CROSS-LINGUISTIC PERCEPTION AND LEARNING OF JAPANESE LEXICAL PROSODY BY ENGLISH LISTENERS

by

IRINA A. SHPORT

A DISSERTATION

Presented to the Department of Linguistics
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

September 2011

DISSERTATION APPROVAL PAGE

Student: Irina A. Shport

Title: Cross-Linguistic Perception and Learning of Japanese Lexical Prosody by English Listeners

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Linguistics by:

Susan Guion Anderson	Chairperson
Melissa A. Redford	Member
Vsevolod Kapatsinki	Member
Kaori Idemaru	Outside Member

and

Kimberly Andrews Espy	Vice President for Research & Innovation/Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded September 2011

DISSERTATION ABSTRACT

Irina A. Shport

Doctor of Philosophy

Department of Linguistics

September 2011

Title: Cross-Linguistic Perception and Learning of Japanese Lexical Prosody by English Listeners

Approved: _____
Susan Guion Anderson

The focus of this dissertation is on how language experience shapes perception of a non-native prosodic contrast. In Tokyo Japanese, fundamental frequency (F0) peak and fall are acoustic cues to lexically contrastive pitch patterns, in which a word may be accented on a particular syllable or unaccented (e.g., *tsúru* 'a crane', *tsurú* 'a vine', *tsuru* 'to fish'). In English, lexical stress is obligatory, and it may be reinforced by F0 in higher-level prosodic groupings. Here I investigate whether English listeners can attend to F0 peaks as well as falls in contrastive pitch patterns and whether training can facilitate the learning of prosodic categories.

In a series of categorization and discrimination experiments, where F0 peak and fall were manipulated in one-word utterances, the judgments of prominence by naïve English listeners and native Japanese listeners were compared. The results indicated that while English listeners had phonetic sensitivity to F0 fall in a same-different discrimination task, they could not consistently use the F0 fall to categorize F0 patterns. The effects of F0 peak location and F0 fall on prominence judgments were always larger for Japanese listeners than for English listeners. Furthermore, the interaction between

these acoustic cues affected perception of the contrast by Japanese, but not English listeners. This result suggests that native, but not non-native, listeners have complex and integrated processing of these cues.

The training experiment assessed improvement in categorization of Japanese pitch patterns with exposure and feedback. The results suggested that training improved identification of the accented patterns, which also generalized to new words and new contexts. Identification of the unaccented pattern, on the other hand, showed no improvement. Error analysis indicated that native English listeners did not learn to attend specifically to the lack of the F0 fall.

To conclude, language experience influences perception of prosodic categories. Although there is some sensitivity to F0 fall in non-native listeners, they rely mostly on F0 peak location in language-like tasks such as categorization of pitch patterns. Learning of new prosodic categories is possible. However, not all categories are learned equally well, which suggests that first language attentional biases affect second language acquisition in the prosodic domain.

CURRICULUM VITAE

NAME OF AUTHOR: Irina A. Shport

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, U.S.A.
Hokusei Gakuen University, Sapporo, Japan
Novosibirsk State University, Novosibirsk, Russia

DEGREES AWARDED:

Doctor of Philosophy, Linguistics, 2011, University of Oregon
Master of Arts, Japanese Language and Pedagogy, 2003, University of Oregon
Bachelor of Arts, Russian Language and Literature, 1997, Novosibirsk State University

AREAS OF SPECIAL INTEREST:

Language acquisition, experimental phonetics, speech perception and production

PROFESSIONAL EXPERIENCE:

Teaching Assistant, University of Oregon, 2000-2010

Instructor of Record, University of Oregon, 2004, Summer Sessions 2009-2011

Research Assistant, 2004-2005, 2009-2011

GRANTS, AWARDS, AND HONORS:

Japanese-Language Programs for Specialists (Researchers and Postgraduate Students), The Japan Foundation, Summer 2008

Graduate Teaching Fellowship, University of Oregon, 2000-2011

Graduate Research Fellowship, 2004-2005, 2009-2011

PUBLICATIONS:

Shport, I. A., & Redford, M. A. (2011). Interactions between lexical and phrasal prosody in school-aged children's speech. In W.-S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 1838-41). Hong Kong.

Guion, S. G., Amith, J. D., Doty, C. S. & Shport, I. A. (2010). Word-level prosody in Balsas Nahuatl: The origin, development, and acoustic correlates of tone in a stress accent language. *Journal of Phonetics*, *38*, 137-166.

Shport, I. A. (2008). Acquisition of Japanese pitch accent by American learners. In P. Heinrich & Y. Sugita (Eds.), *Japanese as foreign language in the age of globalization* (pp. 165-187). München: Iudicium Verlag.

Shport, I. A., & Guion, S. G. (2008). The effect of segmental structure on F0 patterns of words in Tokyo Japanese. *Journal of the Phonetic Society of Japan, 12*(2), 4-16.

ACKNOWLEDGMENTS

Working on this dissertation was like hiking to Phantom Ranch, down and up the Grand Canyon, and I am thankful to colleagues, family and friends who supported me along the way and who made this journey possible. Most of all, I am grateful to my advisor Susan Guion Anderson for being there for me at all stages of the journey, be it choosing a route, inspecting the gear, analyzing the patterns, timing the progress, or having water breaks. Thanks to Susan, I learned the geology of the canyon, from Vishnu Schist to Kaibab Limestone, and I hope to take with her other trails in the area I love. I am also grateful to the members of my committee: Melissa Redford for making me think about the big picture and teaching me to take wide-angle snap shots; Kaori Idemaru for sharing precious water supply and helping me to stay on track; Vsevolod Kapatsinski for showing how to use the gear beyond the first aid kit and making it look so easy.

In addition, I thank my Master Thesis advisor Testuo Harada, who pointed me in the right direction to even consider taking the South Kaibab trail. The University of Oregon community generously assisted my journey, which led to a successful ascend up the Bright Angel Trail: thank you, 129 English and 54 Japanese participants; the instructors at the Japanese language program and American English Institute; the Linguistics Department; the members of the Phon Group and the Beer for Linguists (*aka* BFL) circle. Hanging out with you at the Phantom Ranch kept me going up to the rim.

It takes a lot of love to push on until the end, and I am thankful for love and friendships, centigrade or fahrenheit: my parents Valentina and Anatoly Shport; my friends; and my hiking partner Kathryn Barton who supported me from rim to rim and shared with me the joys of the journey, including the best tasting beers at the top.

TABLE OF CONTENTS

Chapter Page

Chapter Page

Chapter Page

Chapter Page

Chapter Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

Languages vary in their prosodic typology, which includes language-specific lexical prosody, phrasal prosody, and rhythm (Jun, 2005). Infants can distinguish between native and non-native prosody, but they tune their processing to the acoustic information of a native language at a very early stage of language development (e.g., Höhle, Bijeljac-Babic, Herold, Weissenborn, & Nazzi, 2009, and references therein). As a result of this early attunement, adult listeners' perception of prosody in a non-native language is biased by native language experience (Dupoux, Sebastián-Gallés, Navarette, & Peperkamp, 2008; Gandour & Harshman, 1978; Hallé, Chang, & Best, 2004; Xu, Gandour, & Francis, 2006). The native language bias leads to foreign accents, and it may negatively affect speech processing.

The current work contributes to the body of cross-linguistic studies investigating how language experience shapes the processing of prosodic categories, and whether adult listeners can overcome their native language bias in perception of non-native prosody. This area is still underdeveloped in the theories of speech perception (Best, 1995; Best & Tyler, 2007; Flege, 1995a; Kuhl & Iverson, 1995; Strange, 1995; Strange & Shafer, 2008), perhaps due to complexity of prosodic systems and low meta-linguistic awareness for prosodic as compared to segmental categories (Bent, 2005). Perception of lexical prosody by Japanese and English listeners is a promising subject of investigation because these two languages are typologically different in their lexical prosody (pitch accent in Japanese and stress in English), and in the role lexical prosody plays in prosodic organization of speech (the differences are detailed in Chapter 2). In addition, the

phonetics and phonology of the lexical pitch-accent contrast in Tokyo Japanese has been well studied (Beckman, 1986; Kawakami, 1957; Kubozono, 1993; Miyata, 1927; Pierrehumbert & Beckman, 1988; Polivanov, 1976; Poser, 1984; Sugito, 1972; Venditti, 2006, among others), which makes it attractive for a research agenda aiming to better understand the mechanisms of speech processing and learning. The focus of this dissertation is on the processing of the acoustic cues to the pitch-accent contrast, more specifically, the cues to prominence.

Lexical prominence in Tokyo Japanese is expressed with high-low pitch accent, and it is phonetically implemented as a fundamental frequency (F0) peak followed by a sharp F0 fall. The presence of the pitch accent and its location in a word shape accentual patterns that may be lexically contrastive (e.g., the pitch-accent contrast in *tsúru* 'a crane', *tsurú* 'a vine', *tsuru* 'to fish'). Accentual patterns are also the cornerstone in the organization of Japanese phrasal prosody. In contrast, lexical prominence in English is expressed with stress and has the acoustic correlates of duration and intensity. At the phrase level, prominence may be reinforced by another acoustic correlate, F0. English listeners encounter variations in F0, including F0 fall, in phrases of their native language. It is not known, however, whether they can attend to F0 fall as a cue to lexical prominence in a foreign language such as Japanese.

Perception of pitch-accent patterns by native listeners of Tokyo Japanese and other dialects has been extensively investigated (Cutler & Otake, 1999; Hasegawa & Hata, 1992; Kitahara, 2001; Sugito, 1980, 1982a, 1982b; Vance, 1995). However, comparable acoustic studies focusing on perception of patterns by native listeners of languages with different prosodic organization, like English, are scarce (Masaki,

Takasawa, & Arai, 2001). Among existing studies in this area, the majority of work has been done using the Tokyo Japanese Accent Listening Test or a similar methodology (as reviewed in Ayusawa, 2003). The interpretation of the results in such studies appears to be biased by the methodology of the test and an insufficiently defined notion of similarity between native and non-native lexical prosody (as argued in section 2.6.2 below). The question of how non-native listeners perceive the pitch-accent contrast in terms of attention to the relevant acoustic cues remains unanswered, as well as the question of why not all pitch patterns are perceived equally well. Furthermore, a persistent finding of the relative ease of the unaccented pitch pattern for English listeners (Arai, 1997; Hirata, Ayusawa, Nakagawa, & Odaka, 1997; Nishinuma, Arai, & Ayusawa, 1996) invites a comparison with a body of work on perception of lexical tones, in which the relative inattention of English listeners to the F0 change at the word level is consistently reported (Francis, Ciocca, Ma, & Fenn, 2008; Gandour, 1983; Gandour & Harshman, 1978; Guion & Pederson, 2007; Wayland & Guion, 2004).

Extending previous work on Japanese lexical prosody, an experimental approach was used in this dissertation to investigate the perception and perceptual learning of pitch patterns in Tokyo Japanese. Naïve English listeners represented the initial stage of learning of the pitch-accent contrast (Best, 1995). The experimental approach allowed a close examination of listeners' sensitivity to F0 peak location and F0 fall after the peak, which are the primary cues to lexically contrastive patterns. This is the first large-scale study conducted with non-native Japanese listeners that is methodologically comparable to perception studies conducted with native Japanese listeners (Hasegawa & Hata, 1992;

Kitahara, 2001) and to training studies conducted with adult learners of tone languages (Francis et al., 2008; Wang, Spence, Jongman, & Sereno, 1999).

A series of four perception experiments (Chapters 3-6) was designed to examine whether English listeners can attend to the acoustic cues of the pitch-accent contrast. The experiments differed in the number of manipulated acoustic parameters (three vs. two), the number of answer options available to listeners (two vs. three), and the type of the task (categorization vs. discrimination), as all these factors can potentially affect listeners' sensitivity to the cues (Bent, 2005; Eda, 2004; Hallé et al. 2004; Miyawaki et al., 1975; Sakamoto, 2008; Wayland & Guion, 2003). The stimuli were resynthesized F0 patterns in nonwords, modeled after the three-way pitch-accent contrast (1^{st}-syllable accented, 2^{nd}-syllable accented, and unaccented patterns). Importantly, in all categorization experiments, the wording of instructions was similar (and linguistically appropriate) for both non-native and native listeners. The listeners' task was formulated as identification of the most prominent syllable in a word (see section 2.2 in literature review).

In Experiment 1, listeners categorized F0 patterns into the patterns of 1^{st} syllable prominence and 2^{nd} syllable prominence, while three acoustic parameters were manipulated. In Experiment 2, the same task was repeated with just two manipulated parameters. The concluding categorization Experiment 3 offered listeners a *no prominence* answer option in addition to the other two, while using the same stimuli design. Based in previous research, it was predicted that English listeners would use F0 peak location over F0 fall as a cue to prominence, whereas Japanese listeners would use both cues (Gandour, 1983; Gandour & Harshman, 1978). An additional Experiment 4 in

the series was a discrimination task with similar acoustic stimuli. English and Japanese listeners' sensitivity to the F0 fall in a same-different task was examined. The results of all four perception experiments are summarized and discussed in Chapter 7.

The pitch-accent contrast has been shown to be difficult for English listeners of Japanese (e.g., Hirata, 1999; Nishinuma et al., 1996), yet it is important for the organization of Japanese prosody (e.g., Pierrehumbert & Beckman, 1988; Venditti, 2005) and word processing (e.g., Cutler & Otake, 1999; Kitahara, 2001), and as such it constitutes an important aspect of Japanese to be learned. The experiment in Chapter 8 was designed to investigate the effectiveness of training on the formation of categorical representations of the pitch-accent contrast, as well as the difference in learnability of the three pitch patterns in question. The high-variability training paradigm with a classic design including a pretest, a posttest, and generalization tests was used (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Lively, Logan, & Pisoni, 1993; Wang et al., 1999). Participants were trained on naturally produced accentual patterns in sentential contexts. Such a training study, focused exclusively on lexically contrastive pitch patterns in a pitch-accent language, has not yet been conducted (but see pedagogically oriented training on Japanese rhythm and pitch patterns in Hirata, 1999). The main findings of the dissertation as they apply to the perception and learning of second-language prosodic categories are discussed in Chapter 9.

The following Chapter 2 provides a detailed literature review for the reader less familiar with research on lexical prosody specific to Japanese and English, or with research on cross-linguistic perception and learning of lexical prosody in general.

CHAPTER II

LITERATURE REVIEW

2.1. Functions of the Japanese Lexical Pitch Accent

The goal of this section is to describe the functions of the accent in native Japanese speech and the importance of the accent for second-language Japanese speech. A detailed review of literature relevant to functionalism in phonology is available in Beckman's monograph (1986, pp. 11-44). Beckman attributed to Trubetskoy's work the recognition of three functions of the (stress or pitch) accent, namely: 1) signaling difference in lexical meaning (the *distinctive* function), 2) signaling a speech unit that stands out among all others (the *culminative* function), and 3) signaling the boundaries of the unit (the *delimitative* function). She also argued that the culminative and delimitative functions are, in essence, the same: namely, to signal prominent (or salient) units in a phrase and by doing so define the phrase. I will adopt this view in the dissertation.

The distinctive function of the Japanese lexical pitch accent has been apparent for as long as this accent has been described. The accent differentiates the meaning of homophones like *hana* 'nose' and *haná* 'flower'. It has been argued, however, that in modern Japanese the distinctive function of the accent is less important than the culminative function (Beckman, 1986), perhaps, due to the rise of the accentual phrase in Japanese prosodic system (Komatsu, 1989). In addition, Japanese speakers are exposed in their lifetimes to multiple dialects with different accentual systems via mass media and migration. This may explain large individual variability in pitch-accent production and perception by Japanese native speakers. For example, in the experiment by Beckman (1986), in which native speakers of Tokyo Japanese were asked to produce minimal word

pairs in sentences, one of the four speakers failed to produce expected accentual distinctions. This far-from-perfect performance of native speakers is even more apparent in larger scale studies (Ayusawa, 2003; Shport, 2008).

The ease of distinguishing the meaning of homophones in context may also be taken as evidence of a relatively low load of the distinctive function. In principle, a Japanese word can be unaccented or have an accent on any syllable, thus yielding $n + 1$ possible accentual patterns in an n-syllable word. In practice, few words in the Japanese lexicon have all $n + 1$ number of accentual patterns (e.g., *mori* 'woods', *móri* 'guard', *morí* 'leaking'). In many minimal pairs or triplets of homophones, one member of a pair is less frequent than another (e.g., *móri* 'guard'), or it is restricted to mostly written language, or it is a bound morpheme typically appearing in a combination with other morphemes (e.g., *mori-bito* 'protector', *amá-mori* 'leak in the roof'). Due to these reasons and to the pedagogical tradition and time constraints in a second-language classroom, learners of Japanese are almost never explicitly taught accentual patterns of words. As a result, even advanced learners of Japanese rarely acquire the lexical pitch-accent contrast (see the review of studies on this topic in Ayusawa, 2003). Acquisition of these accentual patterns seems to be important, however, for adult second-language learners of Japanese for several reasons.

First, the amount of words distinguished by pitch accent is not completely negligible. Even though few n-syllable words have all $n + 1$ accentual patterns, many shorter words have at least one homophone. Kitahara (2001) has pointed out that the distinctive function cannot be downplayed for modern Japanese, where about 13% of short words are distinguished only by pitch accent, and all accentual pattern types are

equally frequent among 1-2 syllable words. Furthermore, native listeners of Tokyo Japanese use the pitch-accent information in spoken word recognition (Cutler & Otake, 1999; Masuda-Katsuse, 2006; Sekiguchi & Nakajima, 1999). One can imagine that saying *amérikajin* instead of correct *amerikájin* may sound for a Japanese ear the same as *Américán* instead of *Américan* for an English ear. Lexical prosody is an important prosodic characteristic of the language, and processing of non-native speech in cases where lexical prosody is violated should be more difficult (attentionally demanding) for native speakers.

Second, I will review literature showing that accentual patterns of words along with boundary and phrasal tones form larger prosodic units such as accentual phrases and intonational phrases (section 2.3 below). Within these larger prosodic units, the pitch accent has the culminative-delimitative function of signaling prominence. If a learner of Japanese has not acquired lexical patterns, then the prosody of larger units in second-language speech would be distorted, which may affect communication with native listeners. Furthermore, the learner may not be capable of utilizing prosodic cues in processing of Japanese speech.

Japanese lexical pitch accent and intonation of accentual phrases are known to be difficult for learners of Japanese in both production and perception (see a review on this topic in Ayusawa, 2003). American learners tend to produce a) unaccented Japanese words as accented on the first syllable (Shibata, 2005) or on the penultimate syllable (Ueyama, 2000; Yoshimitsu, 1981); b) 2^{nd}-syllable accented disyllabic words as 1^{st}-syllable accented (Shibata, 2005); and c) accented negative verbal morpheme *-nái* as

unaccented (Yoshimitsu, 1981). These production errors may be indexical of perceptual categorization errors.

Perceptual categorization of Japanese accentual patterns by English listeners is investigated in this dissertation under assumption that the pitch accent signals prominence (Beckman, 1986), and thus functionally it is comparable to English stress. The next section establishes the relationship between prominence, accent and stress from a cross-linguistic perspective.

2.2. Accent as Prosodic Prominence

2.2.1. Definitions

The lexical pitch-accent contrast of Tokyo Japanese was construed in perception experiments in this dissertation as prosodic prominence of one-word utterances (i.e., isolated words). The motivation to use the term prominence came from the need to give similar instructions to Japanese and English listeners in the tasks investigating categorization and perceptual learning of patterns varying in their fundamental frequency (F0). Syllables with a relatively high F0 were considered as manifestations of high tones in the experimental stimuli. If we had used the terms suggested by Beckman (1986), these high tones would have been referred to as *stress accents* for English listeners and *pitch accents* for Japanese listeners. This labeling would have been misleading for English listeners because only F0 was manipulated in the stimuli. Using the term pitch accent for both languages would have been confusing because familiarity with this term cannot be expected from listeners naïve to linguistics. The term prominence, on the other hand, allowed us to avoid these pitfalls and to refer to high tones in a similar fashion for all listeners. This term emphasized the general function of pitch accents in both languages,

which is to highlight one syllable in the speech string relative to the rest of syllables (i.e., the culminative-delimitative function in Beckman, 1986). Furthermore, there was no need to specify possible causes of prominence such as tone, stress, pitch accent, focus, or to specify structural prosodic domains of prominence such as a word or a sentence.

This usage of the term prominence is not novel. It is consistent, for example, with the proposal of Sun-Ah Jun, who considered prominence as a category in prosodic typology: namely, the category that "concerns both how the prominence of a lexical item in the language is realized prosodically and how the prominence relation among words is realized postlexically" (2005, p. 442). In this view, the prosodic properties of an utterance emerge from a combination of *word-level* (i.e., lexical) and *sentence-level* (i.e., postlexical) prosody (p. 431). Adopting this view, we can assume that both stress and pitch accent have a common function of marking prominence, although in typologically different ways. Prominent syllables will be marked with acute accent in this dissertation for examples both in English, a stress language, and Japanese, a pitch-accent language – a distinction proposed by Beckman (1986) and widely used in the literature, including Jun's typological overview (2005).

In the broad definition of prominence, both high and low tones may be, in principle, prominent (Pierrehumbert, 1980, among others). However, in the perception experiments in this dissertation, when listeners were asked to identify the most prominent syllable in a word, they were assumed to be listening for a relatively high fundamental frequency. The assumption that high tones are more prominent than low tones is supported by Mo's experimental findings (2009), which showed that the perception of prominence by English listeners naïve to linguistics is correlated with F0 peaks.

Pierrehumbert and Beckman (1988) implemented this assumption in their synthesis model of Japanese intonation, where prominence of any tonal event can be determined mathematically by using several parameters that scale the F0 value of the tone to the pitch range of an utterance. As a result, low tones have lower prominence values than high tones (p. 189).

The broad definition of prominence adopted in this dissertation is different from more narrow or more technical definitions, in which the term refers only to discourse-level categories such as focus (Venditti, Maekawa, & Beckman, 2008), or to relative F0 values associated with a pitch accent (Pierrehumbert & Beckman, 1988), or to heavy syllables containing long vowels, low vowels, high tones, or sonorant codas (Hayes, 1995, pp. 270-276). Some experimental studies also have adopted the broader definition for designing experimental tasks and investigating speech corpora (e.g., Kochanski, Grabe, Coleman, & Rosner, 2005). The types of prominence in English and Japanese, and their language-specific acoustic correlates are discussed in the following two sections.

2.2.2. Prominence in English and Japanese

Early experimental studies were concerned with mapping stress as a complex perceptual phenomenon onto discrete physical properties of sound. In his classic experiments, Fry (1955, 1958) manipulated several physical dimensions such as vowel duration, intensity and F0 in disyllabic noun-verb pairs contrastive only in their stress patterns (e.g., *dígest-digést, óbject-objéct*). The target words were presented to listeners either in isolation or in the carrier sentence *Where is the accent in ___?* In both presentation modes, the words were in a focused position; that is, they were likely to carry nuclear accents characterized by higher F0. Because pitch accents in English land

only on stressed syllables, stress judgments that listeners were asked to make by underlying either DIgest or diGEST were, in fact, judgments of *cumulative prominence*. As a result, the reported effect of F0 on stress perception could have been due to word-level stress, sentence-level pitch accent, or both.

Later studies were careful either 1) to distinguish between lexical (stress) and sentential (pitch accents) components of what listeners might perceive as stress (Sluijter & van Heuven, 1996; Vanderslice & Ladefoged, 1992), or 2) to look at the cumulative effect of the two components on what listeners might perceive as prominent (Kochanski et al., 2005; Silipo & Greenberg, 2000). The first line of research focused on establishing acoustic correlates of lexical stress while controlling the sentential focus. For example, Sluijter and van Heuven (1996) elicited sentence-medial target words in Dutch under two conditions: in the focus position, as in *Wil je kaNON zeggen* 'Will you cannon say (rather than rifle)'; and in no-focus condition, as in *Wil je kanon ZEGgen* 'Will you cannon say (rather than write down).' Interestingly, in contrast to stress production studies, stress perception studies have not always controlled for possible interaction between lexical and postlexical prosody. For example in an article published one year later by Sluijter, van Heuven and Pacilly (1997), all target words were presented to Dutch listeners in the same carrier sentence *Will you ___ say*. Thus, it is unknown whether acoustic cues of stress reported in this study are due to the lexical stress per se, or to the cumulative prominence.

The second line of research investigated the category of prominence, which subsumes the category of stress in English. Terms such as 'nuclear accent,' 'prosodic stress,' and 'rhythmic prominence' refer to cumulative prominence of word-level lexical stress and sentence-level intonational pitch accents that co-occur on metrically strong

syllables. For example, in the sentence *I saw Laura*, the ['lɔ] syllable is perceived as prominent because it contains a full, relatively long vowel at the word level, and it bears a pitch accent at the sentence-level (Ortega-Llebaria, del Mar Vanrell, & Prieto, 2010). This example illustrates how the acoustic cues to prominence at the word and sentence levels overlap. Since pitch-accented syllables are always stressed, the terms pitch-accented and prominent can be used interchangeably for English.

In pitch accent languages like Japanese, word-level prominence is not obligatory, i.e., some words have lexical accents (e.g., *kamí* 'hair'), and some do not (e.g., *kaze* 'wind'). Therefore, sentence-level pitch accents may or may not land on syllables carrying lexical pitch accents. In contrast to English, word-level and sentence-level pitch accents cumulatively mark prominence only in phrases with lexically accented words.

In sum, in both English and Japanese there is a word-level component of prominence (i.e., obligatory lexical stress in English and optional lexical pitch-accent in Japanese), and a sentence-level component of prominence (i.e., obligatory intonational pitch accents). The cross-linguistic difference in acoustic cues coding prominence is discussed below.

2.2.3. Acoustic Manifestation of Prominence

The acoustic manifestation of prominence varies across languages. For stress languages, research indicates that listeners mainly rely on the combination of length and loudness in their perception of stress – in English (Fry, 1955, 1958; Kochanski et al., 2005; Mo, 2008; Ortega-Llebaria et al., 2010; Silipo & Greenberg, 2000; Turk &

Sawusch, 1996), Dutch (Sluijter et al., 1997), Castilian Spanish (Ortega-Llebaria, Prieto, & Vanrell, 2007), and Catalan (Ortega-Llebaria et al., 2010), among other languages.[1]

Although length is uncontroversially represented by the duration measure, the best operationalization of loudness is still debated. Several measures may be used to represent loudness: peak amplitude, root-mean-square amplitude (also known as intensity), average amplitude, spectral tilt (also known as spectral balance, or vocal effort), or subjective loudness on Steven's sone scale. In some studies, when two or more measures of loudness were used, they were found to have similar effect on prominence judgments within one language (for English: Beckman, 1986; Kochanski et al., 2005; Mo, 2008), or their effect on stress judgments was found to vary depending on the language. Perceived loudness is most closely related to root-mean-square amplitude, which was used in this dissertation as the measure of loudness in experimental stimuli.

In contrast to length and loudness, the role of pitch (as measured by F0) has been disputed for perception and production of prominence. Fry (1955) and Terken and Hermes (2000) found that English stress is cued by F0, along with duration and intensity. Beckman (1986) and Mo (2008) made a similar observation with regards to cumulative prominence. These results, however, are inconsistent with results reported in other studies. Sluijter and van Heuven (1996) showed that when words were not in focus in Dutch sentences, F0 was no longer a correlate of stress. Similarly, studies that did not conflate lexical and postlexical prominence in English found that the F0 difference between stressed and unstressed syllables disappears in postnuclear (i.e., after focus)

[1] Vowel quality is considered to be another cue to English stress (Beckman, 1986). However, since only particular kinds of vowels (i.e., open, full) can be stressed, vowel quality appears to be not a consequence of stress, but a condition for it to occur.

position (Huss, 1978; Ueyama, 2000; Ueyama & Jun, 1998). Kochanski et al. (2005) analyzed acoustic characteristics of syllables that two experienced phonologists labeled as prominent. They found that F0 was not correlated even with these cumulative prominence judgments. Similar results were reported by Kochanski and Orphanidou (2008), Silipo and Greenberg (2000), Turk and Sawusch (1996), among others. In sum, prominent syllables in English are undoubtedly coded by longer duration and greater intensity, but they may or may not be coded by higher F0.

In non-stress languages like Japanese, prominent syllables (i.e., syllables with lexical or postlexical pitch accents) are cued primarily by F0 in both perception and production (Beckman, 1986; Beckman & Pierrehumbert, 1986; Sugito, 1982a, 1982b). Furthermore, this role of F0 in the aforementioned work does not seem to depend on the type of materials (e.g., isolated words, context-embedded words) or the F0 measure used (e.g., the third harmonic divided by three, semitones, F0 at vowel midpoints, F0 at peaks and troughs).

Some previous studies suggested that duration and intensity are also relevant to the distinction between accented and unaccented syllables in Japanese. However, when Poser (1984) reviewed the methodology in this work, he disputed the link between Japanese pitch accent and duration (pp. 39-41). Poser noted that the effect of accent on duration was reported only for contrasts between accented and unaccented initial syllables, or words in isolation. Interestingly, studies asserting the link between Japanese pitch accent and intensity also used isolated words (Cutler & Otake, 1999; Neustupný, 1966). And similarly to Poser's argument, experimental studies by Sugito (1982b) and Beckman (1986) disputed the idea that intensity is a correlate of and a cue to pitch accent

in Japanese. Overall then, there seems to be a general agreement in the literature that prominent syllables in Japanese are coded primarily by F0.

Cross-linguistic studies on Japanese and English should be concerned with comparable domains of prominence in both languages. In Japanese, word-level prominence is not obligatory; in English, it is. Furthermore, in Japanese F0 is a correlate of prominence at both word and sentence levels, and in English F0 is probably only relevant at the sentence level. One methodological solution to this problem of comparison is to investigate the effect of the variation in F0 in a priori different domains - the word level for Japanese listeners and the sentence level for English listeners (e.g., Hasegawa & Hata, 1992). Here, I adopt another solution, namely, to admit that even isolated words are one-word utterances with the target word in focus, and to compare the effect of the variation in F0 on the cumulative perception of word- and sentence-level prominence for both English and Japanese listeners. This approach is possible because the effect of F0 is investigated for prominence rather than for stress or pitch accent. It does not undermine, however, cross-linguistic differences between stress and lexical pitch accent. The acoustic manifestation of prominence is language-specific, and word-level prominence is an integral part of language-specific prosodic structures.

2.3. Prosodic Organization in Japanese and English

This section provides a phonological comparison of prosodic organization in Japanese and English and the place of pitch accents in the prosodic systems of these two languages. A comparison that adopts the same theoretical framework and transcription conventions facilitates a better understanding of differences between the systems, and as relevant to the goals of this dissertation, possible challenges that second-language

learners may face in acquisition of non-native lexical prosody. Therefore, the cross-linguistic comparison here is based on the state-of-art analyses of the prosodic systems in Tokyo Japanese (Venditti, 2005) and in Mainstream American English (Beckman, Hirschberg, & Shattuck-Hufnagel, 2005) within the framework of the Autosegmental-Metrical model of intonational phonology. The descriptive application of this framework to different languages and references to seminal papers in the field are available in the volume on prosodic typology edited by Sun-Ah Jun (2005). The Autosegmental-Metrical model has served as a foundation for the development of the Tone and Break Indices (ToBI) transcription system for language prosody, which is becoming a standard tool in the field of prosodic analysis.

Although the Autosegmental-Metrical model per se is not tested in this dissertation, some ToBI conventions will be used throughout to refer to tonal events in intonation patterns. Assumptions of the Autosegmental-Metrical model and ToBI conventions presented below recapitulate the main points of the introduction and the chapter on prosodic typology by Jun (2005), as relevant to the current work.

2.3.1. The Phonological Model of Prosodic Organization

The Autosegmental-Metrical approach assumes that intonation patterns are hierarchically organized sequences of high (H) and low (L) tones and their combinations. Tonal targets are associated with a specific syllable in a word marking the prominence relation in a prosodic grouping (*pitch accents*), or with a specific location in a phrase marking its edge (*phrasal accents* and *boundary tones*). The categorization of tonal targets in ToBI is based on their function (culminative or delimitative), the place of

occurrence in prosodic groupings of an utterance (phrase internal or phrase edge), and shape (single or bitonal).

All languages have prosodic units above the word; however, their types and characteristics are language-specific. Within a language, prosodic groupings are defined by tone events and the degree of juncture among the words that are known to be particular to those groupings. For example, the smallest prosodic grouping in Japanese is the *accentual phrase*, which is defined by a high tone that occurs on the second mora of the phrase or later, and by a low tone that occurs at the right boundary of the phrase (Venditti, 2005, p. 180). Using ToBI conventions, the intonation pattern of the accentual phrase following a pause is %L H- L%, where the hyphen marks a high-tone phrasal accent and the percentage sign marks low-tone boundary tones of the phrase. In English, on the other hand, the smallest prosodic grouping is called the *intermediate phrase*, and it is defined by a high or a low phrasal accent that occurs immediately after the syllable with the most prominent (nuclear) pitch accent in the intermediate phrase (Beckman, et al., 2005, p. 35).[2] The intonation pattern of the intermediate phrase may be, for example, H* H- or L* H-, where the hyphen marks a phrasal accent, and the star marks a prominent pitch accent of the phrase.

Turning to pitch accents, ToBI defines them in terms of their domain and shape. The domain of pitch accents can be a word as in Serbo-Croatian, an accentual phrase as in Tokyo Japanese, an intermediate phrase as in Neapolitan Italian, or an intonational

---

[2] The degree of disjuncture is also important in Mainstream American English ToBI system for identification of intermediate phrases. The phrasal accents H- and L- of intermediate phrases are obligatory at every disjuncture of the relative strength 3 or higher (Table 2.2 in Beckman, et al., 2005). Considering our focus on the lexical pitch accent in the cross-linguistic perspective, the ToBI break indices are not discussed here in relation to prosodic groupings.

phrase as in Chickasaw (Jun, 2005). For Germanic languages such as English and German, whether the domain of pitch accents is an intermediate phrase or an intonational phrase is still a point of ongoing discussion. For the Japanese language, the Autosegmental-Metrical model does not explicitly specify whether the pitch accent is a lexical or postlexical element (Jun, 2005, p. 437). This theoretical ambiguity may complicate a comparison between languages such as Tokyo Japanese and English in that the domains of pitch accents are clearly different in these two languages, but this difference is not consistently reflected in ToBI conventions. In Japanese the pitch accent H*+L is lexical property of a word, and it further interacts with postlexical prosody of an accentual phrase (as detailed and illustrated in the following section). According to ToBI, however, the proposed domain of this pitch accent is an accentual phrase and not a word, which invites a dubious comparison with postlexical pitch accents in English. In this dissertation, lexical pitch accents occurring at the word level (for Japanese) and postlexical pitch accents occurring at the sentence level (for English) are distinguished.

In terms of shape, ToBI distinguishes between single-tone accents H* and L*, on one hand, and their combination resulting in bitonal accents like H*+L and L*+H, on the other hand.[3] Tones in bitonal accents are hierarchically organized in terms of a relative strength relationship: The star notation marks the tone that is relatively more prominent as compared to the other tone in the same accent. Furthermore, the star-marked tone in a

[3] There is some evidence contradicting the distinction between single and bitonal pitch accents. For example, Ladd and Schepman (2003) found in their experiments that both H* and L+H* tones in English had distinct low- and high-tone targets that were realized at the beginning and at the end of the accented syllable, respectively. Dilley, Ladd and Schepman (2005) further showed that the tones in the L+H* bitonal target were aligned relative to segments, not to each other. These findings support the argument that H* and L+H* tones should be regarded as belonging to the same phonological tone sequence.

single or bitonal accent is associated with a particular placeholder in a segmental string, for example, an accented syllable in Japanese or a stressed syllable in English, although the phonetic alignment of the same pitch with respect to segments of a syllable varies in languages. The unstarred tone is leading to or trailing off of the starred tone, and its alignment with the segmental string may vary. In this dissertation, the term prominence refers to the (starred) high tone of the Japanese bitonal pitch accent H*+L.

2.3.2. ToBI-Style Comparison of Tone Events in Japanese and English

The inventories of tone events in Japanese and English are compared by the prosodic grouping in which they occur, that is, intonational phrase, intermediate phrase, and accentual phrase, (Table 2.1). Intonation patterns illustrating the comparison are presented in Figure 2.1 for Japanese and in Figure 2.2 for English.

The highest-level prosodic grouping in both Japanese and English is the intonational phrase, with boundary tones as the only tone events in this grouping. For example in Figure 2.1, the Japanese intonational phrases end either in the explanatory HL% tone (purple contours), or in the question H% tone (yellow contours). Figure 2.2 shows two boundary tones marking the right edge of English intonational phrases – the low tone L% signaling the statement intonation (purple contours), and the high tone H% signaling the question intonation (yellow contours). The examples in Figures include relatively short (3-6 syllable) utterances. On average, the size of intonational phrases varies from seven to ten syllables in both Japanese and English (Jun, 2005, p. 442).

