The Production and Perception of Japanese Pitch Accent
The Production and Perception of Japanese Pitch Accent

By

Yukiko Sugiyama
# Table of Contents

Acknowledgments ................................................................................................. vii

Abstract ................................................................................................................ ix

Chapter One ........................................................................................................... 1
Introduction
  1.1 Japanese Pitch Accent ................................................................................. 3
    1.1.1 Early Studies ......................................................................................... 3
    1.1.2 Japanese Pitch Accent ......................................................................... 9
  1.2 Distribution of Accent Patterns ................................................................. 11
    1.2.1 Accent Pattern of Loanwords ............................................................. 12
    1.2.2 Final-accented and Unaccented Patterns ............................................. 13
  1.3 Acoustic Properties of Accent ................................................................. 14
    1.3.1 Devoicing and Accent ........................................................................ 14
    1.3.2 Other Correlates of Pitch Accent than F₀ ......................................... 16

Chapter Two ....................................................................................................... 19
Production Study
  2.1 Introduction ............................................................................................... 19
    2.1.1 Pierrehumbert and Beckman (1988) ................................................. 20
    2.1.2 Minimal Pairs Produced in Isolation ............................................... 23
    2.1.3 Accented and Unaccented Words in Phrase .................................... 25
  2.2 Production Experiment ............................................................................. 28
    2.2.1 Method .............................................................................................. 28
    2.2.2 Results and Discussion .................................................................... 37
  2.3 Summary of the Production Study .......................................................... 52

Chapter Three .................................................................................................... 55
Perception Study
  3.1 Introduction .............................................................................................. 55
    3.1.1 Minimal Pair Identification Using Natural Speech ......................... 56
    3.1.2 Perception Experiments Using Controlled Stimuli ......................... 61
  3.2 Perception Experiment ............................................................................. 64
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2.1 Method</td>
<td>66</td>
</tr>
<tr>
<td>3.2.2 Results &amp; Discussion</td>
<td>73</td>
</tr>
<tr>
<td>3.3 Summary of the Perception Study</td>
<td>93</td>
</tr>
<tr>
<td>Chapter Four</td>
<td>99</td>
</tr>
<tr>
<td>General Discussion and Conclusions</td>
<td></td>
</tr>
<tr>
<td>4.1 Words in Isolation</td>
<td>100</td>
</tr>
<tr>
<td>4.2 Words in Sentence</td>
<td>100</td>
</tr>
<tr>
<td>4.3 Implications and Future Directions</td>
<td>103</td>
</tr>
<tr>
<td>4.3.1 Production</td>
<td>103</td>
</tr>
<tr>
<td>4.3.2 Perception</td>
<td>105</td>
</tr>
<tr>
<td>4.4 Word Prosody: What is Pitch Accent?</td>
<td>106</td>
</tr>
<tr>
<td>Appendix A</td>
<td>109</td>
</tr>
<tr>
<td>Word List</td>
<td></td>
</tr>
<tr>
<td>Appendix B</td>
<td>111</td>
</tr>
<tr>
<td>Word List with Familiarity Ratings</td>
<td></td>
</tr>
<tr>
<td>Appendix C</td>
<td>113</td>
</tr>
<tr>
<td>Word List in Japanese</td>
<td></td>
</tr>
<tr>
<td>Appendix D</td>
<td>115</td>
</tr>
<tr>
<td>Frame Sentences</td>
<td></td>
</tr>
<tr>
<td>Appendix E</td>
<td>117</td>
</tr>
<tr>
<td>Frame Sentences in Japanese</td>
<td></td>
</tr>
<tr>
<td>References</td>
<td>119</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

As always, a thesis such as this is the result of the efforts of many people over the course of time. I would like to thank my advisors Dr. Karin Michelson and Dr. James Sawusch for their patience and guidance on all phases of this dissertation. I am also grateful to my dissertation committee member, Dr. Tsan Huang, for her advice, encouragement, and suggestions. My dissertation would not have been on this topic if I had not met Dr. Sawusch. His insightful comments have always inspired my curiosity to the wonders of speech. I am also thankful to him for the generous use of his laboratory facility. Many fruitful and stimulating discussions with Dr. Michelson have opened my eyes to the profound questions that theoretical, phonological linguistics tries to address. Dr. Huang provided me with thorough and careful comments on a draft of my dissertation.

I would also like to thank Dr. Doug Roland for his assistance with Perl scripts and Dr. Mitsu Shimojo for helping me recruit native speakers of Tokyo Japanese living in Buffalo.

I owe my thanks to Dr. Haruo Kubozobo at Kobe University for helping me design the production experiment at an early stage of my dissertation research.

I am grateful to my colleagues in the Linguistics department and Speech Research Laboratory in the Psychology department. They supported me emotionally and academically. To my family, I give many thanks for their unending support and love that they have provided with me during my pursuit of a Ph.D. degree far away from home.

The research in this dissertation was supported in part by University at Buffalo College of Arts and Sciences Dissertation Fellowship, Graduate Student Association Mark Diamond Research Fund, and University at Buffalo Department of Linguistics.
ABSTRACT

This thesis reports on the results of production and perception experiments which investigated the nature of lexical accent in Tokyo Japanese (simply Japanese hereafter). Word prosody of Japanese is often labeled as pitch accent, characterized by a steep fall in $F_0$ from the accented mora to the following one. Words that have accent on the final-mora (final-accented words) and words that have no accent (unaccented words) apparently have the same tone sequence, yet they differ in that a particle has a surface low tone after final-accented words while it has a high tone after unaccented words. It has been debated whether the two accent patterns are identical when they occur in isolation, and whether the tone on the following particle is the only acoustic correlate of pitch accent.

