skilled occupation, such as housekeeper, cook/busboy at a franchised restaurant, or custodian. They were using Pohnpeian in their daily life. The Pohnpeian participants’ age ranged from the early 20s to the late 50s.

5.3.2 Stimuli and manipulation

It is difficult for listeners to identify vowel length if there is only a single vowel in the stimulus. Thus, it is better for the stimulus to be two syllables or longer. Furthermore, adding a short vowel at the initial position should provide the participant with a point of reference for vowel length. Consequently, the form VCVCV (or VCV:CV) was chosen as the form of stimuli. Then, I tried to eliminate any lexical influence from either language, as it might affect the participants’ perception. I selected the sound sequence epepe, a combination of bilabial plosive and mid-front vowel, for the form of the stimuli. It is nonsense in both Japanese and Pohnpeian.
The sequence *epepepepepepe* was elicited from a female native speaker of Tokyo Japanese and a female native speaker of Pohnpeian 3 times each. The speakers were asked to recite the material with no pitch movement. The recorded speech sample was digitized with Pitchworks at a sampling rate of 22,050 Hz. The segment *epepe* was extracted from the digitized speech sample for manipulation. The target vowel was the second vowel in the segment. Being aware of the effect of pitch cues, I tried to find a segment in the sample without pitch fall in both languages and a similar pitch range in the speech samples from each language. The pitch ranged from 236.38 Hz to 221.61 Hz in the Japanese stimulus (the difference was 14.77 Hz) and from 209.42 Hz to 201.68 Hz in the Pohnpeian stimulus (the difference was 7.74 Hz).

Some may question whether the stimuli will be processed as *words* or mere sounds. It could be possible that the participants were simply processing each stimulus as a noise and responding to the duration of sounds rather than categorizing vowel length. Their responses may not be categorizations of vowel length. The participants may even process the audio stimuli as nonspeech sounds. Lehnhoff et al. (2004) challenged this issue. They conducted an experiment in which participants were asked to discriminate 30 2-second speech segments (conversational speech from native speakers of French, German, Hebrew, Hindi, Japanese, and Russian) and 30 2-second non-speech events (animal vocalizations, water sounds, and other environmental sounds such as thunder). The results show that participants were able to differentiate speech from non-speech 84 % correctly. In addition, only 12 stimuli (20 %) were categorized at less than chance
levels. That is, we have an ability to differentiate speech sounds from nonspeech sounds whether we understand the meaning of the sounds or not.

The stimuli I used in Experiment 6 were created from actual utterances collected from native speakers of Japanese and Pohnpeian. In addition, the stimuli could be a novel word in both languages; each segment in the stimuli contains phonemes in both languages meeting the phonotactic requirements of both languages. It is reasonable to assume that participants were in fact determining the vowel length not the duration of sounds. We will need to wait, however, for the results from additional/further cross-linguistic perception experiments before we can give a definite answer to this question.

In order to determine the threshold between long and short vowels for Japanese and Pohnpeian speakers, the duration of second vowel ‘e’ was lengthened to 130, 140, 150, 154, 156, 160, 166, 170, 180, and 200 % of the original duration in both languages. The ratios and actual durations of stimuli are shown in Table 5.8.

---

Since I used SoundEdit to manipulate vowel durations, they came out with uneven increment.
Table 5.8: Ratio and vowel duration of stimuli

<table>
<thead>
<tr>
<th>Stimuli</th>
<th>Japanese stimuli</th>
<th>Vowel duration (ms)</th>
<th>Pohnpeian stimuli</th>
<th>Vowel duration (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>JP100</td>
<td>71.9</td>
<td>PN100</td>
<td>85.0</td>
<td></td>
</tr>
<tr>
<td>JP130</td>
<td>94.0</td>
<td>PN130</td>
<td>108.8</td>
<td></td>
</tr>
<tr>
<td>JP140</td>
<td>101.0</td>
<td>PN140</td>
<td>119.8</td>
<td></td>
</tr>
<tr>
<td>JP150</td>
<td>107.8</td>
<td>PN150</td>
<td>127.6</td>
<td></td>
</tr>
<tr>
<td>JP154</td>
<td>110.5</td>
<td>PN154</td>
<td>130.2</td>
<td></td>
</tr>
<tr>
<td>JP156</td>
<td>112.6</td>
<td>PN156</td>
<td>135.0</td>
<td></td>
</tr>
<tr>
<td>JP160</td>
<td>115.0</td>
<td>PN160</td>
<td>138.0</td>
<td></td>
</tr>
<tr>
<td>JP166</td>
<td>119.5</td>
<td>PN166</td>
<td>140.9</td>
<td></td>
</tr>
<tr>
<td>JP170</td>
<td>122.1</td>
<td>PN170</td>
<td>146.2</td>
<td></td>
</tr>
<tr>
<td>JP180</td>
<td>128.5</td>
<td>PN180</td>
<td>154.2</td>
<td></td>
</tr>
<tr>
<td>JP200</td>
<td>143.8</td>
<td>PN200</td>
<td>170.1</td>
<td></td>
</tr>
</tbody>
</table>

Note: JP = Japanese stimuli, PN = Pohnpeian stimuli

The results from the production experiment showed that the mean duration of Japanese long vowels was 1.65 or 1.77 times longer than that of short vowels in accented vowels and 1.93 times or more in unaccented vowels; and the mean duration of long vowels was 1.35 or 1.46 times longer than that of short vowels in Pohnpeian. Thus, if speakers need to rely solely on the durational distinction to determine the vowel length, Pohnpeian listeners will start perceiving the vowel length as long at a point somewhere between 130 and 140% longer than the short vowel; Japanese listeners, somewhere between 160 and 170% longer than the short vowel. In order to find the exact threshold where the listeners shifted across the categories of short and long vowels, the increment between those points were smaller than the other points. A total of 20 stimuli were created from the speech sample in each language (10 manipulated stimuli x 2).
5.3.3 Experiment design

An ABX discrimination test was used to determine how listeners responded to durational cues in vowels. In the ABX test, stimuli were arranged in triads consisting of an A stimulus, a B stimulus, and an X stimulus. Each participant was asked to listen to the triad and categorize the X stimulus as either a member of the category of stimulus A or B in each trial. The A stimulus was identical to the original material *epepe*, and the B stimulus was created by lengthening the medial vowel ‘e’ to double its original duration. Thus, the A stimulus in a triad was either JP100 or PN100 and the B stimulus was JP200 or PN200. The X stimuli were sounds with a target vowel of ambiguous duration; that is, the materials with a medial vowel lengthened to 130, 140, 150, 154, 156, 160, 166, 170, and 180 % of the original length. The A and B stimuli were also used as the X stimuli to detect the accuracy of each participant’s performance. Therefore, there were 22 triads in each block of the experiment – 11 different X stimuli each for Japanese and Pohnpeian.

There were 6 blocks in each complete set of real trials. Each block consists of the same set of triads, but the order of the triads was different in each block. In order to advance to next block, participants needed to press a button, which allowed them to take a brief break between blocks.

Table 5.9 shows the set of triads in each block. In the shaded triads, #1, 11, 12, and 22, the X stimulus was identical to either the A stimulus or the B stimulus. The 22 triads were randomly ordered. In each block, the A stimulus came first half of the time, and the B stimulus came first half of the time. The order of the A and B stimuli was
balanced across blocks, so that the responses should not be biased toward either the A or B stimulus.