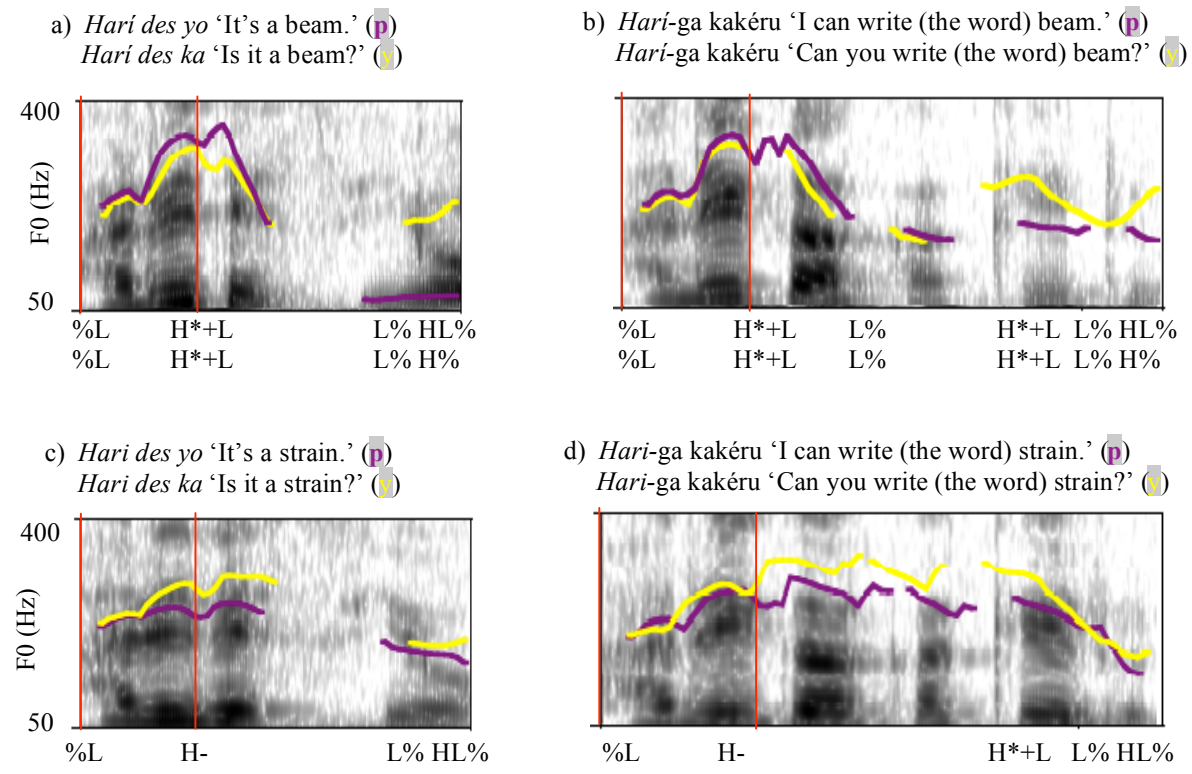


Figure 2.1. F0 contours of utterances containing the 2nd-syllable accented word *harí* 'a beam' (top panels) and the unaccented word *hari* 'strain' (bottom panels) produced by a native Japanese speaker. Vertical lines indicate the word boundaries. Statement contours are in purple (p, top rows); question contours are in yellow (y, bottom rows).

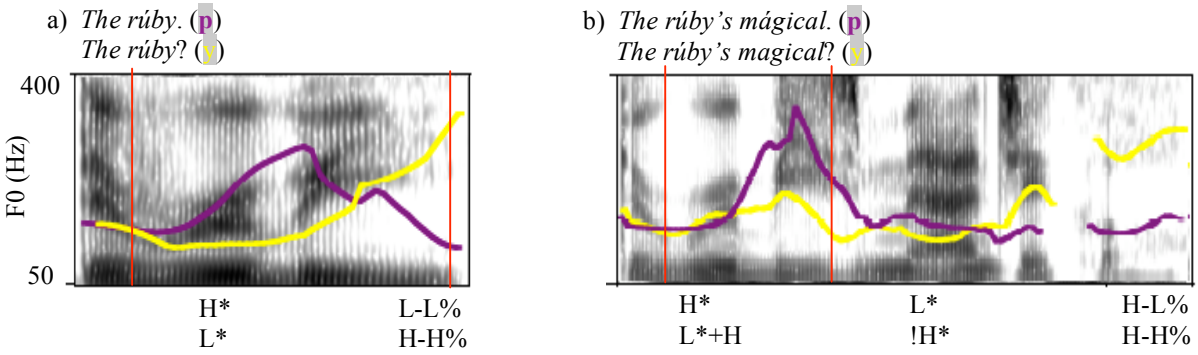


Figure 2.2. F0 contours of utterances containing the 1st-syllable stressed word *rúby* in a focused position produced by a native English speaker. Vertical lines indicate the word boundaries. Statement contours are in purple (p, top rows); question contours are in yellow (y, bottom rows).

Table 2.1

A Comparison of Tone Inventories in Tokyo Japanese and American English

Phrase type	Tone events	Tokyo Japanese	American English
Accentual	Pitch accents	**H*+L**	
	Phrasal accents	H-	_____
	Boundary tones	%L (after a pause) L% (phrase-final)	
Intermediate	Pitch accents	_____	L*, H* L+H*, L*+H, H+!H*
	Phrasal accents		H-, L-
Intonational	Boundary tones	H% (prominence-lending or insisting) LH% (question) HL% (explanation)	H% (continuation rise or question) L% (affirmation) %H (phrase-initial)

Note. In ToBI-style transcription, H = high tone; L = low tone; * = (relatively) prominent tone; % = boundary tone; H-/L- = phrasal accent; ! = downstep. Table presents reorganized inventories from Table 16.1 in Jun, 2005 (p. 434) with additional information in parentheses from Venditti (2005) and Beckman et al. (2005). Pitch-accent domain in MAE is specified tentatively here as an intermediate phrase, to facilitate the comparison.

The next level of prosodic grouping is an accentual phrase in Japanese and an intermediate phrase in English. These types of phrases differ both in terms of the tone events that define them and in their typical size. An accentual phrase in Japanese has an obligatory phrasal accent H- and a boundary tone L% (Figures 2.1c and 2.1d). In addition, the phrase may contain only one word with the lexical pitch accent H*+L (hence, two accentual phrases in Figure 2.1b). This accent also shapes the overall pattern of the phrase and obscures the phrasal accent H- when associated with the same syllable, as in Figures 2.1a and 2.1b. An intermediate phrase in English has an obligatory phrasal

accent (H- or L-) and at least one of the pitch accents shown in Table 2.1.[4] Figure 2.2 shows the phrasal accents following the last pitch accents of the utterances – on the syllable -*by* in Figure 2.2a, and on the syllable -*gi*- in Figure 2.2b.

The pitch-accent inventory is limited in Japanese, whereas English enjoys a wider variety and a larger number of pitch accents that occur at lower-level phrases (Table 2.1). Not surprisingly, the size of lower-level phrases also tends to differ in the two languages. On average, English intermediate phrases are several syllables longer than Japanese accentual phrases (Jun, 2005, p. 446-447, footnote 4). Therefore, the domain size of phrasal accents and pitch accents is likely to be larger in the former than in the latter language.

Considering characteristics of pitch accents, the only Japanese pitch accent is a bitonal lexical accent H*+L. Both high and low tones are two targets within one pitch accent that induce a fall in the fundamental frequency between them. Accentual phrases containing an accented word exhibit a falling F0 movement associated with the accented syllable, regardless of the sentential context in which the word occurs. A comparison of the purple contours showing the F0 movement in and after the 2nd-syllable of the accented word *harí* in Figures 2.1a and 2.1b illustrates this characteristic of the H*+L

---

[4] The inventory or American English pitch accents proposed by Janet Pierrehumbert in her dissertation (1980) and maintained with small changes in her later work (Beckman & Pierrehumbert, 1986; Pierrehumbert & Beckman, 1988; Pierrehumbert, 2000) is slightly different than in the current ToBI version (Beckman et al., 2005). Of interest, one of the originally posited pitch accents was H*+L, which had the same shape as the Japanese lexical pitch accent. Pierrehumbert and Hirschberg (1990) further proposed a meaning distinction between the H*+L and H* accents in English (H*+L: p. 297; H*: pp. 290-291). In the current ToBI version for American English, however, the H*+L accent has been reanalyzed as the H+!H* accent, where the second high-tone is lower than the first one due to downstep (i.e., phrase pitch-range compression after the initial high tone). For the purpose of this dissertation, given either label – H*+L or H+!H* – this pitch accent induces the high-low F0 movement typical for English intonation patterns due to the interpolation between tones.

accent, which is evident from the F0 contours of the unaccented word *hari* in Figures 2.1c and 2.1d. Note that the pitch accent is not obligatory in words, and in accentual phrases formed by those words.

In contrast, English pitch accents are obligatory for an intermediate phrase, which can contain one pitch accent or more, depending on the meaning intended by a speaker. These pitch accents are postlexical, that is, they are not a property of stressed syllables in content words. The comparison of F0 contours corresponding to the same word *the rúby* in Figure 2.2a illustrates this fact. Differently from Japanese, there is no pitch accent of the H*+L shape in the current inventory of English pitch accents (see, though, Footnote 4). However, the fall in fundamental frequency occurs due to the interpolation between separate tone events in a phrase (e.g., the statement intonation pattern H* L-L%), as well as due to the downstep within one tonal event (H+!H*).

Japanese and English also differ in the manifestation of focus, the most prominent event in an utterance. According to the summary of these differences by Venditti, Maekawa and Beckman (2008), there is no specific pitch accent in Japanese associated with focus marking. Instead, to mark focal prominence, Japanese uses 1) pitch expansion of existing pitch accents, 2) postfocal compression of the pitch range or dephrasing, and 3) boundary tones at the end of the focal constituent such as "prominence-lending rise" H% or "rise-fall" HL% (Table 2.1). This focus realization is clearly observed in Figure 2.1a, where the purple contours of the utterance with narrow focus is slightly higher than the yellow contour of the utterance with broad focus in the syllable associated with the H*+L accent (i.e., pitch expansion), and where the purple contour is lower than the yellow one after this accent (i.e., postfocal compression). In contrast to Japanese, English

marks focus by a nuclear pitch accent that occurs on a focused word when the utterance focus is narrow, or on an utterance-final word when the focus is broad. In Figure 2.2, the nuclear pitch accents can be observed in the purple contours of utterances with narrow focus and the yellow contours of utterances with broad focus.

To summarize the comparison between Japanese and English as relevant to pitch accents, both Japanese and English listeners encounter a succession of high and low tones in the intonation patterns of their native language, which results in F0 movement from high to low. However, these sequences tend to occur in larger-size or higher-level prosodic groupings in English (i.e., intermediate and intonational phrases) and in smaller-size or lower-level groupings in Japanese (i.e., accentual phrases). The question therefore arises as to whether English listeners can attend to this movement as contrastive at the lexical level, which is important for recognition of accented words and accentual phrases in Japanese. Another empirical question is whether English listeners can attend to the lack of F0 movement as contrastive, as in Japanese unaccented words and phrases.

2.4. Acoustic Cues to the Japanese Pitch Accent H*+L

The previous section showed how the lexical pitch accent is a building block of phrasal intonation in Tokyo Japanese. Namely, the presence of the lexical pitch accent H*+L changes the prosody of an accentual phrase as compared to phrases consisting of only unaccented words. These prosodic differences are manifested by variation in F0. This section provides the reader with a detailed phonetic description of F0 contours in Japanese accentual phrases containing lexically contrastive words, in order to motivate the modeling of contours in the current work. The focus here is on the acoustic cues to the lexical pitch-accent contrast, although the phrasal context is often necessary for full

manifestation of these acoustic cues.

2.4.1. F0 Fall as a Primary Cue

The F0 fall characteristic of the pitch accent has been known since the beginning of the 20th century, and even earlier in dictionary practice. Sugito (1972) mentions the 1892 dictionary by Yamada, where lexical accents of Tokyo Japanese were marked. In the first chapter of his dissertation, Poser (1984) provides an excellent historical overview of the work on Japanese accents (pp. 29-36), to which the work by Polivanov (1976) should be added. With technological advances and popularization of acoustic studies, experimental evidence supported the observations that the F0 fall is the fundamental acoustic correlate of the pitch accent H*+L in production (Beckman & Pierrehumbert, 1986; Sugito, 1980), and the primary acoustic cue to accent in perception (Sugito, 1980, 1982a, 1982b). As Venditti (2006) summarizes it:

> The presence or absence of a sharp fall in F0… is the hallmark of the accented vs. unaccented distinction in Japanese: accented words display the sharp fall near the end of the accented mora, while unaccented words lack such a fall. (p. 209)

Notably, the emphasis in the definition above is on the F0 *fall* rather than the F0 *peak* (the high tone of the H*+L) or F0 *elbow* (the low tone of the H*+L). This phonetic description focusing on the F0 movement contrasts with the phonological description focusing on the level tonal targets of the accent. However, this accent can be construed in either way, since the F0 fall implies a sequence of the high and low tones and visa versa.

Figure 2.3 provides an illustration of accented and unaccented F0 contours in the one-phrase utterance *Hashi des* 'It's *hashi*' produced by two native speakers of Tokyo Japanese. Three real words *hashi* consist of the same segmental string and differ only in

their lexical prosody. A comparison of F0 contours corresponding to the word *hashi* in Figure 2.3 shows that the accented words 'chopsticks' (green contours) and 'bridge' (blue contours) have F0 maximum in the 1st and 2nd syllable, respectively. After these F0 peaks, a steep fall in F0 occurs either on the following syllable (e.g., *-shi* for 'chopsticks,' top panel), or at the syllable boundary (e.g., between *-shi* and *des* for 'bridge,' top panel), or across two syllables (e.g., *shi* and *des* for 'bridge,' bottom panel). These examples show variability in the timing of the accent, to which we return later in the discussion of accent alignment. For a distinction between accented and unaccented words, the presence or absence of F0 fall is crucial, which is evident in the unaccented (red) F0 contours.

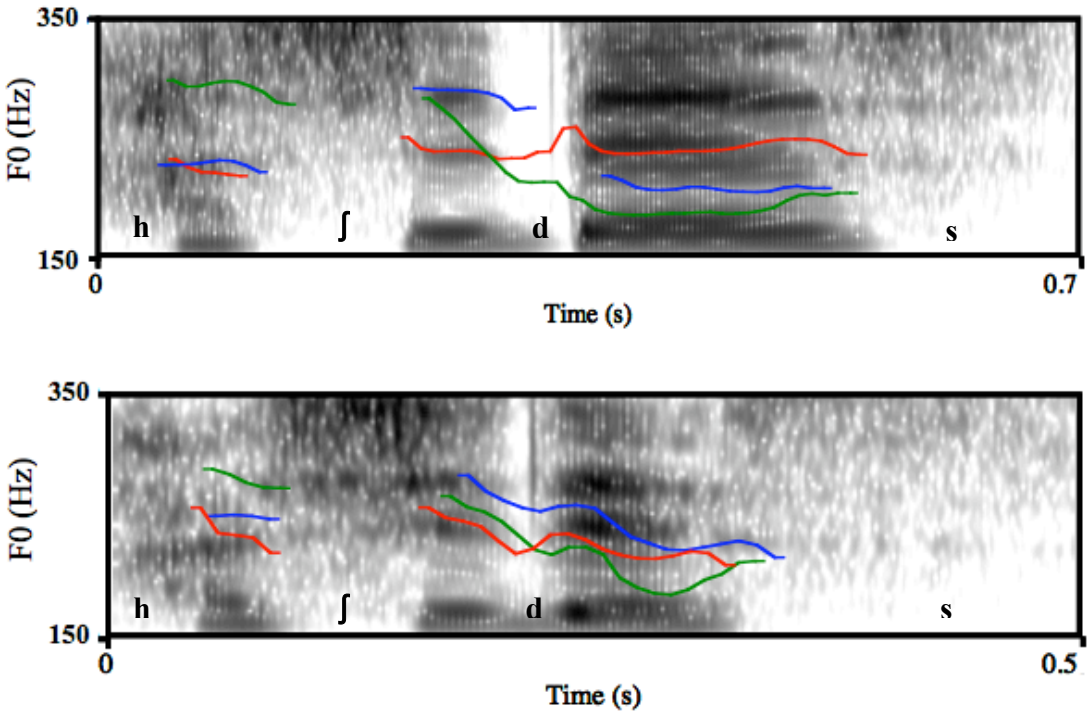


Figure 2.3. F0 contours of three utterances *Hashi des* 'It's *hashi*', where *hashi* was produced as 1st-syllable accented word 'chopsticks' (green contours), 2nd-syllable accented word 'bridge' (blue contours), and unaccented word 'edge' (red contours). Utterances were elicited from female Tokyo Japanese speakers RH (top) and AK (bottom) as responses in scripted dialogues. The F0 contours are overlaid on the spectrograms of the unaccented phrases 'It's an edge.'

The steepness of F0 fall is also important for pitch-accent contrast, especially, in productions of some speakers. Compare the contrast produced by the speakers RH (top panel of Figure 2.3) and AK (bottom panel of Figure 2.3). For RH, the unaccented red contour may be distinguished from the other two accented contours based just on the absence of the F0 fall. For AK, on the other hand, both of the unaccented red contour and 2nd-syllable accented blue contour have F0 maximum on the second syllable, which represents different pitch-accent targets – a single-tone phrasal accent H- and a high-tone part of the bitonal lexical accent H*+L, respectively. The fall in the fundamental frequency from the F0 maximum across *des* also occurs in both phrase types – due to the interpolation between the phrasal H- and the phrase-final boundary tone L% in the unaccented phrase, and due to the interpolation between the high and the low tones in the H*+L of the accented phrase. However, a steeper F0 fall in the 2nd-syllable accented than in the unaccented F0 contour distinguishes these two patterns in AK's productions.

Pierrehumbert and Beckman (1988) have argued that F0 fall across intonational phrases is steeper 1) in accented phrases than unaccented phrases due to the process of *catathesis* (i.e., lowering of all tonal targets following the lexical high tone in a phrase), and 2) in shorter than longer unaccented phrases due to the process of *declination* (i.e., F0 decrease, exponentially related to the amount of segmental material intervening between the phrasal high and boundary low tone targets) (e.g., pp. 70, 90, among other places in the monograph). The difference in the steepness of the F0 fall due to catathesis is evident in Figure 2.3 from phrase-final F0 values (i.e., boundary tones) that are relatively lower in the accented than unaccented phrases as compared to the F0 maximum in the same phrase. Note, that the same difference may be attributed in our short example utterances

to the declination, or to the steepness of the F0 fall in the H*+L accent. Thus, in the case of short one-phrase utterances, the phonological reasons of this falling F0 movement are confounded. Both word-level forces (the pitch-accent contrast) and phrase-level forces (catathesis and declination) may be at work here. Overall, the steepness of F0 fall is the cue to the accented-unaccented distinction, and it is determined by the magnitude of the fall over time, as can be compared to the hypotenuse in a triangle.

Another acoustic cue to the pitch-accent contrast discussed in literature is the height of the F0 maximum (hereafter referred to as F0 peak height). Previous studies suggest that the high tone of the lexical accent has a higher F0 value than the high tone of the phrasal accent, at least for some speakers (Beckman & Pierrehumbert, 1986; Kubozono, 1993; Poser, 1984; Sugito, 1982b; Vance, 1995). This finding is illustrated in both panels of Figure 2.3 – the F0 contour of the syllable -*shi* is higher in the 2nd-syllable accented word than in the unaccented word. This finding is related to the discussion of the F0 fall steepness in that the magnitude of the fall is determined both by the start and end F0 values of the fall, and the start point is, in fact, the same as the F0 peak.

The last cue, F0 rise, has been related in the literature to phrasal-level tone events such as the low boundary tone L%. The fact that the F0 rise is not a part of lexical prosody is evident from longer multi-word phrases such as in Figure 2.1d, where only the phrase-initial word manifests a significant rise in F0. However, Kubozono argued that "accented phrases tended to show a greater degree of initial F0 rise than unaccented counterparts" regardless of such factors as the location of the accent or the position of the phrase in the utterance (1993, p. 88, 91). Therefore, F0 rise may be an acoustic cue to the accented-unaccented distinction. The comparison of the 2nd-syllable accented blue

contours and unaccented red contours in Figure 2.3 confirms this statement for one of the speakers (top panel), but the difference in the rise is not straightforward for the other speaker (bottom panel). In addition, it is also not clear whether the F0 rise cue is truly independent from the F0 peak location (Ishihara, 2006).

In sum, there appear to be three possible acoustic cues to the contrast between accented and unaccented phrases, all of which are related to F0 fall: 1) the steepness of F0 fall (as the magnitude of fall over time), 2) the F0 peak height (as the starting point of the fall), and 3) the degree of initial F0 rise (with the F0 peak as its end point). The challenge in understanding the roles of these cues is that all of them are interrelated. For this practical reason, F0 fall was examined in this dissertation as a primary cue to the Japanese lexical pitch accent; F0 peak was considered to be the beginning of the fall; and the degree of initial F0 rise was determined by F0 peak location.

2.4.2. F0 Peak Location

The accentual pattern of a Japanese word is defined not only by the presence of a pitch accent, but also by the location of the accent in the word. For example, the disyllabic word *hashi* has three lexically-contrastive accentual patterns. The accented-unaccented distinction in these patterns has been described in the previous section. In addition, the 1st-syllable accented versus 2nd-syllable accented distinction is apparent from the comparison of the green and blue contours in Figure 2.3. In the 1st-syllable accented phrase, the F0 peak corresponding to the high tone of the H*+L accent is located in *ha*-, and the F0 fall follows the peak, stretching across -*shi* (for the speaker RH) and even further (for the speaker AK). In the 2nd-syllable accented phrase, the F0 peak is located in -*shi*, and the F0 fall follows the peak stretching across *des*. Thus, both

the location of the accent (1st vs. 2nd syllable) and the presence of the accent (accented vs. unaccented) create the three-way lexical contrast of *hashi*, which is used in this dissertation to investigate listeners' sensitivity to the relevant acoustic cues.

2.4.3. Pitch Accent Alignment and F0 Peak Delay

A description of the F0 contours should also consider delay in the F0 peak of the lexical pitch accent H*+L and the phrasal accent H- because it may affect perception of accentual patterns (Ishihara, 2003, 2006; Kawakami, 1961; Neustupný, 1980; Sugito, 1980). *Peak delay* is the occurrence of the F0 maximum after the phonologically accented syllable. In other words, it is a temporal mismatch between the phonological association of the high tone and its phonetic alignment relative to an accented syllable. Figures 2.2a and 2.2b illustrate this phenomenon for the H*+L accent: In the purple F0 contour corresponding to the 2nd-syllable accented word *haŕi*, the F0 peak (i.e., the start point of the fall) occurs after the accented vowel [i]. Figures 2.2c and 2.2d illustrate peak delay for the H- accent: The F0 peak occurs not on the second syllable of the unaccented word *hari* as some phonological accounts would have predicted (Pierrehumbert & Beckman, 1988), but near the onset of the following syllable.

This variation in the alignment of the high tone of the accent can be attributed to several factors. For example, Hattori suggested (recasted in modern terms) that the peak might be delayed due to an interaction between phrasal prosody and lexical prosody (1960, p. 365). Specifically, unless the phrase is initially accented, the first syllable of the phrase carries a phrase boundary tone, and tends to be lower in F0 and higher in amplitude relative to the following syllables. The interaction between a boundary tone and a high tone of the pitch accent may cause the peak delay in medially-accented words.

Sugito (1980, 1982b) disputed this proposal and suggested that peak delay might be due to segmental effects on fundamental frequency (e.g., the effect of [s] in *ása no kimono* or intrinsically lower F0 of non-high vowels). Kawakami (1995) pointed out that an F0 peak might be also delayed due to: 1) the emotional affect of a speaker such as surprise or uncertainty (e.g., *haná ga des ka* said with the F0 peak on *ga*), or 2) the syllable structure of the accented syllable such as a short voiced portion of the rhyme (e.g., *tas(u)kéru* where the vowel [ɯ] is devoiced).

Experimental evidence provides support for the hypothesis that the syllable structure factor is responsible for F0 peak delay in accented syllables with short sonorant rhymes. In fact, peak delay was observed for both lexical (Ishihara, 2003, 2006; Sugito, 1982a) and phrasal (Shport & Guion, 2008) accents in Tokyo Japanese. Ishihara has shown that the F0 peak associated with the pitch accent H*+L is aligned with the vowel onset of the following syllable in CVCV words as compared to CV:CV and CVNCV words. Shport and Guion have shown that the F0 peak associated with the high phrasal tone H- is also delayed in CVCVCV words as compared with word types where the initial syllable has a longer sonorant rhyme (CV:CV and CVNCV).

The phenomenon of peak delay is not unique to Japanese; it has been also observed in stress languages like English (Silverman & Pierrehumbert, 1990). The interested reader is referred to the first chapter of Ishihara's dissertation (2003) where the peak delay literature is reviewed in detail for a variety of languages. The proposed explanations of earlier peak alignment in English include shorter rhyme duration of the accented syllable, pitch accent clash, intervening word boundary, and slower speech rate. The sonorant rhyme duration factor is reported to have an opposite effect on the peak

alignment in English than in Japanese – a shorter rhyme in the accented syllable correlates with later peaks in Japanese, but earlier peaks in English. (See, however, a counter-example in Figure 2.2, where the peak delay in English appears to be similar to Japanese. That is, in the CVCV word *rúby*, the H* peak in the purple contours is aligned with the beginning of the following syllable *-by*).

For the purpose of this dissertation, the important findings from the peak alignment literature are that: a) the F0 peak alignment may vary within a language, b) it is language-specific, and c) it is influenced by the segmental structure of the accented syllable. Although the examination of F0 peak alignment is out of the scope of this dissertation, a possible effect of peak delay in production on the perception of the pitch-accent contrast should be taken into consideration. Sugito (1980) has shown that both native and non-native Japanese listeners may perceive prominence on a syllable, even when the F0 peak is acoustically manifested in the following syllable. Hasegawa and Hata (1992) and Kitahara (2001) have further investigated the relationship between the perceived pitch accent and the peak delay, as reviewed in the following section.

2.5. Perception of Pitch-Accent Contrast by Native Listeners

In previous experiments on perception of the Japanese pitch-accent contrast by native listeners, both naturally produced stimuli (Sekiguchi & Nakajima, 1999; Sugito, 1980; 1982b) and stimuli with resynthesized F0 contours (Hasegawa & Hata, 1992; Kitahara, 2001; Sugito, 1982a) have been used. An examination of listeners' sensitivity to the acoustic cues to the contrast, which is the objective of this dissertation, requires cue manipulation in the stimuli. As most methodologically relevant to this objective, the designs and findings of two studies will be reviewed below. In order to manipulate the

acoustic cues to pitch accent, Hasegawa and Hata (1992) used a word-based, less complicated resynthesis model, whereas Kitahara (2001) used a sentence-based, more complex model.

2.5.1. Hasegawa and Hata (1992)

To investigate the relationship between the perceived pitch accent and F0 peak delay in *mamama* words, Hasegawa and Hata manipulated:

1) the peak location by varying it in 5 steps within the second vowel of the word;
2) the slope of the F0 fall by keeping the fall's magnitude constant and moving the F0 elbow in 14 steps from the peak.

The F0 values at the beginning and the end of the word were constant. Listeners were asked to judge whether F0 patterns in *mamama* words sounded more like the 1st-syllable accented word *námida* 'tear' or like the 2nd-syllable accented word *okáshi* 'sweets'. Hasegawa and Hata found that the later F0 peak occurred in the second vowel [a], the steeper F0 fall was required for listeners to perceive the accentual pattern as similar to the 1st-syllable accented word. In other words, both of the manipulated parameters affected listeners' perception of accent location in a task contrasting two accentual patterns.

There are several concerns regarding the design of Hasegawa and Hata's study. First, only the perception of accented patterns was tested, and it is not clear what native listeners would do if another answer option representing the unaccented pattern was available to them (e.g., *sakana* 'fish'), especially, when the F0 fall was not steep. Second, the peak delay was manipulated only for the 1st-syllable accented pattern by varying the timing of F0 peak in the second syllable. If the peak was present also in the third syllable of *mamama*, it could have altered listeners' response patterns. Third, the listeners heard

nonwords, but they were asked to respond recalling the accentual patterns of real words, which might have involved additional steps in processing of the stimuli.

2.5.2. Kitahara (2001)

To investigate the relationship between the perceived pitch accent and F0 peak delay in voiced and voiceless syllables, Kitahara had a complex manipulation of stimuli. The manipulation was based on the intonation synthesis model developed for Japanese by Pierrehumbert and Beckman (1988). This model includes a variety of parameters: the high and low tones of the lexical pitch accent H*+L, the phrasal pitch accent H-, phrase-level processes such as catathesis and declination, as well as speaker-specific baseline and upper line of F0 contours. Kitahara presented the target, lexically-contrastive words in his study in a carrier sentence, while modeling the F0 contour of the entire sentence.

In one experiment, Kitahara tested the perception of three accentual patterns in two-way oppositions: a) unaccented versus 1st-syllable accented, b) unaccented versus 2nd-syllable accented, and c) 1st- versus 2nd-syllable accented. In the other experiment, the perception of the three-way opposition was tested. To create the materials in the two-way opposition tests, Kitahara kept constant the F0 values of the low boundary tone at the beginning of the carrier phrase (L%) and the phrasal tone on the second syllable of the phrase (H-). He manipulated:

1) voicing of the final syllable in target words;
2) the peak location of the H*+L accent (the 1st or 2nd syllable of the target words);
3) the peak height of the H*+L accent;
4) the peak alignment of the H*+L accent (the accented or the following syllable);
5) the height of the utterance-final boundary tone L% (scaled to the H*);

6) the alignment of the L%.

The majority of the variables listed above had multiple levels, and they were not fully crossed depending on the two-way opposition in question. In the three-way opposition test, yet another phrasal tone variable was added to the design.

To summarize Kitahara's major findings, he reported that the difference between the H* and L% tones (i.e., the magnitude of the F0 fall) captured the accent-presence detection process (i.e., accented versus unaccented distinction), and the H* alignment captured the accent-location detection process (i.e., 1st- versus 2nd-syllable accented distinction). In contrast to Hasegawa and Hata (1992), Kitahara did not find a compensatory relationship between the F0 peak alignment and the slope of the F0 fall for the perception of the preceding syllable as accented. He interpreted this result as an indication of invariance in the timing of the high and low tones of the H*+L accent relative to each other. This interpretation is consistent with the modeling of the H*+L accent by Pierrehumbert and Beckman (1988), and it assumes no significance of the slope of the F0 fall for the perception of accentedness.

What we can discern from these studies for the purpose of this dissertation is that the perception of the pitch-accent contrast by native listeners of Tokyo Japanese is affected by several parameters in the accentual patterns. These parameters are the location of the F0 peak in the word, the alignment of the F0 peak relative to the accented syllable, and the magnitude of the F0 fall after the peak. The slope of the F0 fall is a complex construct that consists, at least, of the magnitude of the fall and the timing of its endpoint. The timing of the start- and end-points of the F0 fall relative to each other may or may not be important for processing of the pitch accent.

With regard to the integration of lexical and phrasal prosody in a phrase, Kitahara's findings indicate that the preceding context (in particular, the alignment and F0 height of the phrasal accent H-) does not affect the identification of pitch patterns in the target words. Therefore, we can use shorter stimuli focusing only on the effect of variations in F0 contours due to the lexical pitch accent H*+L on the perception of prominence associated with this accent. Shorter stimuli will allow us to simplify the design by reducing the number of manipulated parameters (i.e., phrasal accent H-, catathesis, and declination). Furthermore, shorter stimuli will increase the feasibility of cross-linguistic perception experiments with non-native listeners of Japanese naïve to the language and facilitate interpretation of the results.

2.6. Perception of Pitch-Accent Contrast by Non-Native Listeners

In this section, first, previous studies investigating the perception of Japanese lexical pitch patterns by non-native listeners will be reviewed. Since native language is known to influence the perception of non-native sounds, the description of previous work is limited here to the studies with native English listeners. Then, the discussion will come back to cross-linguistic perception of lexical prosody in general. Previous studies on perception of lexical tones in languages other than Japanese will be reviewed because of the inconsistency of their findings with the findings on perception of pitch-accent contrast, which motivates the approach undertaken in the current dissertation work.

Similarly to research with native Japanese listeners, some studies on non-native perception of the Japanese pitch accent use words with resynthesized pitch patterns, and others use words with naturally produced pitch patterns. Stimuli of perceptual tests largely define what can and what cannot be inferred from the results. Synthesized stimuli

allow a researcher to manipulate the structure of accentual patterns and address the questions about the mechanism of cross-linguistic perception. The work of this type has not yet been done on the perception of the Japanese lexical pitch accent by non-native listeners, with the exception of two studies that in addition to native listeners had a couple of non-native listeners for exploratory purposes (Masaki et al., 2001; Sugito, 1982a).

2.6.1. Findings from Studies with Resynthesized Pitch Patterns

A comparison between native and non-native categorization of F0 patterns is difficult to make when the number of non-native participants is very limited (e.g., one listener in Sugito, 1982a; three listeners in Masaki et al., 2001). The goal of Sugito's study, for example, was to investigate the percept of accent in the disyllabic word *aka* by native listeners of another Japanese dialect, Kinki (Kansai). She manipulated the magnitude of the F0 rise to the peak by varying the initial F0 value. The alighment of the F0 peak and the magnitude of the F0 fall were kept constant. The results suggest that categorization patterns of the listeners formed a continuum: A larger F0 rise before the F0 peak in the 2nd syllable of *aka* was required for the percept of the accent in the 2nd syllable by Kinki listeners as compared to a native Tokyo Japanese listener. The data of one American English listener, which fell on the other end of the continuum as compared to the native Tokyo Japanese listener, were presented to illustrate large individual variability in Kinki listeners' data.

Masaki, Takasawa and Arai (2001) investigated whether the F0 rise or fall in the first syllable of *ata* affected the judgments of accentedness by Japanese listeners (their native dialect was not reported). There also were three non-native listeners – a German, a Spaniard, and a Pole, all naïve to Japanese (their judgments were not detailed). The

authors reported that the overall falling F0 contours induced a robust percept of the accent in the first syllable; the overall rising contours induced less consistent accent judgments. Similar to Sugito's results, a larger F0 rise before the F0 plateau in the second syllable of *ata* was required for the percept of the accent in the 2nd syllable by non-native as compared to native listeners.

In sum, little research has been done investigating the acoustic cues in perception of the Japanese lexical pitch accent by non-native listeners. This dissertation aims to fill this gap. In the current work, resynthesized pitch patterns resembling naturally occurring patterns were used to better control stimuli and the sound structure. The same number of native and non-native participants was recruited to obtain comparable data samples.

2.6.2. Findings from Studies with Naturally Produced Pitch Patterns

An example of a relatively early study that used stimuli with naturally produced pitch patterns was the work by Sugito (1980). Her goal in this work was to investigate the effect of F0 peak delay on the perception of pitch accent by native English listeners. Four American English listeners with varying Japanese language proficiency were asked to mark accented morae in 1st- and 2nd-syllable accented words like *násake* and *anáta*. The answers of one listener were difficult to interpret. The other three English listeners perceived the accent on the first syllable even when the F0 peak was in the second syllable, similar to native Japanese listeners.

In a relatively recent work by Sakamoto (2008), discrimination and categorization of lexical pitch patterns by native and non-native listeners were investigated. First-syllable accented, 2nd-syllable accented and unaccented productions of four nonwords (*mene, noma, mani*, and *nime*) were recorded in the carrier sentence *Sumimasen, ___ mo*

kudasai 'Excuse me, give me ___ too'. Then, the '(target word) *mo*' portions of the recordings were excised to create test stimuli. Three groups of listeners residing in the U.K. participated in this study: Japanese native speakers; relatively experienced learners of Japanese (three years of classroom instructions and one year in Japan); and relatively inexperienced learners of Japanese (two years of classroom instructions). The results of the discrimination experiment revealed that the groups did not differ in their discrimination accuracy (score range: 84-89%), which suggests that all listeners could distinguish the three types of pitch patterns. The results of the categorization experiment revealed that the accuracy of inexperienced learners (74%) was significantly lower than native speakers' accuracy (91%). Sakamoto suggested that the ability of listeners to identify graphically marked pitch-accented syllable (e.g., *méne mo, mené mo*, *mene mo*) correlates with their understanding of the role of the F0 in Japanese pitch-accent contrast.

From her experiments Sakamoto concluded that relatively experienced but not inexperienced learners of Japanese were able to use the variations in F0 as a cue to pitch-accent contrast. However, neither the phonetic characteristics of the stimuli, nor results by pattern were reported. Therefore, it is unclear what the gross accuracy test scores actually mean for the mechanisms of cross-linguistic perception, and whether the listeners discriminated and categorized all three patterns equally well. What is clear from Sakamoto's work is that listeners may categorize the same stimuli differently depending on the nature of the listening task. Same-or-different task with two matching answer options (ABX discrimination) may yield higher response accuracy than which-one-out-of-three task (3AFC categorization). The sensitivity to F0 variations does not necessarily imply the acquisition of non-native prosodic categories such as lexical pitch patterns.

A large group of studies using naturally produced pitch patterns adopted the methodology of the Tokyo Japanese Accent Listening Test. This test was developed as a part of a project funded by the Ministry of Education of Japan in 1994-1998 (Ayusawa, 2003). The purpose of the test was to investigate the acquisition of lexical accent by learners of Japanese with various native language backgrounds. The only available phonetic description of the test stimuli may be found in Nishinuma, Arai and Ayusawa (1996), who reported the F0 maximum and F0 minimum associated with the accented morae of stimuli words. The validation of the test stimuli with native listeners of Tokyo Japanese is not described in the published studies, although presumably the stimuli words had accentual patterns conforming to the dictionary norms, as judged by Japanese researchers. This lack of the methodological description complicates the interpretation of test results as discussed below.

The Tokyo Japanese Accent Listening Test (25 min in total) starts with instructions on the Japanese accentual system, a task explanation, and a ten-minute practice. The test words are written on an answer sheet in Japanese orthography. Listeners hear target words that vary in their length (3-, 4-, and 5-mora words), accent type (unaccented, 1^{st}-, 2^{nd}-, 3^{rd}-, 4^{th}-, and 5^{th}-mora accented words), and presentation mode (in isolation and in carrier sentences). The listeners' task is to mark the accented mora in a word on which they hear "the beginning of a pitch fall" (Nishinum et al., 1996). If listeners do not hear a steep pitch fall, they are instructed to circle the answer "none". Note that, in essence, listeners are asked to do two different tasks for each word in the test: 1) to identify whether a word is accented or unaccented, and 2) in the former case, to identify the location of the accented mora. These task instructions are in accord with

phonological descriptions of pitch-accent patterns. However, several factors possibly affecting listeners' responses are not considered in the test: for example, the phonetic implementation of the pitch accent such as the frequently observed F0 peak delay; word familiarity to non-native listeners; and listeners' strategies in pattern categorization.