In the present study, a computerized database was used to search for all bimoraic and disyllabic minimal pairs of final-accented and unaccented words in Japanese. Because word familiarity is known to influence word recognition and production, only words with a relatively high familiarity rating were used, resulting in 20 minimal pairs. Ten native Japanese speakers (five males and females) produced the 20 pairs in isolation and sentence-medially followed by a particle. The production study found that, when the two types of words were produced in isolation, they were not significantly different in either their $F_0$ peak in the second mora or their $F_0$ rise from the first to second mora. When the words were produced sentence-medially, there was a significant difference within words, both in $F_0$ peak and $F_0$ rise. There was also a greater fall in $F_0$ into the following particle for final-accented words. In the perception experiment, recordings from the production study were used to create three sets of stimuli: (A) final-accented and unaccented words produced in isolation, (B) final-accented and unaccented words excised from a sentence, and (C) words and the following particle excised from a sentence. The listeners (n=23) were not able to identify words under conditions (A) or (B). Thus even though the stimuli type B differed in $F_0$ peak and $F_0$ rise, these acoustic properties were not sufficient for listeners to identify the words. Furthermore, the accuracy was only about 70 percent even when there was a following particle (stimuli type C). However, a strong positive correlation was found between the accuracy and the size of $F_0$ fall difference between the two accent patterns in each pair, suggesting that $F_0$
fall played an important role in perceiving accent. Thus, while accent information was redundantly present in $F_0$ peak, $F_0$ rise, and $F_0$ fall of words produced in sentence context, listeners appear to use only the $F_0$ fall into the following particle for word identification.
CHAPTER ONE

INTRODUCTION

This thesis reports on the results of production and perception experiments which were designed to investigate the nature of pitch accent in Tokyo Japanese.¹ There has been much debate about how to characterize pitch accent languages in terms of a typology of word-level prosody, and whether it is even useful to posit a distinct category of pitch accent languages (e.g. Beckman, 1986; Hyman, 2006; McCawley, 1978; Pike, 1948; Ladd, 1996, just to name a few). The term pitch accent is used in this study, but this use is not based on a claim that Japanese is a pitch-accent language. Rather, it is hoped that this study will make some contribution towards understanding the basis for classifying languages according to their lexical prosody. The working definition of pitch accent in Japanese in the present study refers to the location where there is a steep fall in $F_0$, which will be indicated with a caret ^.² When there is such a fall, the syllable immediately preceding the fall is said to bear the accent. The literature on pitch accent in Japanese dates back to at least as early as Yamada (1892), who first noted the difference between words that have accent on their final mora and words that have no accent. These accent patterns are now known as the final-accented pattern (odaka-gata) and the unaccented pattern (heiban-gata). Since then, much research has been done on pitch accent in Japanese theoretically, instrumentally, and experimentally.³ In spite of a wealth of preceding research on Japanese

¹ The focus of this study is Tokyo Japanese, a variety of Japanese spoken in the Tokyo area. When Japanese is alluded to, it should be understood as referring to Tokyo Japanese. Although Tokyo Japanese is not identical to standard Japanese, the two are very similar.
² The terms pitch and $F_0$ will be used interchangeably when no strict distinction is necessary. For indicating accent, others have used a tick mark or an asterisk.
³ The terms instrumental and experimental are loosely differentiated in describing production studies. Instrumental studies refer to those that used some instrument, such as an oscilloscope, to measure some acoustic properties of pitch accent. Experimental studies refer to studies that conducted research under controlled
pitch accent, there were nevertheless good reasons for conducting the present research. Most previous research that used minimal pairs of final-accented and unaccented words focused on a just few minimal pairs, such as /hana^/ ‘flower’ and /hana/ ‘nose’, and /hasi^/ ‘bridge’ and /hasi/ ‘edge’.4 Now that a computerized database of Japanese words has become available (Amano & Kondo, 1999), it has become possible to search for minimal pairs of the two accent patterns. Furthermore, word familiarity ratings were included in the database so that only words that are familiar to Japanese speakers could be selected as test materials. This is important because recent studies have found that word familiarity has an effect on how people produce and perceive words (e.g. Amano, Kondo, & Kato, 1999; Wright, 1997). Thus, the present study aimed to examine whether acoustic properties of Japanese pitch accent found in earlier studies could be established as general characteristics of Japanese pitch accent. A perception study was conducted following the production study to test if acoustic properties found to be related to pitch accent were actually used by listeners in word identification. Given that speech communication involves a speaker and a listener, although most previous research focused on either production or perception, both aspects need to be examined to fully understand the nature of pitch accent in Japanese.

This thesis comprises four chapters: 1) introduction, 2) production study, 3) perception study, and 4) conclusion. The remainder of this chapter provides an overview of the literature on pitch accent in Japanese, including some earlier theoretical research on how to represent pitch accent, which was important for motivating the instrumental research. Then the distribution of different accent patterns across words of various lengths is discussed, as well as some phonological issues and phonetic phenomena that are related to accent. Chapter 2 introduces and reports on the production study that investigated the acoustics of pitch accent. Chapter 3 discusses a perception study that examined what acoustic properties were used (or not used) by Japanese listeners in word identification using recordings from the production study as stimuli.

---

4 Phonetically, the second consonants in the words /hasi^/ and /hasi/ are realized as voiceless palato-alveolar fricatives. Since the consonant is an allophone of the voiceless alveolar fricative in Japanese, the symbol /s/ will be used to represent it in this dissertation.
Chapter 4 gives a conclusion of the thesis and discusses some remaining issues for future research.

1.1 Japanese Pitch Accent

1.1.1 Early Studies

The word prosody of Japanese has been analyzed tonally and accentually. In short, the early debate over the representation of Japanese word prosody centered around two issues. One issue concerns whether the surface tone patterns of Japanese should be specified in underlying representations in terms of tones or in terms of an abstract property such as accent. The other issue concerns the number of tones needed to describe tone patterns in Japanese.\(^5\) Yamada (1892) noticed that hana is a girl’s name when the tone pattern is HL (high-low), but it means hana ‘flower’ or hana ‘nose’ when the tone pattern is LH. Furthermore, when these two words (i.e. ‘flower’ and ‘nose’) are followed by a particle, it has a different tone depending on which word it follows. When the particle ga, a nominative marker, follows hana ‘flower’, it has a low tone. Thus, the phrase hana ga /hana^ ga/ has a tone pattern LHL. By contrast, when the particle follows hana ‘nose’, the tone pattern is LHH. Based on this observation, Yamada claimed that there are three possible tone patterns (oncho) for bimoraic words and labeled them with different names, as

\(^5\) The term tone is used in the sense Hyman (2006a) uses it: "a language with tone is one in which an indication of pitch enters into the lexical realization of at least some morphemes" (p. 229). Since some words are distinguished only by pitch in Japanese, it is a language with tone, regardless of whether it is classified as a pitch-accent language, tone language, or stress-accent language.
shown in Table 1.1.