**Table 5.9: Set of triads in each block**

<table>
<thead>
<tr>
<th>triad</th>
<th>A stimuli</th>
<th>B stimuli</th>
<th>X stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>JP 100</td>
<td>JP 200</td>
<td>JP 100</td>
</tr>
<tr>
<td>2</td>
<td>JP 100</td>
<td>JP 200</td>
<td>JP 130</td>
</tr>
<tr>
<td>3</td>
<td>JP 100</td>
<td>JP 200</td>
<td>JP 140</td>
</tr>
<tr>
<td>4</td>
<td>JP 100</td>
<td>JP 200</td>
<td>JP 150</td>
</tr>
<tr>
<td>5</td>
<td>JP 100</td>
<td>JP 200</td>
<td>JP 154</td>
</tr>
<tr>
<td>6</td>
<td>JP 200</td>
<td>JP 100</td>
<td>JP 156</td>
</tr>
<tr>
<td>7</td>
<td>JP 200</td>
<td>JP 100</td>
<td>JP 160</td>
</tr>
<tr>
<td>8</td>
<td>JP 200</td>
<td>JP 100</td>
<td>JP 166</td>
</tr>
<tr>
<td>9</td>
<td>JP 200</td>
<td>JP 100</td>
<td>JP 170</td>
</tr>
<tr>
<td>10</td>
<td>JP 200</td>
<td>JP 100</td>
<td>JP 180</td>
</tr>
<tr>
<td>11</td>
<td>JP 200</td>
<td>JP 100</td>
<td>JP 200</td>
</tr>
<tr>
<td>12</td>
<td>PN 100</td>
<td>PN 200</td>
<td>PN 100</td>
</tr>
<tr>
<td>13</td>
<td>PN 100</td>
<td>PN 200</td>
<td>PN 130</td>
</tr>
<tr>
<td>14</td>
<td>PN 100</td>
<td>PN 200</td>
<td>PN 140</td>
</tr>
<tr>
<td>15</td>
<td>PN 100</td>
<td>PN 200</td>
<td>PN 150</td>
</tr>
<tr>
<td>16</td>
<td>PN 100</td>
<td>PN 200</td>
<td>PN 154</td>
</tr>
<tr>
<td>17</td>
<td>PN 100</td>
<td>PN 200</td>
<td>PN 156</td>
</tr>
<tr>
<td>18</td>
<td>PN 200</td>
<td>PN 100</td>
<td>PN 160</td>
</tr>
<tr>
<td>19</td>
<td>PN 200</td>
<td>PN 100</td>
<td>PN 166</td>
</tr>
<tr>
<td>20</td>
<td>PN 200</td>
<td>PN 100</td>
<td>PN 170</td>
</tr>
<tr>
<td>21</td>
<td>PN 200</td>
<td>PN 100</td>
<td>PN 180</td>
</tr>
<tr>
<td>22</td>
<td>PN 200</td>
<td>PN 100</td>
<td>PN 200</td>
</tr>
</tbody>
</table>

Note. JP = Japanese stimulus, PN = Pohnpeian stimulus

There was a 700 ms pause between the A stimulus and the B stimulus and between the B stimulus and the X stimulus. There was a 1000 ms pause between the participant’s response (pressing a button) and the following A stimulus. The reaction time (RT), the period from the end of the X stimulus to the response, was also recorded. Each
participant was asked to respond to the X stimulus by pressing a button. If he/she thought the X was similar to the A stimulus, a green button was pressed; if the X was similar to the B stimulus, a red button was pressed.

![Diagram](image)

**Figure 5.6:** Sequence of an ABX trial

There were 6 practice trials in a practice session – 3 triads were taken from Japanese and the other 3 were taken from Pohnpeian, and the X stimuli were identical to either the A stimulus or the B stimulus. In the practice session, participants got feedback on whether the X stimuli was identical to the A stimulus or the B stimulus. The feedback was a message simply stating, “It was green” or “It was red” depending on which was the correct button.\(^5\) It displayed immediately after the participant’s response. The participants found out if they had responded to the stimulus correctly or not. The following trial started 1000 ms after the feedback was displayed on the screen. The practice session was repeated upon request. The participants were told that there was no feedback in the real trials because some of the X stimuli would be ambiguous, there was no correct answer to each trial, and so their judgments should be based on their own perceptions.

---

\(^5\) The right-side button was red and left-side was green. If the correct response was to press right button, the message read, “It was red,” if the correct responses was to left button the message read, “It was green.”
5.3.4 Procedure

The experiment took place in a quiet room, either a sound attenuated room, a quiet conference room on campus, or a quiet room in a participant's home. Before each participant started the experiment, the procedure was explained in Japanese to Japanese participants, in Pohnpeian to non-student Pohnpeian participants, and in English to the Pohnpeian students. A set of written instructions was also given in English and displayed on the computer screen before the participants started the practice session. After the practice session, participants were asked whether they understood the task.

5.3.5 Hypotheses

Based on the results from Experiments 1, 2, and 5, I constructed the following hypotheses:

(H-14) The durational distinction between long and short is much smaller in Pohnpeian than in Japanese. Therefore, Pohnpeian listeners will start to categorize vowel length as long with a smaller distinction than Japanese listeners will.

(H-15) Since the mean duration of vowels is longer in Pohnpeian than that in Japanese, Japanese listeners will categorize Pohnpeian vowels as long more frequently than Japanese vowels and Pohnpeian listeners will categorize Japanese vowels as short more frequently than Pohnpeian vowels. In other words, the stimulus language will affect the perception of vowel length. Japanese listeners will perceive Pohnpeian vowels as longer

---

6 I am grateful to Mr. Simion Kihleng for his assistance.

152
than Japanese vowels and Pohnpeian listeners will perceive Japanese vowels as shorter than Pohnpeian vowels.

5.3.6 Data analyses

First, I examined the accuracy of each participants’ responses. There were triads where the X stimulus was identical with either the A stimulus or the B stimulus in a block; that is, the target stimulus was identical with either JP100, JP200, PN100, or PN200. Therefore, there were 24 trials where the response was not ambiguous – there were 6 blocks all together and each block had 4 of these triads. If a participant missed more than 5 out of 24 trials; that is, more than 20% of these trials, all of the data from the participant were omitted from the analysis. I was able to use the data from 17 Japanese participants and 11 Pohnpeian participants. The average of their responses for each X stimulus was calculated and I have performed repeated measures ANOVAs to examine categorization patterns of participants. In addition, I conducted Chi-square ($\chi^2$) tests to examine if the categorization of vowel duration was made randomly or not.

5.3.7 Results

I measured RT (response time) for each trial of each participant; however, I had not set a time limit for the trial nor put any time constraint on the participants. Since RT varied too widely among participants, I did not use RT for the analysis.

I performed a repeated measures 2-way ANOVA to see if responses were affected by the difference in stimulus language. Independent variables are stimulus language and listeners’ first language (L1), and listeners were treated as a sample population. Stimulus
language had a significant main effect on listeners' responses ([F(1,26) = 5.060, \( p = .0332 \]). However, there was no significant effect of L1 on responses ([F(1,26) = .366, \( p = .5507 \]) and no significant interaction between these two factors either ([F(1,26) = .793, \( p = .3813 \]). When I looked at responses of Japanese listeners and Pohnpeian listeners separately, stimulus language had a significant effect on Japanese listeners ([F(1,16) = 4.735, \( p = .0449 \]), but it did not have a significant main effect on Pohnpeian listeners ([F(1,10) = 1.585, \( p = .266 \]). In other words, Japanese listeners categorized vowel length differently between Japanese vowels and Pohnpeian vowels, but Pohnpeian listeners’ categorization pattern was consistent whether the stimulus language was Japanese or Pohnpeian (Figure 5.6). The vertical axis indicates the response, where the value 0 indicates the response “short” and 6 indicate the response “long.”