Nishinuma, Arai and Ayusawa (1996) conducted a large-scale study of this type with 54 learners of Japanese whose native language was American English. The students were tested in Japan, where they studied the language after two years of instructions in the U.S. classroom setting. The results showed that the overall identification accuracy of accentual patterns was not high - 63% in isolated words and 59% in sentence-embedded words. Regardless of the presentation mode, the accent type was identified the best in unaccented words, 2nd-syllable accented words, and 1st-syllable accented words, in the order of decreasing accuracy.

In another study by Nishinuma (1997), the previously obtained data from native English listeners (Nishinuma et al., 1996) were discussed in more detail. The author summarized that across all participants, word types, and word presentation modes listeners tended to identify the unaccented pitch pattern better than the accented patterns. Nishinuma proposed that this result might be due to the accented-unaccented prosodic dichotomy that exists in many languages and facilitates the recognition of similar non-native contrasts. The unaccented pitch pattern was suggested to be "simple," probably, meaning that the presence of the accent was a prosodic feature somehow additional to the unaccented pattern and thus increasing the processing effort on the part of listeners. The relatively high identification accuracy for 2nd-syllable accented words (e.g., *tabémono* 'food') was suggested to be due to the similarity in the F0 peak location between this

Japanese pattern and 2nd-syllable stressed pattern in English (e.g., *appárently*). This logic fails though for the 1st-syllable accented words (e.g., *káre kara* 'from him'), which had the lower identification accuracy than the 2nd-syllable accented words. It is unclear why this would be the case, since English certainly has a "similar" 1st-syllable stressed pattern (e.g., *cátegory*). Nishinuma also suggested that the absence of the durational cue to the accent in Japanese could be confusing for English listeners. A possibility of listeners' response bias towards the unaccented pattern was not considered.

Interestingly, a similar result of the relative ease of the unaccented pitch pattern is consistently reported in research based on the methodology of the Tokyo Japanese Accent Listening Test. For example, in two longitudinal studies by Arai (1997) and Hirata, Ayusawa, Nakagawa and Odaka (1997), the change in the identification accuracy of accentual patterns over time was investigated by repeated administration of the test. Both studies were conducted with English native listeners who were studying the language in Japan and who had not have instructions on Japanese accent and intonation prior to their arrival to the country. Arai (1997) administered the test three times with three-month intervals. She found that her participants' accuracy did not change over time (57%, 52%, 57% of correct responses) and that in all three tests the unaccented pattern had the highest accuracy score. Hirata and colleagues administered the test six times with two-month intervals. Again, across all subjects the unaccented pattern was identified the best, at least, in disyllabic words (83% at Test 1, and 86% at Test 6).

The more accurate identification of unaccented words as compared to accented words reported in the above studies is difficult to interpret. Listeners' response biases, the nature of the task in the Tokyo Japanese Accent Listening Test and the scoring procedure

could have influenced results. It is possible that the location of the accented mora (i.e., F0 fall in the test instructions) in accented words is identified less accurately than the absence of the accent in unaccented words, because listeners are sensitive to the phonetic characteristics of the stimuli. For example, if F0 fall is not very steep in accented words, listeners may be biased toward the 'none' response. If F0 peak is delayed in accented words, listeners may tend to mark 1st-syllable accented words as 2nd-syllable accented (e.g., the stimuli 14, 16, and 22 in Nishinuma et al., 1996) and to mark 2nd-syllable accented words as 3rd-syllable accented (e.g., the stimuli 12 and 20 in the same study). In such cases, the confusion of the accented pitch patterns with each other would lead to their lower identification scores as compared to the unaccented pattern. Another possibility is that listeners choose not to mark the location of the accent if they are unsure. The chance level of listeners identifying the absence of the accent is 50% (accented and unaccented answer options), whereas the chance of identifying the location of the accent varies from 50% in three-syllable words (1st- or 2nd-syllable accented) to 20% in five-syllable words (1st-, 2nd-, 3rd-, 4th-, or 5th-syllable accented).

In line with the Nishinuma's argument (1997), Ayusawa also suggested that listeners generally have a higher identification accuracy for Japanese pitch patterns that are similar to prosodic patterns of their native language (2003, p. 54). She argued that these findings constituted the evidence of "perceptual transfer" from the listener's native language to the second-language Japanese. However, the criteria for similarity were not defined in these studies, and the concept of perceptual transfer was not operationalized in any way. Rather, similarity was implied holistically in terms of, for example, prominence location (e.g., the 2nd-syllable accented Japanese pattern and the 2nd-syllable stressed

English pattern). Overall, the studies using the Tokyo Japanese Accent Listening Test are not designed to investigate the mechanisms of perception and categorization in non-native listeners. Thus, the question of why pitch patterns are not perceived equally accurately remains unanswered, and studies on perception of lexical tones in languages other than Japanese may shed light on this question.

2.6.3. Cross-Linguistic Perception of Lexical Tones

2.6.3.1. Perception as determined by pattern similarity. The issue of *similarity* between native and non-native sounds has been addressed in theories of speech perception and learning (e.g., Perceptual Assimilation Model: Best, 1995; Best & Tyler, 2007; Native Language Magnet Model: Kuhl & Iverson, 1995). The basic idea behind these theories is that listeners perceive non-native sounds in relation to native sound categories. Although the current theories of speech perception were originally developed for speech segments, they have begun to be applied to lexical tones as well (Bent, 2005; Hallé et al., 2004; So & Best, 2010).

Bent reviews in her dissertation the methods of assessing similarity, labeling prosodic categories for some assessments methods, and contextual effects on perceived similarity of sounds (2005, pp. 12-17). Assessing the degree of similarity is methodologically more difficult for prosodic than segmental categories due to the cross-linguistic differences in the domain of prosodic categories (e.g., the falling F0 movement in Japanese accented words versus English phrases); cumulative nature of many prosodic categories (e.g., focus); complex structure of prosodic categories (e.g., the debated roles of duration, intensity, spectral slope, vowel quality, and fundamental frequency for English stress); and the lack of metalinguistic awareness of prosodic categories in

listeners (e.g., phrasal pitch accent, pitch-accent contrast), among other factors.

There are no studies to my knowledge that, first, rigorously determine the degree of phonetic similarity between non-native lexical tones and native prosodic categories, and then test the effect of similarity per se on tone perception. Similarity, however, is sometimes inferred from F0 contours associated with prosodic categories. For two tone languages, where the F0 movement has the same domain, the tone inventories may be directly compared. So and Best (2010), for example, described Mandarin and Cantonese tones as similar to each other in terms of the overall F0 contours, at least, in some contexts. In the same work, selected Mandarin and Japanese F0 contours were described as similar, noting that the temporal realization of these similar patterns varies from one syllable in Mandarin (e.g., rising tone 2 and falling tone 4) to at least two syllables in Japanese (e.g., the unaccented pattern and the accented pattern).

For a tone and a non-tone language, the direct comparison of patterns is more complicated. It has been suggested that when prosodic categories at the word level are typologically different in two languages (e.g., Mandarin and French, or Taiwanese and English), there are three possibilities of how non-native listeners may process tones. They may perceive tones as "non-linguistic melodic variations" (i.e., non-speech sounds), or as speech patterns evocative of native language patterns but still uncategorizable (Hallé et al., 2004, pp. 416-418), or as speech patterns evocative of native language patterns and categorizable (So & Best, 2010, p. 289). It is not immediately clear how these alternatives can be tested, given methodological challenges mentioned above. The relationship between the structure of acoustic cues to the pattern, on one hand, and processing of the pattern by listeners, on the other hand, may be investigated more straightforwardly than

the relationship between similarity and processing. Cross-linguistic studies adopting such approach are reviewed below.

2.6.3.2. Perception as determined by acoustic cues. Cross-linguistic differences in the use of acoustic cues that are relevant to perceptual discrimination of segmental and prosodic contrasts have been observed in multiple studies. These differences were argued to be due to linguistic experience that warps perceptual space of listeners (Gandour, 1983; Kuhl & Iverson, 1995; Iverson et al., 2003), and, additionally, due to individual differences among listeners (Chandrasekaran, Sampath, & Wong, 2010; Golestani & Zatorre, 2009; Iverson, Hazan, & Bannister, 2005). With regard to the first factor, Kuhl and Iverson (1995) propose in their Native Language Magnet model that language experience alters the underlying mechanisms of speech perception. Specifically, the perceived distance between sounds in the acoustic space along relevant acoustic dimensions changes in the process of first language acquisition, which has profound consequences for speech perception by mature listeners. For example, Japanese adults are most sensitive to the acoustic cue of F2, irrelevant to the English /ɹ/-/l/ distinction, whereas German adults retain the *sensitivity* to the relevant acoustic cue of F3 (Iverson et al., 2003). This sensitivity is also referred to as *cue weighting*, or *selective attention*, or *selective perception*, the latter, for example, defined as "detection and weighting of particular phonetically relevant information, as a function of perceivers' learned habits of listening" (Strange, in press, p. 4).

In principle, language experience may affect the weighting of multiple acoustic cues to a prosodic category relative to each other, or it may affect the overall sensitivity to the combination of acoustic cues specific to that category. The latter case was

documented in a series of studies by Dupoux and colleagues who found that French native listeners did not hear the difference in stress location in Spanish minimal pairs like *bebé* 'baby' and *bébe* 's/he drinks' (Dupoux, Pallier, Sebastián-Gallés, & Mehler, 1997), or nonwords like *súmi* and *sumí* produced with Spanish stress patterns (Dupoux et al., 2008). The degree of "stress deafness" in French listeners varied with the task: In AX or ABX same-different tasks French listeners had fewer errors than in sequence recall or word lexical access tasks. The researchers concluded that speech perception models cast in terms of perceptual warping and prototype formation failed to account for their data (2008, p. 701). Their results suggest, however, that although adult French listeners do not lose sensitivity to duration and intensity in psychoacoustic tasks, they learn to ignore the combination of these cues specific to stress but irrelevant to their native language prosody. Such result is, indeed, not unique to lexical stress. Similar findings were reported for lexical tones, to which non-native listeners were rather sensitive in discrimination tasks (Bent, 2005; Hallé, et al., 2004; So & Best, 2010).

Let us now consider relative weighting of the acoustic cues to lexical tones. In the seminal work by Gandour and Harshman (1978), multidimensional scaling (MDS) analysis was used to assess the degree of similarity between pairs of tones as percieved by native listeners of Thai, Yoruba, and American English. Thirteen stimulus tones were synthesized for the syllable [wa] to include representative tones of Thai and Yoruba. Level and contour tones varied in their length and pitch height. For falling and rising tones, the magnitude of slope and pitch range were also manipulated. Listeners heard pairs of stimuli in the carrier sentence *Compare the pitch of the following words* ___, and judged the similiraty of the stimuli on 11-point scale.

The degree of percieved similarity was modeled in terms of distance between the stimuli in perceptual space, where greater distance represents greater dissimilarity. In the MDS model, the dimensions of the space are taken to represent acoustic dimensions relevant to the discriminations of sound contrasts; the amount of variance that each dimenstion accounts for is taken to represent listeners' relative sensitivity to the dimensions in the model. Gandour and Harshman found that a five-dimensional solution accounted for 74% of the variance in their data. These dimensions were interpreted as average pitch, direction (rising-level-falling), length (long-short), extreme endpoint, and slope (level and contour). Further analysis showed that all listeners were sensitive to average pitch and length. However, Thai and Yoruba listeners also attached relative importance to the direction of the F0 change and slope, whereas English listeners did not.

Since the 1970s, behavioral and neurological experimental evidence has accumulated to support the hypothesis that native listeners of tone languages attend more to the acoustic dimension of *F0 change* than non-tone speakers do (Chandrasekaran, Krishnan, & Gandour, 2007; Francis et al., 2008; Gandour, 1983; Guion & Pederson, 2007; Kaan, Wayland, Bao, & Barkley, 2007; Wayland & Guion, 2004). Judging dissimilarity in pairs of tones, native speakers of tone languages like Mandarin, Thai, Cantonese, and Taiwanese seem to pay attention to both F0 height and F0 direction of change, whereas native speakers of non-tone languages like English seem to attend more to the F0 height. The perception of pitch change, at least for rising tones, was argued to be categorical for listeners of tone language background as compared to listeners of non-tone language background (Xu et el., 2006). Such cross-linguistic differences in processing of lexical tones were attributed to the small-size window of temporal

integration between lexical tones and syllables, which is evident, for example, from left-hemisphere (more localized) processing of tones by native Mandarin listeners, and bilateral processing of tones by native English listeners (Gandour et al., 2003).

Based on this evidence from research on perception of lexical tones, it is possible that English listeners of Japanese pitch patterns have the same processing deficit as non-native listeners of Mandarin or Thai have, namely, the lack of attention to F0 fall, the fundamental acoustic cue to the pitch-accent contrast. English listeners may attach more importance to F0 height and F0 peak location than to F0 fall, thus missing the distinction between accented and unaccented words in Japanese. This learned processing of phonetic information may affect English listeners' perception of lexical prominence specific to Japanese.

2.7. Perceptual Training for Adult Learners

As described above, attention to F0 fall is necessary for distinguishing Japanese pitch patterns. Previous research has shown that naturalistic (Gottfried & Beddor, 1988; Guion & Pederson, 2007; Wayland & Guion, 2003) or laboratory (Francis, Baldwin, & Nusbaum, 2000; Francis & Nusbaum, 2002; Holt & Lotto, 2006) linguistic exposure can change the relative weightings of acoustic information in speech processing. For example, Francis and Nusbaum (2002) found that native English listeners could be trained to use novel acoustic dimensions for the processing of Korean stops. Likewise, Gottfried and Beddor (1988) found that more advanced English learners of French exhibited more native-like weighting of spectral and duration cues to French vowels than less advanced learners. In the prosodic domain, Guion and Pederson (2007) found that more advanced English learners of Mandarin Chinese weighted the F0 change cue more

heavily than less advanced learners. Thus, native language biases in speech perception by adult listeners can be overcome with exposure. The practical question is to what degree and what type of training facilitates this process.

Many elements of training design may be varied in hopes of better training outcome: for example, the type of a listening task, feedback, and training materials. Identification training improves phonetic discrimination (Flege, 1995b; Wayland & Li, 2008); a mere repetition of an identification test, on the other hand, does not necessarily yield any improvement in identification accuracy (Arai, 1997; Hirata et al., 1997). Teaching practice shows that immediate feedback, which explicitly draws attention of trainees to acoustic differences in sound contrasts, also facilitates learning.

With regard to the type of training materials, target sounds or sound contrasts may be presented in isolation or in context; spoken by one talker or multiple talkers. Training on stimuli that include multiple lexical items, talkers, and sentential contexts has been effective for learning of non-native segmental and prosodic categories. Such training is referred to as high-variability training paradigm (HVTP), and it was first formulated by Logan, Lively and Pisoni (1991). The HVTP has been shown to facilitate identification of non-native segments, for example, English /ɹ/ and /l/ by Japanese listeners (Bradlow, 2008; Bradlow et al., 1997; Lively et al., 1993; Lively et al., 1994; Logan et al, 1991); English /t/-/d/ by Chinese listeners (Flege, 1989); English /ð/-/θ/ by Canadian French listeners (Jamieson & Morosan, 1986, 1989); Korean stops by English listeners (Francis & Nusbaum, 2002); Hindi /t̪ʰ/-/ʈʰ/ (Guion & Pederson, 2007; Pruitt, 1995); English vowels by German and Spanish listeners (Iverson & Evans, 2009).

The HVTP was also successfully used for training on non-native prosodic categories, for example, Japanese syllabic morae (Hirata, 1999; Yamada, Yamada, & Strange, 1996) and Japanese pitch accent (Hirata, 1999) learned by English listeners; Cantonese tones learned by Mandarin and English listeners (Francis et al., 2008); Mandarin tones learned by English listeners (Wang et al., 1999) and Dutch listeners (Leather, 1990). For example, Wang and colleagues (1999) trained eight participants to perceive four Mandarin tones in a two-week training program, with feedback on tone labeling. They reported an overall 21% increase in tone identification accuracy from pretest to posttest. The improvement also generalized to new stimuli and new talkers.

Why the high-variability training with naturally produced sounds is more effective than the low-variability training is the point of ongoing debate. Originally, it was assumed that greater stimulus variability helps learners to focus their attention on phonetic dimensions that serve as the most robust perceptual cues to identifying a foreign sound across various phonetic environments and talkers (Lively et al., 1993). Iverson and Evans argued, however, that high-variability phonetic training encourages listeners to apply more consistently their native-language categories to non-native speech, rather than to re-weight acoustic cues relevant to identification of a non-native speech sound (2009, p. 875). Whatever the processing mechanism encouraged by high-variability training is, its positive effect on speech category learning is well established, and it is consistent with exemplar-based models of speech perception (Bradlow, 2008).

With exception of the work by Hirata (1999) described below, there is a gap in research on training, especially HVTP-type training, as applied to Japanese prosodic categories. In some previous studies, the Tokyo Japanese Accent Listening Test has been

repeatedly administered to the same groups of listeners, with or without separate instructions on Japanese prosody. The review of such longitudinal studies on accent acquisition by learners of different native language backgrounds is available in Ayusawa (2003, p. 54). She cited that in immersion situation (students studying in Japan), the score improvement in the Tokyo Japanese Accent Listening Test varied from 0% in six months to 27% in eight months, 11% in ten months, 6% in a year, and 10% in a year. Thus, the results of these studies were inconsistent. Furthermore, no feedback was typically provided to listeners during the tests, and a large individual variability in the amount of improvement was always observed.

The only true training study on Japanese prosodic categories including pitch accent was conducted by Hirata (1999). She trained American English listeners on rhythm (training 1), and pitch patterns and rhythm (training 2) using immediate computer-mediated feedback and highly variable stimuli. In the rhythm training, participants' task was to count the number of morae in isolated and sentence-embedded words. The goal was to help learners develop the metalinguistic awareness of rhythm-related durational patterns (e.g., moraic nasals, short vs. long vowels and consonants). For the pitch pattern training, the task was to produce isolated and sentence-embedded words according to "prosody graphs" (i.e., graphic representations of words' accentual pitch patterns and rhythmic temporal patterns). Trainees imitated training materials until the F0 contours of their productions matched with those of provided models. The goal was to help learners integrate intonation and rhythm. After ten sessions, the participants' performance in perception of morae improved by 27% in training 1 (6-alternative identification task), and their perception of pitch patterns improved by 18% in training 2

(9-alternative identification task). This study pursued pedagogical goals, and it differed from other HTVP trainings in terms of the training tasks and procedures.

Encouraged by success of previous high-variability training studies on prosodic categories, the training experiment in this dissertation also adopted the HVTP. Target words, sentential contexts and talkers were varied to provide listeners with multiple opportunities to notice phonetic realization of the three-way phonological contrast in the pitch patterns of Japanese words. The training experiment is presented in Chapter 8. It is preceded by a series of perception experiments investigating the processing of pitch-accent contrast by native and non-native listeners of Japanese.

CHAPTER III

EXPERIMENT 1: 2AFC CATEGORIZATION WITH THREE INDEPENDENT VARIABLES

3.1. Introduction

Experiment 1 examined how listeners categorized pitch patterns of words in which the location of the F0 peak, the height of the peak, and the F0 fall were varied. The question addressed here was whether these three parameters had different effects on the perception of prominence by listeners who speak either Tokyo Japanese or American English natively.

For native listeners of English, a stress language, F0 is not a cue to lexical prominence per se, but it may be correlated with cumulative prominence judgments (Chapter II, section 2.2.3). For native listeners of Tokyo Japanese, a non-stress language, F0 is the fundamental cue to prominence. It is possible that the overall sensitivity to variations in F0, as well as sensitivity to the particular F0 cues of the Japanese pitch-accent contrast, vary between the two listeners groups (Chapter II, sections 2.6 and 2.7). In this experiment, the cumulative perception of prominence, that may be attributed to both word and phrase levels, was investigated in one-word utterances consisting of three syllables. The acoustic cues were manipulated to represent three contrastive pitch patterns: 1st-syllable accented, 2nd-syllable accented, and unaccented (Figure 3.1).

3.1.1. F0 Peak Location

The main difference between the 1st-syllable accented and the 2nd-syllable accented patterns lies in the location of the F0 peak, which is defined here as a starting point of the F0 fall. Figure 3.1 illustrates this difference in the minimal pair *mori ga* (the

green and blue contours). These words carry a bitonal lexical pitch accent H*+L (Chapter II, section 2.3.2), where the high tone of the accent (i.e., F0 peak) is associated with the accented syllable, the low tone (i.e., F0 elbow) with the following syllable, and the steep F0 change from high to low (i.e., F0 fall) with the syllable boundary.[5] The contrast between the 1st-syllable accented *móri* and the 2nd-syllable accented *morí* is due to the timing of the F0 peak, which aligns with the ends of the accented syllables of these words.

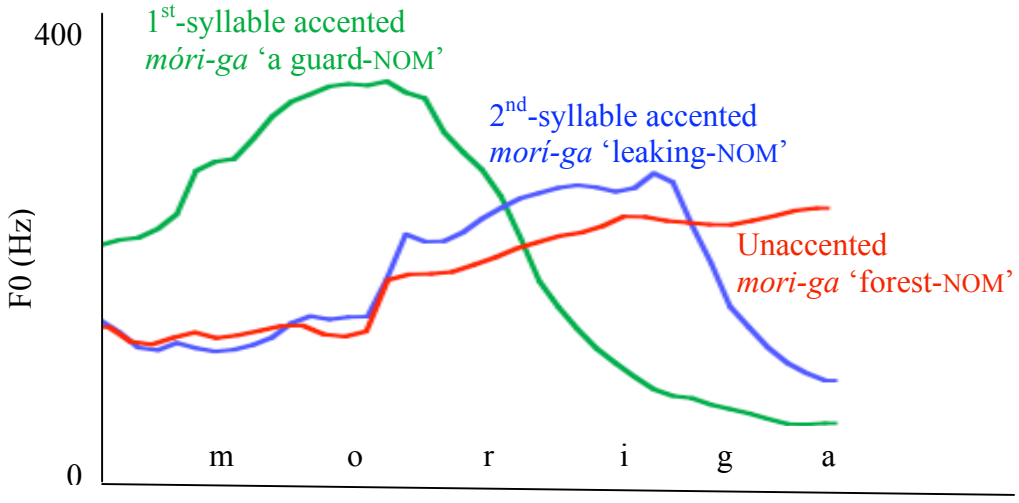


Figure 3.1. Three contrastive accentual patterns of Tokyo Japanese in disyllabic words *mori* followed by the nominative particle *ga*. The F0 contours of the patterns were excised from the utterances *Mori ga ne* 'It's *mori*, isn't it?' produced by a female speaker as responses in short scripted dialogues.

[5] With regard to tone associations, the alternative view is that the high and low tones of a bitonal H*+L accent do not have an independent association with different morae or syllables. In this view, the F0 fall duration is considered to be invariant (Kitahara, 2001; Pierrehumbert & Beckman, 1988). However, as Pierrehumbert and Beckman note, it is difficult to test whether the two tones of the H*+L accent associate with one mora or two successive morae (p. 126). Japanese has a relatively simple syllable structure, which does not allow substantial variation in the amount of segmental material between the accented mora and the following mora such as CVVCCV. With regard to the F0 fall association, Sugito (1982a) showed that the start of the fall does not have to be aligned with the mora boundary in order for the previous syllable to be perceived as accented.

In this experiment, the F0 peak location was modeled to align with the vowel midpoint in the accented syllable. Although the F0 peak delay in the alignment of the accent is well attested (Chapter II, section 2.4.3), the examination of the peak delay effect on the prominence perception was not investigated here. Rather, the location of the peak was grossly varied to be either in the 1st or in the 2nd syllable of a word, to maximize the contrast between these two patterns. The vowel midpoints of the 1st and 2nd syllables were chosen as generic alignment points.

As far as cumulative prominence is concerned, Japanese and English are similar in that the position of a prominent syllable (and, thus, the location of the F0 peak) is not restricted in a word or an utterance. F0 peak may be located in any syllable, as in *móri* versus *morí* in Japanese, and *pérmit* versus *permít* in English.

3.1.2. F0 Fall

A steep F0 fall after the F0 peak defines the difference between accented and unaccented pitch patterns in Japanese (Chapter II, section 2.4.1). As Figure 3.1 illustrates, in contrast to the accented F0 contours (the green and blue lines), the unaccented F0 contour (the red line) does not have a steep F0 fall associated with the lexical pitch accent H*+L. In this experiment, the F0 fall was modeled as present or absent in F0 contours; the steepness of the fall was modeled via the F0 peak height (see the following section).

In English, the F0 fall after the F0 peak associated with a pitch-accented syllable also has been observed. For example, Lehiste and Peterson (1961) reported that for CVC words in a focused position (i.e., *Say the word ___ again*), the average F0 difference between the F0 peak and the offset of the CVC word in five speakers' productions was 38 Hz. Lehiste and Peterson did not analyze their data in the ToBI framework, but this F0

fall was likely to result from the interpolation between the high-tone pitch accent H* marking the CVC word in focus, on one hand, and the low tone of the L+H* accent on *again*, or the phrasal accent L-, or the boundary tone L% marking the end of the utterance, one the other hand (see Table 2.1 and relevant discussion in section 2.3.2 of the introduction). Unlike Japanese, the inventory of pitch accents in English does not include the H*+L accent (Table 2.1, Footnote 4). The domain of the falling F0 contour in English is not a syllable of a word, but a large prosodic grouping such as an intermediate phrase with varying amount of segmental material. Therefore, it was hypothesized that native English listeners would not interpret the F0 fall as a cue to prominence, because they were likely to pay more attention to F0 peak location than to F0 fall, while making prominence judgments (Chapter II, section 2.6.3.2).

3.1.3. F0 Peak Height

The height of the F0 peak is another cue to the distinction between contrastive pitch patterns (Chapter II, section 2.4.1). Cutler and Otake (1999) found that the high and low tones were more different from each other in word-initial than in word-final syllables of Japanese accented words, and that the identification accuracy of contrastive words by native listeners was higher in the former than in the latter cases. In other words, the acoustic difference in F0 peaks was more apparent between 1st-syllable accented and unaccented words than between 2nd-syllable accented and unaccented words. Figure 3.1 illustrates the differences in the F0 peak height for the three accentual patterns.

Since the F0 peak is the starting point of the F0 fall in the H*+L pitch accent, there is a positive correlation between these two acoustic cues to prominence. Namely, a higher F0 peak implies a steeper F0 fall, which may be construed in its turn as the

magnitude of the fall over time. As illustrated in Figure 3.1, the magnitude of the fall is larger in the 1st-syllable accented *móri ga ne* (the green line) than in the 2nd-syllable accented *morí ga ne* (the blue line). However, the time course of the F0 fall is shorter in the 2nd- than in the 1st-syllable accented utterance. Thus, the steepness of the F0 fall is contingent on the difference between the values of the preceding F0 peak and the following F0 elbow, which appears to be mainly due to the F0 peak height in Figure 3.1. It is also contingent on the alignment of the F0 elbow in relation to the F0 peak, but this time interval, over which the fall occurs, is limited by the available segmental material in *mori ga ne.*

In the current experiment, various degrees of F0 fall steepness were represented via manipulation of the F0 peak height. The value of the following F0 elbow and the duration of the fall were kept constant due to the limited amount of segmental material in three-syllable stimuli.

3.1.4. Predictions

Perception of prosodic prominence by English and Japanese listeners was investigated by using resynthesized F0 contours in a three-syllable word, in which the location of the F0 peak, the F0 fall, and the height of the F0 peak were systematically varied. It was hypothesized that the English listeners may be in general less sensitive to variations in F0 than the Japanese listeners. English listeners may especially under-attend to the F0 fall cue and its steepness as manifested by the F0 peak height, while their sensitivity to the F0 peak location may be similar to that in Japanese listeners. In such a case, the F0 peak location would have the largest effect on the judgments of prominence, as compared to the other two cues.

3.2. Method

3.2.1. Participants

3.2.1.1. Participant characteristics common for Experiments 1-4. For all perception experiments, native English listeners were recruited from the Human Subjects Pool at the University of Oregon. They received course credit for their participation. Native Japanese listeners were recruited from the American English Institute classes at the University of Oregon and by word of mouth. They participated as volunteers.

Participants in both groups studied foreign languages as a part of their school curriculum: typically, in high school in the U.S. (i.e., 15 years old and later) and in middle and high school in Japan (i.e., 12 years old and later). None of the participants had learned a second language in childhood. English listeners had never studied Japanese or other tone languages, and they were not actively using their other second languages outside of the foreign language classroom. Japanese listeners were actively using English in the study abroad setting when the experiments were conducted. However, their exposure to spoken English was limited by the length of their residence in the U.S., which did not exceed one year. Thus, it was assumed that the task performance was not heavily affected by listeners' fluency in second languages.

3.2.1.2. Participant groups in Experiment 1. The listener group characteristics in Experiment 1 are summarized in Table 3.1. There were 16 listeners in each group: 14 females and 2 males. According to language background questionnaires, English listeners' dialect was West Coast American English; Japanese listeners' dialect was Tokyo Japanese. At the time of the study, the mean length of residence of the Japanese participants in the U.S. was two weeks. Their age of arrival in the U.S. was 18 or older.

Table 3.1

Listeners' Background Information: Experiment 1 (2AFC Categorization)

	Japanese (n = 16)	English (n = 16)
Dialect	Tokyo Japanese	West Coast American English
Age	18-23 (M_{year} = 20.1)	18-20 (M_{year} = 18.9)
Sex	14 females, 2 males	14 females, 2 males
Languages learned as adults for more than a year	English, German, French, Spanish, Mandarin (M_{number} = 1.6; M_{year} = 8.9)	Spanish, French, German (M_{number} = 1.3; M_{year} = 3.6)
Life-abroad experience	AoA: 18-23 (M_{year} = 20.1) LoR: 0.5 (M_{month} = 0.5)	

Note. AoA = age of arrival to the U.S.; LoR = length of residence in the U.S.

3.2.2. Materials

3.2.2.1. Stimuli characteristics common for Experiments 1-4. In all perception experiments reported in Chapters III-VI, test materials were variations of a nonword *nenema*, and practice materials were four variations of a nonword *nimiga*. A native speaker of Japanese trained in linguistics naturally produced *nenema* and *nimiga* words as unaccented. The digital recordings with a 22,050 sampling rate were made in a sound-attenuated booth. Test and practice stimuli were created from the digital audio files in Praat using the Pitch Synchronous Overlap and Add resynthesis routine (PSOLA function, Boersma & Weenink, 2010). Prior to the F0 manipulation, durations and amplitude of the segments in the base productions were examined (Table 3.2).

In order to eliminate the duration cue to prominence that English listeners are likely to use, the durations of the first two syllables in *nenema* were normalized using the mean duration of two [n] segments (58.5 ms) and the mean duration of two [e] segments

(124.5 ms). The durations of the segments in the first two syllables of the practice word *nimiga* were normalized in a similar fashion.

Table 3.2

Duration and Amplitude of Segments in the Stimulus Word Used for F0 Manipulation

Segment	Duration (ms)		RMS amplitude (dB)
	Original	Normalized	Original
n	57	58.5	75.9
e	111	124.5	81.4
n	60	58.5	78.9
e	138	124.5	79.8
m	104	104	82.5
a	137	137	80.3

Intensity is also a cue to prominence in English (Chapter II, section 2.2.3), therefore, it was examined for each segment of *nenema* using the root-mean-square (rms) amplitude measure. The rms amplitudes of the first syllable [ne] and the second syllable [ne] were similar (both 79 dB). The amplitude differences between syllables were even smaller than just noticeable difference of 1.5 dB found for 150-milisecond monosyllables in a discrimination task (Nishinuma, di Cristo, & Espesser, 1983). Based on this experimental evidence, the decision was made to preserve the original amplitude of individual segments.

The original durations and amplitudes of the [m] and [a] segments were not manipulated because test tasks did not involve consideration of the third syllable.

3.2.2.2. F0 manipulation in Experiment 1. Three acoustic parameters were manipulated in the design of the current experiment, namely, the location of the F0 maximum (Peak Location), the height of the F0 maximum (Peak Height), and the F0 fall after the peak (F0 Fall). The orthogonal design of the experiment yielded 28 different F0 contours: Peak Location (2) x Peak Height (7) x F0 Fall (2).

As Figures 3.2a and 3.2b illustrate, the two levels of Peak Location were the midpoint of the first vowel [e] ($E1_{50\%}$) and the midpoint of the second vowel [e] ($E2_{50\%}$). Peak Height was manipulated in 10 Hz step increments, ranging from 200 to 270 Hz. This step size was chosen because of the evidence that a difference in about 10 Hz (1.5 semitones) is sufficient to cause a difference in the perception of prominence (Rietveld & Gussenhoven, 1985). Since the F0 of the word onset and offset was fixed at 200 Hz in all stimuli, the magnitude of the rise and fall in F0 contours varied with the height of the F0 maximum. The two levels of F0 Fall were the *F0 fall* condition (blue contours in Figure 3.2) and the *no F0 fall* condition (red contours in Figure 3.2).

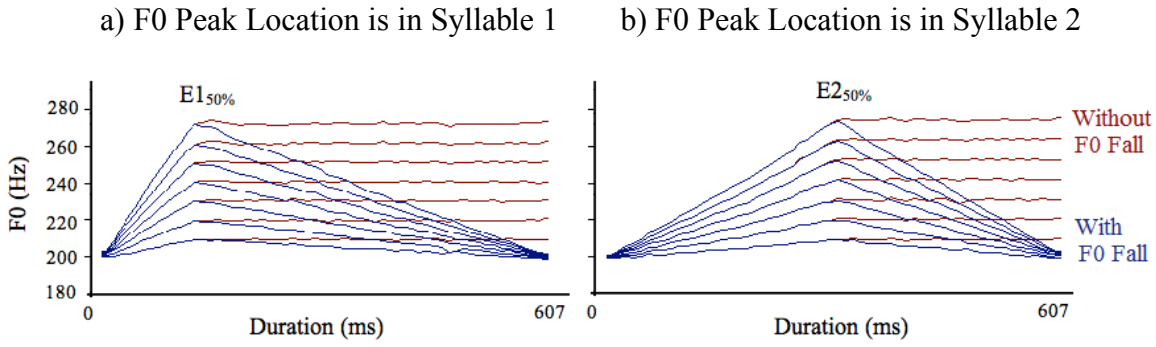


Figure 3.2. Manipulation of F0 peak location, F0 peak height and F0 fall in *nenema*.

Due to the experimental design (i.e., manipulation of the three parameters while controlling for other variables), the resynthesized F0 contours were not exactly the same as they would have been in natural productions of real Japanese words. For example, Cutler and Otake (1999), among others, observed that a) the F0 peak is higher in the 1st- than 2nd-syllable accented words (258 Hz vs. 211 Hz), and that b) the F0 of the first syllable is consistently higher in accented than unaccented words of similar segmental structure (211 Hz vs. 195 Hz). In current manipulation, the F0 peak was the same regardless the location of the peak (i.e., the first or the second syllable) and presence of the F0 fall (i.e., fall and no fall conditions). F0 range, on the other hand, was informed by previous research. Cutler and Otake (1999) found that in productions of 1st- or 2nd-syllable accented CV.CV words by three female speakers of Tokyo Japanese, the mean F0 fall associated with the pitch accent was 71 Hz. The female speaker, whose recordings were used for F0 manipulation in this study, had even larger (probably, more exaggerated) F0 fall in her productions of real three-syllable words – around 100 Hz. The F0 maximum in this study varied from 210 to 270 Hz, thus, introducing variation in the F0 fall natural for a female voice (10-70 Hz).

3.2.3. Procedure

Each stimulus shown in Figure 3.2 was repeated 10 times in the experiment, yielding a total of 280 trials. Trails were presented to listeners in random orders. For each variation of the *nenema* word that they heard, listeners were asked to decide whether the first or the second [ne] was more prominent. In other words, listeners were forced to choose which syllable stood out the most for them as compared to the other syllables in the same word while given two alternatives – *1st syllable NE* and *2nd syllable NE*. The

instructions of this two-alternative forced-choice (2AFC) task were given to listeners in their native language.

3.2.3.1. Methodology common for Experiments 1-4. Several methodological points in task instructions and procedure were common for all perception experiments carried out in this dissertation. First, listeners had instructions in their native language with no reference to stress accent or pitch accent. The key words *prominent syllable* / *kenchona onsetsu* (顕著な音節) and *stand out* / *medatte kikoeru* (目立って聞こえる) have similar meanings in English and Japanese. Second, treatment of *nenema* as a word was encouraged by the explanation of its phonological plausibility in English (for English listeners) and Japanese (for Japanese listeners).

Prior to each test, listeners had a short practice with variations of the *nimiga* word that were manipulated similar to the *nenema* stimuli. During this practice, listeners familiarized themselves with the task format, asked clarifying questions, and adjusted the listening level on headphones when needed. Every listener performed the task individually in a sound-attenuated booth. Using the Multiple Forced Choice experiment paradigm in Praat (MFC, Boersma & Weenink, 2010), the answer options for each trial were presented in the listeners' native language as text buttons on a computer screen. Stimuli were presented binaurally over headphones. Listeners responded by clicking on answer buttons and self-paced the stimulus presentation. They were encouraged to make their best guess when unsure, and were required to take breaks after approximately every 50 trials to fill a language background questionnaire section by section.