This three-way distinction is reflected in Polivanov’s (1925) representation of these words. When Polivanov measured $F_0$ movements of final-accented and unaccented words produced in isolation, the two types of words were indistinguishable. However, because they had different realizations when they were produced with a following particle, they were represented differently. Polivanov indicated the contrast among the triplet shown in Table 1.1 as shown in Table 1.2. Polivanov’s representation distinguishes words that have accent and ones that do not have accent, as seen in the use of two types of symbols. For words that have accent, the location of accent was indicated with $\lceil$. Words that have no accent were represented with a line above them. Polivanov referred to these words as toneless words. The representation also shows that he considered the key element of Japanese word prosody was accent, the location where the pitch drops, rather than actual tones.

By contrast, Sakuma (1929) took a tonal approach. Sakuma’s representation was crucially different from Polivanov’s and Yamada’s in terms of how to represent unaccented words. Sakuma argued that it was inaccurate to represent \textit{hana} ‘nose’ as having a flat tone pattern. He measured $F_0$ movements of monomoraic minimal pairs of \textit{ho}, which means ‘ears of wheat’ when accented and ‘sails of a ship’ when unaccented, followed by a particle to, a citation marker. While the $F_0$ peak reached 169 Hz for the accented word, it was much lower for the unaccented word, reaching only 112 Hz (in male voice). Sakuma indicated the tone on the accented word as high (H) and the tone on the unaccented word as middle (M). (Sakuma also indicated the second tone of the girl’s name as M, as will be discussed later.) This idea was applied to final-accented words and unaccented words in general. According to this account, the triplet words are represented as shown in Table 1.3. For comparison, Yamada’s

<table>
<thead>
<tr>
<th>Representation</th>
<th>hana</th>
<th>hana</th>
<th>hana</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>‘girl’s name’</td>
<td>‘flower’</td>
<td>‘nose’</td>
</tr>
<tr>
<td></td>
<td>[hana ga]</td>
<td>ha</td>
<td>na ga</td>
</tr>
</tbody>
</table>

---

6 In Polivanov’s system, the beginning of the mora that has a drop was marked, instead of the actual location where the drop occurs, which is the end of the mora.
7 Fundamental frequency was extracted by measuring the vibration of vocal folds from outside the subject’s throat.
representation is also shown at the bottom. The representation of *hana* ‘flower’ is the same for Sakuma and Yamada, but the representation of the other two words are different. Specifically, for the girl’s name, the low tone after the accented mora in Yamada was replaced by a middle tone in Sakuma for reasons that are not clear. For ‘nose’, the unaccented high tone in Yamada was also replaced by a middle tone in Sakuma. In this tripartite system, Sakuma was able to show the rising movement of $F_0$ on unaccented words as well as on final-accented words, while still showing that the $F_0$ peak on unaccented words was lower than that of final-accented words. In contrast to Polivanov, Sakuma represented Japanese words in terms of tones specified for each mora. The notion of accent and the distinction between accented words and unaccented words are not present in Sakuma’s representation. Sakuma’s proposal that Japanese has three tones is the basis of a long debate among Japanese researchers as to whether Japanese requires three tone levels in order to account for final-accented and unaccented patterns, or just two tones.

Sakuma’s three tone approach had theoretical and empirical problems. Theoretically, the three tone system was not symmetrical (Tamaru, 1918). That is, while the low tone appears only word initially, the middle tone appears only word medially and finally. Since the two tones are in complementary distribution, they should be categorized as a single tone in the phonological representation. Tamaru noted that although the second mora of *hana* ‘flower’ may have a higher $F_0$ than the second mora of *hana* ‘nose’, this phonetic difference should be disregarded in (phonological) representations of tone patterns. The same criticism was also made by Arisaka (1941) and Saeki (1932). Basically, Sakuma’s representation was too phonetic and it missed a generalization. In addition, while Sakuma’s representation was phonetically accurate in showing that final-accented words had a higher $F_0$ peak than unaccented words, some middle tones apparently contradicted phonetic facts. The tone pattern of the girl’s name is represented as a high tone followed by a middle tone. However, as Poser

<table>
<thead>
<tr>
<th>Examples</th>
<th>hana ‘girl’s name’</th>
<th>hana ‘flower’</th>
<th>hana ‘nose’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tones</td>
<td>HM</td>
<td>LH</td>
<td>LM</td>
</tr>
<tr>
<td>Yamada’s</td>
<td>HL</td>
<td>LH</td>
<td>LH</td>
</tr>
</tbody>
</table>

Table 1.3: Accent Patterns Classification by Sakuma
(1984, p. 30) notes, the tones following an accented mora (as in the girls’ name) tend to be lower than initial low tones (as in ‘flower’ and ‘nose’).\footnote{Throughout this dissertation, references are indicated with relevant page numbers for convenience whenever possible, even when there is no direct citation.}

Miyata (1927) is different from earlier researchers such as Polivanov (1915), Sakuma (1929), and Yamada (1892) in that he explicitly stated that the fall of pitch is the essential aspect of Japanese word prosody, not the tone of each mora. This is based on the observation that the location of pitch fall is the only unpredictable property of Japanese words. Under Miyata’s analysis, each Japanese word only needs to be represented with the location of pitch fall, or accent, when there is one. Although some words have a high tone initially while other words have a low tone, this is predictable once the accent information is given. Thus, the tone patterns of entire words can be determined from rules like the following (adopted from Poser, 1984, p. 32).

\begin{enumerate}
\item Make everything up to and including the accent high.
\item Make everything following the accent low.
\item Make the first mora low if the first syllable is light and the following mora is high.\footnote{Since one syllable can contain more than one mora, strictly speaking, syllables and moras must be distinguished. Light syllables refer to those that consist of one mora. Further discussions of moras and syllables follow later in this chapter.}
\end{enumerate}

Miyata also claimed that nouns and a following particle should be taken as a unit of accent, unless the noun is produced in isolation. Since nouns alone are not the entire accentual unit, accent is not fully manifested in isolation. The claim that the domain of accent consists of a noun and a following particle is common to other researchers such as Arisaka (1941) and Hattori (1954), who are discussed next.