![Interaction Line Plot for Responses](image)

**Figure 5.7:** Interaction line plot in Experiment 6
Japanese listeners categorized Pohnpeian vowels as long slightly more often than they categorized Japanese vowels as long. The mean response of Japanese listeners’ to Japanese stimuli was 3.326 and to Pohnpeian stimuli 3.727. Pohnpeian listeners’ mean response to Japanese stimuli was 3.587 and to Pohnpeian stimuli 3.760. The comparison of the results from Experiments 2 and 5 showed that the mean duration of Pohnpeian short vowels was longer than that of Japanese short vowels. The fact that Japanese listeners’ responses to Pohnpeian stimuli were slightly biased toward long might have been due to the influence of the differences in the mean duration of vowels, as I expected in (H-15).

Results of Experiment 6 indicate that in order for each language’s listeners to perceive a long vowel as long, a certain amount of durational difference between long and short vowels was required. In both languages, production experiments show that the mean duration of long vowels was significantly longer than that of short vowels regardless of speech rate; thus, it is not surprising for listeners to require a certain amount of durational distinction in order to discriminate vowel length.

I examined the participants’ responses according to their L1 and the stimulus language, since Japanese listeners responses were different between the two languages’ stimuli. The patterns of responses to the stimuli are shown in Figures 5.7-10. Since I manipulated the original stimuli (JP100 and PN100) to create the target stimuli 130, 140, 150, 154, 156, 160, 166, 170, 180, and 200 to find out the exact threshold for categorization of vowel length, there are no values for the stimuli 110, 120, and 190. The solid line indicates a response of “short;” the dotted line, a response of “long.”
The crossover points in the length categorization by Japanese listeners was at the 150% stimulus, but earlier than the 150% stimulus by Pohnpeian listeners, which indicates that Japanese listeners require greater durational distinction between long and short vowels than Pohnpeian listeners, as I expected in (H-14).

Figure 5.8: Japanese listeners’ responses to Japanese stimuli

Figure 5.9: Japanese listeners’ responses to Pohnpeian stimuli
Figure 5.10: Pohnpeian listeners’ responses to Japanese stimuli

Figure 5.11: Pohnpeian listeners’ responses to Pohnpeian stimuli

To facilitate the comparison, I created two additional figures below. Figure 5.10 shows each language listeners’ categorization of Japanese stimuli as long and Figure 5.11 shows the categorization of Pohnpeian stimuli as long. Japanese listeners’ categorization
was slightly biased towards short vowels in comparison to the Pohnpeian listeners’ categorization of both Japanese and Pohnpeian stimuli.

Table 5.10: Mean responses to each stimulus as long (unit: %)

<table>
<thead>
<tr>
<th>Stimulus Duration</th>
<th>Japanese Listener</th>
<th>Pohnpeian Listener</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Japanese stimuli</td>
<td>Pohnpeian stimuli</td>
</tr>
<tr>
<td>100</td>
<td>3.92</td>
<td>0.98</td>
</tr>
<tr>
<td>130</td>
<td>10.78</td>
<td>11.76</td>
</tr>
<tr>
<td>140</td>
<td>19.61</td>
<td>29.41</td>
</tr>
<tr>
<td>150</td>
<td>48.04</td>
<td>49.02</td>
</tr>
<tr>
<td>154</td>
<td>54.90</td>
<td>62.75</td>
</tr>
<tr>
<td>156</td>
<td>37.25</td>
<td>50.98</td>
</tr>
<tr>
<td>160</td>
<td>50.98</td>
<td>51.96</td>
</tr>
<tr>
<td>166</td>
<td>46.08</td>
<td>69.61</td>
</tr>
<tr>
<td>170</td>
<td>62.75</td>
<td>69.61</td>
</tr>
<tr>
<td>180</td>
<td>55.88</td>
<td>66.67</td>
</tr>
<tr>
<td>200</td>
<td>65.69</td>
<td>68.63</td>
</tr>
<tr>
<td></td>
<td>12.12</td>
<td>15.15</td>
</tr>
<tr>
<td></td>
<td>37.88</td>
<td>31.82</td>
</tr>
<tr>
<td></td>
<td>43.94</td>
<td>43.94</td>
</tr>
<tr>
<td></td>
<td>75.76</td>
<td>56.06</td>
</tr>
<tr>
<td></td>
<td>72.73</td>
<td>81.82</td>
</tr>
<tr>
<td></td>
<td>59.09</td>
<td>60.61</td>
</tr>
<tr>
<td></td>
<td>71.21</td>
<td>65.15</td>
</tr>
<tr>
<td></td>
<td>59.09</td>
<td>75.76</td>
</tr>
<tr>
<td></td>
<td>69.70</td>
<td>83.33</td>
</tr>
<tr>
<td></td>
<td>74.24</td>
<td>83.33</td>
</tr>
<tr>
<td></td>
<td>81.82</td>
<td>87.88</td>
</tr>
</tbody>
</table>

Note: JP = Japanese, PN = Pohnpeian

Figure 5.12: Responses to Japanese stimuli as long
Note: JP = Japanese, PN = Pohnpeian

**Figure 5.13:** Responses to Pohnpeian stimuli as long

I conducted a chi-square ($\chi^2$) test for responses to each stimulus to examine if categorizations were made randomly. The results are in Table 5.11.

**Table 5.11:** Result of $\chi^2$ tests in Experiment 6

<table>
<thead>
<tr>
<th>Stimulus Duration</th>
<th>Japanese Listener (n = 17)</th>
<th>Pohnpeian Listener (n = 11)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Japanese stimuli $\chi^2$</td>
<td>Pohnpeian stimuli $\chi^2$</td>
</tr>
<tr>
<td>100</td>
<td>38.333</td>
<td>.001</td>
</tr>
<tr>
<td>130</td>
<td>29.333</td>
<td>.022</td>
</tr>
<tr>
<td>140</td>
<td>19.667</td>
<td>.236</td>
</tr>
<tr>
<td>150</td>
<td>17.333</td>
<td>.364</td>
</tr>
<tr>
<td>154</td>
<td>27.333</td>
<td>.038</td>
</tr>
<tr>
<td>156</td>
<td>7.667</td>
<td>.958</td>
</tr>
<tr>
<td>160</td>
<td>13.000</td>
<td>.673</td>
</tr>
<tr>
<td>166</td>
<td>11.000</td>
<td>.809</td>
</tr>
<tr>
<td>170</td>
<td>27.000</td>
<td>.041</td>
</tr>
<tr>
<td>180</td>
<td>25.667</td>
<td>.059</td>
</tr>
<tr>
<td>200</td>
<td>29.000</td>
<td>.024</td>
</tr>
</tbody>
</table>

$p$ values are boldfaced when responses were made without confusion.

Note: $n =$ the number of participants, boldfaced numbers indicate responses made without confusion.
Japanese listeners categorize vowel length without confusions when the target stimulus was shorter than 140 % stimulus in Japanese and 130 % in Pohnpeian. They were not able to categorize vowel duration without confusion even at the point of 180 % stimulus in Japanese and 200 % in Pohnpeian. On the other hand, Pohnpeian listeners started to display confusion about vowel length at 140 % stimulus in Japanese and 130 % stimulus in Pohnpeian. However, they categorized vowel length without confusion when the durational distinction was greater than 166 % in Japanese and 180 % in Pohnpeian.

5.3.8 Discussion

I must also discuss one obvious factor affecting the results of Experiment 6. When an experiment involves the analysis of the participants' performance, there is always the potential to observe ceiling effects and floor effects (Shearer, 1997). Ceiling effects may be observed when the task is too easy and it is impossible to identify the effects of factors. Floor effects indicate the task is unexpectedly too difficult for some participants. It is obvious that the performance of Pohnpeian participants particularly reflects floor effects. There were 26 Japanese participants and 25 Pohnpeian participants, yet I was able to use the responses from only 17 Japanese and 11 Pohnpeian participants.