3.2.4. Analysis

The percentage of Syllable 1 Prominent responses to ten repetitions of each

stimulus type was calculated for each listener. Syllable 2 Prominent responses were not reported because they complemented Syllable 1 Prominent responses up to 100%. To investigate the differences between Japanese and English listeners in categorization of the F0 contours, percentages of Syllable 1 Prominent responses were submitted to two separate analyses of variance (ANOVA) with the within-subject factors of Peak Location, F0 Fall, and Peak Height.[6] When the ANOVA's assumption of similar covariance in observations across conditions was violated as indicated by significant outcomes of Mauchly's tests of sphericity, the Greenhouse-Geisser corrections for the degrees of freedom were used (Box, 1954). An alpha level of .05 was used as a threshold of significance; exact *p*-values are reported below for all statistical tests.

3.3. Results

Two separate analyses of variance revealed that a three-way interaction between Peak Location, F0 Fall, and Peak Height was significant in the Japanese data [$F(6, 55.41) = 4.07$, $p = .007$, $\eta_p^2 = .21$], but not in English the data [$F(6, 90) = 1.12$, $p = .360$, $\eta_p^2 = .07$]. The effects of the factors and their interactions are reported below. Figures 3.3 and 3.4 illustrate the influence of the three acoustic cues on response patterns of Japanese and English listeners.

[6] Analyses of variance, where a dependent variable is percentage (or proportion) of 'X' category responses and independent variables are acoustic cues of interest, have been used in categorization experiments in previous research (e.g., Bradlow et al., 1997; Francis et al., 2008; Holt & Lotto, 2006), though mixed-effects logistic regression analysis would have been a better statistical procedure for such data (Baayen, Davidson, & Bates, 2008). The findings in categorization experiments in this dissertation were consistent with each other, supporting the conclusions drawn from the ANOVA analyses in the current work.

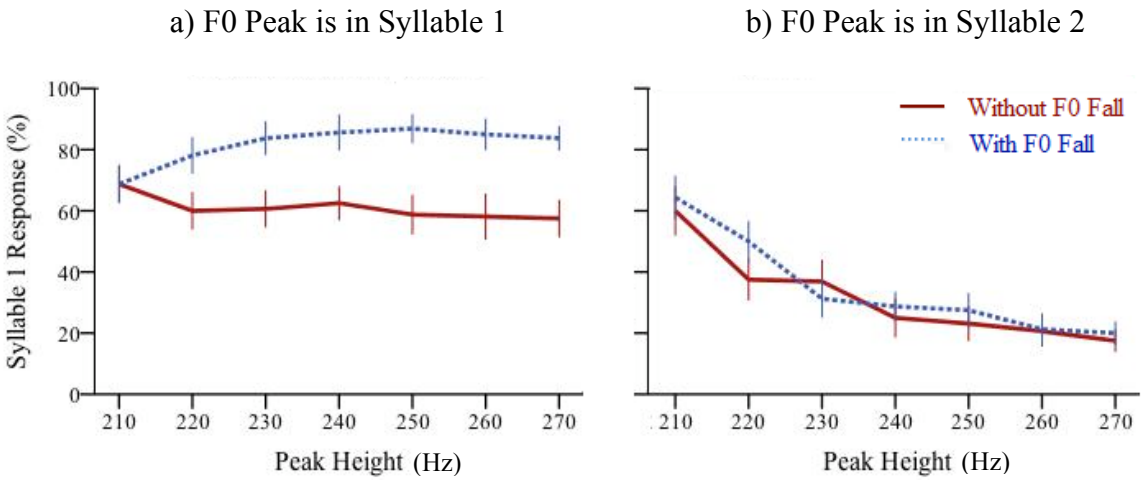


Figure 3.3. Syllable 1 Prominent responses by Japanese listeners as a function of F0 peak location, F0 fall, and F0 peak height. Error bars indicate ± 1 *SE*.

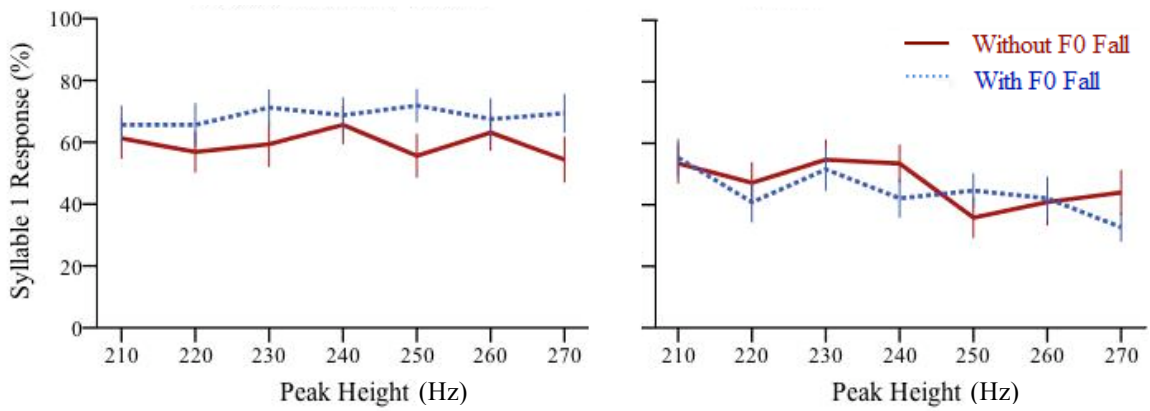


Figure 3.4. Syllable 1 Prominent responses by English listeners as a function of F0 peak location, F0 fall, and F0 peak height. Error bars indicate ± 1 *SE*.

A comparison of responses to stimuli with the F0 peak in the 1st syllable (Figure 3.3a and Figure 3.4a) and in the 2nd syllable (Figure 3.3b and Figure 3.4b) reveals an effect of peak location on listeners' responses in both groups (Japanese: $F(1, 15) = 30.08$, $p < .001$, $\eta_p^2 = .67$; English: $F(1, 15) = 9.21$, $p = .008$, $\eta_p^2 = .38$). Namely, Japanese

listeners had more Syllable 1 Prominent responses when the F0 peak was on the first than second syllable (M = 70.1%, SD = 2.6 vs. M = 30.3%, SD = 2.8, respectively). The difference in Syllable 1 Prominent response patterns due to the peak location in English listeners was smaller (M = 60.4%, SD = 2.6 vs. M = 40.5%, SD = 2.8, respectively). The interactions between Peak Location and the other two factors are detailed below.

The effect of F0 Fall was significant in the Japanese group [$F(1, 15) = 18.82$, $p = .001$, $\eta_p^2 = .56$], but not in the English group [$F(1, 15) = .90$, $p = .359$, $\eta_p^2 = .06$]. The interaction between F0 Fall and Peak Location was significant in both groups, although it had a larger effect size for the Japanese listeners [$F(1, 15) = 21.04$, $p < .001$, $\eta_p^2 = .58$] than for English listeners [$F(1, 15) = 7.70$, $p = .014$, $\eta_p^2 = .34$]. A comparison of responses to stimuli with F0 fall (dotted lines) and without F0 fall (solid lines) shows that in the Japanese data the judgment of prominence location in the 1st syllable was facilitated by the fall (Figure 3.3a), but there was no such effect of the fall for the 2nd syllable (Figure 3.3b). In the English data, a similar trend was observed (Figures 3.4a and 3.4b), although the difference between the dotted line and the solid line in Syllable 1 was smaller. These results suggest that the Japanese listeners were more sensitive to the F0 fall as a cue to prominence when the F0 peak was in the first syllable as compared to the English listeners.

The effect of Peak Height was also significant in the Japanese group [$F(6, 90) = 6.06$, $p < .001$, $\eta_p^2 = .29$], but not in the English group [$F(6, 90) = 1.46$, $p = .200$, $\eta_p^2 = .09$]. The interaction between Peak Height and Peak Location was significant in both Japanese data [$F(6, 90) = 21.15$, $p < .001$, $\eta_p^2 = .59$] and English data [$F(6, 90) = 2.56$, $p = .025$, $\eta_p^2 = .15$], although the effect size was larger in the former than the latter group.

As is apparent from Figures 3.3b and 3.4b, the effect of Peak Height was observed in the 2nd syllable for both listener groups: namely, the higher the Peak Height, the more prominent the 2nd syllable. These results suggest that the Japanese listeners were more sensitive to Peak Height as a cue to prominence than English listeners when the F0 peak was in the second syllable.

To further examine the effect of Peak Height in the 2nd syllable, paired-sample *t*-tests were conducted to compare the difference in Syllable 1 Prominent responses at the highest peak of 270 Hz with responses at other six peaks. In the Japanese data, the 210-240 Hz peak height levels were significantly different from the 270 Hz level ($p \leq .017$), however, the 250-270 Hz levels were not significantly different from one another ($p \leq .137$). In the English data, only the 210 and 230 Hz levels were significantly different from the 270 level ($p \leq .050$); the other levels did not differ from one another ($p \leq .577$). These results suggest that in the 2nd syllable, peak height facilitates the perception of the syllable as prominent, when it ranges from 250 Hz to 270 Hz for Japanese listeners, and from 240 to 270 Hz for English listeners.

The last interaction between F0 Fall and Peak Height was significant in the Japanese data [$F(6, 90) = 2.44$, $p = .031$, $\eta_p^2 = .14$], but not in the English data [$F(6, 90) = 1.67$, $p = .139$, $\eta_p^2 = .10$]. Figure 3.3a illustrates the significant interaction: Namely, when the Peak Height was 210 Hz, the responses of the Japanese listeners to stimuli with F0 Fall and without F0 fall were the same. However, as Peak Height increased, the role of the F0 fall in identification of prominence on the 1st syllable also increased. This finding suggests that a fall larger than 10 Hz facilitates the percept of prominence in the 1st syllable by Japanese listeners.

3.4. Discussion

3.4.1. General Conclusions

Experiment 1 showed that linguistic experience affected perception of prosodic prominence. When listeners were forced to decide whether the 1st syllable or the 2nd syllable stood out the most for them in a word, all three cues of F0 peak location, F0 fall and peak height together influenced the decisions of Japanese listeners. The percept of syllable prominence was achieved when the F0 peak was located in that syllable; an F0 fall larger than 10 Hz enhanced this percept in the 1st syllable; and peak height higher than 240 Hz enhanced this percept in the 2nd syllable.

Statistical tests showed a somewhat similar but much weaker response pattern in English listeners. The factor that affected English prominence judgments the most was the location of F0 peak, which in turn interacted with the peak height and F0 fall. The interpretation of the latter interaction is not straightforward. For example, it is not apparent from Figure 3.4 how peak height affected the responses of English listeners differently when the F0 peak was in the 1st syllable and when it was in the 2nd syllable. In contrast to Japanese listeners, English listeners were less sensitive to the presence of F0 fall and the height of F0 peak. The interaction among all three cues was not significant for English data.

In sum, for F0 contour manipulation in Experiment 1, the unambiguous location of the F0 peak at the vowel midpoint could not create by itself a robust percept of prominence in the respective syllable. The presence of the F0 fall after the F0 peak, and the magnitude of the fall (as determined by the peak height) facilitated the percept of prominence more for Japanese than English listeners. A smaller sensitivity to the F0

variations in English listeners may be because duration and loudness are primary cues to word-level prominence in English (Chapter II, section 2.2.3), and because F0 contours of the stimuli in Experiment 1 may not resemble typical F0 contours in English one-word utterances (Chapter II, section 2.3.2).

3.4.2. F0 Slope

The differential use of acoustic cues for 1st syllable prominence (i.e., F0 fall) and 2nd syllable prominence (i.e., F0 peak height) by listeners in Experiment 1 may be epiphenomenal to the manipulation of the acoustic cues in the current design. As was detailed in Chapter II, section 2.4.1, the steepness of F0 fall is the most important cue to the pitch-accent contrast in Japanese. The steepness of the fall in Experiment 1 was construed by varying F0 peak height. In this design, the magnitude of a peak height in Hz was equal to the magnitude of a fall; however the slope of the fall in Hz/ms was not controlled for. If the slope of the fall is calculated as its magnitude over duration, then the slopes corresponding to 210-270 peak heights in the 1st syllable varied from 0.02 Hz/ms to 0.14 Hz/ms. The slopes corresponding to the same seven peak heights in the 2nd syllable varied from 0.03 Hz/ms to 0.23 Hz/ms. Thus, given the same height of the F0 peak, the F0 fall was always slightly steeper when the peak was in the 2nd syllable than when it was in the 1st syllable. This may explain why listeners were more sensitive the F0 fall cue to prominence when the F0 peak was in the 1st syllable.

In intonational phonology, starting, probably, with the dissertations of Bruce Gösta (1977) and Janet Pierrehumbert (1980), the role of F0 slopes between target pitch accents has been commonly understood as varying under time pressure imposed by segmental material with which tonal targets are aligned (see also Footnote 5 for the

bitonal H*+L). Although, the steepness of F0 rises and falls may influence listeners' perception, it is not considered a phonological target, and thus, is not expected to have an independent effect on pattern categorization. In Experiment 1, however, significant interactions between F0 fall and F0 peak location in both listener groups may be interpreted as indicating importance of F0 fall slope for judgments of prominence.

Previous experimental evidence regarding the role of F0 fall steepness for perception of accentedness by native Japanese listeners is inconsistent. Based on a word identification task with real word minimal pairs, Kitahara (2001) argued that the magnitude of F0 fall affected accent-presence detection in stimuli like *shíro* 'white' and *shiro* 'castle,' or *haná* 'flower' and *hana* 'nose', but F0 slope did not. Hasegawa and Hata (1992), on the other hand, argued that F0 slope affected accent-location detection in nonwords. In their experiment, later F0 peaks in the second vowel of *mamama*, required a steeper fall slope for accent perception in the 1st syllable. For example, when the peak was at the midpoint of the second [a] (130 ms vowel duration), the 1st syllable tended to be perceived as accented at F0 slope of approximately 0.7 Hz/ms or larger. At less steep slopes, the 2nd syllable was perceived as accented.

In Experiment 1 of this dissertation, the F0 fall after the peak in the 2nd syllable was considerably less steep than in the study by Hasegawa and Hata (1992), but it seemed to have an opposite effect on the responses of Japanese listeners. Namely, a steeper F0 fall (corresponding to a higher F0 peak) in the 2nd syllable facilitated perception of the 2nd (not the 1st) syllable as prominent. This inconsistency of the results may be due to differences in the manipulation of F0 slope. Differently from Hasegawa and Hata, in this experiment 1) peak height (not peak location) was varied within the

second vowel [e] of *nenema*, and 2) the slope extended until the offset of the word (not the offset of the vowel). It is possible, however, that manipulation of peak location even within one vowel may affect listeners' judgments. In addition, all studies showed that the location of the F0 peak and the F0 fall (magnitude or/and slope) jointly contribute to the perception of prominence in a particular syllable of the word by native Japanese listeners.

Consistently with Hasegawa and Hata (1992), the slope of F0 fall affected prominence judgments by American English listeners, as may be inferred from the interaction between peak location and F0 fall. Hasegawa and Hata's materials were utterances with manipulated focus (*This is MY net* vs. *This is my NET*). The F0 maximum was aligned with the onset of [e]; the F0 fall was approximately 50 Hz; the slope of the fall varied from 0.44 Hz/msec to 2.4 Hz/msec. The researchers concluded from their results that prominence judgments by English listeners were affected by extreme values of the F0 fall slope. Very gradual slopes facilitated the perception of prominence on *net*; very steep slopes facilitated the perception of prominence on *my*.

The methodological differences between the Hasegawa and Hata's study (1992) and Experiment 1 (use of nonwords vs. real sentences; identification of prominence task vs. matching with more appropriate questions task) complicate the comparison of the results. However, it is interesting to note that in Experiment 1 the F0 fall slopes were relatively gradual (0.23 Hz/ms at maximum), and according to the suggestion by Hasegawa and Hata this acoustic information might have been interpreted by listeners as ambiguous in terms of the accent-location detection. It is possible that a more extreme F0 fall slope is required for native English listeners to consider the fall as a cue to pitch accent.

3.4.3. Limitations

Experiment 1 has several limitations. Firstly, peak height and F0 fall are not completely independent parameters even though they were treated as such in the design and analyses. These parameters are positively correlated: namely, the higher the F0 peak, the more pronounced the F0 fall (Figure 3.2). This confound complicates the interpretation of significant interactions between the three acoustic cues for Japanese listeners. For example, it is unclear whether peak height and F0 fall had different effects on the perception of prominence by Japanese listeners, since the magnitude of the height and the magnitude of the fall are essentially the same thing. It is also possible that due to the F0 fall-height confound, the response patterns of Japanese and English listeners appear to be more similar than they really are. Even if one group is more sensitive to the F0 fall and the other group is more sensitive to the F0 height, we have no way of knowing it with the current experiment design.

Secondly, as discussed above, the characteristics of F0 fall also varied with F0 peak location: namely, the later the peak, the steeper the F0 fall. In the Japanese data, the F0 fall may have mattered only when the F0 peak was located in the 1st syllable, because the listeners had more time to process it, whether the slope of the fall was steep or not. F0 peak height, on the other hand, may have mattered mostly when the F0 peak was located in the 2nd syllable because it compensated for the limitation of the shorter (in time) F0 fall duration. If this interpretation of the results is correct, then we should see that with a gradient manipulation of the peak location in the word, the effect of the F0 fall would vary. We would expect a greater F0 fall effect when the F0 maximum is aligned earlier in a word.

Lastly, in the current stimuli, the F0 at the offset of word was equal to the F0 at the onset of word, namely 200 Hz. However, in pitch patterns of real accented words, the F0 at the end of the word may be much lower than at the beginning of the word (e.g., the green contours of the 1st-syllable accented phrases and words in Figures 2.3 and 3.1).

In the next two experiments, the concerns discussed above were addressed in the following way. First, F0 peak height was kept constant at 60 Hz above word onset to disentangle the effects of F0 peak height and F0 fall. The 60 Hz value was chosen because it unambiguously facilitated the identification of prominence for native Japanese listeners in Experiment 1 (Figure 3.2). Second, the slope of F0 fall was controlled by having the same magnitude of the fall over the same duration. In other words, the F0 elbow associated with the low tone of the pitch accent H*+L was equidistant from the F0 peak associated with the high tone of the accent. Third, a steeper slope of the F0 fall variable was introduced in addition to the other two levels. At this level, F0 at the word end was lower than at the word beginning, which is similar to naturally occurring F0 contours of accented words. Last, F0 peak location was manipulated gradiently throughout the syllable to investigate the effect of peak alignment on perception of prominence.

CHAPTER IV

EXPERIMENT 2: 2AFC CATEGORIZATION WITH TWO INDEPENDENT VARIABLES

4.1. Introduction

Experiment 2 examined how listeners categorized pitch patterns of words in which the location of the F0 peak and the magnitude of the F0 fall were varied. The current experiment was designed to tease apart the roles of the peak location and the fall, which were confounded in Experiment 1. In addition, the role of the peak delay in perception of prominence was investigated. The question addressed here was whether Japanese and English listeners differed in their sensitivity to the acoustic cues signaling prominence in Tokyo Japanese, on average and individually.

4.1.1. Peak Delay

It has been demonstrated that both native and non-native listeners may perceive prominence on a syllable even when the F0 peak occurs in the syllable following the accented one (Sugito, 1980). This observation is consistent with reports of the F0 peak delay in experimental production data (Ishihara, 2006; Neustupný, 1966; Shport & Guion, 2008; Sugito, 1980, 1982a). For example, Ishihara found that in initially-accented CV-sequence words varying in their segmental and morphological structure, the pattern of accent alignment was always uniform. Namely, the F0 peak occurred after the end of the accented syllable, just around the onset of the vowel in the following syllable (p. 30, 39). In medially-accented CV-sequence words, the F0 peak occurred earlier, namely around the end of the accented mora (pp. 65-66). In unaccented CV-sequence words, the F0 peak tended to align with the onset of the vowel in the 3rd syllable (p. 70). Thus, in all

three conditions the F0 peak is aligned past the syllable with which the high tone of the lexical pitch accent H*+L or the phrasal accent H- is associated.

Similar to Japanese, a peak delay tendency is reported for English (Chapter II, section 2.4.3). Borrowing English examples from Pierrehumbert, when the words *límb, límo* and *límousine* are said with the same intonation, the alignment of the high-tone pitch accent H* associated with the prominent syllables in these words varies depending on the amount of segmental material available in these words (2000, pp. 12-13). Pierrehumbert observed the F0 peak in the beginning of [i] in *límb*, around the end of [i] in *límo*, and well into the [m] in *límousine*. Experimental evidence suggests that on average the F0 peak of high-tone pitch accents in English aligns with the end of the accented syllable (Beckman & Pierrehumbert, 1986; Ladd & Schepman, 2003). Taking into account the relationship between syllable duration and F0 peak alignment, Silverman and Pierrehumbert (1990) proposed that the alignment would be better expressed as a proportion of the duration of the accented vowel rather than an absolute measure.

There might be an auditory motivation for the later alignment of the F0 peak such as segmental effects on F0 (Sugito, 1980, among others). Khouw and Ciocca (2007) showed that in Cantonese, a tone language, the sections near the beginning of the vocalic segment in CV syllables did not distinguish Cantonese tones either in perception or production, at least, when the C was an obstruent (i.e., when listeners might have expected segmental effects on F0). However, the section around 6/8 and 7/8 of the vocalic segment facilitated the tone distinction. It may be the case that listeners, being aware of F0 perturbation caused by obstruents, interpreted F0 change at the ends of accented syllables as tonal targets.

4.1.2. Predictions

From previous studies investigating the relationship between the tonal target alignment and perception of that target, we may predict that in CVCVCV words, like *nenema*, Japanese listeners will perceive the first syllable as prominent when the F0 peak occurs anywhere until the midpoint of the vowel in the second syllable. The second syllable will be perceived as prominent when the F0 peak occurs in later locations. English listeners may also interpret later F0 peaks as a peak delay; however, the categorical boundary between the 1st syllable prominent and 2nd syllable prominent words may occur earlier in the word for English than Japanese listeners (Chapter II, section 2.4.3). In the current experiment, F0 peak location was manipulated gradiently to investigate cross-linguistic differences in the effect of peak alignment on perception of prominence.

Similar to the predictions of Experiment 1, both groups were expected to be sensitive to F0 peak location, but differ in their sensitivity to F0 fall, more specifically, the magnitude and slope of the fall.

4.2. Method

4.2.1. Participants

There were 20 listeners in each group: 15 females and 5 males (Table 4.1). According to language background questionnaires, English listeners' dialect was West Coast American English; Japanese listeners' dialect was Tokyo Japanese. At the time of the study, the mean length of residence of the Japanese participants in the U.S. was 4.5 months. Their age of arrival was 19 or older.

Table 4.1

Listeners' Background Information: Experiment 2 (2AFC Categorization) and Experiment 4 (AB Discrimination)

	Japanese (n = 20)	English (n = 20)
Dialect	Tokyo Japanese	West Coast American English
Age	19-45 (M_{year} = 22.4)	18-31 (M_{year} = 21)
Sex	15 females, 5 males	15 females, 5 males
Languages learned as adults for more than one year	English, German, French, Spanish, Mandarin, Korean, Polish (M_{number} = 2; M_{year} = 7.9)	Spanish, French, German (M_{number} = 1; M_{year} = 3.1)
Life-abroad experience	AoA: 19-45 (M_{year} = 22.2) LoR: 0.25-10 (M_{month} = 4.5)	

Note. AoA = age of arrival to the U.S.; LoR = length of residence in the U.S.

4.2.2. Materials

Test materials were variations of the nonword *nenema* described below, and practice materials were four variations of the nonword *nimiga*. Segmental characteristics of *nenema* are detailed in the Methods section of Chapter III.

Differently from Experiment 1, only two acoustic parameters – the location of the F0 peak and the magnitude/slope of the F0 fall after the peak – were manipulated in the design of this experiment (Figure 4.1). Alignment of the F0 peak was varied in eight steps throughout the first two syllables of *nenema*: 10%, 50% and 90% in the first vowel [e]; 50% in the second consonant [n]; 10%, 50% and 90% in the second vowel [e]; and 50% in the third consonant [m]. These eight locations of the F0 peak were determined in terms of proportion of duration into sonorant segments rather than absolute duration into the segments. In terms of the distance in milliseconds from the word onset, the peak

locations were at 71 ms ($E_{10\%}$), 121 ms ($E_{50\%}$), 171 ms ($E_{90\%}$), 212 ms ($N_{50\%}$), 254 ms ($E_{10\%}$), 304 ms ($E_{50\%}$), 354 ms ($E_{90\%}$), and 418 ms ($M_{50\%}$) into the word.

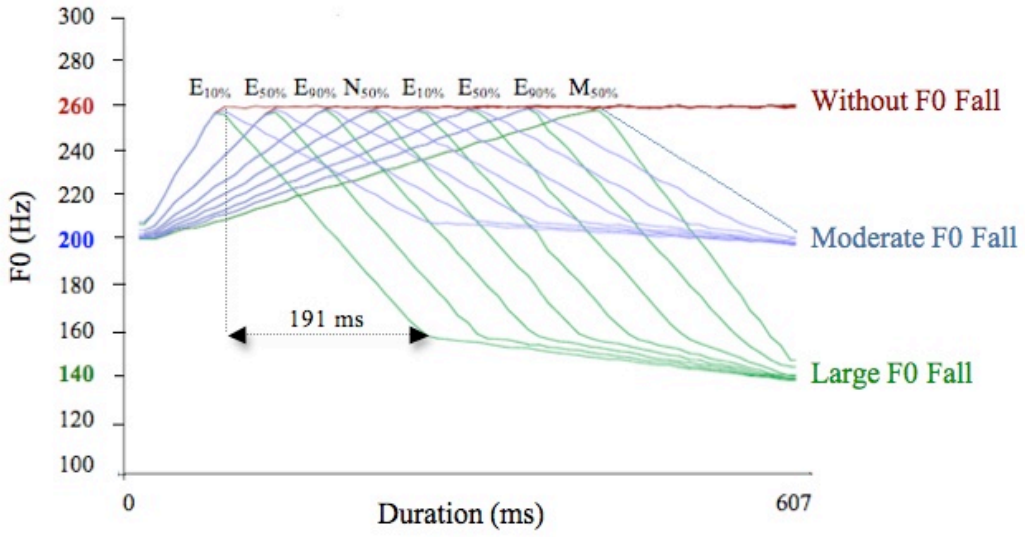


Figure 4.1. Manipulation of F0 peak location and the magnitude/slope of the F0 fall in *nenema*: Experiment 2 (2AFC categorization) and Experiment 3 (3AFC categorization).

The F0 rise from the word onset to the F0 peak was 60 Hz. In other words, the F0 peak had a similar height of 260 Hz in all stimuli, the height that facilitated the identification of prominence for native Japanese listeners in Experiment 1. At each location of the F0 peak, the overall magnitude of the F0 fall from the peak to the word offset was manipulated in three steps (Figure 4.1): a large fall of 100 Hz (green contours), a moderate fall of 50 Hz (blue contours), or no F0 fall (red contours). Except for the stimuli with the F0 peak at the midpoint of [m], the F0 trailed down another 20 Hz (the large-fall series) or 10 Hz (the moderate-fall series) between the F0 elbow and the word offset. The orthogonal variation of the two parameters in question yielded 24 different F0 contours (i.e., 8 peak locations x 3 levels of fall).

In contrast to the design of Experiment 1, the slope of the F0 fall was kept constant at each level of the F0 fall manipulation by controlling the time course of the F0 fall. As described in the previous Chapter, the durations of vowels and consonants in *nenema* were different (Table 3.2); therefore, the temporal distance between adjacent F0 peaks also varied from 41 to 64 milliseconds. To control the slope of the F0 fall, its endpoint (i.e., the location of a low tone, or elbow) was always 191 ms after the peak (Figure 4.1). This distance was determined by the distance between the midpoint of /m/ (the last temporal location of the F0 peak) and the offset of the word. This manipulation allowed for the 100 Hz magnitude and 0.524 Hz/ms slope of the F0 fall in the large-fall series, and for the 50 Hz magnitude and 0.262 Hz/ms slope in the moderate-fall series. The slope in the large-fall series was similar to the 0.544 Hz/ms and 0.435 Hz/ms slopes in the experimental design of Kitahara (2001, p. 36). The slope in the moderate-slope series was smaller than 0.311 Hz/ms, the smallest slope in Kitahara.

In current design, the moderate-fall series resembled the manipulation of the F0 fall in Experiment 1 in that the F0 at the word onset and offset were the same. The large-fall series, on the other hand, approximated naturally occurring F0 patterns in Japanese accented words, where the F0 is lower at the word offset than the word onset. Similar to Experiment 1, the steepness of the slope of the F0 fall between series was modeled by varying the magnitude of the fall, rather than its time course.

4.2.3. Procedure

Each stimulus was repeated 10 times resulting in the total number of trials equal to 240. The listeners were instructed to choose the most prominent syllable in the word

nenema, while given two alternatives – *1st syllable NE* and *2nd syllable NE* (the same 2AFC task as in Experiment 1).

4.2.4. Analysis

The percentage of Syllable 1 Prominent responses to ten repetitions of each stimulus type was calculated for each listener and used as a dependent variable in 1) examination of group differences, and 2) examination of individual differences in sensitivity to the acoustic cues. Both types of analysis are detailed below.

To investigate the differences between Japanese and English listeners in categorization of the F0 contours, two separate repeated measures analysis of variance (ANOVA) with the within-subject factors of Peak Location and F0 Fall were conducted on Syllable 1 Prominent responses. Similar to the analyses in Experiment 1, the Greenhouse-Geisser corrections for the degrees of freedom were used when the equality of the variances of the differences between levels of the repeated-measure factors was not warranted as indicated by significant Mauchly's tests (Box, 1954).

A logistic regression analysis was also conducted for each listener to predict the probability of the Syllable 1 Prominent response as a function of three predictors - F0 Peak Location (L), F0 fall (F), and their interaction (L x F). This analysis allows us to compare listener's reliance on several acoustic cues to a phonological contrast. Escudero, Benders and Lipski (2009) and Morrison (2005, 2007) have described in detail the logic and application of this statistical procedure for categorization of vowels. Adopting a similar analysis here, a logistic regression expressed by the equation (1) predicted log odds of Syllable 1 Prominent responses (i.e., the natural logarithm of the chance that a

participant selects Syllable 1 Prominent response divided by the chance that s/he selects Syllable 2 Prominent response), given three predictors.

(1) Full regression model:

Log(odds(Syllable 1)) =

$\ln(p(\text{Syllable 1})/p(\text{Syllable 2})) = \alpha + \beta_L \text{ x L} + \beta_F \text{ x F} + \beta_{LF} \text{ x LF}$

Several statistics from each individual analysis were recorded.

1. Nagelkerke R^2 in the full model summary: This statistic shows variance in listeners' answers explained by the regression model that included the manipulated predictors (Equation 1).
2. The significance of chi-square: This statistic shows whether the full model (Equation 1) explained more variance in the odds of choosing the Syllable 1 Prominent answer than the null model without any predictors.
3. β-coefficients for predictors: A β-coefficient shows to what extent a one-step difference in a predictor causes a change in the log odds of Syllable 1 Prominent response by a listener.
4. The significance of β-coefficients (Wald statistics): A significant β-coefficient of a predictor indicates that the predictor significantly aided in the choice of Syllable 1 response at $p < .05$.

Logistic regression analyses allowed us to examine which cues were used significantly for the categorization of F0 patterns within each group, and whether the use of the cues was uniform across listeners sharing the same language background.

Escudero et al. (2009) and Morrison (2007) pointed out that a comparison among β-coefficients for predictor variables is only justified if all predictors are measured on the

same scale with the same range.[7] This was not the case in our design, where the eight peak locations were measured on the duration scale ranging between 71-418 ms into the word, and the three levels of F0 fall were measured on the frequency scale ranging between 0-100 Hz.

In general, the magnitude of the regression coefficients depends upon the scales and ranges of independent variables. For example, given the larger scale of Peak Location than F0 Fall in the Experiment 2 design, the magnitude of the β_L is going to be smaller than the magnitude of the β_F. In order to test whether the difference in the coefficient magnitude between the two factors could be attributed to the difference in the number of levels of the factors, a Monte Carlo simulation was carried out. 10,000 replications of the actual analysis were created by drawing a random sample from the binomial distribution with the same number of observations as the sample drawn from each of the participants (i.e., the data from 240 trials). This procedure was repeated twenty times per replication, thus, replicating the number of subjects in the experiment (i.e., twenty listeners per group). For each replication we ran a Wilcoxon's test on the difference between the coefficients for the two factors. The resulting *p*-values were centered on .52 suggesting that the number of levels in a factor barely influenced the coefficients' size. Furthermore, the minimum expected *p*-value was always larger than the observed *p*-value in the experiment ($p = .39 \times 10^{-3}$ and $p = .85 \times 10^{-6}$, respectively). Since 10,000 replications were run, it can be assumed that the probability of getting a *p*-

[7] Gelman and Hill (2007) propose that the logistic regression coefficients of independent variables measured on different scales can be made comparable by centering the scores on each independent variable and normalizing them by dividing the centered scores by two times the standard deviation. This procedure would be impossible in the current experiment because the eight levels of the Peak Location are on an ordinal scale (although, in principle, they can be described as 71 ms, 418 ms, etc.).

value as low as was observed in the current experiment by chance is less than 1/10000 ($p < .0001$). Therefore, the comparison among the β-coefficients of F0 Peak Location (β_L) and F0 fall (β_F) for each listener group was justified in this experiment.

4.3. Results

4.3.1. Group Differences

An ANOVA on Syllable 1 Prominent responses by Japanese listeners revealed a significant effect of Peak Location [$F(7, 456) = 33.99$, $p < .001$, $\eta_p^2 = .34$], F0 Fall [$F(2, 456) = 57.84$, $p < .001$, $\eta_p^2 = .20$], and interaction between them [$F(14, 456) = 4.68$, $p < .001$, $\eta_p^2 = .13$]. The results indicated what is evident in Figure 4.2a; namely, the F0 fall affected responses when the F0 peak was located earlier in the word, but not when it was located later in the word. The same analysis on the English data revealed a significant effect of Peak Location only [$F(7, 456) = 11.86$, $p < .001$, $\eta_p^2 = .15$]. English listeners did not appear to be sensitive to the F0 fall, which can be inferred from the overlap of solid red, dotted blue, and dashed green lines in Figure 4.2b.

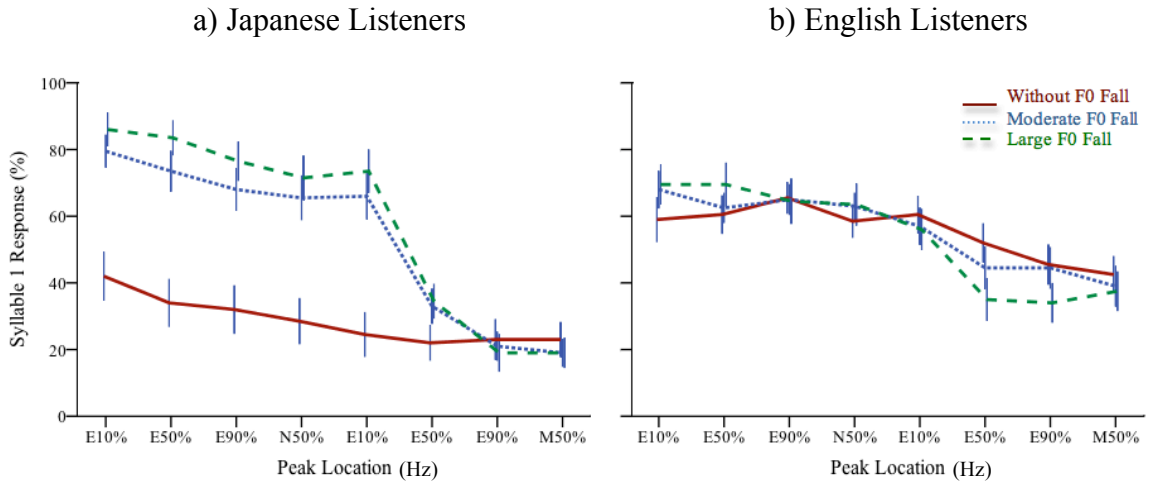


Figure 4.2. Syllable 1 Prominent responses as a function of F0 peak location and the magnitude of the F0 fall. Error bars indicate ± 1 *SE*.

The effect of F0 Fall in the Japanese group was examined further by pair-wise comparisons of each level of an ordinal variable to the mean of the subsequent levels (i.e., Helmert contrasts with Bonferroni corrections). A comparison of responses to the F0 contours without F0 fall (solid red lines in Figure 4.2a) to the mean of responses to the F0 contours with the moderate and large F0 falls (dotted blue and dashed green lines in Figure 4.2a) yielded a significant result, [$t(456) = -2.69$, $p < .001$]. The difference between dotted and dashed lines was not significant, [$t(456) = -0.48$, $p = .105$].

The effect of Peak Location in both groups was also examined further by pair-wise comparisons of each peak location to the following one (i.e., repeated contrasts with Bonferroni corrections). Among seven comparisons in the Japanese data, a significant difference between adjacent locations of F0 peaks was observed only between the F0 peak aligned 10% and 50% into the *second* vowel [e], [$t(456) = 2.48$, $p < .001$]. A similar result was obtained in the English data, [$t(456) = 1.40$, $p = .003$].

Figure 4.3 offers another illustration of the group difference in categorization of the stimulus space. Overall, the Japanese listeners were more consistent in their responses than the English listeners, which is evident from the larger circle sizes in Figure 4.3a as compared to Figure 4.3b. For the Japanese listeners, the boundary between the 1st-syllable prominent and 2nd-syllable prominent categories was determined by variation in both Peak Location and F0 Fall. The most consistent perception of the first syllable as prominent (filled circles) occurred when the F0 peak was in the first syllable, and it was followed by the F0 fall. The most consistent perception of the second syllable as prominent (dotted circles) occurred at later peak locations (i.e., $E_{50\%}$, $E_{90\%}$, and $M_{50\%}$)

regardless of the F0 fall. In contrast, the English listeners did not show much agreement in their answers, because for any stimulus type in the experiment (i.e., a circle in Figure 4.3b) their averaged response did not exceed 75%. Only the location of F0 peak influenced their judgments.