Arisaka (1941) took Miyata’s (1927, 1928) analysis a step further, distinguishing the actual realization of words from the abstract representation of words. In modern terms, this is a distinction between surface representations and underlying representations. When words surface as actual utterances, accent information in underlying representations is translated into sequences of high and low tones. Arisaka also observed that the contrast between final-accented words and unaccented words is neutralized when they are produced in isolation because words alone are not the domain for realizing the acoustic properties of accent. Arisaka
referred to this domain as *bunsetsu* (equivalent to Jakobson’s *syntagma*). Therefore, even though final-accented and unaccented words are distinct in underlying representations, the distinction does not surface when the two types of words are produced without having the minimum unit of *bunsetsu*. A study by Han (1962a, pp. 112-115), which will be reviewed in the next chapter, was conducted to confirm this claim.

Miyata’s and Arisaka’s analyses deserve a few remarks since they brought to light at least two important aspects of Japanese word prosody. First, they note that tone patterns of Japanese words are completely predictable once the location of accent is known. This characteristic of Japanese words is missed under the analysis where tone is specified for each mora. Furthermore, Arisaka made the distinction between underlying representations, where the accent location is marked abstractly, and surface representations, where the accent is realized with physical content. This idea has been adopted in various accentual analyses of Japanese. However, it should also be noted, as the analyses are raised to a higher level of abstraction, the phonetic difference that Sakuma (1929) found was forgotten. The identity of the final-accent pattern and the unaccented pattern at the phonological level was somehow extended to be taken that the two accent patterns are also identical at the phonetic level.

Hattori (1951, 1954) further extended Miyata’s (1927, 1928) work by stating explicitly what should be represented in the phonology of a language. He claimed that only those phonetic properties that are distinctive (explanations to follow) should be included in phonological representations. In Japanese, the relation between the final high tone and the low tone that follows is always constant. Hattori called the location where a high tone is (potentially) followed by a low tone an accent nucleus (*akusentō no kaku*), and defined it as a distinctive property of Japanese words. Hattori proposed that there are two distinctive properties in Japanese: 1) whether a word/phrase has an accent nucleus, and 2) if it does, where it is. He also observed that the role that tone plays in Japanese is different from that in tone languages like Chinese. While tone must be specified for each syllable in tone languages, this is not the case in Japanese. In Japanese, it is sufficient to specify only the location of tone change from high to low, when there is one. In other words, whether a mora has a low tone or a high tone is a phonetic realization of the accent nucleus and not a phonological property in Japanese. As will be seen next, this account is different from Kindaichi’s (1947, 1957) account, which holds that Japanese words consist of sequences of tones similar to Chinese. Hattori’s account of Japanese words contains no redundancy in the
underlying representation and was relatively abstracted from the actual tone patterns of words. In the lexicon, words are marked with only the location of accent nucleus, if they are accented. On the other hand, there is no clear account of how surface forms are derived from the abstract representations he proposes.

Kindaichi’s account is similar to Arisaka (1941) and Miyata (1927, 1928) in that there are only two tones in Japanese, rather than three tones as Sakuma (1929) proposed. However, it is similar to Sakuma in taking a tonal approach. He claimed that prosody of Japanese words consists of high and low tones, or tonemes. He described the characteristics of tones in Japanese as follows.

\[(2)\]

1. Japanese words consist of two tonemes of high and low.
2. One mora consists of one toneme.
3. The first and second moras (of a word) always have a different toneme.
4. The high tonemes are never contrastive with low tonemes.

Apparently, Kincaichi’s characterization of Japanese words is very different from Miyata, Arisaka, and Hattori’s. However, his further characterization of Japanese suggests that his tonal account is not very different from an accentual account. He observes first that, the location where the pitch drops is important in Japanese, and, second, the tone patterns that can occur in Japanese are fairly limited. In fact, he relates these properties of Japanese to those of stress-accent languages such as English. While Kindaichi’s account and accentual account are conceptually very distinct in what they regard to be the organizing element of Japanese word prosody, the difference becomes less obvious as their theories are applied to the prosodic system of Japanese.

There is one clear difference, however, between Kindaichi and Miyata, Arisaka, and Hattori. Kindaichi (1947) included only those properties that were present at the word level in the phonological representations. Based on the subjective impression that the difference between *hana* ‘flower’ and *hana* ‘nose’ appeared only when there was a following word, the difference between the two types of words was not regarded as a phonological one. As will be reviewed in Chapter 2.1, this claim was a motivation for Sugito’s (1968) study that examined bimoraic final-accented and unaccented words produced in isolation.
1.1.2 Japanese Pitch Accent

This section gives a summary of the discussions on Japanese word prosody so far and presents some additional details. Pitch accent in Japanese is a lexical property that has a “contrastive function” (Martinet, 1960). That is, one locates accent within the lexical word domain, as opposed to identifying the tone of tone bearing units. While lexical contrasts are realized by different tone patterns in Japanese, the possible tone patterns are highly limited. For words with n number of syllables, there are only (n + 1) possible tone patterns. In particular, the F$_0$ fall occurs at most once within words. Unlike typical tone languages such as Chinese, in which each syllable is specified for a tone, in Japanese the tone patterns are specified for each word, regardless of the number of syllables. While prosodic patterns are described in terms of a lexical pitch accent in this study, note that the descriptions are compatible with a tonal analysis by replacing accent with a H tone that is linked to syllable immediately before a drop in pitch. Accent patterns of Japanese can be described in terms of two parameters: 1) whether a word has accent or not, and 2) if it does, where it is.\textsuperscript{10} The important characteristic of the tone patterns of Japanese is that the only unpredictable property is the location of a F$_0$ fall, or pitch accent, so that only the location of such a fall, if there is one needs to be marked. In order to describe accent patterns in Japanese, it is necessary to take both moras and syllables into consideration, unless syllables are monomoraic. In addition, syllables are classified into light and heavy syllables, which are further divided into short and long syllables, following Pierrehumbert and Beckman (1988) and Poser (1984).\textsuperscript{11} The light versus heavy distinction depends on the number of moras in a syllable. Light syllables contain only one mora whereas heavy syllables contain two. Heavy syllables are further distinguished by the type of phoneme that occurs in the second mora. Short syllables contain only one sonorant mora, i.e. a short vowel followed by the part of a geminate obstruent (VQ).\textsuperscript{12} Long syllables contain two sonorant moras, i.e. a long