Several possibilities may account for this floor effect. First, it was simply a difficult experiment. This might be true, as more than one third of Japanese participants and more than a half of Pohnpeian participants performed poorly. There also was a possibility that the stimuli and the task might not have been natural. Moreover, several participants (both Japanese and Pohnpeian participants) requested to retry the practice session. Second, the environment where Pohnpeian participants participated might have
affected their performance. Although the experiment was conducted in a quiet room, the
effect of the setting in which experiment took place might have been different from the
atmosphere of a computer laboratory on campus. Third, the majority of Pohnpeian
participants were not students at any institute but ordinary residents of Hawai‘i. They
might have felt intimidated about using a computer. It could be possible to improve
experimental design of Experiment 6. The issue of how and where to conduct an
experiment is always an important matter for a researcher to consider.

Results of Experiment 6 show the differences in vowel length categorization
between Japanese and Pohnpeian listeners. Japanese listeners require greater durational
distinctions to discriminate vowel length than Pohnpeian listeners do. Pohnpeian listeners
are sensitive to durational differences; even though the durational distinction is small,
they are able to perceive vowel length as long. As far as durational cues are concerned,
the categorization patterns of vowel length in Experiment 6; that is, when the listeners
perceive vowel length as long, correspond to the results of Experiments 1, 2, and 5.

5.4 General discussion

Results of Experiment 5 showed that Pohnpeian speakers produce vowels
differently from Japanese speakers. The mean duration of Pohnpeian vowels was longer
than that of Japanese vowels; however, durational distinctions between long and short
were much smaller than for Japanese vowels. I conducted Experiment 6 to investigate if
these differences would be reflected in the perception of vowel length. Results of
Experiment 6 indicate that Japanese listeners did categorize vowel length differently from Pohnpeian listeners.

In order for listeners to perceive a vowel as long, the duration of the vowel does not have to be twice as long as a short vowel. Many researchers have already confirmed that the duration of long vowels is often not twice that of short vowels, phonetically. The present perception experiment illustrated that as far as the duration of vowels alone is concerned, when the duration of a vowel is more than 140% of the duration of a short reference vowel, neither Japanese nor Pohnpeian listeners perceive vowel length as short. When the duration of a vowel is longer than 160% stimulus, listeners perceived the vowel as long in most cases. The difference in responses between Japanese listeners and Pohnpeian listeners was that Japanese listeners perceived all stimuli as short more often than Pohnpeian listeners did. Since there were no pitch fall nor quality differences in the stimuli, and the stimuli were semantically neutral in both languages, the listeners had to judge the vowel length only by the durational information in the stimulus. The results from the production experiments of Japanese vowels (Experiment 1 and 2) show that native speakers of Japanese use both duration and pitch fall to mark the vowel length of accented vowels and that when they produce unaccented vowels, in which they cannot use pitch fall, the durational distinction is magnified. Furthermore, the results from the perception experiments (Experiment 3 and 4) indicate that the presence of pitch fall affects the Japanese listeners' perception of vowel length. Considering these results and the results of Experiment 6, it is plausible to conclude that Japanese listeners use both durational cues and pitch cues to categorize vowel length. When pitch fall is not present,
they tend to perceive vowels as short unless duration of a vowel is excessively long. The effect of spectral cues on Pohnpeian listeners’ perception of vowel length needs to be investigated in the future.

The differences in the perception patterns of vowel length between Japanese speakers and Pohnpeian speakers indicate that the speakers of Japanese and Pohnpeian must have developed the phonological knowledge to categorize vowel length by being exposed to their own first language. If so, speakers of a language that does not distinguish phonological vowel length, such as English, do not develop this phonological knowledge. It would be interesting to investigate how speakers of such languages respond to vowel length distinctions.

5.5 Conclusion

In this chapter, I have examined the phonetic properties of Pohnpeian long and short vowels and the cues for vowel length used by Japanese and Pohnpeian listeners. The results from Experiment 5 investigating Pohnpeian vowels show that the distribution of moras in long vowels is phonetically far from isochronous, which complemented the evidence from Japanese vowels. The results from Experiment 6 also suggest that the duration of vowels need not necessarily be twice as long as the duration of short vowels for listeners to categorized vowel length as long. For Tokyo Japanese listeners, it is not just duration but also pitch fall that plays an important role in categorization of vowel length. In the absence of pitch fall, Japanese listeners require longer durations to categorize a vowel as long than Pohnpeian listeners do. Pohnpeian listeners are sensitive
to durational distinction between long and short vowels, as the durational distinction between long and short vowels is stable across various speech rates. They do not require as large durational distinction to perceive vowels as long.

The results of the production experiments using Japanese (Experiments 1 and 2) and Pohnpeian (Experiment 5) reveal that the phonetic properties of long vowels differ between these two languages; even though the vowels are phonologically categorized into long and short in both languages. The results from the cross-linguistic perception experiment (Experiment 6) appear to reflect these acoustic differences in the vowels of each language. Both Japanese and Pohnpeian listeners require a certain amount of durational distinction to categorize vowel length, but Japanese listeners seem to need a greater distinction between long and short to accurately determine the vowel length.
6.1 Summary

In this dissertation, I investigated the phonetic properties of long and short vowels in Japanese and Pohnpeian, focusing on duration and pitch movement and how speakers of each language use these acoustic cues for the categorization of vowel length. In both Japanese and Pohnpeian, vowel length is phonologically contrastive. Acoustic investigations revealed that the way speakers produce vowels was different between Japanese and Pohnpeian. Perceptual investigations showed that the way speakers perceive vowel length reflects the acoustic properties of their own language.

Japanese speakers use both duration and pitch-fall to mark vowel length when they produce accented vowels. The two phonological features – vowel length and pitch-accent – are realized acoustically as vowel duration and adequate amount and position of pitch-fall. At various speech rates, Japanese speakers are able to control these two phonetic properties to mark vowel length correctly. When Japanese speakers produce unaccented long vowels in which F0 values should not be declined, the speakers increase the durational distinction between long and short vowels. This durational distinction increases as speech rate decreases. In slow speech, the duration of all segments in an utterance may increase. Japanese speakers increase the durational distinction between long and short vowels to a greater degree in slower speech, as if to ensure that long
vowels will be perceived as long. Results of Experiments 1 and 2 suggest that both duration and F0 are important acoustic properties in Japanese long vowels.

Pohnpeian speakers, on the other hand, do not use F0 information to discriminate vowel length. However, the durational difference between short and long vowels is much smaller, compared to the distinction in Japanese vowels. Whether Pohnpeian speakers use other phonetic properties, such as spectral cues, to mark vowel length is a question that requires further investigation.

The phonetic properties of long vowels in each language were reflected in the patterns of listeners' perception of vowel length. Results from two perception experiments (Experiments 3 and 4) using accented vowels in Japanese showed that durational cues alone were not robust enough for Japanese listeners to determine vowel length. The Japanese listeners took longer to categorize the vowel length of ambiguous vowels whose duration was equivalent to long vowels and whose pitch movement was identical to that of short vowels. Moreover, their categorization of ambiguous vowels was randomly made. If the listeners had discriminated vowel length based entirely on durational cues, all ambiguous vowels would have been categorized as long. Likewise, if the listeners had categorized vowel length based solely on pitch cues, all ambiguous vowels would have been categorized as short. However, their responses were at the chance level. Japanese listeners require an adequate pitch contour to judge vowel length as long in accented long vowels. Results of Experiments 3 and 4 suggest that pitch-fall is an important cue for Japanese listeners' perception of vowel length. Furthermore, Experiment 6 was to investigate the way listeners categorize vowel length based solely
on durational cues. The stimuli did not have pitch cues that could mark vowel length. The results show that Japanese listeners were not able to categorize vowels without confusions even when vowel duration was more than 180% longer than that of a short reference vowel, which suggests that Japanese listeners have a tendency to perceive vowel length as short without an adequate pitch contour or an enhanced durational distinction.