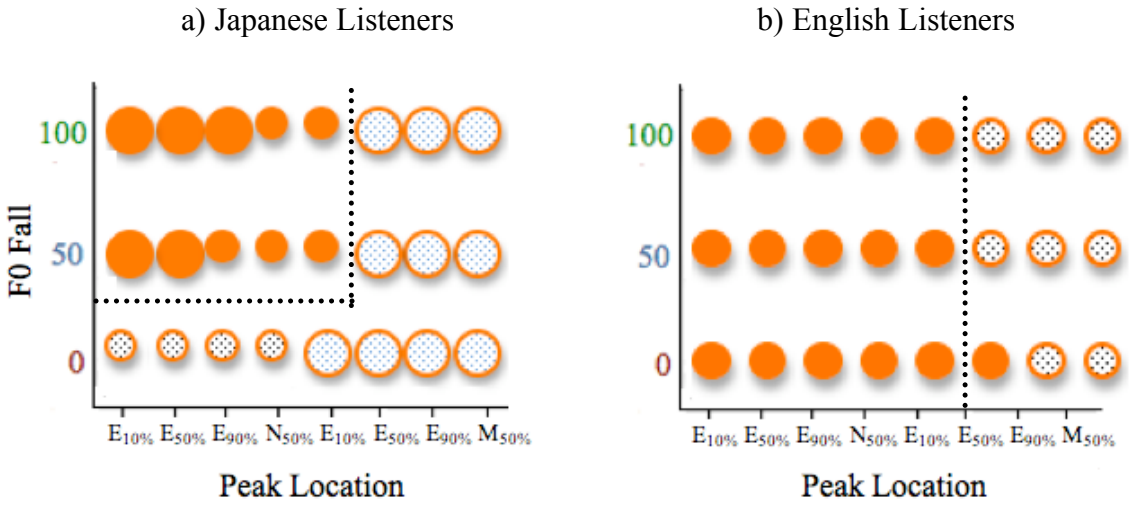


Figure 4.3. Above chance responses in the 2AFC task by group (Syllable 1 Prominent: filled circles; Syllable 2 Prominent: dotted circles). Smaller circles indicate 50-75% of responses; larger circles indicate 75-100% of responses.

The boundary between the two response categories along the *y*-axis dividing the stimuli in a larger 1st-syllable prominent and a smaller 2nd-syllable prominent spaces suggests that both groups perceived the 1st syllable [ne] as prominent even when the F0 peak was located in the 2nd syllable [ne]. This can be interpreted as an expectation of the peak delay. The boundaries between the two response categories in the Japanese data also show an interaction between the two acoustic parameters, creating a more complex concept of prominence. In the bottom circle row for the stimuli without F0 Fall in Figure 4.3a, a comparison of the smaller dotted circles at earlier peak locations and the larger dotted circles at later peak locations suggests that the Japanese listeners could have

categorized the earlier location stimuli differently, if they had another answer option such as No Prominence (i.e., unaccented words).

4.3.2. Individual Differences

Forty individual logistic regression analyses on the log odds of Syllable 1 Prominent responses were conducted to examine within-group individual differences in employing the peak location (L) and F0 fall (F) information during the categorization task. On average, the full regression model explained a similar amount of variance in answers of the Japanese listeners (M = 72%, SD = 23) and the English listeners (M = 69%, SD = 19). Within-listener comparisons of the full model (with independent variables) and the null model (without independent variables) revealed that the inclusion of the predictors significantly improved the fit of the model for 19 out of 20 Japanese listeners and 14 out of 20 English listeners ($p < .05$). Thus, the decisions of 95% of the Japanese listeners and 70% of the English listeners were influenced by at least one manipulated parameter.

To investigate which predictors influenced the answer probability, β-coefficients obtained from 19 Japanese and 14 English full individual regression analyses were examined. A significant β-coefficient showed that a predictor had a unique contribution to the explanation of the variance in answers by the full model $\alpha + \beta_L$ x L+ β_F x F + β_{LF} x LF. As Table 4.2 summarizes, for the majority (ten) of the Japanese listeners, F0 Fall and the interaction between F0 Fall and Peak Location had the best predictive power in a full regression model. The regression analyses of these listeners yielded an average Nagelkerke R^2 = .83, showing that the full model explained 83% of variance in the answers. A larger Nagelkerke R^2 was yielded only for two Japanese listeners who, in

addition, relied on Peak Location. As for English listeners, no pattern in significant β-coefficients emerged. The largest Nagelkerke R^2 explained 90% of variance in the answers of two English listeners, who relied mostly on the Peak Location information.

In sum, it appears that the probability of Syllable 1 Prominent responses by Japanese listeners was best explained in terms of F0 Fall and its interaction with peak location. The best predictors of responses in the English group varied across individual listeners.

Table 4.2

Within-Group Summary of Significant β-Coefficients in Individual Logistic Regressions

Significant predictor	Japanese (N = 19)		English (N = 14)	
	Listeners (#)	Variance explained % (*SD*)	Listeners (#)	Variance explained % (*SD*)
None	2	22 (19)	1	29
β_L	2	62 (22)	2	90 (7)
β_F	2	48 (15)	3	73 (20)
$\beta_L + \beta_F$	0		1	69
β_{LF}	1	73	2	31 (21)
$\beta_L + \beta_{LF}$	0		0	
$B_F + \beta_{LF}$	**10**	**83 (15)**	2	57 (11)
$\beta_L + \beta_F + \beta_{LF}$	2	97 (4)	3	77 (5)

Note. The summary includes only listeners, for whom the full regression model, Log(odds(Syllable 1 Prominent)) = $\alpha + \beta_L x\ L + \beta_F x\ F + \beta_{LF} x\ LF$, was a significantly better fit than the null model. The threshold of significance is $p < .05$; L = Peak Location; F = F0 Fall.

The β-coefficients obtained from forty individual logistic regressions without the interaction term (i.e., $\alpha + \beta_L x\ L + \beta_F x\ F$) on the log odds of Syllable 1 Prominent

responses are plotted in Figure 4.4. The β-coefficients are taken to indicate the relative reliance of listeners on the two acoustic cues – F0 Peak Location and F0 Fall (Morrison, 2005). Only the coefficients averaged across the five earlier peak locations are shown in Figure 4.4, because F0 fall did not influence the listeners' responses for the three later peaks (see Figures 4.2 and 4.3), and because the overall reliance on F0 fall and peak location was impossible to interpret in presence of a significant interaction between these cues.

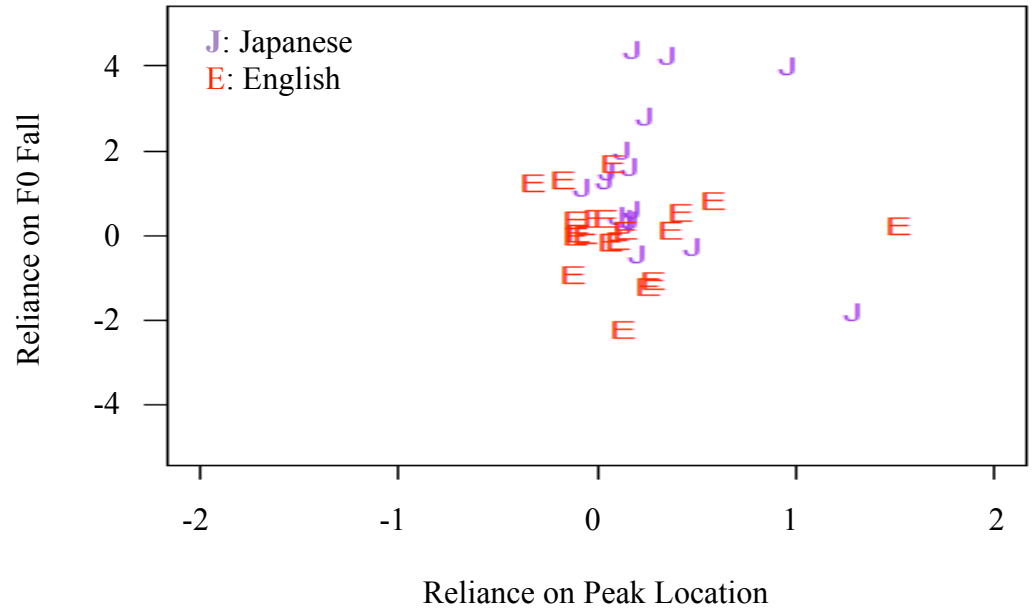


Figure 4.4. Reliance of individual listeners on F0 fall and peak location in categorization of the stimuli. The reliance is indicated by β-coefficients from forty individual logistic regressions, Log(odds(Syllable 1 Prominent)) = $\alpha + \beta_L x\ L + \beta_F x\ F$, for five earlier peak locations in *nenema* (ranging from $E_{10\%}$ in the first vowel to $E_{10\%}$ in the second vowel).

Positive *x*-values in Figure 4.4 indicate that the earlier the F0 peak, the greater the probability of choosing Syllable 1 as prominent. Negative *x*-values indicate that the later the peak, the greater the probability of choosing Syllable 1 as prominent. The figure

suggests that for the first half of the *nenema* word, both Japanese and English listeners did not differ substantially in their reliance on the Peak Location. Namely, they tended to associate an earlier alignment of the F0 peak with Syllable 1 prominence, especially, one Japanese and one English listener, whose β_L coefficients were larger than 1. It also appears that there were slightly more negative β_L coefficients in English listeners than in Japanese listeners, which may be interpreted as individual expectation of a greater peak delay by English listeners. These listeners might have expected the F0 peak, associated with the prominence of the first syllable, to occur after the first syllable's boundary.

Likewise, positive *y*-values in Figure 4.4 indicate that the larger the F0 fall (its fall and slope), the greater the probability of choosing Syllable 1 as prominent. This cue is the primary cue to the Japanese pitch accent. Negative *y*-values indicate a reverse prediction. The figure suggests that more Japanese listeners relied on the F0 fall than English listeners. The data point distribution in the figure shows that some Japanese listeners were exceptional in their reliance on the F0 fall cue. In particular, the *y*-values for five Japanese listeners look substantially larger than the *y*-values of other listeners in the same group. This raised a question of whether this ¼ of the Japanese listeners drove the group mean. A skewed distribution of the β_F-coefficients would support this hypothesis.

Figure 4.5 below illustrates that the distribution of β_F-coefficients for the F0 fall cue was, indeed, highly skewed in the Japanese data, but not in the English data. This indicates that the quarter of the Japanese listeners was the driving force behind the big between-group difference in the effect of the F0 fall on the perception of prominence.

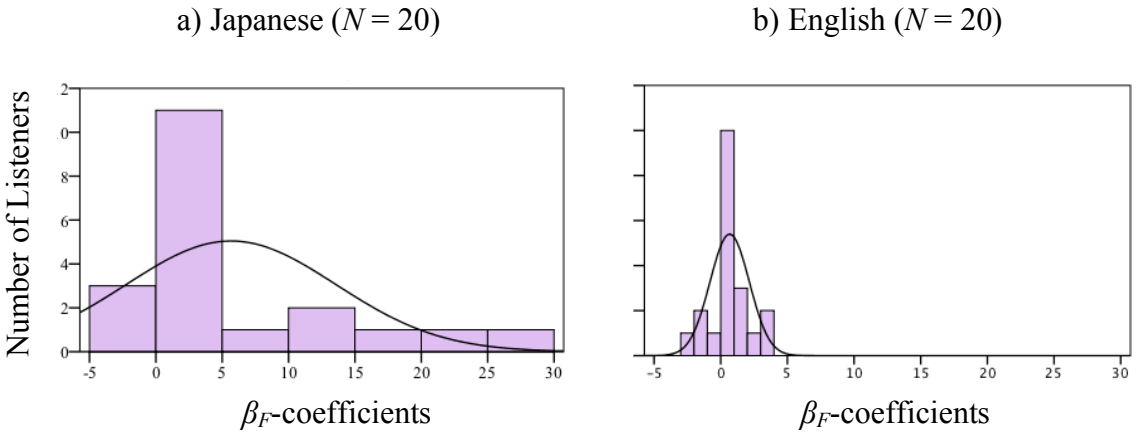


Figure 4.5. By group distributions of β_F-coefficients indicating the reliance of listeners on F0 fall in their prominence judgments.

4.4. Discussion

4.4.1. General Conclusions

Consistent with the previous experiment, the results of Experiment 2 supported the prediction that native-language experience influences the perception of prosodic prominence. The Japanese listeners were more sensitive to the variations in F0 than the English listeners, as indicated by a) the effect size of manipulated parameters, that is, the Peak Location and F0 Fall, on listeners' prominence judgments; b) the significant interaction between the parameters resulting in a complex response pattern; and c) a larger number of individual listeners whose judgments were influenced by at least one parameter.

For both Japanese and English groups, the boundary between the 1st syllable prominent and 2nd syllable prominent responses lay in the beginning of the second [e] vowel of the *nenema* word. More specifically, the shift between Syllable 1 Prominent and Syllable 2 Prominent responses was observed between 10% alignment and 50%

alignment of the peak into the vowel. This result suggests a similar expectation of peak delay in CVCVCV words for both Japanese and English listeners on average, at least in this two-way categorization task. This finding is in accordance with the reports of peak delay in the production of pitch patterns by native Japanese speakers (Ishihara, 2006; Neustupný, 1966; Shport & Guion, 2008; Sugito, 1980, 1982a), and reports of compensation for peak delay in perception (Sugito, 1980). The compensation means here that the percept of the 1st syllable as prominent is achieved even when the F0 peak and a subsequent F0 fall align as late as approximately the middle of the following syllable. As for the English listeners, the prediction that the categorical boundary between the 1st syllable prominent and 2nd syllable prominent patterns may occur earlier in the word than for the Japanese listeners was not borne out.

In both experiments, the Japanese and English listeners were similar in their reliance on the F0 peak location cue; however, they differed dramatically in their reliance on the F0 fall cue. A complex response pattern of the Japanese listeners with regard to the stimulus space in Figure 4.3a suggests that in addition to the Syllable 1 Prominent and Syllable 2 Prominent response options, the Japanese listeners could have used a No Prominence response option for categorization of the stimuli (see also Chapter II, section 2.4.1. for the description of the acoustic cues to Japanese unaccented pitch patterns). Experiment 3 in the following chapter tested this hypothesis.

4.4.2. F0 Fall and Cue Interaction

For the majority of the Japanese listeners, the F0 fall, especially in its interaction with an earlier peak location, facilitated the judgment of prominence location on the 1st syllable. There was no effect of the F0 Fall in the 2nd syllable. This result replicated the

finding of the previous experiment (Chapter III). Namely, for Japanese listeners' perception of prominence in three-syllable words, the F0 fall matters only when the F0 maximum occurs relatively early in the word, regardless of the slope of the F0 fall as determined by the fall magnitude (compare the large fall series and moderate fall series in Figure 4.2a).

It is unclear why the F0 fall did not affect the prominence judgments of the Japanese listeners with later alignment of the F0 peak. It was suggested in the discussion of Experiment 1 that this result might be due to a time limitation on the processing. Namely, the F0 fall might have mattered only when the F0 peak was located in the 1st syllable, because the listeners had more time to process it. In Experiment 2, though, the time course of the F0 fall within each series of stimuli was controlled: The fall always occurred over 191 ms after all peak locations. Therefore, the interpretation based on the time limitation in processing (and just noticeable differences in the F0 fall) has to be abandoned.

An alternative plausible interpretation of the lack of the F0 fall effect in the second syllable is that the interaction between the acoustic cues occurred due to the expectation of F0 peak delay. Namely, the F0 fall might not have mattered when the peak was located in the 2nd CV syllable, because the listeners would have expected a delayed peak for the percept of prominence in that syllable (Ishihara, 2006; Neustupný, 1966; Shport & Guion, 2008; Sugito, 1980, 1982a). Since the latest F0 peak location was at the midpoint of the [m] segment, which followed the second syllable in *nenema*, the hypothesis that the F0 fall would matter at later peak locations, such as the last vowel [a], could not be tested.

Another alternative explanation is that once Japanese listeners heard F0 fall in the first syllable (and chose the Syllable 1 Prominent answer), they did not need to attend to this information in the second syllable in the 2AFC task. Listeners may pay attention to the F0 fall in the second syllable, if the third answer option such as No Prominence is introduced to the task. This hypothesis was tested in Experiment 3. In the context of the 3AFC task reported in the next chapter, Japanese listeners are forced to keep listening for the F0 fall, even if they have not heard it in the first syllable, in order to make a choice between Syllable 2 Prominent and No Prominence answers. These two categories represent the 2nd-syllable accented and unaccented pitch patterns, which are modeled in this dissertation as F0 contours with and without F0 fall in the 2nd syllable, respectively (Chapter II, section 2.4.1).

As compared to Experiment 1, the new finding in Experiment 2 was that with gradient manipulation of F0 peak alignment the effect of the F0 fall changed. The prediction of a greater F0 fall effect when the F0 maximum was aligned earlier in a word was supported by the data. The effect of the F0 fall itself, however, appeared to be not gradient, but categorical for Japanese listeners. Although Figure 4.2 suggests that the percentage of Syllable 1 Prominent responses slightly decreased from the peak location at 10% in the first [e] to the peak location of 10% in the second [e], these slight differences did not reach significance in the Japanese or English data.

4.4.3. Group Differences and Individual Variation

Within-group individual variation is always to be expected, especially, for non-native listeners of a foreign language. In Experiment 2, for example, the English listeners varied in their reliance on the F0 fall and peak location cues, however, no pattern in this

variation emerged (Table 4.2). A substantial variation in performance was observed also in native Japanese listeners, and this pattern can be summarized as follows: The majority of native listeners relied on F0 fall and its interaction with peak location. Furthermore, a quarter of the listeners had an extremely high reliance on F0 fall, so that they caused a positively skewed distribution among the Japanese listeners. These exceptional listeners also appeared to drive the magnitude of the between-group difference. A large dispersion and skeweness in the reliance on the F0 fall by Japanese listeners explains why some group means look similar, for example, the average amount of variance explained by the model with the two predictors (M = 72%, SD = 23 for the Japanese and M = 69%, SD = 19 for the English listeners).

CHAPTER V

EXPERIMENT 3: 3AFC CATEGORIZATION WITH TWO INDEPENDENT VARIABLES

5.1. Introduction

Experiment 3 was designed to investigate whether the availability of three answer options instead of two would affect listeners' categorization patterns and their reliance on the acoustic cues of the F0 peak location and F0 fall. In two previous experiments, the F0 fall cue was available to listeners, but in principle listeners could have used only the F0 peak location cue or only the F0 fall cue in the first syllable in order to distinguish between the categories of 1st-syllable prominence and 2nd-sylable prominence. The question addressed in Experiment 3 was whether Japanese and English listeners were more sensitive to the F0 fall cue at later peak locations in a word, when they had to consider the unaccented pattern as a possible response option in the 3AFC task.

In all perception experiments in this dissertation, the F0 contours of the stimuli were resynthesized to represent three prosodic categories – 1st-syllable accented, 2nd-syllable accented, and unaccented patterns. The two-way categorization task (2AFC) in Experiments 1 and 2 was utilized in consideration of native English listeners. In English, prominence (a stress accent attracting a pitch accent) is obligatory for any word uttered in isolation. Therefore, for English listeners the unaccented category would have been meaningless at both the word- and phrase-levels, and their responses were not expected to vary in the 2AFC and 3AFC tasks.

In contrast, the three-way categorization task (3AFC) would be more appropriate for native Japanese listeners, since it involves three naturally occurring pitch patterns in

the Japanese language. In Experiments 1 and 2, Japanese listeners were forced to choose between the answers Syllable 1 Prominent and Syllable 2 Prominent, even though they could have perceived the unaccented contours (without the F0 fall) as not fitting either of these categories. In the current experiment, the response patterns of Japanese listeners were expected to be altered, when a third answer option representing the unaccented pitch pattern was introduced into the categorization task.

5.1.1. Two-Way Versus Three-Way Categorization in Kitahara (2001)

In Kitahara's dissertation on perception of Japanese lexical pitch patterns by native speakers, both 2AFC and 3AFC categorization tasks were conducted. As the only such study, to our knowledge, on Japanese prosody, it is discussed in some details here. Kitahara has argued that a three-way categorization task is harder for Japanese listeners than a two-way categorization task for two reasons. First, more categories are involved in decision-making. Second, listeners have to detect both the presence of the accent (i.e., accented vs. unaccented distinction), and the location of the accent (1^{st}- vs. 2^{nd}-syllable accented distinction). Although the current study does not have the objective of distinguishing between accent-presence detection and accent-location detection, these two terms are useful in referring to two processes that may be independent or integrated during pattern categorization by listeners.

In his experiments, Kitahara resynthesized F0 patterns in real Japanese words (e.g., *háshi* 'chopsticks', *hashí* 'bridge', *hashi* 'edge') and embedded them in the carrier sentence *Ano __ kara* 'From that __.' Listeners were asked to identify a word from a possible duplet or triplet set. Comparing pair-wise responses in the 2AFC task and triplet-wise responses in the 3AFC task, Kitahara found that a larger magnitude of the F0 fall

was required for listeners to distinguish unaccented patterns from accented during the three-way categorization (e.g., p. 77). Based on these results, he inferred that the sensitivity to the accent presence was decreased in a 3AFC task as compared to the 2AFC task. He followed to conclude that between accent-presence detection and accent-location detection, the latter process is more fundamental. Furthermore, these two processes are independent, and possibly interfering with each other.

Another interpretation of Kitahara's results is plausible, however. The larger F0 fall magnitude for unaccented-accented distinction in the 3AFC task might have been driven by the *interaction* between the F0 fall and peak location. Experiments 1 and 2 in this dissertation showed that Japanese listeners' sensitivity to the F0 fall varies with F0 peak location. In the 3AFC task, the interaction of the cues may be even more important due to similarity of the F0 contours between the unaccented and 2nd-syllable accented patterns, and due to substantial F0 peak delay in the 1st-syllable accented patterns. Among other results, Kitahara reported that in the 3AFC task listeners had more uncertainty distinguishing between the unaccented and 2nd-syllable accented patterns with the late F0 peak alignment, which is possibly explained by acoustic similarity of the two patterns. Furthermore, the distinction between the unaccented and 1st-syllable accented categories was also less sharp in the 3AFC task. He noted that this latter result was in line with interpretation that both accent-location detection and accent-presence detection were processed in an integrated way (p. 77). In terms of stimulus manipulation in this dissertation, this suggests a larger effect of the cue interaction in the three-way than in the two-way categorization task.

5.1.2. Categorical Boundaries

The degree of interaction between the F0 fall and peak location may also affect the sharpness of the categorical boundary, resulting in 'grey areas' between categories. The boundary sharpness is defined here as consistency of listeners' categorization patterns, where more consistent responses result in a sharper boundary. For example, in Figure 4.3 of the previous chapter, the boundary between the 1st-syllable prominent and 2nd-syllable prominent responses of the Japanese listeners appears to be sharper than for the English listeners as indicated by the percentage of responses (circle sizes) in the figure. This observation is in line with many previous findings that between-category boundaries are more categorical (sharp, crisp) in native listeners as compared to non-native listeners.

A technical assessment of boundary crispness has been suggested by Morrison (2007) and further applied by Escudero et al. (2009, pp. 457-458). The boundary crispness may be conceptualized as "the rate of change from one category to the other in the direction perpendicular to the orientation of the boundary" in the stimulus space (Morrison, 2007, p. 232). The proposed polar-coordinate magnitude measure assesses boundary crispness using β-coefficients obtained from a logistic regression analysis (see this analysis in Chapter III). For example, we could have calculated this measure for every listener in Experiment 2 by using the formula $\sqrt{(\beta_L^2 + \beta_F^2)}$. The data in Experiment 3, however, were not binomial. The dependent variable consisted of three response types; therefore, a binary logistic regression could not be conducted, unless the data were split for pair-wise comparisons (e.g., 2nd-syllable accented vs. unaccented patterns, etc.). Since no straightforward comparison of category crispness between Experiments 2 and 3 was

possible, the polar-coordinate magnitude measure was not used. The terms *sharpness* and *distinctiveness* are used here to avoid the confusion with the boundary *crispness*, as defined by Morrison (2007).

5.1.3. Predictions

Experiment 3 did not attempt to distinguish between the processes of accent-location detection and accent-presence detection. Neither was this experiment designed to determine the difficulty of the 3AFC task as compared to the 2AFC task. The goal of Experiment 3 was to investigate whether the presence of three answer options encouraged higher sensitivity to the manipulated acoustic cues (especially, F0 fall) and changed the distinctiveness of response patterns. From a relative distinctiveness of response patterns in the two types of tasks with the same stimuli space, we may infer the task effect on perceptual categorization of F0 contours. Another goal of Experiment 3 was to replicate the results of Experiment 2, which would strengthen our interpretation of listeners' sensitivity to the acoustic cues of the lexical pitch accent.

The results of the previous experiments suggested that Japanese listeners either could not categorize the stimuli without F0 fall as having prominence on the first or second syllable (Experiment 1), or tended to perceive them as being more similar to the 2nd-syllable accented than the 1st-syllable accented category (Experiments 1 and 2). Thus, it still remains unclear whether native listeners can distinguish between the 2nd-syllable accented and unaccented patterns in our experimental paradigm. English listeners tended to be insensitive to the F0 fall suggesting that they did not perceive F0 patterns without the fall as an independent prosodic category.

Based on previous findings, it was predicted that the group difference between Japanese and English listeners in reliance on the F0 fall and F0 peak location would be similar to those in Experiments 1 and 2. It would be especially evident with regard to the new answer option representing the unaccented category, which is unfamiliar to English listeners. Japanese listeners would categorize the stimuli with no F0 fall using this third answer option. It was also hypothesized that for Japanese listeners the interaction between the two acoustic cues would be larger in the 3AFC task than in the 2AFC task, and the category boundaries would be less sharp due to phonetic similarity between the 2nd-syllable accented and unaccented patterns and due to expectation of F0 peak delay.

5.2. Method

5.2.1. Participants

Five Japanese and five English listeners participated in this experiment (Table 5.1). General characteristics of the participants were similar to the participants from

Table 5.1

Listeners' Background Information: Experiment 3 (3AFC Categorization)

	Japanese (n = 5)	English (n = 5)
Dialect	Tokyo Japanese	West Coast American English
Age	19 (M_{year} = 19)	19-22 (M_{year} = 20)
Sex	5 females	2 females, 3 males
Languages learned as adults for more than a year	English, German, Mandarin (M_{number} = 1.6; M_{year} = 6.8)	Spanish, Italian (M_{number} = 1; M_{year} = 4.3)
Life-abroad experience	AoA: 18-19 (M_{year} = 18.8) LoR: 4 (M_{month} = 4)	

Note. AoA = age of arrival to the U.S.; LoR = length of residence in the U.S.

Experiments 1-2 (as detailed in the Method section of Chapter II). Average background of both groups is given in a table above.

5.2.2. Materials

Test materials were identical to those used in Experiment 2 (Chapter IV, Figure 4.1.). For the stimuli without F0 fall, peak location was defined as the end of the initial F0 rise, that is, a point of change in F0 from the rise to a flat contour.

5.2.3. Procedure

Each stimulus was repeated ten times resulting in the total number of trials equal to 240. The listener's task was to decide for every F0-contour variation of the *nenema* word whether the first syllable sounded prominent, the second syllable sounded prominent, or none of the syllables in the word sounded prominent. The latter response type was thought to represent the unaccented pitch pattern. In principle, listeners might have perceived the third syllable as prominent. However, since the category of the 3rd-syllable accented word (with the F0 fall after the third syllable) could not be represented in the stimuli due to the absence of segmental material over which the F0 fall would have been realized, only three answer options were given to the listeners.

Differently from the instruction in Experiments 1-2, the Japanese listeners were given examples of real 1st-syllable accented, 2nd-syllable accented and unaccented words as representative of the answer options Syllable 1 Prominent, Syllable 2 Prominent, and No Prominence. These examples were given in the hope of facilitating categorization of the stimuli in terms of Japanese pitch-accent contrast. Since English listeners were naïve to Japanese, no language examples were provided for them. During debriefing, none of

the Japanese or English listeners mentioned that they wished they had had the 3rd-syllable prominent answer option.

5.2.4. Analysis

The percentages of three response types (i.e., Syllable 1 Prominent, Syllable 2 Prominent, No Prominence) to ten repetitions of each stimulus were calculated for each listener. Then, the data were normalized using arcsine transformation (e.g., Francis et al., 2003; Lee, Tao, & Bond, 2009). Because the effect of Group was consistent in Experiments 1 and 2, the Japanese and English data were analyzed separately in the current experiment. To investigate the effect of manipulated variables on categorization of the F0 contours in each group, three separate repeated measures ANOVAs were conducted on Syllable 1 Prominent, Syllable 2 Prominent, and No Prominence responses (arcsine transformed), with within-subject factors of Peak Location and F0 Fall. Significant main effects are not reported in presence of significant interaction between the factors, which were further explored in pair-wise comparisons with Bonferroni corrections – repeated contrasts for the Peak Location (seven comparisons between adjacent peak locations), and Helmert contrasts for the F0 Fall (two comparisons between the unaccented and accented, and between the two accented patterns). The results of these analyses shall be interpreted with caution given that each group consisted of only five participants.

The distinctiveness of the category boundaries was conceptualized as percentage of similar responses patterns to particular F0 contours. Low percentage of responses to a stimulus pattern (weak response pattern) was interpreted as less sharp category boundary, and vice versa.

5.3. Results

5.3.1. Japanese Listeners

Figure 5.1 illustrates Japanese listeners' categorization of the stimuli by response type. The repeated measure ANOVAs on the Japanese data yielded significant interactions between Peak Location and F0 Fall on Syllable 1 Prominent responses [$F(14, 56) = 13.80$, $p < .001$, $\eta_p^2 = .78$]; Syllable 2 Prominent responses [$F(14, 56) = 6.32$, $p < .001$, $\eta_p^2 = .61$]; and No Prominence responses [$F(14, 56) = 6.79$, $p < .001$, $\eta_p^2 = .63$]. The interactions were further examined in the tests of within-subject contrasts.

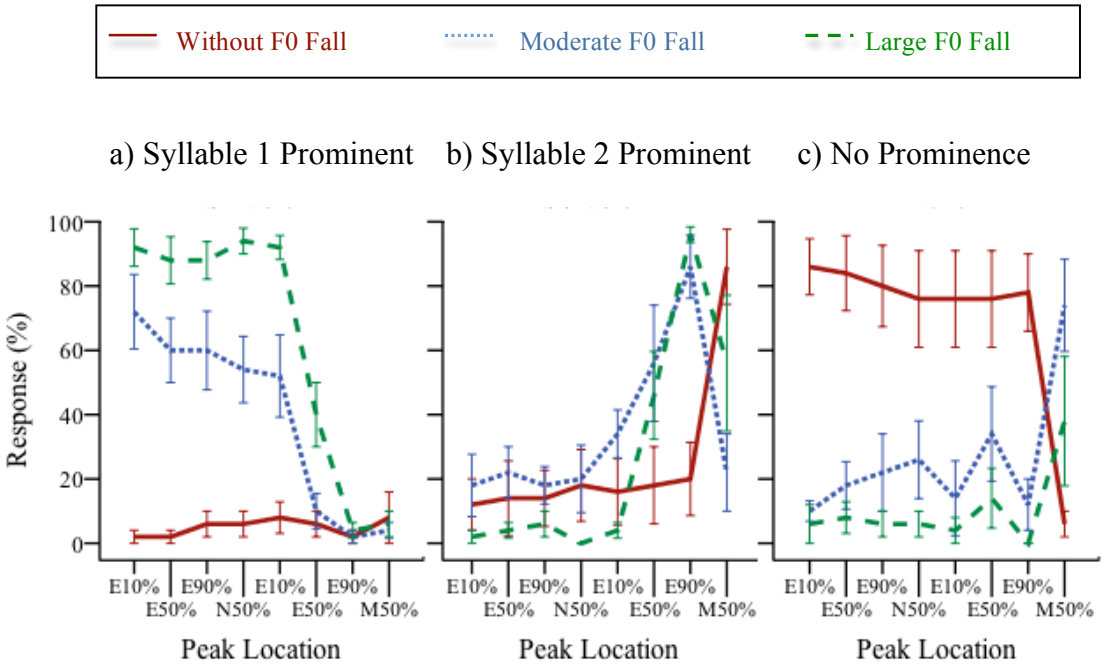


Figure 5.1. Three-way categorization of F0 patterns by Japanese listeners as a function of F0 peak location and the magnitude of the F0 fall. Error bars indicate ± 1 *SE*.

For Syllable 1 Prominent responses, the comparison between the unaccented and accented patterns yielded a significant result at the level of the peak locations $E_{10\%}$ and $E_{50\%}$ in the second syllable [ne], [$F(1, 4) = 39.38$, $p = .003$, $\eta_p^2 = .91$]; the comparison

between the two accented patterns was significant between the $E_{10\%}$ and $E_{50\%}$, [$F(1, 4) = 7.88$, $p = .048$, $\eta_p^2 = .66$], as well as between $E_{50\%}$ and $E_{90\%}$, [$F(1, 4) = 10.96$, $p = .030$, $\eta_p^2 = .73$]. These results, as displayed in Figure 5.1a, indicated that the Japanese listeners categorized the stimuli with the F0 fall (dashed green and dotted blue contours) as having prominence on the first syllable more than the stimuli without F0 fall (solid red contours) for the peak locations from the beginning of the word until the beginning of the second vowel [e]. After that temporal point, no differences among the three series of the stimuli were perceived as related to prominence on the first syllable. In addition, the stimuli with a large fall (dashed green contours) were perceived more as having prominence on the first syllable than the stimuli with a moderate fall (dotted blue contours) until the F0 peak was located in the second vowel [e]. In sum, replicating the results in Experiment 2 (Figure 4.2a), the larger F0 fall conditioned more Syllable 1 responses than the smaller fall in the first half of the three-syllable word.

For Syllable 2 Prominent responses, the comparison between the unaccented and accented patterns was significant between the peak locations of $E_{50\%}$ and $E_{90\%}$ of the second syllable [ne], [$F(1, 4) = 61.84$, $p = .001$, $\eta_p^2 = .94$], and between $E_{90\%}$ and $M_{50\%}$, [$F(1, 4) = 21.25$, $p = .010$, $\eta_p^2 = .84$]. The comparisons between the two accented patterns were not significant at any peak location for this response type. These results, as displayed in Figure 5.1b, indicated that the Japanese listeners perceived stimuli with the F0 fall as having prominence on the second syllable more than the stimuli without F0 fall, when the F0 peak was located at the end of the second vowel [e], or later. The magnitude of the F0 fall did not matter for perception of prominence on the second syllable of the three-syllable word.

For No Prominence responses, the only comparison that reached significance was the comparison between the unaccented and accented patterns between the peak locations of $E_{90\%}$ and $M_{50\%}$, [$F(1, 4) = 14.57$, $p = .019$, $\eta_p^2 = .79$]. This result, as displayed in Figure 5.1c, indicated that the absence of the F0 fall for the most part of the three-syllable word was perceived by listeners as the absence of prominence. The magnitude of the F0 fall did not affect the listeners' responses.

5.3.2. English Listeners

Figure 5.2 illustrates English listeners' categorization of the stimuli by response type. The repeated measures ANOVAs on the English data yielded marginally significant interactions between Peak Location and F0 Fall for Syllable 1 Prominent responses [$F(14, 56) = 1.82$, $p = .058$, $\eta_p^2 = .31$], and No Prominence responses [$F(14, 56) = 1.83$, $p = .057$, $\eta_p^2 = .31$]. The effect of Peak Location was significant for Syllable 1 Prominent responses [$F(7, 28) = 3.08$, $p = .016$, $\eta_p^2 = .44$], and Syllable 2 Prominent responses [$F(7, 28) = 3.45$, $p = .009$, $\eta_p^2 = .46$]. The effect of F0 fall was not significant, which is evident from the overlapping dashed green and dotted blue contours in Figure 5.2.

Marginally significant interactions were not further statistically analyzed, although Figure 5.2a suggests that, similar to the Japanese listeners, the F0 fall (dashed green and dotted blue contours) may condition Syllable 1 responses in the English listeners, when the F0 peak is in the first half of the word. Figure 5.3c also suggests that the English listeners may tend to categorize the stimuli without the F0 fall as having no prominence, when the F0 peak was in the first half of the word.