\textsuperscript{10} Accented words and unaccented words differ clearly in terms of the presence versus absence of downstep when they are followed by an accented word within the same accentual phrase. Accentual phrases will be discussed in detail in Chapter 2.1.1.

\textsuperscript{11} While superheavy syllables consisting of a long vowel and a geminate exist in Japanese, they are not very frequent.

\textsuperscript{12} In Japanese phonology, the symbol Q is used to represent a moraic obstruent,
vowel (V:), diphthong (V₁V₂), or a vowel followed by a moraic nasal (VN). The classification of mora and syllable types in Japanese is summarized in Table 1.4. In assigning accent, the bearer of accent is the syllable. Heavy syllables are never contrastive in the location of accent. When heavy syllables are accented, the pitch fall always occurs between the first and second mora. This property becomes more evident in loanwords, which are discussed in Section 1.2.1. When an initial syllable is light, (C)V, the tone patterns can be described as the following.

1. The moras from the second mora to the accented mora are high.
2. The moras after the accent are low.
3. The first mora is low unless it is accented.

It follows from this description that words that have the accent on the first mora have a high tone on the first mora and the remaining moras are all low. It also follows that the first and second moras never have the same tone. The tone patterns of the first two moras are either LH or HL. For unaccented words, the first mora is low and the rest of the moras are high until the end of the word. Table 1.5 shows examples of accent patterns with trimoraic words. One situation where the tone pattern diverges from (3) is when the initial syllable is long and unaccented. When an initial syllable is long and unaccented, there is no low tone initially. The tone is high from the beginning until the accented mora. When the word is unaccented, the entire word has a high tone. For example, the unaccented word toomin ‘hibernation’ has the tone pattern HHHH. In other cases when an initial syllable is heavy, that is, when an initial syllable is short but unaccented, and when an initial syllable is heavy and accented, the tone pattern follows namely the first element of a geminate.
3. In order to avoid tone pattern differences that arise from heavy syllables, only words that consist of monomoraic syllables were used in this study.\(^\text{13}\)

## 1.2 Distribution of Accent Patterns

As discussed in the preceding section, the possible accent patterns are much more limited than those encountered in a typical tone language. In addition to the existence of a highly constrained set of accent patterns in Japanese, some of the accent patterns occur more frequently than others (Kitahara, 2001; Tanaka & Kubozono, 1999). Tanaka and Kubozono (1999, pp. 58-59) show the distribution of accent patterns of nouns that are monomoraic, bimoraic, trimoraic, and four-mora long, as summarized in Table 1.6.\(^\text{14}\) Kitahara (2001, pp. 87-88) shows the distribution of accent patterns for bimoraic, trimoraic, and four-mora words for all parts of speech based on the information provided in Amano and Kondo (1999). Since Kitahara illustrates the distribution of accent patterns with bar graphs, actual percentages are not known. However, the distribution shown in Tanaka and Kubozono and Kitahara show similar patterns. For monomoraic and bimoraic words, the unaccented pattern is not common. The most common accent pattern for bimoraic words is initial accent.

### Table 1.5: Possible Accent Patterns for Trimoraic Words

<table>
<thead>
<tr>
<th>Accent</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Words</strong></td>
<td>sakana</td>
<td>makura</td>
<td>kokoro</td>
<td>yasumi</td>
</tr>
<tr>
<td><strong>Tones</strong></td>
<td>LHH (H)</td>
<td>H^LL (L)</td>
<td>LH^L (L)</td>
<td>LHH^ (L)</td>
</tr>
<tr>
<td><strong>Gloss</strong></td>
<td>fish</td>
<td>pillow</td>
<td>heart</td>
<td>vacation</td>
</tr>
</tbody>
</table>

*Note. The number n in the accent row indicates that accent is on the nth mora. The number 0 refers to the unaccented pattern. The letters in parentheses in the tone row indicate the tone on the following particle when there is one.*

(Kitahara, 2001; Tanaka & Kubozono, 1999). Tanaka and Kubozono (1999, pp. 58-59) show the distribution of accent patterns of nouns that are monomoraic, bimoraic, trimoraic, and four-mora long, as summarized in Table 1.6.\(^\text{14}\) Kitahara (2001, pp. 87-88) shows the distribution of accent patterns for bimoraic, trimoraic, and four-mora words for all parts of speech based on the information provided in Amano and Kondo (1999). Since Kitahara illustrates the distribution of accent patterns with bar graphs, actual percentages are not known. However, the distribution shown in Tanaka and Kubozono and Kitahara show similar patterns. For monomoraic and bimoraic words, the unaccented pattern is not common. The most common accent pattern for bimoraic words is initial accent.

\(^\text{13}\) For example, the distinction between final accented and unaccented phrases is realized phonetically in the following words: /senseˆi/ ‘doctor, professor’ vs. /sensei/ ‘despotism.’

\(^\text{14}\) Accented monomoraic words can be seen as having either initial-accent or final-accent. Judging from the distribution of initial-accent and final-accent patterns in bimoraic and trimoraic words, however, it would be more consistent to treat them as having initial-accent.
Comparing final-accented and unaccented patterns, final-accent is more common than unaccented. For trimoraic words, the unaccented pattern is the most common, but initial accent is still relatively common. For four-mora words, the unaccented pattern becomes by far the most common pattern than any other pattern. What Table 1.6 does not show and Kitahara does is how the distribution of accent patterns interacts with familiarity. According to Kitahara, there is a general trend for all accent patterns to be somewhat equal as the familiarity becomes higher.