As far as the duration of vowels is concerned, Pohnpeian speakers do not produce as large a durational distinction between long and short vowels as Japanese speakers do. Results of Experiment 6 showed that Pohnpeian speakers did not perceive vowel length as short when the duration of vowels were 140% longer than a short reference vowel, which roughly corresponds to the durational distinction observed in Experiment 5 (ratios of the mean duration of long vowels to short vowels was 1.30 : 1.00 in fast speech, 1.47 : 1.00 in slow speech).

6.2 Residual questions

This series of experiments brought out issues that are worth exploring empirically in order to deepen our understanding of the Japanese and Pohnpeian phonological systems. I have investigated the phonetic properties of the mora as realized in vowels in Japanese and Pohnpeian. Results from the production experiments indicate that Japanese speakers use both duration and amount of F0 fall to mark the length distinction of vowels. The mora is also realized in coda consonants in Japanese. Some researchers have investigated the durational aspect of Japanese coda consonants (Aoyama, 2001; Sato,
1998). However, the interaction of duration and pitch in Japanese consonants has not been the subject of an investigation. It would be interesting to see if Japanese speakers change the way they produce coda consonants according to the presence of pitch-accent, such in the minimal pair *set.ta* (HH.L) ‘compete-PAST’ vs. *set.ta* (HH.H) ‘Japanese slippers.’

Second, the question of whether pitch-accent makes the duration of Japanese vowels longer remains unanswered. Two opposite views have been suggested on this issue, as I discussed in Chapter 3. In order for speakers to produce an accented vowel correctly, speakers must lower F0 by a certain amount, and at the right position within a vowel. It may be articulatorily natural that speakers take longer to produce an accented vowel, because they need to make the pitch drop. However, some researchers argue that pitch-accent has no effect on the duration of vowels. Results from Experiments 1 and 2 show that pitch-accent has an effect on the duration of vowels. The mean duration of accented short vowels was longer than the mean duration of unaccented short vowels. If we compare only the mean duration of short vowels, it seems that pitch-accent makes vowel duration longer. However, the mean duration of unaccented long vowels in slow speech was longer than the mean duration of accented long vowels in slow speech. The relative duration of unaccented long vowels was especially increased and the duration of unaccented short vowels was reduced in slow speech. Although the result of statistical analyses indicates that pitch-accent affects the mean duration of vowels, due to the effect of speech rate on the duration of vowels, I cannot conclude that pitch-accent makes the
duration of vowels longer if vowel length is short and shorter if vowel length is long. This question also remains unanswered.

As far as the Pohnpeian mora is concerned, the question of spectral cues for vowel length is worth a thorough investigation. The origin of long vowels, at least in nouns, in Pohnpeian is partly related to the stress assignment system that the language used to have. It is a well-accepted fact that stress changes the quality and duration of a vowel in English, for example. The length distinction in Old English vowels shifted to a quality distinction phonologically. In addition, Gussenhoven & Driessen (2004) report that long high vowels are in a process of diphthongization in dialects in the Dutch and the Belgian province of Limburg. Durational distinctions in vowels could be changed to quality distinctions, or vice versa. Therefore, it is reasonable to expect to see quality differences between long vowels and short vowels in Pohnpeian.

6.3 Implication for the phonetics-phonology interface

The notion of the mora and thus moraic theory has been developed to represent prosodic phenomena. The mora is a prosodic unit commonly used to represent phonological length, the site of stress/pitch-accent, and the tone/pitch-bearing unit. The series of experiments in this dissertation revealed differences in phonetic details of the mora realized in vowels in Japanese and Pohnpeian. At the prosodic level, the mora in Japanese has two functions – as a unit of phonological length and as a potential pitch-bearing unit. In Pohnpeian, it is a unit of phonological length. The mora is used to represent phonological length in both languages; however, the distribution of moras is not
only far from isochronous, but also the way listeners perceive mora counts varies, depending on the language. When pitch cues are available, Japanese speakers use not only duration but also pitch cues in their perception of the mora count. For Pohnpeian speakers, the perception of the mora count in vowels may relate to quality differences as well as duration.

The mora is an important prosodic unit in other Oceanic languages, as well. Many Oceanic languages are mora-timed languages. In Tongan, a Polynesian language, the duration of long vowels in citation speech is twice as long as that of short vowels. Moreover, when a vowel bears definitive accent that denotes the attached noun phrase is referential, the duration of short vowels doubles, and the duration of long vowels increases by 30% (Anderson & Otsuka, 2003). In Tongan, the duration of the mora is affected by a syntactic function (definitive accent), just as the duration of the Japanese mora is influenced by a lexical function (pitch-accent).

Since *The Sound Pattern of English (SPE)* (Chomsky & Halle, 1968), a handful of scholars have been arguing about the importance of phonetics in phonology (Stampe, 1972; Ohala, 1974, 1990; Donegan & Stampe, 1979; Pierrehumbert, 1980; Keating, 1988, 1991; Browman & Goldstein, 1992; Myers, 2000). Their central claim is that phonetics and phonology are not autonomous or independent fields; rather these two fields must be integrated. In her *Evolutionary Phonology*, Blevins (2004) argues that common sound patterns typically result from common phonetically motivated sound change and that sound patterns in languages are learned from language-specific phonetic signals. In this dissertation, I have described some differences in phonetic properties between Japanese
and Pohnpeian vowels. The mora, a phonological notion, is psychologically real in mora-
timed languages. However, the physical realization of the mora varies depending on the
language; and even within a language, the physical realization of the mora varies
according to the functions that the mora carries. These findings support the claim that our
phonological knowledge emerges from or is learned by being exposed to a speech
environment. In order to understand how speakers develop phonological knowledge
internally, we must give our full attention to phonetic properties.
APPENDIX A: Materials used for Experiment 1

Japanese accented vowels

### Long vowel /a/

- さっき R と言いました。
  *sakki aaru to iimashita*  
  ‘I said aaru a while ago’
- そのカード ふいて。
  *sono kaando fuite*  
  ‘Please you clean that card’
- 私は ナースが好きです。
  *watashi-wa naasu ga sukidesu*  
  ‘I like nurses’
- 昨日おばあさんに会った。
  *kinoo obasan-ni atta*  
  ‘I met my grandmother yesterday’
- この人がおかあさんです。
  *kono hito-ga okaasan desu*  
  ‘This is the mother’

### Short vowel /a/

- さっき 有ると言いました。
  *sakki aru to iimashita*  
  ‘I said aru a while ago’
- その角 ふいて。
  *sono kado fuite*  
  ‘Please wipe that corner’
- 私は なすが好きです。
  *watashi-wa nasu ga sukidesu*  
  ‘I like eggplants’
- 昨日おばあさんに会った。
  *kinoo obasan-ni atta*  
  ‘I met my untie yesterday’
- この人が 岡さんです。
  *kono hito-ga okasan desu*  
  ‘This is Mr. Oka.’