For Syllable 1 Prominent responses, the pair-wise comparisons for the main effect of Peak Location on the perceived patterns of prominence (i.e., repeated contrasts)

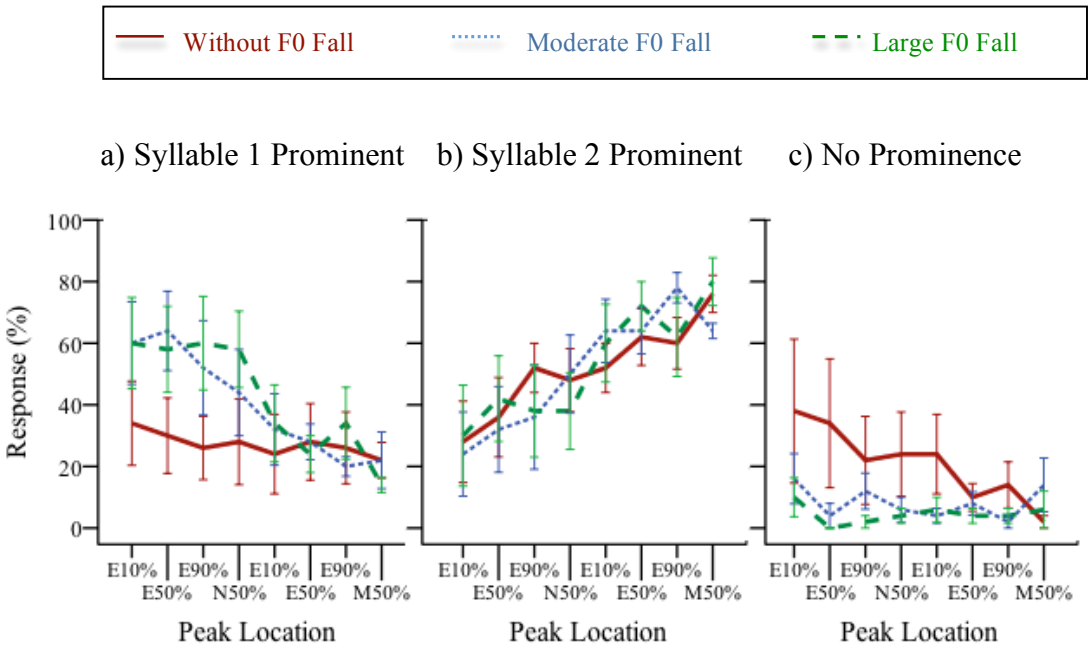


Figure 5.2. Three-way categorization of F0 patterns by English listeners as a function of F0 peak location and the magnitude of the F0 fall. Error bars indicate ± 1 *SE*.

yielded a significant result for the comparison between $N_{50\%}$ and $E_{10\%}$, [$F(1, 4) = 10.97$, $p = .030$, $\eta_p^2 = .73$]. Similarly, the pair-wise tests on the *Syllable 2 Prominent* responses yielded a significant effect of the Peak Location for the comparison between $N_{50\%}$ and $E_{10\%}$, [$F(1, 4) = 16.86$, $p = .015$, $\eta_p^2 = .81$]. Taken together, these results indicated, as displayed in Figures 5.2a and 5.2b, that the English listeners perceived the stimuli as having prominence on the first syllable, when the F0 peak was located anywhere until the beginning of the second vowel [e]. The stimuli with later peak locations were heard as having prominence on the second syllable.

The results described above also suggest that the categorical boundary between Syllable 1 Prominent and Syllable 2 Prominent patterns may be located slightly later in the word for the Japanese listeners than for the English listeners. Figures 5.1a and 5.1b

suggest that in the Japanese data, the boundary was associated with the first half of the vowel [e] in the second syllable of *nenema* (i.e., $E_{10\%}$ and/or $E_{50\%}$). Figures 5.2a and 5.2b suggest that in the English data, the boundary might have been associated with the beginning of the second syllable, around the midpoint of the second [n] ($N_{50\%}$). However, the change in the English listeners' responses as a function of peak location appeared to be more gradual (i.e., less categorical) than in the Japanese listeners' responses.

Lastly, it is worthwhile noticing in Figure 5.2 that the percentage of No Prominence responses is in general much lower that the percentage of Syllable 1 and Syllable 2 Prominent responses. This may indicate that the English listeners were not comfortable with the No Prominence category, and hence, by our assumption, with the unaccented category.

Another illustration of group differences between the Japanese and English listeners in categorizing the same stimuli by the means of the three responses is presented in Figure 5.3. First, the vertical boundary between Syllable 1 Prominent response space and Syllable 2 Prominent response space did not change for Japanese listeners as compared to Experiment 2 – it was still between the $E_{10\%}$ and $E_{50\%}$ peak locations in the second vowel [e]. In contrast, this boundary became less clear for English listeners. Second, the horizontal boundary in Figure 5.3a clearly separates stimuli without the F0 fall from stimuli with the fall (except for the last peak location) in Japanese response space. In English response space (Figure 5.3b), this boundary cannot be drawn in the absence of preferred No Prominence responses for any of the stimuli. Last but not least, there were clusters of mixed responses, larger in the English response space than in the Japanese one. These grey areas indicate stimuli space for which no strong preference in

responses was apparent, that is, the stimuli that yielded the most inconsistent response patterns.

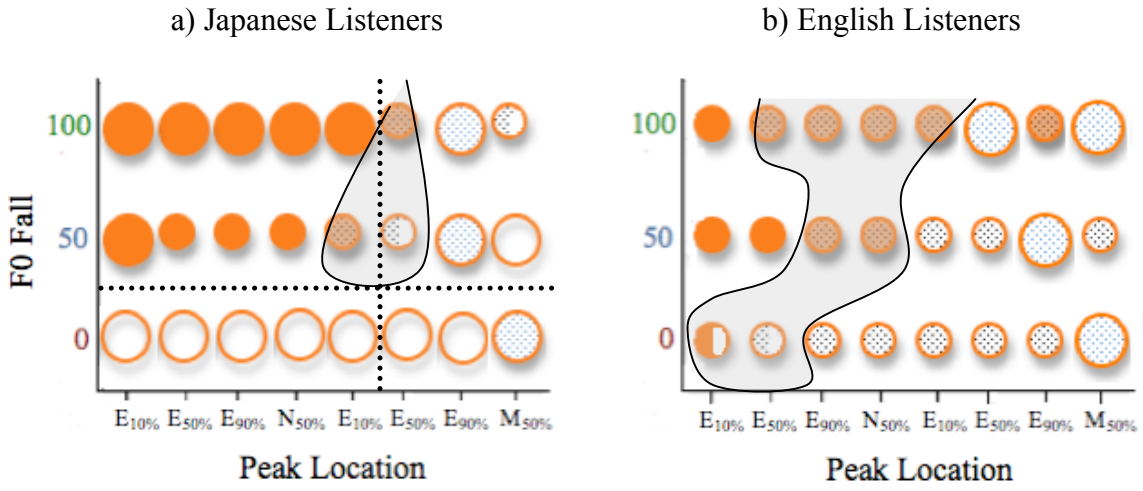


Figure 5.3. Above chance responses in the 3AFC task by group (Syllable 1 Prominent: filled circles; Syllable 2 Prominent: dotted circles; No Prominence: empty circles). Smaller circles indicate 34-66% of responses; larger circles indicate 67-100% of responses. Grey areas show clusters of stimuli for which the responses were mixed (as also indicated by mixed circle type patterning).

5.4. Discussion

The goal of the current experiment was to examine whether the number of categories in the 3AFC task as compared to the 2AFC task affected categorization of the same stimuli by native and non-native listeners of Japanese. The discussion below will focus, first, on the distinction between the accented and unaccented categories (i.e., prominent syllable vs. no prominence responses), and, then, on the distinction between the 1st-syllable accented and 2nd-syllable accented categories (i.e., Syllable 1 Prominent vs. Syllable 2 Prominent responses).

5.4.1. Accented Versus Unaccented

Consistent with the results of the two previous experiments, the manipulated acoustic cues affected Japanese listeners' prominence judgments in the current experiment. Moreover, the sizes of these effects were considerably larger for Syllable 1 Prominent responses in the 3AFC task than in the 2AFC task (Peak Location: $\eta_p^2 = .87$ vs. $\eta_p^2 = .34$; F0 Fall: $\eta_p^2 = .86$ vs. $\eta_p^2 = .20$, respectively). Differently from Experiment 2, the Japanese listeners were sensitive not only to the presence, but also to the magnitude of the F0 fall for this prominence pattern. These findings suggest that the cues were relevant to the distinction of categories available to listeners. As a result, listeners' sensitivity to the gradual variation in both acoustic cues was heightened.

Although the Japanese listeners were sensitive to each of the acoustic cues, the cues' combination was the most meaningful to them in perception of prominence, as indicated by the significant cue interactions in all categorization experiments. The effect of the interaction was stronger in the 3AFC task ($\eta_p^2 = .61$-$.76$ for the three response types) than in the 2AFC task ($\eta_p^2 = .13$ for Syllable 1 Prominent responses). This finding supports the view that integrated perception of the acoustic cues to the prosodic contrast is characteristic of native listeners, under an assumption that integrated cue perception is reflected in cue interaction. It also suggests that accent-presence detection and accent-location detection are integrated perceptual processes, contrary to the proposal that these two processes are separable in native listeners' performance (Kitahara, 2001).

Surprisingly, English listeners also showed some sensitivity (marginally significant) to the interaction between the Peak Location and F0 fall, in contrast to the 2AFC task of the two previous experiments. This finding may be interpreted as evidence

for the accent-presence detection (i.e., attention to F0 fall) and accent-location detection (i.e., attention to F0 peak location) as two separable tasks in non-native listeners' performance. While categorizing the F0 contours in the 2AFC task, English listeners paid attention to the peak location only; that is, they performed the accent-location detection task only. While categorizing the F0 contours at peaks located in the first half of the word in the 3AFC task, listeners preferred Syllable 1 Prominent responses in the presence of the F0 fall and No Prominence responses in the absence of the fall; that is, they additionally performed the accent-presence detection task. It appears that the peak location information may be sufficient for the choice between Syllable 1 Prominent and Syllable 2 Prominent in the 2AFC task, but in the 3AFC task listeners are forced to pay more attention to the F0 fall information in order to choose among the three answer options. This suggests that the cue use is encouraged by task demands, and the two cues of F0 peak location and F0 fall do not have to be used together by non-native listeners for any task related to categorization of pitch patterns.

Lastly, with regards to the categorization of the stimuli without the F0 fall that represented the unaccented category in Experiment 3, the data supported our prediction. Namely, the Japanese listeners clearly perceived the absence of the F0 fall as the absence of prominence, and the percentage of their No Prominence responses to such stimuli was relatively high (solid red contours in Figure 5.1c). The English listeners also returned No Prominence answers to stimuli without an F0 fall, but the percentage of such answers was relatively low, just slightly above the chance level of 33% (solid red contours in Figure 5.2c). This point is clearly illustrated in Figure 5.3, where the Japanese listeners appear to have a robust unaccented category, but the English listeners do not. This result provides

another post-hoc justification for using the 2AFC task, as meaningful for non-native listeners of Japanese who do not have a phonological representation of the unaccented category at the word- or phrase-level.

5.4.2. 1st-Syllable Accented Versus 2nd-Syllable Accented

Similar to Experiment 2 results, the Japanese listeners switched between Syllable 1 Prominent and Syllable 2 Prominent responses between the peak locations of 10% and 50% in the vowel of the second syllable. This supports our previous conclusion that in terms of the F0 peak (or a high tone of the H*+L pitch accent), the boundary between the 1st- and 2nd-syllable accented categories in CVCVCV words is associated with the second syllable, thus, conforming to the expectation of the peak delay.

Although the shift between these two response types appears to be categorical in both tasks judging by the abruptness of the shift (compare Figures 4.2a and 5.1a), the 2AFC task did not yield a "grey area" of mixed response patterns as the 3AFC task did (compare Figures 4.3a and 5.3a, which are also juxtaposed in Figure 7.1, Chapter VII). This observation seems to support our prediction of a more distinct (sharp) category boundary for Japanese listeners in the 2AFC task than in the 3AFC task. The boundary itself does not shift, but a more natural (and a more complex) task in terms of naturally existing categories (i.e., 1st-syllable accented, 2nd-syllable accented, and unaccented patterns) enhances the cue integration in perception and the gradiency of responses. An alternative explanation of a less distinct category boundary in the 3AFC than in the 2AFC task may be fewer participants in the former experiment than in the latter one.

For English listeners as well, the boundary between the 1st- and 2nd-syllable accented words in terms of the F0 peak location was considerably more sharp in the

2AFC than in the 3AFC task (compare Figures 4.2b and 5.3b, which are also juxtaposed in Figure 7.2 of Chapter VII). Similar to the Japanese data, in the 2AFC task the boundary was observed in the beginning of the second [e] vowel of the *nenema* word. In the 3AFC task, on the other hand, the boundary appeared to associate with slightly earlier peak location, namely, the midpoint of the second [n]. Given the difference in the number of participants in the 2AFC and 3AFC tasks, no far-going interpretation of these slight differences can be made. However, this observation resonates with the finding of Atterer and Ladd (2004) that the high tone of the pitch accent in English seems to be aligned earlier than in other languages (e.g., German).

5.4.3. Summary and Limitations

Overall, the results of Experiment 3 suggest that in a more naturalistic 3AFC task Japanese listeners relied in their judgments of prominence on F0 peak location, the presence of F0 fall, the magnitude of F0 fall, and the interaction of these acoustic cues. English listeners relied mostly on peak location, and only for some stimuli, on the presence of F0 fall. Nevertheless, the capability of English listeners to use the F0 fall information was surprising given that the fall is not a cue to lexical prominence in English (Chapter II, section 2.2). This capability was further explored in a discrimination task of Experiment 4 and reported in the following Chapter VI.

A practical implication of current results is that English listeners may be able to use the F0 fall information, at least for categorization of some F0 contours in the 3AFC task as designed in this dissertation. Therefore, the presentation of all three possible categories of Japanese pitch patterns (i.e., 1st-syllable accented, 2nd-syllable accented and unaccented) may be more beneficial for training than the pair-wise presentation of the

categories (especially, only 1st- versus 2nd-syllable accented). Such training task is reported as Experiment 5 in Chapter VIII. A possibility that a training focusing only on the two most phonetically similar categories (i.e., 2nd-syllable accented and unaccented) may be equally beneficial for learners cannot be excluded.

Lastly, the boundaries dividing the stimuli space, and presumably, the categories of F0 patterns, appeared to be less sharp during 3-way categorization than during 2-way categorization. This general observation is reminiscent, in part, with the results reported by Kitahara (2001). However, this observation does not necessarily mean that for processing of F0 contours the detection of accent location is more fundamental than the detection of accent presence, or that these two processes are independent of each other. First, it is unclear what the term fundamental means here: the relative timing of detection or its robustness. Second, the evidence for such hierarchical representation (or binary implementation) of the two detection processes would be, for example, faster response time in identification of accented than unaccented patterns in the 3AFC task, or vice versa. Less sharp boundaries between categories in a task with a larger amount of answer options may be simply due to the interaction of the cues in native Japanese listeners, and confusion of native English listeners with regard to the unaccented category.

The major limitation of Experiment 3 is a small number of participants in each of the listener groups. Due to a different sample size in Experiments 2 and 3, the comparisons of listeners' performance in the 2AFC and 3AFC tasks should be also considered with caution. Although the results of these two experiments seem to complement each other, a comparable number of participants should be recruited in future work.

CHAPTER VI

EXPERIMENT 4: AX DISCRIMINATION WITH TWO INDEPENDENT VARIABLES

6.1. Introduction

The goal of Experiment 4 was to examine the sensitivity of listeners specifically to the F0 fall, a primary cue to the Japanese pitch accent contrast (Chapter II, section 2.4.1). English listeners were sensitive to the manipulation of the F0 peak location in all categorization experiments reported so far. However, the results with regard to the F0 fall perception by non-native listeners were mixed. Perhaps, due to task demands, English listeners showed some sensitivity to this acoustic dimension in the 3-way categorization in Experiment 3, but not in the 2-way categorization in Experiment 2. The current experiment was hoped to clarify these results using an AX discrimination task. The specific question addressed here was whether English listeners could discriminate between patterns with and without F0 fall after the peak. To facilitate interpretation of results, materials in the current discrimination experiment were as similar as possible to the categorization experiments: *nenema* words, with two levels of F0 fall and eight levels of F0 peak location.

6.1.1. Categorization Versus Discrimination

In second language research, a stimulus categorization task is thought to reflect language-specific patterns of perception that maximize group differences due to linguistic experience, whereas a same-different discrimination task is thought to assess basic auditory sensory capabilities that "do not differ as a function of specific experience with a particular phonological system" (Strange & Shafer, 2008, p. 164).This task effect, related

to phonological versus auditory and/or phonetic mode of processing, has been reported in research on prosodic categories as well (for stress: Dupoux et al. 1997, 1998; for lexical tones: Hallé et al., 2004; for boundary tones: Eda, 2004). In general, non-native listeners perform better in discrimination tasks than in categorization tasks. Furthermore, when predicted from categorization, discrimination accuracy is usually underestimated, presumably due to the different modes of processing involved in these tasks (Xu et al., 2006).

Not surprisingly, native listeners outperform non-native listeners in discrimination tasks. However, the sensitivity of non-native listeners to tones appears to be relatively high, despite a well-known difficulty in perception of the F0 movement for some listener groups (Chapter II, section 2.6.3.2). For example, in Hallé et al. (2004) experiment on three continua of Mandarin tones, naïve French listeners were relatively successful in the discrimination task (74% correct), although their overall accuracy was lower than the accuracy of Taiwanese listeners (88% correct). Bent (2005) found that both Mandarin and naïve English listeners were very accurate in discrimination of Mandarin tones in a task with ten possible pairings of trisyllabic words, and the average accuracy in both groups was no less than 90%. Similarly for the discrimination of Thai middle versus low tones, Wayland and Guion (2003) found that all listeners displayed a good sensitivity to the contrast as manifested in the A' score of 0.92 in native Thai listeners, A' of 0.86 for learners of Thai, and A' of 0.78 for naïve English listeners, where the score of 1 indicated perfect discrimination. Francis and Ciocca (2003) found that English listeners were as sensitive to Cantonese tone contrasts as native Cantonese listeners as manifested in the d' scores of 2.60 for native and 2.53 for non-native listeners (i.e., above 85% accuracy).

Previous work on Japanese pitch accent contrasts and intonation contrasts yielded similar results. Sakamoto (2008) conducted an ABX discrimination task on three lexical pitch patterns of Japanese nonwords [mene], [noma], [mani], and [nime]. The discrimination accuracy did not vary among native Japanese speakers (89%), relatively more experienced learners of Japanese (88%), and relatively less experienced learners (84%). Eda (2004) conducted an AX discrimination task on intonation contrasts of Japanese sentences as indicated by boundary tones such as information question versus insisting (Chapter II, Table 2.1). She found that the second language proficiency of learners of Japanese did not correlate with discrimination accuracy. In both studies, however, the between-group differences emerged in categorization tasks.

Based on previous findings, it was predicted in Experiment 4 that both native and naïve non-native listeners would be sensitive to the F0 fall in a discrimination task, perhaps, Japanese listeners more so than English listeners. The phonetic mode of processing was thought to be encouraged by utilizing an AX type of task. This task is likely to require the least working memory load, stimulus uncertainty, and stimulus order effect as compared to the ABX, AXB, and Oddity tasks (Strange & Shafer, 2008, p. 161). A concern that AX discrimination may increase the chances of guessing (response bias) by listeners (Wayland & Guion, 2003), was addressed by using bias-free sensitivity measure developed in the Signal Detection Theory (Macmillan & Creelman, 2005). It was also hypothesized that English listeners might be more sensitive to the F0 fall for earlier locations of the peak, similar to the 3AFC categorization task.

6.2. Method

6.2.1. Participants

The same Japanese native listeners, who participated in the 2AFC categorization task, also volunteered to participate in the current experiment (Chapter IV, Table 4.1). A new group of English native listeners (15 females and 5 males) was recruited from the Human Subjects Pool at the University of Oregon. The profile of this group was very similar to the English group described in Table 3.1, with only a slight difference in age: range 18-20, $M = 19$ in Experiment 4 versus range 18-31, $M = 21$ in Experiment 2.

6.2.2. Materials

Test materials were variations of a nonword *nenema*, and practice materials were four variations of a nonword *nimiga*. Segment durations of the first two syllables in *nenema* were normalized as described in the Methods section of Chapter III.

Similar to Experiments 2-3, only two acoustic parameters – the location of the F0 peak and the F0 fall after the peak – were varied in the design of the current experiment (Figure 6.1). The location of the F0 peak was varied in eight steps, and the alignment of the F0 elbow was always 191 ms after the peak (a more detailed description is provided in the Methods section of Chapter IV). At each location, the F0 either fell or did not fall after the peak. The eight locations of the F0 peak were not equidistant from each other, but the controlled alignment of the elbow (i.e., endpoint of the F0 fall) ensured the same magnitude (40 Hz) and slope (2.209 Hz/ms) of the fall at each level of the location manipulation. After the elbow, the F0 trailed down another 10 Hz until the offset of the word, except for the stimuli with the F0 peak at the midpoint of [m] and 191 ms distance

between the peak and the word offset. Thus, as in real Japanese accented words, the F0 at the word offset was lower than the F0 at the word onset.

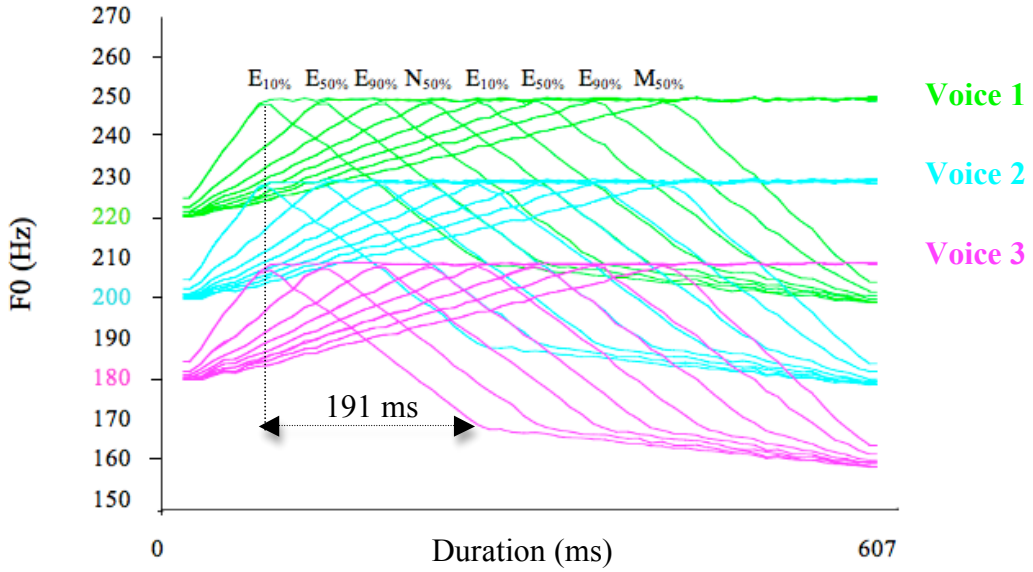


Figure 6.1. Manipulation of F0 peak location and F0 fall in the *nenema* word in the AB discrimination task.

6.2.3. Procedure

A pilot study with F0 contours manipulated as described above showed that English listeners were sensitive to the F0 fall contrast, and they achieved a high accuracy in this task. To avoid a ceiling effect, three 'voices' were introduced to the stimuli design (Figure 6.1). As compared to Voice 2 (blue contours), the F0 range of the stimuli was shifted 20 Hz up for Voice 1 (green contours) and 20 Hz down for Voice 3 (pink contours).

Listeners were presented with two stimuli at a time, and their task was to decide whether the pitch patterns of the two *nenema* words were the same or different. The instructions explained that words were spoken in three voices with different overall pitch:

One voice had a higher pitch, another voice had a mid pitch, and another voice had a lower pitch. In each pair of words, the voices were going to be different, but the pitch patterns of the words might be the same or different. Thus, the AB discrimination task was transformed from categorical to categorial; that is, stimuli in a pair were always physically different, and listeners had to determine whether they were of the same category (Strange & Shafer, 2008). It was emphasized that listeners should not confuse the overall pitch of the voice with a pitch pattern of the word.

Stimuli were paired in the following way for presentation. Voices were different in each pairing – three voices yielded six possible combinations. Each combination was repeated twice. The F0 peak location was the same in each pairing (and combination). Considering the F0 fall variable, there were four possible orders of presentation of the paired stimuli, namely, fall-fall, flat-flat, fall-flat, and flat-fall. For two of these orders the correct answer was 'same,' and for the other two orders the correct answer was 'different.' Following these pairing principles, the total number of word pairs was 384: voice combination (6) x repetition (2) x F0 peak location (8) x within-pair presentation order (4). The inter-stimuli interval between the words in a pair was one second. Pairs of stimuli were presented in a random order.

Before the beginning of the experimental trials, listeners had a short practice with the pairs of *nimiga* words. During the practice, they familiarized themselves with the task format, asked clarifying questions, and adjusted the listening level when needed. Each listener performed the task individually in a sound-attenuated booth using a computer and headphones. Responses were recorded when the Same or Different button on the computer screen was clicked. Response time was not limited. Selection was followed by

a 500-ms pause, and then a new trial began. Listeners were encouraged to make their best guess when unsure, and to take breaks after every 50 trials to fill a section of a language background questionnaire.

6.2.4. Analysis

Listeners' responses were analyzed by calculating a proportion of correct answers to assess the overall accuracy, and calculating d' scores to assess sensitivity to the F0 fall. The latter measure was developed within the framework of Signal Detection Theory (Macmillan & Creelman, 2005) and is expressed in Equation (2) below. Sensitivity of a listener to the F0 fall was constructed as a measure of the discrepancy between the proportion of correct responses to pairs of stimuli that are different in F0 fall (i.e., hits) and the proportion of incorrect responses to pairs of stimuli that are similar in F0 fall (i.e., false alarms). Thus, sensitivity depends on both the ability to recognize a difference between stimuli and the ability to recognize same pairings. In this way, the d' measure controls for individual response biases, in other words, the listeners' willingness to say Different rather than Same.

(2) d' score (z is the inverse of the normal distribution function):

$$d' = z(\text{hits}) - z(\text{false alarms})$$

To calculate d' scores, the following steps were taken.

1. The proportion of hits (Different responses for fall-fall and flat-flat pairs) and false alarms (Different responses for fall-flat and flat-fall pairs) were calculated out of 24 trials taken by a listener for each F0 peak location.
2. In order to obtain normally distributed z-scores and to avoid infinite values of d' scores, false alarm rates of 0 and hit rates of 1 were adjusted to 0.0021 and

0.9792, respectively. For this adjustment, the proportions of 0 and 1 were converted to 1/(2N) and 1 - 1/(2N), where N was the number of targets (i.e., 24), 1/N was the smallest possible false alarm rate, and 1/(2N) was representing a half of a false alarm (Macmillan & Creelman, 2005, p. 10).

3. The difference between a hit *z*-score and a false alarm *z*-score for the same location of the F0 peak constituted a *d′* score.

Higher *d′* scores reflect higher sensitivity to the F0 fall. A *d′* score of 1.0 represents 69% correct responses for both different and same trials; a *d′* score of 1.35 represents 75% correct responses (Macmillan & Creelman, 2005). A *d′* score of 4.65 represents the rate of hits equal to .99, and the rate of false alarms equal to .01, therefore, this score is interpreted as an effective ceiling in listeners' perfomance.

For statistical analysis, *d′* scores calculated for each listener were entered into a repeated-measure ANOVA with a between-subject factor of Group, and a within-subject factor of Peak Location.

6.3. Results

On average, Japanese listeners achieved 85% accuracy and English listeners 72% accuracy. The difference in accuracy between the groups was 13%. This difference could have reflected listener biases, such as varying confidence threshold for the Different answer. To eliminate this possibility, *d′* scores were calculated for each listener. Japanese listeners had an average sensitivity to the F0 fall of *d′* = 2.53 (*SD* = 1.16; range -0.21-4.90). English listeners' sensitivity was found to be *d′* = 1.43 (*SD* = 1.15; range -0.43-4.90). Thus, non-native listeners' sensitivity was smaller than native listeners' sensitivity

in terms of both response accuracy and *d′* scores. However, both groups performed in the discrimination task well above the chance level of the *d′* = 0.7. Furthermore, it is worth noticing, that the Japanese and English listeners overlapped in their sensitivity to F0 fall, which is illustrated in Figure 6.2.

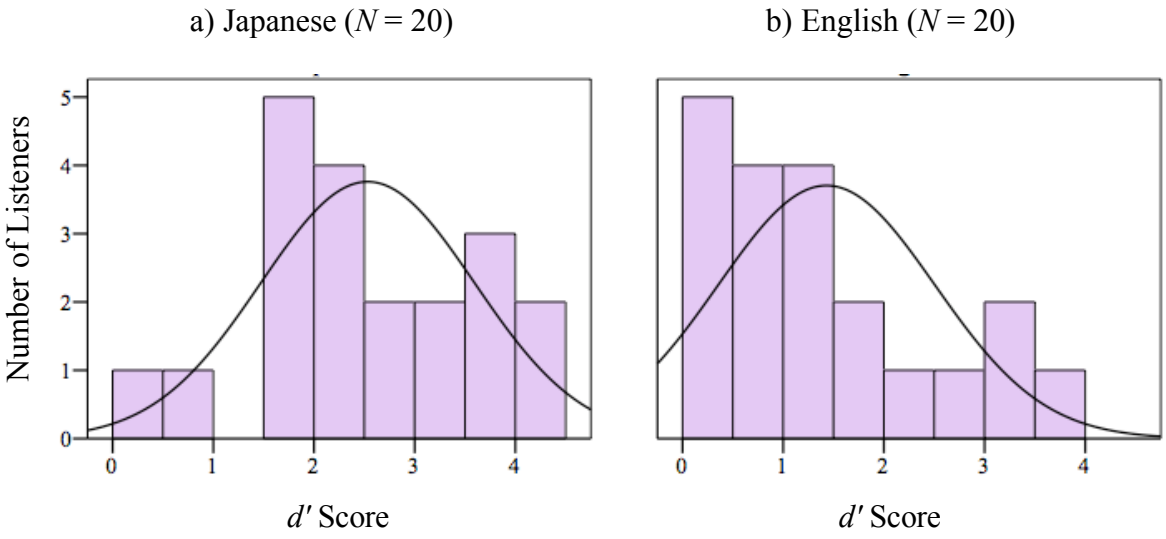


Figure 6.2. The distributions of *d′* scores in two groups, averaged across eight peak locations in individual listeners.

A repeated measures ANOVA on the individual *d′* scores yielded a significant effect of Group [$F(1, 38) = 10.64$, $p = .002$]. Neither the main effect of Peak Location, nor its interaction with Group were significant, [$F(7, 266) = 1.33$, $p = .239$] and [$F(7, 266) = 1.17$, $p = .318$], respectively. These results confirmed what is evident in Figure 6.3. Namely, Japanese listeners had higher sensitivity to F0 Fall than English listeners; and in each group, the level of sensitivity was similar across all F0 peak locations.

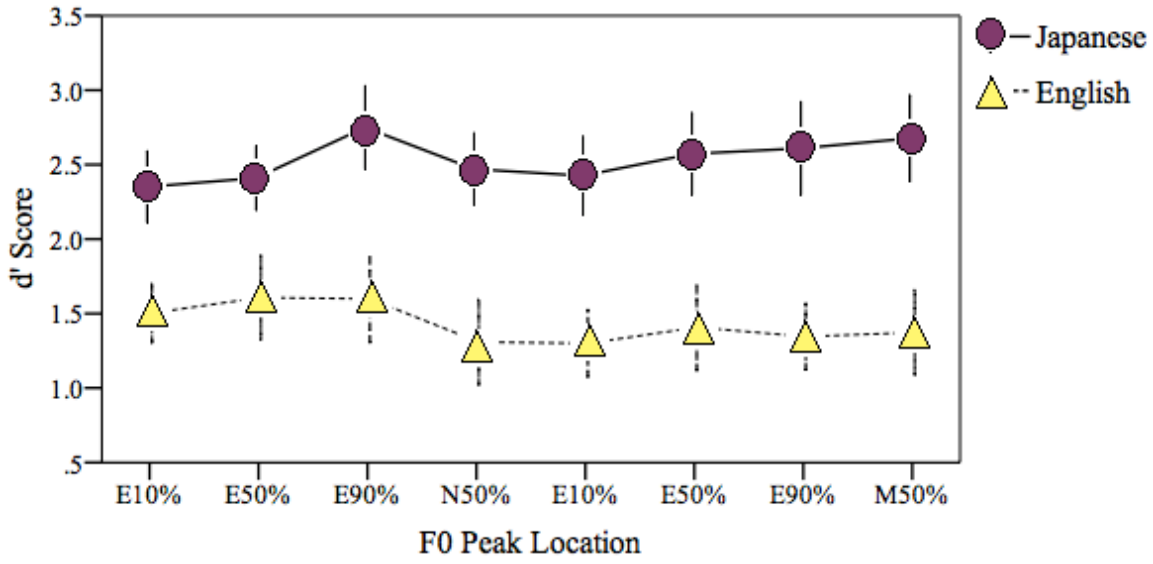


Figure 6.3. Mean sensitivity to F0 fall. *d′* score of 1.0 corresponds to 69% correct responses for both different and same trials. Error bars indicate ± 1 *SE*.

6.4. Discussion

Experiment 4 investigated whether English listeners retained basic auditory sensitivity to F0 fall as assessed in an AX discrimination task and indexed by a bias-free *d′*-score measure. The English listeners were less sensitive to the F0 falling movement than the Japanese listeners, however, they discriminated the stimuli with and without the F0 fall at above-chance level. Both listeners groups exhibited similar discrimination accuracy across the eight F0 peak locations.

In comparison with another discrimination experiment where both percentage correct and *d′*-score measures were used to assess listeners' responses (Bent, 2005), the discrimination accuracy was lower in the experiment reported here for both non-native listeners (72% vs. 90% correct; $d' = 1.43$ vs. $d' = 3.46$) and native listeners (85% vs. 97% correct; $d' = 2.53$ vs. $d' = 4.01$). This overall poorer performance of listeners in

Experiment 4 as compared to Bent's experiment may be attributed to the manipulation of the F0 contours. Three voices in the current experiment design were introduced to avoid a ceiling effect observed in a pilot study on non-native listeners' discrimination (Figure 6.1). These voices had overlapping F0 contours, which were likely to introduce a higher cognitive load to the task by requiring differentiation between voices and patterns in each stimulus pair. In contrast, recordings of six native speakers were used in Bent's study, but the words in each stimulus pair seemed to have been produced by the same talker.

Apart from stimulus design and presentation, the lower discrimination accuracy in Experiment 4 may be explained by individual variation in attention to the F0 fall cue among native listeners of Japanese (Figure 6.2a). The results of Experiment 2 demonstrated such individual variation in the same listeners who participated in Experiment 4 (Chapter IV, Figure 4.5a). Only a quarter of the listeners were heavily reliant on the F0 fall cue in the categorization task. This reliance might have affected the results of the discrimination task.

A comparison with Eda's (2004) study suggests that the lower discrimination accuracy in Experiment 4 was not just due to unfortunate sampling of participants. Eda reported a mean 89% accuracy in native Japanese listeners for her ABX discrimination task on lexical pitch patterns of nonwords, which is not much different from 85% accuracy for the AX discrimination task in Experiment 4. Both results are not at ceiling, as could have been expected from native listeners of a language.

The relatively high auditory sensitivity of English listeners to the F0 fall in Experiment 4 could be explained by psychoacoustic mode of stimulus processing in discrimination tasks (Strange & Shafer, 2008; Strange, in press). Strange advocates the

hypothesis that "a phonetic mode of perception must be used at the beginning stages of L2 learning when processing phonetic sequences that include non-native segments or segments that occur in both L1 and L2, but are produced differently in native and non-native languages" (2011, p. 5). This theory may be also applied to the processing of non-native prosodic categories by naïve listeners.

Strange (in press) further suggests that perception in the phonetic mode requires a greater attentional focus than perception in the phonological mode. In other words, the cognitive demand is larger for non-native than native listeners during experimental tasks. This proposal appears to be supported by findings such as an effect of the ISI duration on discrimination accuracy (Wayland & Guion, 2003; Werker & Tees, 1984). For example, Wayland & Guion (2003) found that a shorter ISI yielded a higher accuracy in learners than a longer ISI, which suggests that, compared to Thai native listeners, English learners relied more on working memory, and that a Thai tone contrast might have a less robust representation in their long-term memory. The ISI in Experiment 4 was one second, which is ambiguous between the ISI used to encourage auditory/phonetic processing (250 ms in Francis et al., 2008; 500 ms in Wayland & Guion, 2003, and in Xu et al., 2006), and the ISI used to encourage phonological processing (1500 ms in Wayland & Guion, 2003). Future work may include an examination of the ISI effect on discrimination accuracy in native and non-native listeners of Japanese, an experimental manipulation that has not been explored yet for the Japanese language.

Further investigation of the cognitive demand of a task on performance accuracy, as manipulated via longer ISI, stimuli in noise, presentation order, or more complex stimuli, is theoretically interesting for cross-linguistic perception of Japanese lexical

prosody. The relatively high sensitivity of English listeners to the F0 fall in the current experiment can be explained not only by the auditory/phonetic mode of processing (as discussed above), but also by the listeners' ability to apply prosodic categories of their native language to the task at hand. The F0 fall is prosodically meaningful in both Japanese and English, albeit in different domains - lexical in Japanese and phrasal in English (Chapter II, section 2.3.2). For a similar cross-linguistic perception study on three continua of Mandarin tones, Hallé et al. (2004) concluded that while French listeners were sensitive to Mandarin tones, they did not process the tones as phonological categories. The discrimination accuracy of French listeners did not vary across the continua (including near category boundaries), providing support for this conclusion. Results reported by Bent (2005) also echoed this conclusion in the case of English listeners. Thus, it appears that familiarity with F0 variations in the intonation domain of the native language may be applied to F0 variations in the lexical domain of non-native language, at least for less cognitively demanding tasks.

This line of thought engages with the Perceptual Assimilation Model (Best, 1995; Best & Tyler, 2007). Specifically, non-native listeners may process tones as evocative of native language but uncategorizable speech patterns (Hallé, et al., 2004, pp. 416-418), or as evocative of native language and categorizable (So & Best, 2010, p. 289). This hypothesis may be empirically tested by asking English listeners to rate the similarity of Japanese lexical pitch patterns and English phrasal intonation contours as similar or dissimilar in stimuli of different length (see, however, the discussion in Francis et al., 2008, on how mappings between native and non-native categories may be insufficient to account for the data).