Words that have five or more moras tend to be compound words. Although the words analyzed were limited to nouns, Sato (1993) found that the unaccented pattern was not very common for compound words. In addition, they tended to have accent somewhere in the middle of words, rather than initially or finally.

1.2.1 Accent Pattern of Loanwords

Japanese vocabulary consists of three kinds of words: native Japanese words (yamato words), Chinese words, and loanwords. Chinese words were borrowed from Chinese more than a thousand years ago, and Sino-Japanese words now comprise an essential part of Japanese vocabulary. Loanwords are words that have been introduced from foreign languages relatively recently. Among the three kinds of words, loanwords are known to have highly regular accent patterns. In general, the accent falls on the antepenultimate mora, such as in /ba^nana/ ‘banana’, /ore^Nji/ ‘orange’, /aisukuri^:mu/ ‘ice cream’ (Tanaka & Kubozono, 1999). However, when the antepenultimate mora happens to be a special mora ‘Q’ (see below), or a second mora in a heavy syllable (the first consonant of


### Table 1.6: Distribution of Accent Patterns (%)

<table>
<thead>
<tr>
<th>Unaccented</th>
<th>First</th>
<th>Second</th>
<th>Third</th>
<th>Fourth</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30&lt;</td>
<td>&lt;70</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>65</td>
<td>&lt;20</td>
<td>–</td>
</tr>
<tr>
<td>3</td>
<td>50</td>
<td>&lt;40</td>
<td>&lt;10</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>&lt;70</td>
<td>&lt;10</td>
<td>10&lt;</td>
<td>&lt;10</td>
</tr>
</tbody>
</table>

1.2.1 Accent Pattern of Loanwords

Japanese vocabulary consists of three kinds of words: native Japanese words (yamato words), Chinese words, and loanwords. Chinese words were borrowed from Chinese more than a thousand years ago, and Sino-Japanese words now comprise an essential part of Japanese vocabulary. Loanwords are words that have been introduced from foreign languages relatively recently. Among the three kinds of words, loanwords are known to have highly regular accent patterns. In general, the accent falls on the antepenultimate mora, such as in /ba^nana/ ‘banana’, /ore^Nji/ ‘orange’, /aisukuri^:mu/ ‘ice cream’ (Tanaka & Kubozono, 1999). However, when the antepenultimate mora happens to be a special mora ‘Q’ (see below), or a second mora in a heavy syllable (the first consonant of

In this dissertation, the symbol /r/ will be used to denote the alveolar tap.
geminates, the moraic nasal, or the second part of long vowels or
diphthongs), the accent shifts to the fourth mora from the last. For example,
the word ‘soccer’ is pronounced as /sa^Qka:/, and not /saQ^ka/. Since the
antepenultimate mora is a special mora, /Q/, the accent moves a mora
earlier. Similarly, ‘calendar’ is pronounced as /kare^Nda:/ Since the
antepenultimate mora /N/ is a moraic nasal, accent moves to the initial
mora of the syllable that contains the antepenultimate mora. These accent
shifts show that, although the mora is counted for assigning the accent, the
bearer of accent is the syllable in Japanese. Referring to this phenomenon,
McCawley (1978, p. 119) described Japanese as a "mora-counting syllable
language". The accent assignment of loanwords in Japanese can be de-
scribed as:

(4) Place accent on the syllable containing the antepenultimate mora.

While a majority of loanwords conform to the regularity described in (4),
about ten percent of loanwords are known to be unaccented (Kubozono,
1996; Tanaka & Kubozono, 1999). These loanwords have three
characteristics almost without exception: 1) they are four moras long, 2)
the last two syllables are monomoraic, and 3) the final vowel is a non-high
vowel, /a/, /e/, or /o/.16 Examples of these words include words such as
that have accent on the final mora or on the penultimate mora are very
limited (Tanaka & Kubozono, 1999).17

1.2.2 Final-accented and Unaccented Patterns

This section briefly discusses some interesting properties of the two
accent patterns compared in this study. The final-accent and unaccented
patterns are compared because this is the only environment in which word
prosody differs only in the presence or absence of accent. However, as
already reviewed in Section 1.1.1, traditionally, the contrast between
final-accented and unaccented words has normally been said to be lost
when they are produced in isolation. The neutralization is also said to
occur when the two types of words occur sentence-finally (Vance, 1995).

16 One exception is /honoruru/ ‘Honolulu’, which ends in a high vowel.
17 Examples of loanwords with penultimate accent are /ko:hi*/ ‘coffee’ and /buru*: ‘blue.’
In addition, the two types of words were collapsed under a single accent pattern in past experimental studies. One interesting aspect about the two accent patterns is that even if it is the case that they are neutralized in certain environments, it is not clear to which accent pattern they are neutralized to. To the author’s knowledge, no studies investigated this question. Nor is it certain which accent pattern would be preferred on theoretical and empirical grounds, as neither seems to be preferred. On the one hand, given that it is more common for words to have some prominence than not to have any prominence, the final-accent pattern would seem to be the pattern expected to appear in neutralization. In fact, as will be discussed in Chapter 3.1.1, past perception studies found that listeners tend to respond that they heard accented words rather than unaccented. On the other hand, there is some evidence that it is not common for a language to realize prominence word finally (Hyman, 1977). For stress-accent languages, while penultimate stress is very common, final stress is not. Furthermore, languages that remove stress from final position often move stress to penultimate position, but the opposite process is not common. If stress-accent and pitch-accent in Japanese are thought to behave similarly, neither the final-accented nor unaccented pattern would seem to be preferred, which makes it uncertain which accent pattern would appear in neutralization.

One phenomenon observed in Japanese suggests that the unaccented pattern may not be disfavored in Japanese. There is a trend, especially among younger generations, to produce words as unaccented even though they are not originally unaccented (Sugito, 1983).