### /e/

- かわいいベルだね。
  *kawai beru na ne*  
  ‘What a cute bell!’
- たくさん映画見たい。
  *takusan eega mitai*  
  ‘I want to watch a lot of movies’
- これは ‘K’です。
  *kore-wa kee desu*  
  ‘this is a letter k’
- 私が 零時に行く。
  *watashi-ga reeji-ni iku*  
  ‘I will go there at midnight’
- 去年も生徒を訪ねた。
  *kyonen-mo seeto-o tazuneta*  
  ‘I went to visit a student last year, too’
このビールが高い。
kono biru-ga takai
‘This beer is expensive’

あのチーズが欲しい。
ano chizu-ga hoshii
‘I want that cheese’

* 去年も飯坂屋に行った。
kyonen-no iizakaya-ni itta
‘We went to izakaya last year, too’

昨日 兄さんが 6 と答えた。
kinoo nisan-ga roku-to kotaeta
‘my brother said six yesterday’

* おじいさんが来るよ。
ojisan-ga kuru yo
‘Grandpa is coming’

これが 同期の桜です。
kore-ga doki-no sakura desu
‘This is my former schoolmate’

それが好意のあらわれです。
sore-ga koi-no araware desu
‘That is her/his kindness’

あれが 大野さんです。
are-ga Oono-san desu
‘That is Mr. Ohno’

山道を老婆が歩いている。
yamamichi-o roba-ga aruiteiru
‘An old lady is walking on a mountain path’

その通りです。
sono tori desu
‘That is right’

* これが 土器の桜です。
kore-ga doki-no sakura desu
‘These are earthenware cherry blossoms’

それが恋のあらわれです。
sore-ga koi-no araware desu
‘That is the love’

* あれが 小野さんです。
are-ga Ono-san desu
‘That is Mr. Ono’

山道をロバが歩いている。
yamamichi-o roba-ga aruiteiru
‘A donkey is walking on a mountain path’

* その鳥です。
sono tori desu
‘It is that bird’
きれいな空気だね。
kireina kuuki da ne
‘It’s clean air, isn’t it?’

今 クールと言いました。
ima kuuru to iimashita
‘I said cool now’

ジョンが 数理にやられた。
Jon-ga suuri-ni yarareta
‘John failed in Math theory’

あの人が ゆうかさんです。
ano hito-ga Yuuka san desu
‘That person is Yuuka’

君を ‘ゆうき’ と呼ぼう。
kimi-o Yuuki-to yoboo
‘Let’s call you Yuuki’
Appendix B: Material sentences used for Experiment 2

**Japanese unaccented vowels**

Carrier sentence: 今____と言いました。

*ima ______-to iimashita* ‘I said ______ now.’

<table>
<thead>
<tr>
<th><strong>Long vowel</strong></th>
<th><strong>Short vowel</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/</td>
<td>/a/</td>
</tr>
<tr>
<td>赤鉾 (akaari)</td>
<td>‘red ants’</td>
</tr>
<tr>
<td>持握 (haaku)</td>
<td>‘to grasp’</td>
</tr>
<tr>
<td>カーキ色 (kaaki-iro)</td>
<td>‘kahki’</td>
</tr>
<tr>
<td>邪悪 (jaaku)</td>
<td>‘evil’</td>
</tr>
<tr>
<td>スーパーマーケット (suupaamaaketto)</td>
<td>‘grocery store’</td>
</tr>
<tr>
<td>/e/</td>
<td>/e/</td>
</tr>
<tr>
<td>成功 (seekoo)</td>
<td>‘success’</td>
</tr>
<tr>
<td>カレーライス (kareeraisu)</td>
<td>‘rice curry’</td>
</tr>
<tr>
<td>傾向 (keekoo)</td>
<td>‘tendency’</td>
</tr>
<tr>
<td>メーキャップ (meekyappu)</td>
<td>‘makeup’</td>
</tr>
<tr>
<td>トレーナー (toreena)</td>
<td>‘sweatshirt’</td>
</tr>
<tr>
<td>/i/</td>
<td>/i/</td>
</tr>
<tr>
<td>瞅展目 (hiikime)</td>
<td>‘with favor’</td>
</tr>
<tr>
<td>消印 (kishiin)</td>
<td>‘post mark’</td>
</tr>
</tbody>
</table>

175
<table>
<thead>
<tr>
<th>Character</th>
<th>Pinyin</th>
<th>English Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>新潟</td>
<td>niigata</td>
<td>‘Niigata prefecture’</td>
</tr>
<tr>
<td>敷設</td>
<td>shiku</td>
<td>‘to breed’</td>
</tr>
<tr>
<td>置き石</td>
<td>okiishi</td>
<td>‘mile stone’</td>
</tr>
<tr>
<td>スキー場</td>
<td>sukijio</td>
<td>‘ski slope’</td>
</tr>
<tr>
<td>放送</td>
<td>hoosoo</td>
<td>‘broadcast’</td>
</tr>
<tr>
<td>狹率</td>
<td>kooritsu</td>
<td>‘efficiency’</td>
</tr>
<tr>
<td>孔子</td>
<td>kooshi</td>
<td>‘Confucius’</td>
</tr>
<tr>
<td>総会</td>
<td>sookai</td>
<td>‘general meeting’</td>
</tr>
<tr>
<td>東海</td>
<td>tookai</td>
<td>‘(a district name)’</td>
</tr>
<tr>
<td>盲目</td>
<td>moomoku</td>
<td>‘blind’</td>
</tr>
<tr>
<td>風潮</td>
<td>fuuchoo</td>
<td>‘trend’</td>
</tr>
<tr>
<td>空地</td>
<td>kuuchi</td>
<td>‘empty lot’</td>
</tr>
<tr>
<td>通信販売</td>
<td>tsuushin hanbai</td>
<td>‘mail order’</td>
</tr>
<tr>
<td>数式</td>
<td>suushiki</td>
<td>‘math formula’</td>
</tr>
<tr>
<td>誘拐</td>
<td>yuukai</td>
<td>‘kidnap’</td>
</tr>
</tbody>
</table>
APPENDIX C: Materials used for Experiment 3

Existing words

Long vowel

1. sakki aaru to iimashita
   ‘I said aaru while ago’

2. sono kaado fuite
   ‘Please you clean that card’

3. watashi-wa naasu ga sukidesu
   ‘I like nurses’

4. kawai beeru na ne
   ‘What a cute bell!’

5. kore-wa kee desu
   ‘this is a letter k’

6. watashi-ga reeji-ni iku
   ‘I will go there at midnight’

7. kyonen-mo seeto-o tazuneta
   ‘I went visit a student last year, too’

8. kono biiru-ga takai
   ‘This beer is expensive’

9. ano chiizu-ga hoshii
   ‘I want thatcheeze’

10. kinoo nisan-ga roku-to kotaeta
    ‘my brother said six yesterday’

Short vowel

1. sakki aru to iimashita
   ‘I said aru while ago’

2. sono kado fuite
   ‘Please wipe that corner’

3. watashi-wa nasu ga sukidesu
   ‘I like egg-plants’

4. kawai beru na ne
   ‘What a pretty veil!’

5. kore-wa ke desu
   ‘this is ke’

6. watashi-ga reji-ni iku
   ‘I will go to the cashier’

7. kyonen-mo Seto-o tazuneta
   ‘I went to Seto last year, too’

8. kono biru-ga takai
   ‘This building is high’

9. ano chizu-ga hoshii
   ‘I want that map’

10. kinoo nisan-ga roku-to kotaeta
    ‘I said two times six was yesterday’

11. kore-ga dooki-no sakura desu
    ‘This is my former schoolmate’

12. sore-ga koi-no araware desu
    ‘That is her/his kindness’

13. yamamichi-o rooba-ga oruiteiru
    ‘An old lady is walking in a mountain path’
14. sono toori desu
   'That is right'

15. kireina kuuki da ne
    'It’s clean air, isn’t it?'