CHAPTER VII

PERCEPTUAL PROCESSING OF F0 PATTERNS: SUMMARY OF FOUR EXPERIMENTS

7.1. Use of Acoustic Cues to Prominence

In the three categorization experiments reported so far, the lexical pitch-accent contrast was construed as a contrast in patterns of prosodic prominence (see Chapter II, section 2.2). F0 contours of the stimuli were resynthesized to represent the Japanese accentual patterns in disyllabic words, namely, the 1st-syllable accented, 2nd-syllable accented, and unaccented patterns. Across all experiments, F0 peak location and F0 fall, which are fundamental cues to the pitch-accent contrast, were manipulated. In addition, F0 peak height was manipulated in the first experiment of the series. The effect sizes of the manipulated cues on listeners' categorization of the F0 contours and the cue interactions are summarized in Table 7.1 by experiment and by listener group.

A comparison of the results across Experiments 1-3 shows that they were largely consistent with each other. Processing differences between the Japanese and English listeners emerged in several areas. For example, the effect sizes of both the F0 peak location and the F0 fall cues on listeners' decisions were consistently larger for the Japanese group than for the English group. This indicates a stronger relationship between these acoustic cues and categorization responses of native listeners, as compared to non-native naïve listeners. Most importantly, however, the effect of the interaction between the cues was always significant and relatively large for the Japanese listeners, but not for the English listeners. This interaction suggests complex and integrated processing of the cues to prominence, fundamental for perception of the pitch-accent contrast. This

integrated processing was observed in categorization tasks (Experiments 1-3), which resemble natural speech tasks better than discrimination. In the discrimination task (Experiment 4), no interaction between the peak location and F0 fall was observed, perhaps, due to an auditory/phonetic mode of stimuli processing by the listeners.

Table 7.1

Effect Sizes (η_p^2) of Variables in the Three Categorization Experiments

Factor	Experiment 1 2AFC-3IV		Experiment 2 2AFC-2IV		Experiment 3 3AFC-2IV	
	Japanese $N = 20$	English $N = 20$	Japanese $N = 20$	English $N = 20$	Japanese $N = 5$	English $N = 5$
Peak location (L)	.67*	.38*	.34*	.15*	.87*/.87*/.19	.44*/.46*/.31
F0 fall (F)	.56*	.06	.20*	.00	.86*/.17/.71*	.18/.03/.33
L x F	.58*	.34*	.13*	.02	.76*/.61*/.63*	.31/.27/.31
Peak height (H)	.29*	.09				
H x L	.59*	.15*				
H x F	.14*	.10				
H x L x F	.21*	.07				

Note. *$p < .05$. In Experiments 1-2, the dependent variable was Syllable 1 Prominent response. In Experiment 3, the dependent variables were Syllable 1 Prominent, Syllable 2 Prominent, and No Prominence responses, the effect sizes for which are listed in the respective order.

Next, let us consider the effects of the two manipulated variables on listeners' responses separately (Table 7.1). In the preceding three categorization experiments, both listener groups were sensitive to the manipulation of F0 peak location. The effect size of the peak location tended to be largest in comparison with the other variables and their interactions, and it was also larger in the Japanese data than in the English data. In

addition, Experiment 3 clarified that manipulation of the peak was relevant to the detection of accent location, namely, the prominence of the 1st versus 2nd syllable in the *nenema* word, as indicated by significant effects on Syllable 1 Prominent and Syllable 2 Prominent responses in the table.

As for F0 fall, the Japanese group consistently relied on this cue in their prominence judgments (with the only exception of the Syllable 2 Prominent judgments in Experiment 3), in contrast to the English group. The effect of F0 fall was significant whether or not the third variable, F0 peak height, was added to the design (compare Experiments 1 and 2), and whether or not the third answer option was added to the task (compare Experiments 3 and 2). Experiment 3 suggested that the F0 fall manipulation was most relevant to the detection of the accent presence in the 1st-syllable accented pattern, and the detection of the accent absence in the unaccented pattern as indicated by significant effects on Syllable 1 Prominent and No Prominence responses in the table.

Syllable 2 Prominent responses were not affected by the magnitude of F0 fall. A possible explanation may be a tendency among native speakers of Tokyo Japanese to neutralize the distinction between the 2nd-syllable accented and unaccented patterns in utterance-final position in both perception and production (Sugito, 1982b; Vance, 1995). Cutler and Otake (1999) have shown that the high and low tones are more different from each other in word-initial than in word-final syllables, which corresponds to higher identification accuracy of words by native listeners when the pitch-accent contrast is in word-initial compared to word-final syllables. Considering Experiment 3 in this dissertation, perhaps, once the No Prominence answer option became available, listeners failed to distinguish consistently between the 2nd-syllable accented and unaccented

patterns using the F0 fall information (see also mixed responses in Figure 7.1). Another explanation of the result may be due to the interplay of the later peak alignment associated with the prominent CV syllable, and the absence of a substantial peak delay (for example, in the beginning of the [a] vowel of the third syllable) that the listeners might have expected.

Interestingly, the F0 fall manipulation appeared to have no effect on the answers of the English listeners in the two-alternative forced choice task. However, the listeners showed an effect of this cue at certain F0 peak locations in the three-alternative forced choice task, although the interaction was rather small ($\eta_p^2 = .31$) and only marginally significant ($p \leq .058$). This may suggest that the English listeners were not completely insensitive to the F0 fall (this conclusion was supported by the results of the discrimination Experiment 4). Rather, they did not apply the falling F0 cue to the 1st-syllable prominent versus 2nd-syllable prominent categorization. Once given another answer option, they started to associate the presence of the F0 fall with the prominence of the 1st syllable and the absence of the F0 fall with no prominence. Overall, however, the English listeners exhibited non-categorical and less consistent responses than the native listeners in the 3AFC task, which is the closest task to the natural speech situation where all three F0 contours (i.e., 1st-syllable accented, 2nd-syllable accented and unaccented) may be encountered.

The experimental materials used in this dissertation were deliberately confined to simple designs in which the cross-linguistic perception of F0 contours could be easily tested. First, a nonword *nenema* was used as the base for all stimuli. Second, a more simple manipulation of F0 patterns was utilized here than in the intonation synthesis

models by Fujisaki and Sudo (1971) or by Pierrehumbert and Beckman (1988). These methodological choices impose some limitations on generalization of the findings to other types of speech materials. It would be optimal to confirm these findings in naturally produced utterances, where words contrastive in their lexical pitch patterns are integrated in the intonation patterns. This integration would result in additional acoustic cues to the contrast available for listeners, such as downstep (lowering of tones following the pitch accent), declination, longer initial rise to the high phrasal tone than to the high initial pitch accent, etc.

7.2. Categorization of the Stimuli Space

Figures 7.1 and 7.2 in this section illustrate response patterns as a function of the task in the Japanese and English listeners, respectively. Although the sample size was different in the two-alternative and three-alternative categorization tasks, the comparison of the results appears to have some explanatory power.

Figure 7.1 shows the F0 contour categorization by the Japanese listeners reported in Experiments 2 and 3. It is apparent that an addition of No Prominence response option to the task procedure did not greatly change the answers of the Japanese listeners with regards to the Syllable 1 Prominent response. In both tasks, the 1st syllable of the *nenema* word was perceived as prominent when the F0 peak was aligned earlier than the middle of the second vowel, and when the magnitude of the F0 fall was relatively large. The interaction between the two cues appeared to be larger in the 3AFC task (compare also the effect sizes in Table 7.1). With regards to Syllable 2 Prominent response, it is clear that in the absence of No Prominence response alternative, F0 contours without the F0 fall were categorized as 2nd-syllable prominent (Figure 7.1a). Otherwise, these patterns

were perceived as unaccented (Figure 7.1b). In sum, in the 3AFC task the Japanese listeners demonstrated increased sensitivity to accent-presence (note the robust unaccented category in Figure 7.1b), but not accent location (note a less clear vertical boundary between 1st and 2nd syllable responses in Figure 7.1b as compared to Figure 7.1a). This observation further supports the finding that peak location is processed in integration with F0 fall.

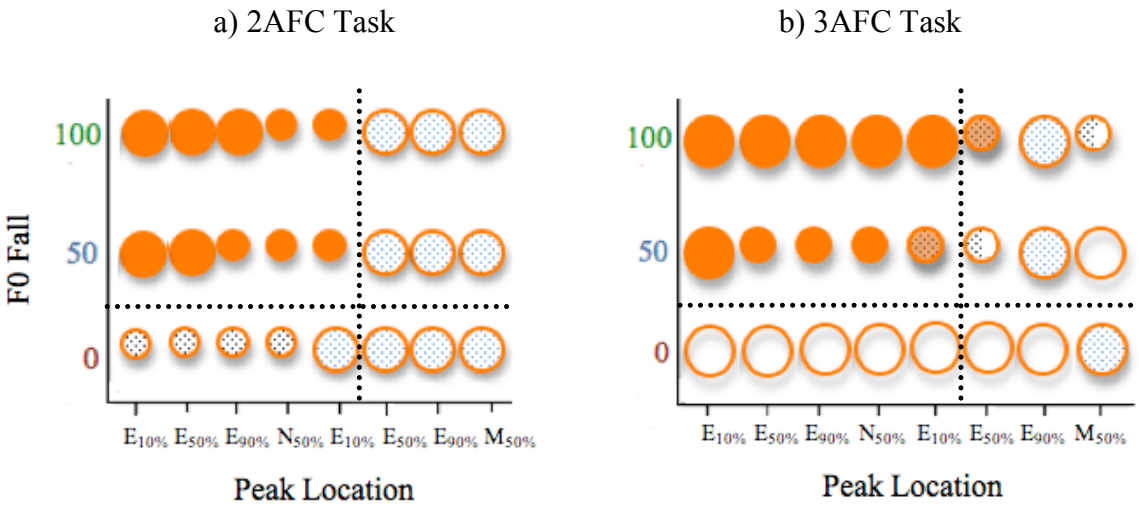


Figure 7.1. Above chance responses of Japanese listeners to the same stimuli space in the 2AFC (20 participants) and 3AFC (5 participants) experiments. Circle pattern indicates response type (Syllable 1 Prominent: filled circles; Syllable 2 Prominent: dotted circles; No Prominence: empty circles; mixed responses: mixed patterns). Circle size indicates the percentage of responses (smaller circles: 50-75% for 2AFC, 34-66% for 3AFC; larger circles: 75-100% for 2AFC, 67-100% for 3AFC).

Figure 7.2 shows that the addition of the No Prominence answer option in the three-alternative forced choice task dramatically changed the response pattern of the English listeners, in contrast to the Japanese listeners. The English listeners did not categorize F0 patterns as having no prominence, with exception of slightly above-chance

responses for the series without F0 fall and the earliest peak location (i.e., two half-empty circles in Figure 7.2b). However, the space of Syllable 2 Prominent responses (dotted circles) vastly expanded in Figure 7.2a as compared to Figure 7.2b. This result suggests that the unaccented pattern does not represent a category for the English listeners. They may apply the stress categories of their native English language to the F0 contours, so that each contour is perceived as having a prominent syllable due to the obligatoriness of stress in English lexical prosody.

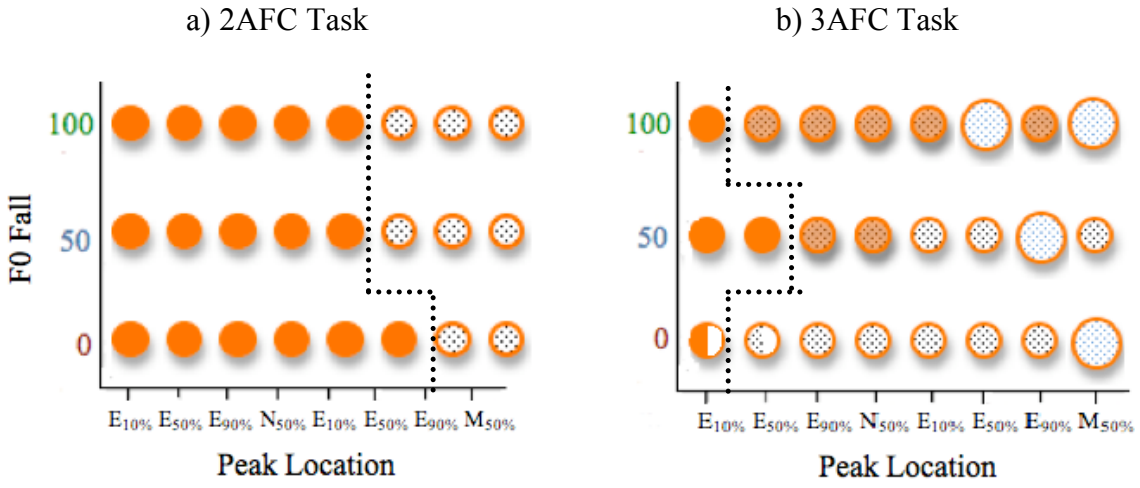


Figure 7.2. Above chance responses of English listeners to the same stimuli space in the 2AFC (20 participants) and 3AFC (5 participants) experiments. Circle pattern indicates response type (Syllable 1 Prominent: filled circles; Syllable 2 Prominent: dotted circles; No Prominence: empty circles; mixed responses: mixed patterns). Circle size indicates the percentage of responses (smaller circles: 50-75% for 2AFC, 34-66% for 3AFC; larger circles: 75-100% for 2AFC, 67-100% for 3AFC).

7.3. Implications for Learning of the Pitch-Accent Contrast

The results of four perception experiments indicated that the perception of the fundamental cues to the Japanese pitch-accent contrast is influenced by the native language experience of the listeners. While English listeners had phonetic sensitivity to

the F0 fall, and used the F0 peak location for categorization of prominence patterns, they could not consistently apply their phonetic experience with F0 variations to categorization of the no-prominence pattern. Future work should examine whether English listeners actually map their native pitch accents and intonation contours on non-native lexical prosody. Such mapping between native and non-native categories may occur in their entirety (Best, 1995; Best & Tyler, 2007; So & Best, 2010), or with regard to particular acoustic dimensions of a category (Gandour, 1983; Gandour & Harshman, 1978; Iverson & Kuhl, 1995; Iverson et al., 2003). Depending on the outcome of this investigation, the leaning of Japanese lexical prosody by English learners may require holistic attention to the new prosodic pattern as a specific combination of acoustic cues, or selective attention to a particular acoustic cue in the pattern structure.

Three-way categorization of F0 contours by the English listeners did not emerge from mere exposure to the stimuli in the 3AFC perception experiment. The unaccented category is likely to be the most difficult to categorize, as this category resembles English phrase-level F0 contours the least from the point of view of prominence and culminative/delimitative functions of accent (Chapter II, section 2.1). Hence, the main challenge in the second-language acquisition of the Japanese pitch-accent contrast may lie in learning that the absence of the H*+L accent is linguistically meaningful. The 3AFC task may be the most promising for the learning purpose, since this is the only task in this dissertation where English listeners were somewhat sensitive to the F0 fall, at least at some F0 peak locations. In addition, identification tasks have been argued to be more language-like than discrimination tasks in that they are more representative of what listeners are doing while processing language stimuli (Strange & Shafer, 2008, p. 160).

The 3AFC task was used in the following training experiment in order to investigate whether English listeners could improve in their categorization of the same three pitch patterns after a short training with explicit feedback and auditory reinforcement. The assumption of the training experiment was that English listeners could learn new prosodic categories holistically, without explicit attentional orientation toward the F0 fall cue.

CHAPTER VIII

EXPERIMENT 5: TRAINING TO CATEGORIZE LEXICAL PITCH PATTERNS

8.1. Introduction

The series of perception experiments reported in the previous chapters demonstrated that naïve English listeners are sensitive to F0 fall in the discrimination task, but that they rely mostly on F0 peak location in more language-like categorization tasks. Thus, it appears that at the initial stage of exposure to word-level F0 variations, English listeners may under-attend the most important perceptual cue to the Japanese pitch accent, namely, the F0 fall. The question arises as to whether English listeners can improve in their categorical judgments of Japanese lexical pitch patterns with training.

To investigate perceptual learning of lexical pitch patterns, a training experiment was conducted using the 3AFC categorization task. English listeners were trained to identify the 1st-syllable accented pattern (F0 fall after the first syllable), the 2nd-syllable accented pattern (F0 fall after the second syllable), or the unaccented pattern (no F0 fall) in real Japanese words.

8.1.1. First Language Bias

As reviewed in section 2.6.3.2 of the introduction, first language *attentional biases* affect perception of segmental and prosodic contrasts in second languages. An attentional bias means that a learned way of processing of one's native language (i.e., long-term categorical representations) influences perceptual attention to acoustic characteristics of a second language sound pattern and its categorization. Sensitivity to physical acoustic characteristics of sound patterns is not lost in adult listeners, which is evident from discrimination experiments using speech stimuli (Hallé et al., 2004; Iverson

et al., 2003; Kuhl & Iverson, 1995; Strange & Shafer, 2008), and from experiments using nonspeech stimuli (Xu et al., 2006). Rather, adult listeners learn to ignore acoustic information irrelevant to sound patterns of their native languages, especially for more language-related and complex tasks such as sound categorization.

The findings in the perception experiments of this dissertation summarized in Chapter VII corroborate this interpretation for the perception of sound patterns in the prosodic domain as influenced by first language attentional biases. Namely, the findings revealed attentional biases of naïve English listeners in perception of pitch patterns such as insufficient attention to the change in F0 that differentiates Japanese accented and unaccented patterns (in conjunction with the alignment of F0 maximum). A similar attentional bias for native listeners of stress languages, like English, during perceptual processing of tone contrasts has been reported in previous research on tone languages other than Japanese (as reviewed in Chapter II). Native listeners of stress languages attend less to the acoustic dimension of F0 change than native listeners of tone languages. This bias may negatively affect the acquisition of Japanese lexical and phrasal prosody by English learners.

Current theories of speech perception postulate that not all segmental categories in a second language are percieved and acquired equally well by learners (Best, 1995; Flege, 1995a; Kuhl & Iverson, 1995). Furthermore, in a phonemically contrastive pair of sounds, one member of the pair tends to be learned slightly more sucessfully than the other member. A classic example is the /ɹ/-/l/ segmental contrast in English, in which the /ɹ/ sound is identified by Japanese listeners better than the /l/ sound (e.g., Bradlow et al., 1997), presumably due to a greater dissimilarity between the first language /ɹ/ and the

second language /ɾ/ than between /l/ and /ɾ/ (Aoyama, Flege, Guion, Yamada, & Akahane-Yamada, 2004). Thus, predictions of relative learnability of a segment are drawn from similarities and dissimilarities between first language and second language categories.

Section 2.6.3.1 described why assessing similarity between native and non-native categories is more challenging, and maybe even impossile, for prosody as compared to segments. What is possible, however, is assesing acoustic-phonetic similarity of the categories belonging to the same prosodic contrast in a second language. Consider three contrastive pitch patterns in disyllabic Japanese words (Figures 2.3 and 3.1; also Chapter II, section 2.4). The 1st-syllable accented and 2nd-syllable accented words differ at least in the location of F0 maximum, and in F0 peak height or F0 range. They are similar in the presence of a steep F0 fall after the peak. The 1st-syllable accented and unaccented patterns have the most dissimilar F0 contours. The 2nd-syllable accented and unaccented patterns have the most similar F0 contours. One may predict that these latter patterns would be the most confusable with each other among the three two-way comparisons. The 1st-syllable accented pattern would be identified the best among the three patterns due to the absence of close competitors in terms of acoustic similarity.

Previous research on cross-linguistic perception of the Japanese pitch-accent contrast was reviewed in Chapter II, section 2.6. The majority of the studies used the methodology of the Tokyo Japanese Accent Listening Test, or similar tasks and procedures. A common finding across these studies was that in multi-syllabic words non-native listeners identified the unaccented pattern with relatively higher accuracy than a variety of accented patterns, that is, the patterns with the lexical pitch accent H*+L

associated with syllables from one to five. However, this result can be explained by the nature of the task and listeners' strategies, as was suggested in the second chapter. This result was not predicted in the current experiment because target words were not varied in their length, and because the task did not bias listeners to select the unaccented pattern as their default answer.

8.1.2. Other Factors Relevant to Learning of Second Language Prosody

Apart from first language attentional biases, other factors influence perceptual learning of a new prosodic contrast. Such factors include a) listeners' metalinguistic awareness of prosodic structures as related to percieved similarities or differences; b) language-specific domains of prosodic categories; c) physical acoustic similarity between prosodic categories of a second language, and d) individual characteristics of learners such as linguistic experience, music experience, memory, and attention. These factors are reminiscent of stimulus variables and subject variables outlined by Strange and Shafer for consideration in perception experiments (2008, pp. 182-183). The first two factors were discussed elsewhere in the dissertation. The acoustic similarity factor is discussed in the following section in relation to predictions about learnability of Japanese picth patterns. A subject variable, which was expected to correlate with task performance of English listeners, is reviewed below.

Individual variability among listeners sharing the same first language and similar experiences in learning of a second language has been persistently reported in previous research (Bent et al., 2006; Chandrasekaran et al., 2010; Golestani & Zatorre, 2009; Iverson et al., 2005; Iverson & Evans, 2009; Nishinuma et al., 1996; Wang et al., 1999). This variability may correlate with a variety of non-linguistic experiences, as well as

general cognitive abilities. One such experience is training to process pitch in music.

The relationship between processing of pitch in music and in a non-native tone language has been investigated in previous research. A general finding is that musicians are better in processing language-related pitch than non-musicians. Even more specifically, Rø, Behne and Wang (2006) have suggested that instrumentalists and singers are more accurate than non-musicians in identification of linguistic and hummed tones. This finding is interpreted as sharpened sensitivity to pitch variation, which is acquired during musical training and applied to pitch perception in a second language (Behne et al., 2006; Rø, Behne, & Wang, 2006; Wayland, Herrera, & Kaan, 2010). Considering this finding, the information about musical, instrumental and singing experience of participants was collected in the current training experiment. Other information describing demographic, academic and language learning background of participants was also collected via questionnaires. None of these subject variables was manipulated in the experiment design, and their examination was exploratory.

8.1.3. Effective Training: Variability in Training Materials

The current study followed a high-variability training paradigm, in which listeners identify sound categories of interest across a variety of stimuli and receive feedback on their responses (Chapter II, section 2.7). High-variability in training materials may include multiple target words, contexts, and speakers. Such training leads to successful learning of non-native contrasts; availability of rich training stimuli promotes rather than interferes with second language learning (Bradlow, 2008).

Sentential context of stimulus words has been shown to facilitate perception (Lee et al., 2009; Nishinuma et al., 1996) and learning (Hirata, 1999, 2003, 2004) of sound

patterns by adult non-native listeners. Specifically for the Japanese lexical pitch accent, Nishinuma et al. reported that pitch pattern identification accuracy was greater for stimuli presented in carrier sentences as compared to stimuli presented in isolation or excised from their sentential contexts. This facilitative effect of context may be due to fully manifested word-level acoustic cues to the target prosodic contrast. For instance, the F0 fall that distinguishes the unaccented and 2nd-syllable accented patterns in disyllabic words may be not fully realized in the utterance-final pattern (Haraguchi, 1999; Sugito, 1982b; Vance, 1995). In such cases, the acoustic difference between these two patterns is partially neutralized. Therefore, using carrier sentences where target words are in initial or medial positions would ensure that contrastive acoustic cues are preserved at their fullest.

Providing a larger sentential context for lexical pitch patterns is also important, because lexical patterns are integrated in intonation contours of higher prosodic units such as an accentual or intonation phrase. Intonational grouping of words into phrases facilitates speech recognition. As a by-product of grouping, context provides listeners with additional, sentence-level, acoustic cues to lexical pitch patterns. For example, the rise from the phrase-initial low boundary tone to the high tone is steeper in 2nd-syllable accented words (where the high tone is a part of the lexical pitch accent H*+L) than in unaccented words (where the high tone is the phrasal accent H-). This rise is virtually absent in phrases starting with 1st-syllable accented words. Although these additional cues were neither controlled nor manipulated in this study, they were certainly available for attentive listeners.

Successful training implies not only a higher accuracy at posttest as compared to

pretest, but also generalization to new stimuli. For Japanese lexical pitch patterns, Ayusawa (2003) suggested to test the generalization of training experience with declarative sentences on new interrogative sentences. Differently from statements, sentence-final lexical pitch patterns in interrogatoves are integrated with question boundary tones. Following this suggestion, interrogative sentential contexts were included in the materials of the generalization tests for the current experiment.

8.1.4. Predictions

Experiment 5 focused on perceptual learning of the Japanese pitch-accent contrast in disyllabic words. The implicit training objective was two-fold: first, to provide adult English listeners with various examples of the target prosodic categories, and second, to engage them in a language-related task. This experience was presumed to encourage a search for acoustic-phonetic cues, which differentiate the pitch-accent patterns, and a formation of categorical representations of the patterns for the purpose of the task. This is the first training experiment focused exclusively on lexically contrastive pitch patterns in a pitch-accent language. It contributes to the body of work on learning of second-language prosodic categories.

The experiment included pretest, training, posttest and generalization tests. It was predicted that categorization skills would improve with training. In other words, the difference between pretest and posttest scores was expected to be greater in participants who underwent training, as compared to participants who only took the tests. Furthermore, training-induced learning was expected to generalize to new test stimuli spoken by a familiar speaker (generalization test 1, or gen 1 hereafter), and to new stimuli spoken by a new speaker (generalization test 2, or gen 2 hereafter).

The identification accuracy and perceptual learning was expected to vary among the 1st-syllable accented, 2nd-syllable accented, and unaccented pitch patterns. The unaccented pattern was predicted to be the most difficult because English listeners under-attend the acoustic dimension of F0 fall, which distinguishes the unaccented and accented patterns. Especially, the 2nd-syllable accented and unaccented patterns were predicted to be highly confusable with each other due to similarity of their F0 contours.

8.2. Method

8.2.1. Participants

In previous training experiments, the total number of participants in both training and control groups was about 16-24 people (Bradlow et al., 1997; Francis et al., 2008; Jamieson & Morosan, 1986; Wang et al., 1999). In this experiment, 42 native speakers of American English with no reported speech or hearing impairments were recruited from psychology and linguistics classes at the University of Oregon. The students from psychology classes participated for a research credit, and the students from linguistics classes participated for a class extra-credit. They were randomly assigned to either control or training group. In addition, six native speakers of Tokyo Japanese were recruited by word of mouth. They volunteered to take a pretest and two generalization tests in one session.

A preliminary analysis of the pretest scores revealed that their distribution was bimodal in the control group, and close to bimodal in the training group (Figure 8.1). A 60% correct response threshold was used to divide both of the groups into two levels: low-scoring level and high-scoring level. Thus, according to the pretest results, the participants were placed in one of four groups - Control Low, Control High, Training

Low, or Training High (Table 8.1). One subject was excluded from the Control Low group because he failed to respond to 12% of the pretest trials.

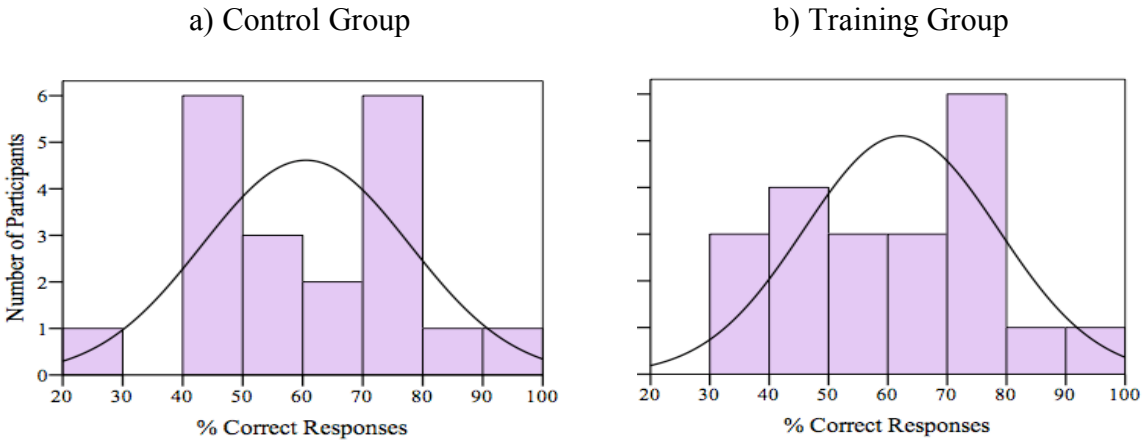


Figure 8.1. Distribution of pretest scores in English listener groups.

Individual characteristics of the trainees and controls are detailed in Appendix A and Appendix B, respectively, and summarized by Group in Table 8.1. All participants with one exception had studied a foreign language from age twelve or older (i.e., in middle school, high school, or university). With regard to tone language experience, four participants in the Training High group had studied Japanese and/or Mandarin as a second language in high school or university, but for no more than one year. One Control Low participant had studied Japanese for six years in middle and high school, and one Control High participant had studied Japanese for two years in university. All six participants self-reported their proficiency level in tone languages as beginners.

Previous research has shown that native-language experience with tones may be advantageous in learning another tone language. For example, Wayland and Li (2008) found that Chinese listeners learning Thai tones outperformed English listeners in

Table 8.1

Four Listener Groups Formed in Accordance with Pretest Scores

Group	Level	Pretest score %		Tone lang. learners	Linguistic majors	GPA	Instrument players	*n*
		mean	median					
Control	Low	45	45	1	5	3.4	5	10
	High	76	75	1	8	3.5	6	10
	Total	61	62	2	13	3.5	11	20
Training	Low	46	47	0	2	3.4	6	10
	High	74	74	4	8	3.4	5	11
	Total	61	61	4	10	3.4	11	21

identification and discrimination tasks. However, there are no research reports showing that *any* type of second-language experience with tones (such as foreign language instructions typical in the U.S. classroom setting) has the same facilitative effect. To examine whether in the current experiment the participants familiar with tone languages as second-language learners outperformed at pretest the participants without such experience, a linear regression analysis on the pretest scores was conducted. In this analysis, Group (controls vs. trainees) was an independent variable; experience with a tone language (yes vs. no), the number of tone languages learned, major (linguistic vs. non-linguistic), GPA, and the number of musical instrument played were covariates. The analysis returned significant effects of major ($p = .001$) and GPA ($p = .013$); the regression model explained 22% of variance in the pretest scores. These results indicated that linguistic experience with a tone language in the second language classroom setting

did not increase participants' response accuracy at pretest.[8] Therefore, the data of these participants were included in further analysis. The number of participants who were linguistics majors and who had a higher GPA was balanced across the control and training groups.

8.2.2. Materials

Stimuli consisted of 36 words contrastive in their pitch patterns (Table 7.2). The word triplets were chosen from a textbook by Imada (1989), and confirmed with the NHK accent pronunciation dictionary (1998). Six of these word triplets (Set 1) served as stimuli for pretest and posttest; the other six triplets (Set 2) served as stimuli for two generalization tests. Both of the sets represent common types of (C)VCV words, and include low and high vowels that have intrinsically different fundamental frequency (Whalen & Levitt, 1995), duration (Peterson & Lehiste, 1960), and intensity (Homma, 1973; Lehiste & Peterson, 1959). The number of words where high vowels could be devoiced was similar: *aki, hashi, kaku* and *waki* in Set 1, and *kaki*, *maku*, *mushi*, and *yoku* in Set 2 (devoicable vowels are underlined).

According to the NHK dictionary, three of the target words had a double pronunciation norm in terms of lexical pitch patterns: *tsurú* or *tsúru* 'vine'; *kakú* or *kaku* 'stroke'; and *yoku* or *yóku* 'next' (pitch accent is marked with acute accent). For each of these words, speakers were asked to use the pattern that fit the stimuli design in Table 8.2. All speakers reported to know the words, and examination of the F0 contours of their productions showed expected patterns.

---

[8] Similar regression analyses were performed on the posttest, gen1 and gen2 scores to examine whether familiarity with a tone language was advantageous in the course of exposure to Japanese accentual patterns and training. The effect of the previous tone language exposure was non-significant in all analyses.

Table 8.2

Word Triplets Contrastive in Their Lexical Pitch Patterns

Set	Word	1st-syllable accented	2nd-syllable accented	Unaccented
I Pretest Posttest Training	*aki*	秋 autumn	飽き weariness	空き opening
	hashi	箸 chopsticks	橋 bridge	端 edge
	kaku	核 core	画 stroke	格 status
	tsuru	鶴 crane	つる (蔓) vine	釣る to fish
	umi	海 sea	膿み pus	産み giving birth
	waki	和気 (藹々) harmonious	脇 side	沸き boiling
II Gen1 Gen2	*hari*	針 needle	梁 beam	張り straining
	kaki	牡蠣 oyster	垣 (根) fence	柿 persimmon
	maku	まく(播く) to sow seeds	膜 membrane	巻く to wrap
	mori	守 guard	漏り leaking	森 woods
	mushi	無視 to ignore	蒸し steaming	虫 insect
	yoku	良く well	欲 greed	翌 next

Note. Gen1 = test of generalization 1; Gen2 = test of generalization 2.

Three male and three female native speakers of Tokyo Japanese recorded materials for this experiment.[9] Speakers CH, AK and KC recorded the materials for pretest and posttest, gen1, and gen2, respectively (Table 7.3). Pitch-accent contrast

[9] Originally, seventeen native speakers of the Tokyo Japanese dialect who participated in Experiments 2 and 4 were asked to record materials for Experiment 5. These speakers were not professionally trained to follow the pronunciation norms as teachers or mass media workers. F0 contours in their productions were visually inspected in Praat (Boersma & Weenink, 2010), and categorized by the author as matching or not matching the accentual norms in the NHK dictionary (1998), using well-known acoustic-phonetic criteria (Chapter I, sections 1.4.1-1.4.3). Six speakers, who naturally produced the majority of the target words with accentual patterns consistent with the dictionary norms, were invited back to the laboratory to record the final version of training materials. Such selectiveness in participants in studies on Tokyo Japanese pitch accent is not uncommon; perhaps, it is due to a merger of accentual patterns in progress. For example, in his dissertation work, Ishihara (2006) selectively analyzed productions of only seven speakers out of eighteen.

produced by these speakers is displayed in Appendix C. As the graphs in Appendix C show, the speakers clearly produced the three-way contrast, though its phonetic realization slightly varied by speaker. Speakers RS, AK, TN and TM recorded the materials for training.

Table 8.3

Variability in Sentential Contexts and in Speakers

Words		Sentential context		
(Table 8.2)	#	Sentence	Word position	Speaker
I Pretest Posttest Training	1	__ *ga kakemas.* '(I) can write __.'	initial	CH-m (pretest & posttest, excluding #4)
	2	*Watáshi wa __ ga kakemas.* 'I can write __.'	medial	
	3	__ *ga ne*. 'It's __, you know.'	initial	RS-f, TM-m AK-f, TN-m (training)
	4	__ *ni chúuishite kudasai!* 'Please, pay attention to __!'	initial	
II Gen1 Gen2	5	__ *o kuríkku shite!* 'Click on ___.'	initial	AK-f (gen1)
	6	__ *ga hatsuon shiyasui?* 'Is ___ easy to pronounce?'	initial	KC-f (gen2)
	7	*Are wa __ na no?* 'Is that __?'	medial	

Note. Gen1 = test of generalization 1; Gen2 = test of generalization 2; m = male; f = female. Target words are followed by the particles (nominative *ga*, accusative *o,* dative *ni*, topical *wa*) or the copula (present tense *na*) that do not alter the location of the pitch accent in the target word. Pitch accent is marked with acute accent.

Seven carrier sentences were used to provide a sentential context (Table 8.3). Sentences 1-3 served as contexts at pretest and posttest for the target words of Set I

(Table 8.2). All of them were non-interrogative sentences with words followed by the nominative particle *ga*. Training materials utilized these three sentences and an additional Sentence 4. In Sentence 4, the structure was slightly varied in that the target words were followed by a new particle *ni*. In addition, the words were further from the end of the sentence than in other contexts, creating a larger memory load during the identification task.

Sentences 5-7 served as contexts at gen1 and gen2 tests for the target words of Set II (Tables 8.2 and 8.3). Besides the familiar particle *ga* (#6), the target words were immediately followed by the unfamiliar particle *o* (#5), and copula *na* (#7). Sentences 6 and 7 were interrogative. Sentence 7 was expected to be the most difficult context because the pitch patterns of the target words were immediately followed by the boundary tones, and thus closely integrated with the rising boundary pitch movement.

Appendix D and Appendix E show how lexical pitch patterns of target words were integrated into sentence intonation contours, using as an example the fully voiced target word *mori* spoken by one of the female speakers, AK. The sentences differed in the number of accentual phrases and the type of boundary tones (Chapter II, Table 2.1). Sentences containing more than one accented word could have had more than one accentual phrase. Sentences 3, 4 and 5 were likely to end in the boundary pitch movement called "insisting rise"; Sentence 6 in "information (top-range) question rise"; Sentence 7 in "incredulity (mid-range) question rise" (Eda, 2004; Venditti, Maeda, & van Santen, 1998). These boundary tones did not neatly correspond to the difference between declarative and interrogative sentences marked by the final pitch accent, which was exploited in previous studies (Ayusawa, 2003; Nishinuma et al., 1996). The variation in

prosodic phrasing and boundary pitch movements was impossible to control during the recording of stimuli, although speakers were asked to avoid within-sentence pauses.

In addition to structural differences in carrier sentences, another source of variability in F0 contours was likely to be due to individual speakers. Appendix A illustrates that the realization of the pitch-accent contrast in sentences may slightly vary by speaker. For example, in unaccented phrases some speakers do not show a consistent declination pattern, and males appear to have smaller declination rates than females, perhaps, due to the differences in pitch range (Poser, 1984; Pierrehumbert & Beckman, 1988, p. 70). Speakers with most declination may have about 10 Hz/sec decrease in the fundamental frequency after the phrasal high tone.