### 1.3 Acoustic Properties of Accent

#### 1.3.1 Devoicing and Accent

In many varieties of Japanese including Tokyo Japanese, the high vowels /i/ and /u/ typically become voiceless, or devoiced, when they occur between voiceless consonants (Maekawa, 1989; Vance, 1987). For example, in the word /kita/ ‘north’, the first vowel is normally produced with no voicing and the vocal fold vibration is delayed until the word final vowel. In addition to vowel quality and the phonological environment in which vowels occur, pitch accent has been said to have an influence on devoicing, although it is not clear how substantial this effect is. On the one hand, studies show that accented high vowels are devoiced less often than
unaccented ones (Lovins, 1976; Maekawa, 1989). Lovins examined whether tone (high or low) or accent (always high) affects devoicing and found that unaccented vowels devoiced more frequently than accented vowels while tone had little effect.\(^{18}\) On the other hand, studies show that accented high vowels are often realized through *ososagari* (delayed pitch fall), whereby accented vowels are devoiced and the pitch peak is realized on the following syllable (Neustupný, 1966; Sugito, 1972; Yoshida, 2002).

Yoshida (2002, p. 41) discusses three possible realizations of accented high vowels that occur between voiceless consonants. Among them, (5.1) is thought to be most common (Sakurai, 1998).

\[
\begin{align*}
(5) & \quad 1. \text{Devoicing occurs, moving accent to the following mora.} \\
& \quad 2. \text{Devoicing is avoided to realize accent on that mora.} \\
& \quad 3. \text{Devoicing occurs with accent staying on that mora.}^{19}
\end{align*}
\]

In order to examine the relation between devoicing and accent, Yoshida created 24 trimoraic pseudowords that looked like foreign place names. Based on the statistical trend that antepenultimate accent is common for words that are relatively short, it was assumed that talkers would place accent on the initial mora. For all the pseudowords, initial syllables consisted of a voiceless obstruent and a high vowel, which was followed by a voiceless consonant. This phonetic environment guaranteed that the vowels would be completely voiceless when the pseudowords were produced in the middle of a frame sentence. When accent location and devoicing were examined for all the utterances recorded, Yoshida found that, for about 70 percent of the utterances, devoicing occurred on the initial mora so that the accent moved to the second mora, consistent with the proposal in (5.1). For 20 percent of the utterances, devoicing did not occur and accent was realized on the initial mora, consistent with the proposal in (5.2). Although Yoshida’s data were based on pseudowords, they at least suggest that devoicing is likely to occur relatively frequently even when high vowels are accented.

---

\(^{18}\) The judgment of devoiced or voiced vowels was apparently based on auditory assessment.

\(^{19}\) Presumably, this is assessed by the absence of F\(_0\) peak on any other moras.
1.3.2 Other Correlates of Pitch Accent than $F_0$

In Japanese, pitch accent manifests itself most consistently and clearly in $F_0$ (Beckman, 1986; Weitzman, 1970). Other acoustic properties such as intensity and duration are normally not found to be correlated with accent. Duration is not likely to be a correlate of accent because length is phonemic in Japanese (Cutler & Otake, 1999; Kaiki, Takeda, & Sagisaka, 1992; but see Mori, 2001). Intensity was once believed to be a correlate of pitch accent, but its status is not certain. In addition, it is difficult to compare intensity across talkers and studies because it is subject to variability. However, some studies show some possible correlation between accent and duration or intensity, even though it is not strong. This section mainly reviews research that investigated other properties than $F_0$ as correlates of pitch accent. In the studies discussed below, it has to be kept in mind that final-accented and unaccented patterns were not distinguished except in Neustupný (1978). The initial accent pattern $H^A L$ was compared with either the final accent pattern $L H^A$ or the unaccented pattern $L H$. Therefore, duration and intensity measures were also influenced by positional differences. Research that investigated the $F_0$ properties will be reviewed in detail in the next chapter.

Neustupný (1966) was the first to discuss the delayed pitch fall, the phenomenon that the $F_0$ fall after accent does not always occur on the accented mora. This lead him to suspect that $F_0$ may not be the only correlate of pitch accent, and that there might be other correlates of accent. Neustupný analyzed the $F_0$ and intensity of 181 words produced in isolation by four male talkers. He found that, for some of the words, accented moras showed a rise in $F_0$ but not in intensity. For other words, accented moras did not show a rise in $F_0$ but did in intensity. Based on this finding, he concluded that accent in Japanese was realized by some inconsistent set of interacting features, and that $F_0$ and intensity were among the correlates of pitch accent. In a later study, Neustupný (1978) replicated the results with monomoraic and bisyllabic minimal pairs comprising final-accented and unaccented patterns. However, the detailed analyses of intensity by Weitzman (1970) and Beckman (1986) suggest that accent and intensity are not as strongly correlated as Neustupný claimed. Weitzman (1970) analyzed the intensity of six bisyllabic words,

---

20 For example, /siˆ/ ‘city’ and /siˆi/, a kind of tree that grows in Japan, differ only in vowel length, and /kaˆko/ ‘past’ and /kaˆkko/ ‘parenthesis’ differ only in whether a medial consonant is a single consonant or a geminate.
while controlling vowel quality. He discusses six different measures for intensity as well as their problems, and used the measures of phonetic speech power (the maximum value of speech power) for a vowel and the area measure of vowel intensity (Weitzman, 1970, pp. 75-78). Intensity as measured by phonetic speech power of the vowel was not consistently associated with accent patterns. Area measures also showed some inconsistency, but they were better correlated with accent than phonetic speech power. Weitzman states that “to the extent that some degree of correlation exists, both intensity and the area of the intensity curve may be said to contribute to the acoustic quality of accent” (p. 92), but concludes that intensity is only a secondary cue to word accent.