16. ima kuru to iimashita
    'I said cool now'

17. Jon-ga suuri-ni yarareta
    'John failed in Math theory'

18. ano hito-ga Yuuka san desu
    'That person is Yuuka'

Visual stimuli

1. R  有る
2. カード 角
3. ナース なす
4. ペル ペール
5. K  け
6. 零時 レジ
7. 生徒 瀬戸
8. ビール ビル
9. チーズ 地図
10. 兄さん 2 x 3
11. 同期 土器
12. 好意 恋
13. 老婆 ロバ
14. 通り 鳥
15. 空気 茎
16. クール 来る
17. 数理 スリ
18. ゆうか ゆか
### Practice trials

<table>
<thead>
<tr>
<th>Audio stimuli</th>
<th>Visual stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. <em>kinoo uma-ni norimashita</em></td>
<td>馬</td>
</tr>
<tr>
<td>‘I did a horse-back riding yesterday’</td>
<td><em>uma</em> ‘a horse’</td>
</tr>
<tr>
<td>2. <em>sore-wa kyuuri desu</em></td>
<td>キーワイ</td>
</tr>
<tr>
<td>‘It’s a cucumber’</td>
<td><em>kiiwui</em> ‘kiwi’</td>
</tr>
<tr>
<td>3. <em>ima ii-to iimashita</em></td>
<td>E</td>
</tr>
<tr>
<td>‘I said E now’</td>
<td><em>ii</em> ‘letter e’</td>
</tr>
<tr>
<td>4. <em>koko-wa pattaa-de kimemasu</em></td>
<td>パター</td>
</tr>
<tr>
<td>‘I will finish it with a good patter’</td>
<td><em>bataa</em> ‘butter’</td>
</tr>
</tbody>
</table>

### Fillers

#### Type 1: Geminate vs. Singleton
(pitch contour is HL and HLL)

<table>
<thead>
<tr>
<th>Audio stimuli</th>
<th>Visual stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. <em>knarazu kitte kudasai</em></td>
<td>切って</td>
</tr>
<tr>
<td>‘Please don’t forget to cut (it).’</td>
<td><em>kitte</em> ‘cut’</td>
</tr>
<tr>
<td>2. <em>sore-ga kata to omotta-no</em></td>
<td>勝った</td>
</tr>
<tr>
<td>‘I thought it’s a shoulder.’</td>
<td><em>katta</em> ‘win-PAST’</td>
</tr>
<tr>
<td>3. <em>suteeji-ni makku ga oritekita</em></td>
<td>幕</td>
</tr>
<tr>
<td>‘Mac came down to the stage.’</td>
<td><em>maku</em></td>
</tr>
<tr>
<td>4. <em>ima sakku-to iimashita</em></td>
<td>咲く</td>
</tr>
<tr>
<td>‘(I) like this slope.’</td>
<td><em>saka</em> ‘bloom’</td>
</tr>
<tr>
<td>5. <em>kono saka-ga sukidesu</em></td>
<td>坂</td>
</tr>
<tr>
<td>‘(I) like this slope’</td>
<td><em>saka</em> ‘slope’</td>
</tr>
<tr>
<td>6. <em>ima futa to iimashita</em></td>
<td>ふた</td>
</tr>
<tr>
<td>‘I said a lid now.’</td>
<td><em>futa</em> ‘lid’</td>
</tr>
</tbody>
</table>

#### Type 2: Voicing
(pitch contour is HL)

<table>
<thead>
<tr>
<th>Audio stimuli</th>
<th>Visual stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. <em>ima futa to iimashita</em></td>
<td>豚</td>
</tr>
<tr>
<td>‘I said a lid now.’</td>
<td><em>futa</em> ‘lid’</td>
</tr>
</tbody>
</table>
7. *Ichiro-ga ginmedaru-o totta*  
Ichiro won the silver.'

8. *sokono kushi totte*  
Pass the comb'

9. *kore-wa pataa de kimemasu*  
'Finish with a patter'

10. *tsui guchi-ga deta*  
'(I) complained unintentionally.'

<table>
<thead>
<tr>
<th>Type 3: Pitch-accent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Audio stimuli</strong></td>
</tr>
</tbody>
</table>
| 11. *kinoo kaki-o tabeta*  
'(I) ate oysters yesterday' | 牡蠣  
kaki-o [HLL]  
'persimmon-NOM' |
| 12. *kanojyo-wa kan-ga ii*  
'She favors can (beer)' | 勸  
kan-ga [HHH]  
'sixth sense-NOM' |
| 13. *sono hashi-o mite kudasai*  
'Please look at the edge.' | 端  
hashi-o [HHH]  
'bridge-ACC' |

<table>
<thead>
<tr>
<th>Type 4: Contrasting segment</th>
</tr>
</thead>
<tbody>
<tr>
<td>(same pitch contour)</td>
</tr>
<tr>
<td><strong>Audio stimuli</strong></td>
</tr>
</tbody>
</table>
| 14. *ano oka-de hana-o [LHL] mita*  
'(I) saw a flower on the hill’ | 穴  
an 'hole'  
hana 'flower' |
| 15. *sono naka-ni an-ga [HLL] haitte imasu*  
'There is a read-bean paste inside.' | 判  
han 'stamp'  
an 'read bean paste' |
| 16. *kaigan-o hadashi-de [LHHH] aruita*  
'(I) walked on the beach with bear foot.' | はだか  
hadaka 'naked'  
はだし 'bear foot' |
17. *sore-wa okashii desu*
   'That is funny.'

18. *Taro-wa mushi-ga [LHH] suki desu*
   'Taro likes insects'
## APPENDIX D: Materials used for Experiment 4

### Nonsense words

<table>
<thead>
<tr>
<th>Long vowel</th>
<th>Short vowel</th>
<th>Visual stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/</td>
<td>/a/</td>
<td></td>
</tr>
<tr>
<td>1. baapi</td>
<td>1. bapi</td>
<td>バーピ</td>
</tr>
<tr>
<td>2. daamu</td>
<td>2. danu</td>
<td>ダーヌ</td>
</tr>
<tr>
<td>3. gaage</td>
<td>3. gage</td>
<td>ガーゲ</td>
</tr>
<tr>
<td>4. made</td>
<td>4. made</td>
<td>マーデ</td>
</tr>
<tr>
<td>5. nade</td>
<td>5. nade</td>
<td>ナーデ</td>
</tr>
<tr>
<td>6. zage</td>
<td>6. zage</td>
<td>ザーデ</td>
</tr>
<tr>
<td>/e/</td>
<td>/e/</td>
<td></td>
</tr>
<tr>
<td>7. beepo</td>
<td>7. bepo</td>
<td>ベーポ</td>
</tr>
<tr>
<td>8. deedo</td>
<td>8. dedo</td>
<td>デード</td>
</tr>
<tr>
<td>9. geebi</td>
<td>9. gebi</td>
<td>ゲービ</td>
</tr>
<tr>
<td>10. meego</td>
<td>10. mego</td>
<td>メーボ</td>
</tr>
<tr>
<td>11. neepi</td>
<td>11. nepi</td>
<td>ネーピ</td>
</tr>
<tr>
<td>12. zeete</td>
<td>12. zete</td>
<td>ゼーテ</td>
</tr>
<tr>
<td>/i/</td>
<td>/i/</td>
<td></td>
</tr>
<tr>
<td>13. biida</td>
<td>13. bida</td>
<td>ビーダ</td>
</tr>
<tr>
<td>14. giizo</td>
<td>14. gizo</td>
<td>ギーツ</td>
</tr>
<tr>
<td>15. jiimo</td>
<td>15. jimo</td>
<td>ジーモ</td>
</tr>
<tr>
<td>16. jiipe</td>
<td>16. jipe</td>
<td>ジーペ</td>
</tr>
<tr>
<td>17. mide</td>
<td>17. mide</td>
<td>ミーデ</td>
</tr>
<tr>
<td>18. niima</td>
<td>18. nima</td>
<td>ニーマ</td>
</tr>
<tr>
<td>/o/</td>
<td>/o/</td>
<td></td>
</tr>
<tr>
<td>19. boopu</td>
<td>19. bopu</td>
<td>ボーブ</td>
</tr>
<tr>
<td>20. doode</td>
<td>20. dode</td>
<td>ドーデ</td>
</tr>
<tr>
<td>21. goome</td>
<td>21. gome</td>
<td>ゴーメ</td>
</tr>
<tr>
<td>22. mooso</td>
<td>22. moso</td>
<td>モーヌ</td>
</tr>
<tr>
<td>23. noozo</td>
<td>23. nozo</td>
<td>ノーヌ</td>
</tr>
<tr>
<td>24. zoope</td>
<td>24. zope</td>
<td>ゾーブ</td>
</tr>
<tr>
<td>/u/</td>
<td>/u/</td>
<td></td>
</tr>
<tr>
<td>25. guumo</td>
<td>25. gumo</td>
<td>グーモ</td>
</tr>
<tr>
<td>26. muapa</td>
<td>26. mupa</td>
<td>ムーパ</td>
</tr>
<tr>
<td>27. muopo</td>
<td>27. mupo</td>
<td>ムーポ</td>
</tr>
<tr>
<td>28. mude</td>
<td>28. mude</td>
<td>ムーデ</td>
</tr>
<tr>
<td>29. muuchi</td>
<td>29. nuichi</td>
<td>ヌーチ</td>
</tr>
<tr>
<td>30. zuuma</td>
<td>30. zuma</td>
<td>ツーマ</td>
</tr>
</tbody>
</table>
### Practice trials

<table>
<thead>
<tr>
<th>Audio stimuli</th>
<th>Visual stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>aaru</td>
<td>aaru aru</td>
</tr>
<tr>
<td>betta</td>
<td>beta betta</td>
</tr>
<tr>
<td>paze</td>
<td>paze pase</td>
</tr>
<tr>
<td>kufu</td>
<td>kufu kusu</td>
</tr>
</tbody>
</table>

### Fillers

**Type 1: Geminate vs. Singleton**

(pitch contour: HL or HLL)

<table>
<thead>
<tr>
<th>Audio stimuli</th>
<th>Visual stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>mapa</td>
<td>mapa mappa</td>
</tr>
<tr>
<td>nakki</td>
<td>naki nakki</td>
</tr>
<tr>
<td>zake</td>
<td>zake zakke</td>
</tr>
<tr>
<td>mecci</td>
<td>mecci mecci</td>
</tr>
<tr>
<td>nekka</td>
<td>nekka neka</td>
</tr>
<tr>
<td>zete</td>
<td>zete zette</td>
</tr>
<tr>
<td>mipo</td>
<td>mipo mipo</td>
</tr>
<tr>
<td>nicci</td>
<td>nicci nicci</td>
</tr>
<tr>
<td>zippe</td>
<td>zippe zippe</td>
</tr>
<tr>
<td>Audio stimuli</td>
<td>Visual stimuli</td>
</tr>
<tr>
<td>---------------</td>
<td>----------------</td>
</tr>
<tr>
<td>1. bate</td>
<td>パデ bade</td>
</tr>
<tr>
<td>2. dapa</td>
<td>タバ tapa</td>
</tr>
<tr>
<td>3. gago</td>
<td>ガゴ gago</td>
</tr>
<tr>
<td>4. bepa</td>
<td>ベパ beba</td>
</tr>
<tr>
<td>5. gepi</td>
<td>デベ gepi</td>
</tr>
<tr>
<td>6. tepe</td>
<td>テペ depe</td>
</tr>
<tr>
<td>7. bipe</td>
<td>ビベ pipe</td>
</tr>
<tr>
<td>8. gizo</td>
<td>ジソ gizo</td>
</tr>
<tr>
<td>9. zisu</td>
<td>シス shisu</td>
</tr>
</tbody>
</table>

**Type 2: Voicing**

(pitch contour: HL)
<table>
<thead>
<tr>
<th>No.</th>
<th>Japanese</th>
<th>Kanji</th>
<th>Romaji</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.</td>
<td>buge</td>
<td>ブゲ</td>
<td>fuge</td>
</tr>
<tr>
<td>11.</td>
<td>gude</td>
<td>グデ</td>
<td>gute</td>
</tr>
<tr>
<td>12.</td>
<td>mose</td>
<td>モゼ</td>
<td>mose</td>
</tr>
<tr>
<td>13.</td>
<td>noza</td>
<td>ノザ</td>
<td>nosa</td>
</tr>
<tr>
<td>14.</td>
<td>sope</td>
<td>ソペ</td>
<td>zope</td>
</tr>
<tr>
<td>15.</td>
<td>zupu</td>
<td>ズブ</td>
<td>supu</td>
</tr>
</tbody>
</table>
APPENDIX E: Material words used for Experiment 5

**Pohnpeian vowels**

Carrier sentence: *Ia wehwehn _____ ni lokaihn Pohnpei*

/\ja wehwe: n _____ ni lokai:hn pompej/  
‘What is the meaning of _____ in Pohnpeian?’.

<table>
<thead>
<tr>
<th>Long vowel</th>
<th>Short vowel</th>
</tr>
</thead>
<tbody>
<tr>
<td>/a/</td>
<td>/a/</td>
</tr>
<tr>
<td>kahng</td>
<td>kang</td>
</tr>
<tr>
<td>/kaŋ/</td>
<td>/kaŋ/</td>
</tr>
<tr>
<td>mahs</td>
<td>aramas</td>
</tr>
<tr>
<td>/mas/</td>
<td>/aramas/</td>
</tr>
<tr>
<td>pahpa</td>
<td>pap</td>
</tr>
<tr>
<td>/papa/</td>
<td>/pap/</td>
</tr>
<tr>
<td>sahl</td>
<td>lal</td>
</tr>
<tr>
<td>/sail/</td>
<td>/lal/</td>
</tr>
<tr>
<td>/e/</td>
<td>/e/</td>
</tr>
<tr>
<td>pehi</td>
<td>pei</td>
</tr>
<tr>
<td>/pej/</td>
<td>/pej/</td>
</tr>
<tr>
<td>/et/</td>
<td>/e/</td>
</tr>
<tr>
<td>neh</td>
<td>ne</td>
</tr>
<tr>
<td>/ne/</td>
<td>/ne/</td>
</tr>
<tr>
<td>/i/</td>
<td>/i/</td>
</tr>
<tr>
<td>ihd</td>
<td>id</td>
</tr>
<tr>
<td>/it/</td>
<td>/it/</td>
</tr>
<tr>
<td>kihd</td>
<td>kid</td>
</tr>
<tr>
<td>/kit/</td>
<td>/kit/</td>
</tr>
<tr>
<td>lih</td>
<td>lil</td>
</tr>
<tr>
<td>/lii/</td>
<td>/lii/</td>
</tr>
<tr>
<td>pilh</td>
<td>pil</td>
</tr>
<tr>
<td>/pil/</td>
<td>/pil/</td>
</tr>
<tr>
<td>/p&quot;i:k/</td>
<td>'pig'</td>
</tr>
</tbody>
</table>

186
| /dohl/       | ‘mountain’       | /dol/       | ‘to mix’       |
| /to:l/       |                 | /tol/       |                 |
| /kohri/      | ‘ice’           | /korila/    | ‘gorilla’      |
| /koril/      |                 | /litok/     | ‘hen’          |
| /tohto/      | ‘many’          | /litok/     |                 |
| /tohtol/     |                 |             |                 |
| /poahd/      | ‘individual planting’ | /poad/ | ‘to be planted’ |
| /poahrd/     | ‘to wipe’       | /poar/      |                 |
| /poahr/      |                 |             | (classifier)   |
| /poar/       | ‘wound’         | /soan/      | ‘aligned’      |
| /soahn/      |                 | /soan/      |                 |
| /poahsoan/   | ‘foundation’    | /poasen kaung/ | ‘capital’ |
| /poasen/     |                 |             |                 |
| /lus/        | ‘to lose’       | /lus/       | ‘to jump’      |
| /lusu/       |                 | /p’wung/    | ‘correct’      |
| /p’wuhng/    | ‘rights’        | /p’wen/     |                 |
| /uluwl/      | ‘pillow’        | /lul/       | ‘to flame’     |
|             |                 |             |                 |
REFERENCES