Beckman (1986) examined $F_0$, duration, and amplitude of six minimal pairs of disyllabic words. Out of the six pairs, three pairs (the first three word types in Table 1.7) showed a contrast between initial accent and unaccented patterns. The other three minimal pairs contrasted initial accent and final-accented patterns. All the words were produced in the same frame sentence. For each word, the peak amplitude ratio, and average amplitude ratio, duration ratio, and $F_0$ ratio between the first and the second vowels were computed to perform separate $t$-tests on each pair.

Beckman argues that the accent patterns do not have significant correlates in peak or average vowel amplitude, or in duration. Rather, the contrast in accent patterns is realized as differences in the $F_0$ patterns. While her conclusions seem reasonable overall, it is worthwhile examining how the data were analyzed more closely. Beckman performed separate $t$-tests on each pair of words, instead of performing ANOVAs with all of the six pairs of words. This procedure was taken presumably because the phonological structures of the words were not uniform across the six pairs. On the other hand, this resulted in the loss of statistical power with each of the $t$-test consisting of a comparison of only 12 tokens (three repetitions × four speakers), as Levi (2005, p. 93) discusses. In addition, the level of significance was set at 0.01 for multiple comparisons. In order to investigate how the choice of statistical procedures affects statistical results, Levi analyzed the Turkish data that she collected using both separate $t$-tests and ANOVAs for duration. She indeed found that the $t$-test

---

21 Since some of the words contained heavy syllables, syllables are not interchangeable with moras here.

22 The $F_0$ was measured at the obvious peak if there was one, and otherwise midway through the syllable nucleus, which was defined as including at least the duration of the voiced part of the vowel.
analyses yielded fewer significant differences than the ANOVAs.\textsuperscript{23} Given this finding, it is possible that Beckman’s data on amplitude and duration may be subject to somewhat different interpretations.\textsuperscript{24} Yet, it would still be the case that the $F_0$ is the most clear and consistent correlate of pitch accent. In this respect, Beckman’s results and Weitzman’s results are fairly consistent. Although Weitzman did not measure duration, both results show that duration and intensity are not strongly correlated with accent.

Some researchers do not seem to hold that duration or intensity is a possible correlate of accent. Based on the data from Beckman (1986) and Weitzman (1970), it may be too strong a claim to say that there are no correlates other than $F_0$ for pitch accent. Intensity and duration are correlated with accent to some extent, but the correlations are not strong.

\textsuperscript{23} The significance level of 0.05 was used for both types of analyses.

\textsuperscript{24} See Levi (2005, pp. 92-93) for further discussions of possible factors for lack of significance found for the amplitude and duration data in Beckman.
CHAPTER TWO

PRODUCTION STUDY

2.1 Introduction

Pitch accent in Japanese has traditionally been described in terms of a fall in the $F_0$ (fundamental frequency) between the accented syllable and the following syllable when there is one (e.g. Hattori, 1954; Kindaichi, 1947; Kubozono, 1993; McCawley, 1977; Miyata, 1927; Pierrehumbert & Beckman, 1988; Poser, 1984; Vance, 1987). When a particle follows a word that has accent on its final syllable, as in /hana^/ + /ga/ ‘flower’ + nominative. By contrast, when a particle follows a word that has no accent, as in /hana/ + /ga/, ‘nose’ + NOMINATIVE, the phrase has an LH H tone pattern. Words that have accent on their final syllable (final-accented words) and words that have no accent (unaccented words) have the same tone pattern of LH, but the $F_0$ drops when a particle follows a final-accented word while it stays high when a particle follows an unaccented word.\(^1\) Although this difference in $F_0$ on the following particle has been consistently reported in the literature as what distinguishes the two accent types, it has been debated if this is the only acoustic difference between the two accent types, and if there is any difference between final-accented words and unaccented words themselves.

Based on the impressionistic observations, most researchers of Japanese phonology before Pierrehumbert and Beckman (1988) held that final-accented and unaccented words can only be distinguished by the pitch on the following syllable. As reviewed in Chapter 1, Sakuma (1929) is an exception to claim that the two types of words are not realized the same. The finding that final-accented words had a higher $F_0$ peak than unaccented words led Sakuma to represent the high tone of accented words as having a high tone and the high tone of unaccented words as having a

---

\(^1\) This distinction is neutralized in certain environments. Final-accented nouns of more than one mora lose their accent when they are followed by the possessive particle no. See McCawley (1977, p. 267).
middle tone. Sakuma’s three tone system once gave rise to much controversy as to how to represent pitch accent system in Japanese, but it died out as the idea that final-accented and unaccented words can only be distinguished by the tone on the following particle and the two types of words are identical within the words gained more support. However, Poser’s (1984, pp. 198-208) experimental work showed that this impressionistic claim was not tenable. A systematic comparison between the accented high tone and unaccented high tone found that the accented high tone had a higher $F_0$ peak than unaccented high tone when they occurred phrase-medially. This finding was replicated in Pierrehumbert and Beckman (1988) and reflected in their theory on Japanese tone structure. Pierrehumbert and Beckman (1988, p. 121) stipulate and represent that the high tone of accented words is higher than that of unaccented words. While this may lead one to believe that bimoraic final-accented words should have a higher $F_0$ peak than bimoraic unaccented words, it is not certain if this would be true for bimoraic words because of some theory-internal reasons. This issue will be discussed in detail in Section 2.1.1.

The remaining part of this chapter is organized as follows. First, Pierrehumbert and Beckman’s theory will be introduced briefly in order to understand how the high tone of lexical accent and the high tone of unaccented phrases are differentiated in their theory and why bimoraic words pose a potential problem. Next, previous instrumental studies that examined final-accented and unaccented words produced in isolation, phrase, and sentence will be reviewed. Then, the production study conducted in this study will be introduced and discussed. The present study aims to obtain general characterization of Japanese pitch accent by comparing bimoraic final-accented and unaccented minimal pairs that exist in Japanese by conducting a thorough search in an electronic database of Japanese words.

2.1.1 Pierrehumbert and Beckman (1988)

Pierrehumbert & Beckman’s (1988) theory predicts that final-accented words and unaccented words will be different even within the word. Because of theoretical assumptions adopted in their work, however, it

---

2 The high tone of unaccented words is equivalent to the phrasal high tone in Pierrehumbert and Beckman.