Recordings of the target words produced in the carrier sentences were made with a 22,050 Hz sampling rate on a flash card using Marantz PMD670/U1B recorder and a Shure SM81 microphone in a sound-attenuated booth. Then, the root-mean-squared amplitude of the sound files was equated to 75 dB.

8.2.3. Procedure

The task throughout the whole experiment was a three-alternative forced-choice categorization of pitch patterns. The presentation of the stimuli was in Romanized transcription for English participants and in Japanese orthography for Japanese participants. In each trial, first, a target word was displayed in isolation on the computer screen. The next slide showed the word in a carrier sentence along with three schematic diagrams of possible pitch patterns for English listeners (Figure 8.2), or along with three words of a triplet set for Japanese listeners (Table 8.2). The sound file of the sentence was played with a one-second delay after the visual display. English listeners were

instructed to identify the pattern of the target word by clicking the mouse on one of the three diagrams. Japanese listeners were instructed to identify the word that they heard by clicking the mouse on one of the words in a triplet set.

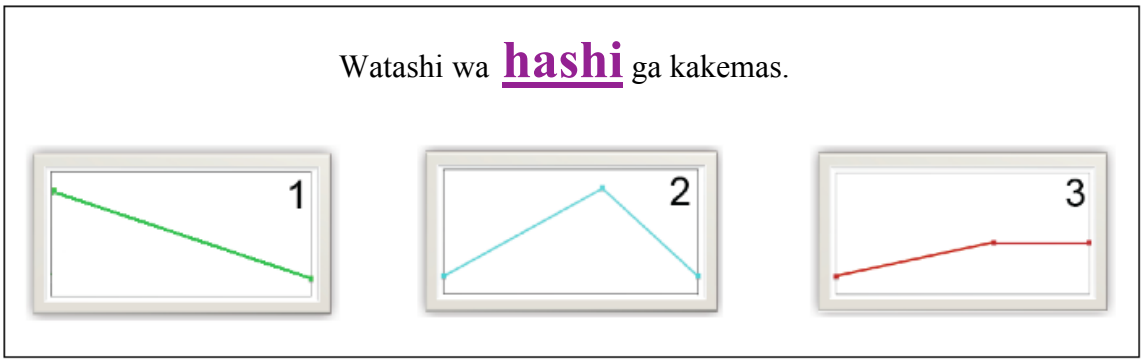


Figure 8.2. Visual presentation of stimuli to English listeners in the 3AFC task.

Response time was limited to four seconds in order to reduce individual variability in test taking strategies. One second after the click, the next target word was displayed on the screen (i.e., ISI = 1 s). Each trial lasted about 8-10 seconds. No breaks were allowed within a block, but listeners could take a short optional break between the blocks. The experiment was conducted using E-Prime software (Schneider, Eschman, & Zuccolotto, 2007) in a quiet room, one listener at a time. The audio was presented binaurally through the headphones at a comfortable listening level.

During the training, the first part of each trial was the same as described above, but in the second part English listeners received feedback on their answer choice (*Correct! The right answer was* __ or *Incorrect. The right answer was* __). The feedback sentence was displayed on the next slide along with a diagram (or a word) representing

the correct answer. After a one second delay, listeners heard the audio file of the trial one more time.

The experiment was conducted within one week in three sessions for the English training group and two sessions for the English control group (Table 8.4). Sessions were at least a day apart. At the beginning of the experiment, the distinctive function of pitch patterns in the three-way contrast was briefly described to the listeners. The participants took a pretest on the first day of the experiment, and a posttest, gen1 and gen2 tests on the last day. The order of the gen1 and gen2 was counterbalanced. Due to the limitations of recruitment procedures, the training component was relatively short. The total duration of training did not exceed one hour carried out over the course of three days; therefore, only a modest improvement in response accuracy of trainees was expected. The Japanese participants took a pretest, gen1 and gen2 in one day (in counterbalanced order) since they did not undergo training.

Table 8.4.

Overview of the Experiment Procedure for the Control and Training Groups

Session (day)	Procedure	Controls	Trainees	Sentence (Table 7.3)	Trial x Block (*n*)	Duration (min)
1	Pretest	✓	✓	1, 2, 3	54 x 3 (162)	22
	Training 1		✓	1, 2	72 x 2 (144)	20
2	Training 2		✓	3, 4	72 x 2 (144)	20
3	Training 3		✓	1, 2, 3, 4	18 x 4 (72)	10
	Posttest	✓	✓	1, 2, 3	54 x 3 (162)	22
	Gen1	✓	✓	5, 6, 7	54 x 3 (162)	22
	Gen2	✓	✓	5, 6, 7	54 x 3 (162)	22

Material presentation was blocked by sentential context; within each block the sentences were randomized (Table 8.4). The blocks were ordered in a way that allowed a gradual increase in difficulty as determined by sentence length and proximity of pitch accents to boundary tones. Tests consisted of three sentence blocks, each of which had 54 trials (1 speaker x 6 triplets x 3 accentual patterns x 3 repetitions). The first and second trainings had two sentence blocks made of 72 trails each (4 speakers x 6 triplets x 3 accentual patterns x 1 repetition). The third training served as a review of all training materials: In each of the four sentential blocks, listeners heard 18 trials (6 triplets x 3 accentual patterns). A short practice section preceded each block to familiarize the listeners with the context.

In addition to a standard language background questionnaire, information about musical experience of the participants was collected, that is, the number of years they played musical instruments, the number of musical instruments played, the number of years they sang in a choir or a band. In the end of the experiment, the listeners were also asked to rate the degree of their attentiveness during the tests as very attentive, attentive, occasionally not attentive, or not attentive. Finally, the listeners ranked the three speakers who produced the test materials – CH, AK, and KC (see also Appendix C) – in terms of perceived difficulty of performing the task.

8.2.4. Analysis

The percentage of correct responses (test: $N = 162$; pattern: $n = 54$; context: $n = 3$) was calculated by listener. These accuracy scores were used as dependent variables in all following analyses. The data of six Japanese listeners (pretest, gen1 and gen2 tests only) are presented for rough comparisons in figures and tables. These Japanese data were not

used in statistical analyses due to a relatively small number of the native listeners, and due to the focus on cross-linguistic perceptual training.

To investigate the effect of training on overall improvement in identification accuracy, an ANOVA was conducted on accuracy scores with between-subject factors of Group (control, training) and Level (high and low pretest scores), and a within-subject factor of Test (pretest, posttest, gen1, and gen2). Greenhouse-Geisser corrections for the degrees of freedom were used when the sphericity assumption was violated as indicated by significant Mauchly's tests (Box, 1954). Main effects were not reported in the presence of significant interactions, which were further investigated in within-subject contrast analyses. These tests examined the effect of Group or Level on the paired score at pretest and the other tests (i.e., three possible contrasts for each interaction).

To investigate the improvement by pitch pattern from pretest to posttest, an ANOVA was conducted with between-subject factors of Group (control, training) and Level (high, low), and within-subject factors of Test (pretest, posttest) and Pattern (1^{st}-syllable accented, 2^{nd}-syllable accented, unaccented). Significant interactions were further examined in tests of within-subject contrasts. In similar analyses, the difference in generalization to new stimuli spoken by a familiar speaker as compared to new stimuli spoken by a new speaker was investigated. Next, an error analysis was conducted for the pretest and posttest data. It included a detailed inspection of pattern-by-pattern responses for all participant groups, and chi-square tests for trainees only.

Lastly, to investigate what personal experiences correlated with the degree of individual improvement in trainees, two-tailed Pearson correlation analyses were conducted for pretest-posttest and pretest-gen1 score difference, and subject variables

such as GPA, experience with foreign languages and musical instruments, and self-reported attentiveness during the tests.

8.3. Results

8.3.1. Overall Improvement and Generalization

Figure 8.3 displays the mean percentages of correct responses by the Control Low, Control High, Training Low, and Training High participants at four tests. In addition, the right panel of the figure shows mean accuracy by six native listeners, who scored on average 75% correct responses at pretest, 84% at gen1 test, and 78% at gen2 test. As is evident from the comparison of Figures 8.3b and 8.3c, the identification accuracy was similar in the high-scoring groups of native English listeners, on the one hand, and in the native Japanese listeners, on the other hand, except for the gen2 test.

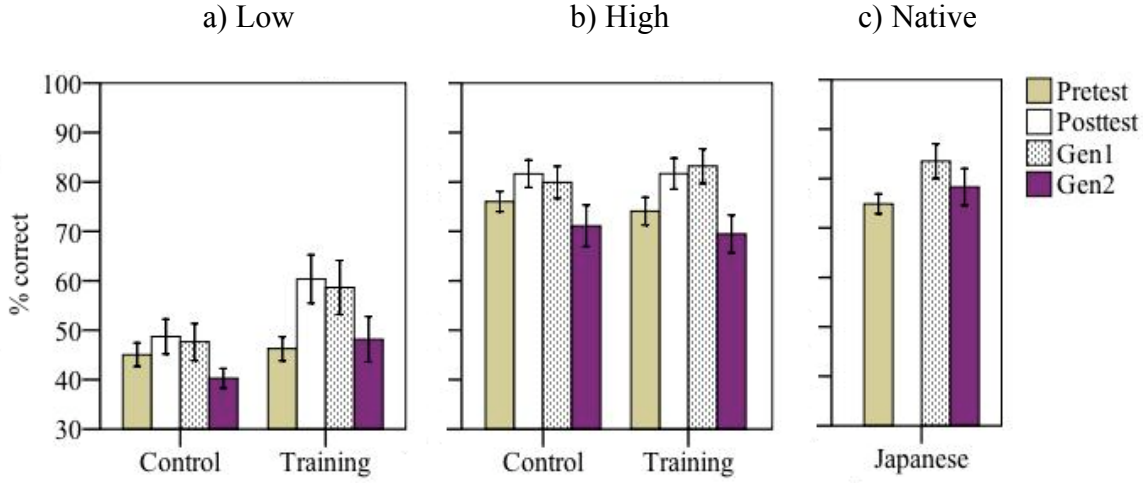


Figure 8.3. Identification accuracy in low-scoring and high-scoring English listeners (four tests), and in Japanese listeners (three tests). Error bars indicate ± 1 *SE*.

The accuracy scores of the English participants' were examined in a (2) Group x (2) Level x (4) Test repeated measure ANOVA. The analysis yielded significant main

effects of Level [$F(1, 37) = 81.97$, $p < .001$, $\eta_p^2 = .69$], and Test [$F(2.27, 111) = 26.57$, $p < .001$, $\eta_p^2 = .42$]. Level did not interact either with Group or Test. The interaction between Group and Test was marginally significant [$F(2.27, 111) = 2.61$, $p = .072$, $\eta_p^2 = .07$], and it was further examined in tests of within-subject contrasts. The tests showed that the score difference between controls and trainees (collapsed across Level) was significant in the comparisons of pretest and posttest [$F(1, 37) = 6.15$, $p = .018$, $\eta_p^2 = .14$], and pretest and gen1 test [$F(1, 37) = 5.92$, $p = .020$, $\eta_p^2 = .14$,], but not pretest and gen2 test [$F(1, 37) = .86$, $p = .360$, $\eta_p^2 = .02$].

These statistical analyses confirmed what is evident in Figures 8.3a and 8.3b. First, the high-scoring participants had higher accuracy in all tests compared to the low-scoring participants. Second, although the pretest scores were similar in the training and control groups (61% accuracy in each group, collapsed across level), the improvement in accuracy from pretest to posttest was greater in trainees than in controls (72% vs. 65%, respectively). This result suggests that a greater improvement is due to training rather than to mere exposure. Especially, low-scoring trainees appear to benefit from training according to Figure 7.3, though this result may be obscured in the omnibus ANOVA reported above, which did not yield a significant interaction between Level and Group. Third, trainees also outperformed controls at gen1 test (72% vs. 64%, respectively), which indicates that training on Japanese pitch patterns generalized to new words in sentential contexts spoken by a familiar speaker. However, there was no effect of training or exposure on listeners' accuracy at gen 2 test (trainees: 59%; controls: 56%). This indicates that relatively short experience with the pitch-accent contrast did not generalize

to new words in sentential contexts spoken by an unfamiliar speaker, which was arguably the most difficult test in the battery.

Given similar improvements between pretest-posttest and pretest-gen1, only pretest and posttest data of English listeners were compared in further statistical analyses of identification accuracy by accentual pattern and by sentential context.

8.3.2. Pretest-Posttest Improvements in Identification of the Three Pitch Patterns

Figure 8.4 shows the mean percentages of correct responses by the English listeners for each accentual pattern at pretest and posttest. In addition, the mean accuracy of six Japanese listeners at pretest and gen1 test is also shown for comparison. Visual inspection of the figure suggests that the unaccented pattern (Figure 8.4c) was relatively more difficult to identify correctly than the accented patterns (Figures 8.4a and 8.4b), even for native listeners of Tokyo Japanese.

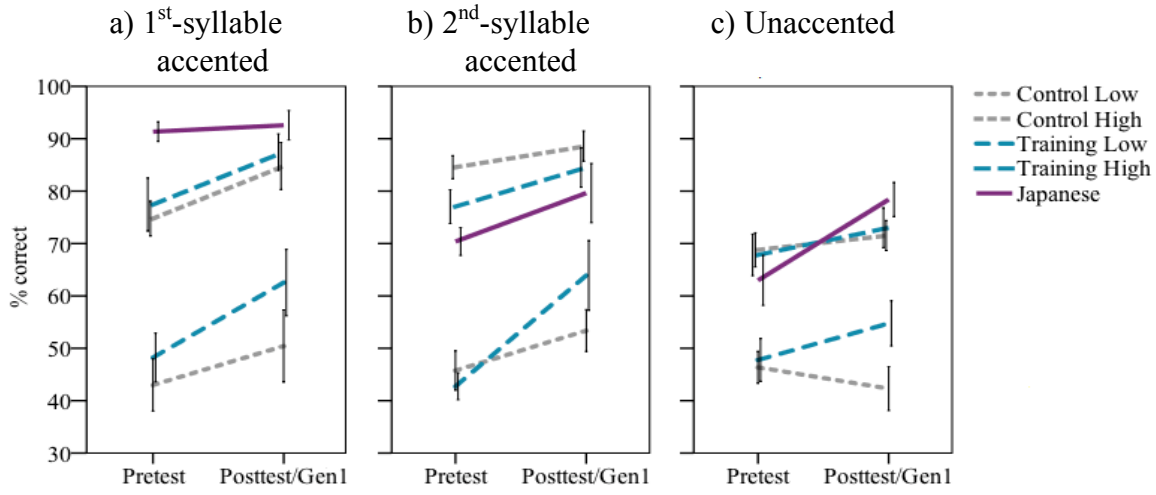


Figure 8.4. Identification accuracy by pattern in English listeners (pretest, posttest) and in Japanese listeners (pretest, generalization test 1). Error bars indicate ± 1 *SE*.

A (2) Group x (2) Level x 2 (Test) x (3) Pattern repeated measure ANOVA on percentage of correct responses by the English participants yielded only three significant interactions: namely, between Test and Group [$F(1, 37) = 6.17$, $p = .02$, $\eta_p^2 = .14$], Pattern and Test [$F(2, 74) = 3.94$, $p = .024$, $\eta_p^2 = .10$], and Pattern and Level [$F(2, 74) = 3.38$, $p = .040$, $\eta_p^2 = .08$]. The first interaction replicated the result reported in the previous section. The interactions between Pattern and the other two factors were further examined in the tests of within-subject contrasts.

For the Pattern and Test interaction, the effect of Test was significant for the comparison between the 1st-syllable accented and unaccented patterns [$F(1, 37) = 5.73$, $p = .022$, $\eta_p^2 = .13$], and between the 2nd-syllable accented and unaccented patterns [$F(1, 37) = 5.97$, $p = .019$, $\eta_p^2 = .14$]. These results taken together with the inspection of Figure 7.4 suggest that the three patterns were identified equally well at pretest; however, at posttest, the two accented patterns were identified with greater accuracy than the unaccented pattern.

As for the Pattern and Level interaction, the effect of Level was significant only for the comparison between the 2nd-syllable accented and unaccented patterns [$F(1, 37) = 8.94$, $p = .005$, $\eta_p^2 = .20$], but not for the 1st-syllable accented and unaccented patterns [$F(1, 37) = 2.87$, $p = .099$, $\eta_p^2 = .07$]. Figure 7.4 suggests that the high-scoring listeners were better in identification of the 2nd-syllable accented than the unaccented pattern, as compared to the low-scoring listeners.

Next, confusion matrices of the three accentual patterns were analyzed in order to examine the types of errors made by the listeners. Table 7.5 shows which patterns were confused with which in the five groups of participants. At pretest and gen1 test, the

Japanese listeners mostly confused the 2nd-syllable accented and unaccented patterns with each other. Additionally, the 1st-syllable accented pattern was sometimes misidentified as 2nd-syllable accented (6% in both tests). The unaccented pattern tended to have the lowest percentage of correct responses.

Table 8.5

Confusion Matrices of the Accentual Patterns at a) Pretest and b) Posttest (Gen1) for English (and Japanese) Listeners

Group	a) Pretest response (%)				b) Posttest (Gen1) response (%)			
	Pattern	1st acc.	2nd acc.	0 acc.	Pattern	1st acc.	2nd acc.	0 acc.
Japanese native	1st acc.	**91**	6	3	1st acc.	**93**	6	1
	2nd acc.	4	**70**	26	2nd acc.	4	**80**	17
	0 acc.	3	34	**63**	0 acc.	2	20	**78**
Control Low	1st acc.	**43**	34	23	1st acc.	**51**	29	21
	2nd acc.	29	**46**	26	2nd acc.	26	**53**	21
	0 acc.	25	28	**46**	0 acc.	26	32	**42**
Control High	1st acc.	**75**	14	12	1st acc.	**85**	9	7
	2nd acc.	6	**85**	9	2nd acc.	7	**89**	5
	0 acc.	3	29	**69**	0 acc.	2	26	**72**
Training Low	1st acc.	**48**	30	22	1st acc.	**62**	24	14
	2 acc.	35	**43**	23	2 acc.	24	**64**	12
	0 acc.	25	28	**48**	0 acc.	17	28	**55**
Training High	1st acc.	**77**	15	8	1st acc.	**88**	6	7
	2nd acc.	15	**77**	8	2nd acc.	7	**85**	8
	0 acc.	6	26	**68**	0 acc.	4	23	**73**

Note. 1st acc. = 1st-syllable accented; 2nd acc. = 2nd-syllable accented; 0 acc. = unaccented.

For the English listeners, pattern confusability seemed to have been affected by Level and Test. At pretest, listeners with a pretest score less than 60% correct (i.e., Control Low and Training Low) had much confusion among all three patterns. In

contrast, listeners with a pretest score larger than 60% (i.e., Control High and Training High) had the majority of errors misidentifying the unaccented pattern as 2nd-syllable accented. Similar to the Japanese data, the unaccented pattern had the lowest percentage of correct responses in the high-scoring groups. Differently from the Japanese data, the confusion between the two patterns did not seem to be mutual, that is the unaccented pattern was misidentified as 2nd-syllable accented at least three times more frequently than the other way around.

Table 8.5 shows that at posttest the English listeners improved in their identification accuracy of all patterns, except for the unaccented pattern in the Control Low group (46% and 42%, respectively). The difference in the confusion patterns between the high-scoring and low-scoring groups also seemed to be reduced in the participants who underwent training. To statistically examine this observation, the pretest and postest responses by the Training Low and Training High groups were tabulated to generate (3) Pattern x (3) Response x (2) Group contingency tables. Chi-square tests were conducted to assess the association between the expected and observed responses. The analyses showed an effect of Group on identifying the 1st-syllable accented as unaccented, [$\chi^2(17) = 28.33$, $p = .04$]; the 2nd-syllable accented as unaccented, [$\chi^2(18) = 28.19$, $p = .06$]; the 2nd-syllable accented as 1st-syllable accented, [$\chi^2(20) = 36.30$, $p = .01$]; the unaccented as 1st-syllable accented, [$\chi^2(19) = 35.30$, $p = .01$]. However, at posttest these differences disappeared between the Training Low and Training High groups (none of the chi-square tests yielded significant results).

The results presented so far suggest that identification of the unaccented pattern was the most challenging for the listeners, and that a short exposure or training was not

sufficient to improve its perception accuracy. This pattern tended to be confused with the 1st- and 2nd-syllable accented patterns, which indicates that the English listeners were not sensitive to the F0 fall that distinguishes the unaccented pattern from the accented ones. The mutual confusion of these two patterns by Japanese listeners suggests that the source of confusion for this participant group was not restricted to the acoustic cue of F0 fall. A conclusive interpretation, however, cannot be drawn from a small sample of six Japanese listeners who participated in this experiment.

Lastly, identification of accentual patterns in specific sentential contexts was examined for trainees only. The right panel of Figure 7.5 shows that at pretest and posttest the task was especially challenging for the unaccented pattern in two contexts, where the target word occurred in the sentence-initial position (i.e., squares and triangles).

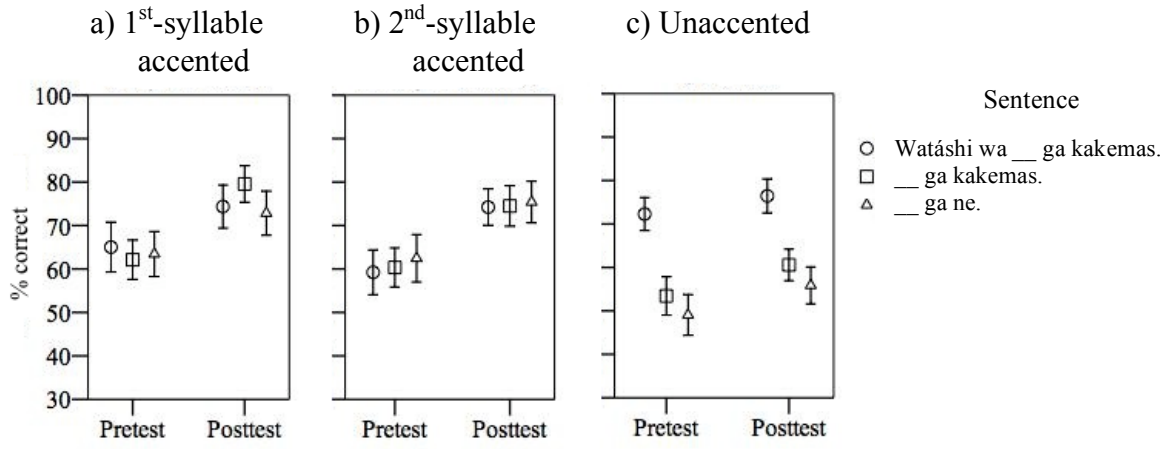


Figure 8.5. Trainees' identification accuracy in three sentential contexts (pretest, posttest). Error bars indicate ± 1 *SE*.

As compared to the *Watáshi wa ___ ga kakemas* context, unaccented words were identified significantly less accurately in the ___ *ga kakemas* context [$t(82) = 4.40$, $p <$

.01], and in the ___ *ga ne* context [$t(82) = 5.25$, $p < .01$]. This result may be attributed to additional acoustic cues that aid identification of the unaccented pattern in the sentence-medial position of *Watáshi wa ___ ga kakemas.* These cues may include a relatively small F0 reset in the beginning of the second accentual phrase of the *Watáshi wa ___ ga kakemas* sentence, when it starts from an unaccented word (compare the F0 maxima in the three patterns in Appendix D2).

8.3.3. Generalization by Pattern and Sentential Context in Trainees

Figure 8.6 provides another visualization of the identification accuracy by pattern, for trainees only. The controls' data is not reported separately here, because no interaction between Group and Pattern was found in section 8.3.2 above. Therefore, improvements by pattern were assumed to be similar in the controls and trainees.

To investigate the effects of level and pattern on trainees' accuracy at two generalization tests, a (2) Level x (3) Pattern x 2 (Test) repeated measure ANOVA was conducted on the gen1 and gen2 scores of the trainees. The analysis yielded significant main effects of Level [$F(1, 19) = 14.46$, $p = .001$, $\eta_p^2 = .43$], Pattern [$F(2, 38) = 8.48$, $p = .001$, $\eta_p^2 = .31$], and Test [$F(1, 19) = 88.30$, $p < .001$, $\eta_p^2 = .82$]. The interaction between Pattern and Test was also significant [$F(2, 38) = 4.26$, $p = .021$, $\eta_p^2 = .18$]. The interaction between Pattern and Level did not reach significance, although Figure 8.6 suggests that the high-scoring trainees were better in identification of both accented patterns than the low-scoring trainees.

Tests of within-subject contrasts for the Pattern and Test interaction showed that the only significant effect of Test was observed for the comparison between the 2nd-syllable accented and unaccented patterns [$F(1, 19) = 8.20$, $p = .010$, $\eta_p^2 = .30$]. Figure

8.6 suggests that the unaccented pattern was identified less accurately than the 2nd-syllable accented pattern at gen2 as compared to gen1 test. This may suggest a smaller benefit of training for generalizations to the unaccented pattern.

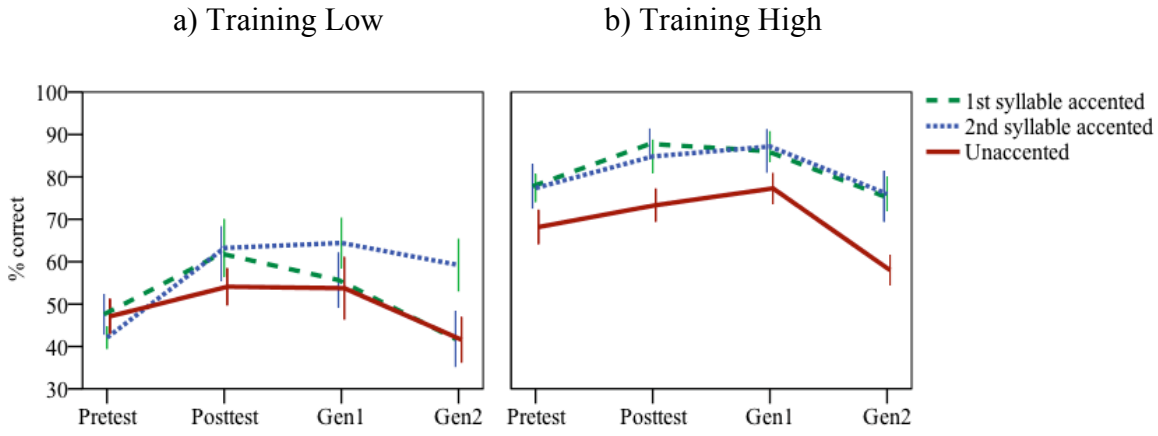


Figure 8.6. Identification accuracy by pattern in trainees across the four tests. Error bars indicate ± 1 *SE*.

Generalization to new sentential contexts was further examined for the trainees' data. Figure 8.7 shows that, in general, the identification accuracy in interrogative contexts (filled circles and squares) was comparable with non-interrogative contexts (open symbols and filled triangles). This suggests that training generalized to questions, where the presence of the boundary tone could have been distracting for the listeners. The mere presence of a rising boundary tone did not seem to affect the listeners' judgments.

Our prediction was that the context *Are wa _ na no?* might be the most difficult for the listeners because of the proximity of the rising boundary tone to the target word. Overall, this did not seem to be the case, which is evident from the comparison of filled circles with other symbols in Figure 8.7. Among the three patterns, the lowest accuracy

score in this context was observed for the unaccented pattern (right panel, filled circle, 67% correct).

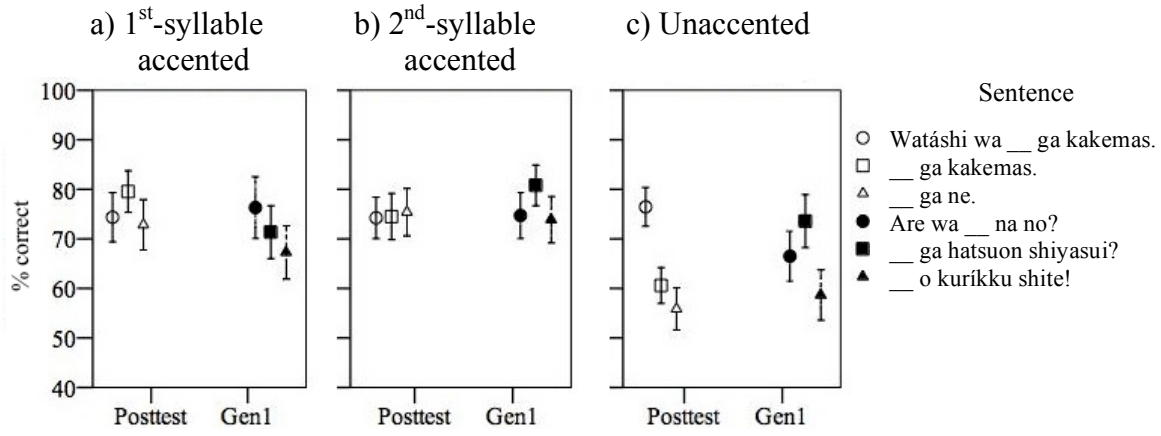


Figure 8.7. Trainees' identification accuracy in six sentential contexts (posttest, generalization test 1). Error bars indicate ± 1 *SE*.

Figure 8.7 displays that at both posttest and gen1 test, context mattered the most for the identification of the unaccented pattern (right panel) than the accented patterns (left and middle panels). It is interesting to note that in the contexts that yielded the highest accuracy in the right panel, the target word was in the sentence-medial position, separated from the final boundary tone by a relatively large amount of segmental material, and immediately followed by the unaccented words (*kakemas*: open circle, 77% correct; *hatsuon*: filled square, 74% correct). A comparison of F0 contours in Appendix D2 and E2 suggests that in these two sentential contexts, the sentence with the unaccented target word may be the most different over time from the sentences with the two accented target words.

The identification accuracy of the unaccented pattern was especially low in the context __ *o kuríkku shite!* (filled triangle, 59% correct). A comparison of F0 contours in the panel E1 with the other two panels in Appendix E suggests that over the course of the sentence, the unaccented pattern in this particular context is the most similar to the accented patterns. Notably, the target word in this context is almost immediately followed by another lexical pitch accent in the accented word *kuríkku* (the particle *o* has a very short duration, and the first vowel [ɯ] tends to be devoiced).

Taken together, these results suggest that the context, which follows an unaccented word, may influence the perception of the word's accentual pattern. The unaccented word pattern in relatively long unaccented accentual phrases (__ *ga hatsuon shiyasui?*) may be easier to categorize than the unaccented word pattern in accented accentual phrases (__ *o kuríkku shite!*). Although the context __ *ni chúuisite kudasái!* Is not included in Figure 8.5 (it was used for the training sessions only), it may be predicted that this context would be as relatively difficult as __ *o kuríkku shite!* for identification of unaccented word patterns.

7.3.4. Individual Differences in Trainees

The analysis of the pretest scores reported in the Method section revealed that there were large individual differences among listeners from the very onset of the experiment, which resulted in a subsequent subdivision of the listeners into high-scoring and low-scoring groups. The pretest scores significantly correlated only with the participants' major and GPA. No correlations between the scores and other subject variables were found at pretest (i.e., self-assessed learning and test-taking skills; the number of foreign languages studied and the length of study; the number of musical

instruments practiced and the length of practice; the number of years singing in a choir; the number of months spent abroad).

Another question regarding individual differences that begged exploration was whether individual improvement of the participants over the course of experiment correlated with any of these subject variables. Significant correlations would indicate which personal experiences were likely to be important for learning of the pitch accent patterns. Pearson correlation analyses explored this question for the trainees only, presuming that the difference between the posttest-pretest and gen1-pretest scores was indicative of improvement.

Table 8.6 below summarizes the percentage of individual improvement at posttest and gen1 test as compared to pretest, and some individual experiences of the trainees. Posttest and gen1 scores were significantly correlated [$r = .80$, $p < .01$], indicating that listeners who were better at pattern identification at posttest were also better at gen1 test. All individual experiences that significantly correlated with posttest-pretest improvement also correlated with gen1-pretest improvement, therefore only the latter ones are reported here. The gen1 test scores showed greater improvement for trainees who had studied more foreign languages [$r = .41$, $p = .03$], who played more musical instruments [$r = .47$, $p < .02$] for longer time [$r = .45$, $p < .02$], and who self-reported being more attentive during the tests [$r = .58$, $p < .01$]. A regression analysis on gen1-pretest score change with the number of foreign languages studied, the number of musical instruments played and attention during the tests as predictors showed that these factors combined together explain 44% of variance in the score gain.

Table 8.6

Relationship Between Individual Experiences and the Degree of Improvement in Identification of Pitch Patterns by Trainees

Group	Subject	Improvement (%) post-pre	Improvement (%) gen1-pre	Pretest (%)	GPA	Tonal l. (years)	FL* (#)	FL (years)	Instrument* (#)	Instrument* (years)	Attention* (degree)
Training Low	219	+5	-2	39	3.4		2	3	2	3	occasionally not
	216	+4	+1	43	3.5		1	3			occasionally not
	211	+5	+2	45	3.1		1	1			attentive
	218	+15	+2	49	3.2		1	8	3	13	very attentive
	215	+13	+7	49	3.1		1	8	1	5	attentive
	217	+13	+7	56	3		1	1			attentive
	220	+4	+12	34	3.4		1	3			attentive
	234	+21	+28	53	3.8		2	6	2	8	very attentive
	212	+14	+28	56	3.4		1	2	2	7	attentive
	223	+48	+41	39	3.9		3	4	2	6	very attentive
Training High	229	+1	-1	91	3.9		1	4			attentive
	221	+2	0	61	3.6	1	2	2			attentive
	222	+8	+3	77	3.1		1	2			very attentive
	235	+5	+3	61	3.8		1	8			attentive
	214	+2	+9	71	2.9		2	5			attentive
	230	+8	+10	77	2.4		2	7			attentive
	228	+7	+12	87	3.4	1	3	3	2	5	attentive
	226	+14	+12	77	3.8	0.5	3	2	1	3	attentive
	224	+14	+13	71	3.5	1	3	5	2	8	attentive
	213	+9	+18	74	3.5		1	5	3	10	attentive
	236	+13	+22	69	3.7		3	6	1	13	attentive

Note. FL = foreign language. Stars (*) mark experiences that correlate with post-pre and gen1-pre accuracy improvement.

8.4. Discussion

8.4.1. Learning on Average

The goal of this training experiment was to examine whether English listeners can learn to identify contrastive lexical pitch patterns in disyllabic Japanese words embedded in carrier sentences. The question of tone pattern learning by native speakers of non-tonal languages is of linguistic and pedagogical interest, and this experiment was the first to examine it in a pitch-accent language using the high-variability training paradigm. The high variability in training materials of Experiment 5 originated from three different sources – target words, sentential contexts, and speakers. During the experiment, listeners encountered sentences varying in their prosodic organization, acoustic implementation of the target prosodic contrast, and acoustic integration of the lexical contrast in higher-level prosodic groupings.

This high-variability approach proved to be effective for the English trainees largely unfamiliar with the Japanese pitch-accent contrast. The identification accuracy from pretest to posttest increased by 11% in the training group and by 5% in the control group, which was a significant difference between the groups. Similar improvement was observed from pretest to the gen 1 test, with new materials produced by a speaker familiar to listeners from the training sessions. This change was indicative of generalized learning. The finding that trainees significantly outperform controls was consistent with other HVTP studies. The magnitude of this improvement, however, was slightly smaller in the current experiment (Hirata, 1999: 18% for $[CV]^n$ words with contrastive pitch patterns; Wang et al., 1999: 21% for four Mandarin tones; Francis et al., 2008: 17% for six Cantonese tones). This is not surprising considering the limited length of one hour

training in this study (Hirata, 1999, and Wang et al., 1999: five hours; Francis et al., 2008: ten hours).

Training generalized to new stimuli spoken by a familiar speaker (gen 1 test), but not to new stimuli spoken by an unfamiliar speaker (gen 2 test). Two interpretations of this result are possible. First, short training was not sufficient for generalization to new voices; or second, the new talker, KC, had a less intelligible pronunciation of the target prosodic contrast. Although both interpretations are speculative, the latter one is supported by informal evidence. Namely, in a follow-up questionnaire, half of the English participants reported that the voice of KC was the most challenging for the task compared the voices of the other two test speakers, CH and AK.

The variation among native Japanese speakers in conveying the lexical pitch contrast has been reported before in the literature. For example, Cutler and Otake observed that one of the speakers in their experiment produced lexical pitch contrast in less differentiated way than the other two speakers, which corresponded to a lower identification accuracy of test words by native Japanese listeners (1999, p. 1881). Cutler and Otake speculated that this finding might be due to the relatively higher voice of the most difficult talker. In our training experiment, KC did not have the highest F0 range (Appendix C). However, her 1st-syllable accented pattern appeared to have a shorter high F0 plateau, as compared with the other speakers in Appendix C. Manipulation of the training time and speaker voices would be necessary to examine what factors influence the intelligibility of native speakers for non-native listeners.

8.4.2. Learning New Prosodic Categories

The participants in this training study learned to identify the 1st-syllable accented pattern (F0 fall after F0 peak in the first syllable), the 2nd-syllable accented pattern (F0 fall after F0 peak in the second syllable), and the unaccented pattern (no F0 fall). All three of these patterns were new prosodic categories for the English listeners, because English does not have lexical pitch-accent contrast manifested in F0 variations. Not all of these patterns were perceived and learned equally well. The current theories of second language speech perception would explain this result by similarity or dissimilarity of native English and non-native Japanese prosodic categories, (Best & Tyler, 2007; Flege, 1995; Kuhl & Iverson, 1995). Other types of explanation would be based on acoustic-phonetic similarity among the three patterns. In the following discussion, both types of explanation are attempted.

The 1st-syllable accented pattern was expected to be the easiest for listeners, because it was likely to map on the 1st-syllable stressed English pattern (Chapter II, section 2.2). However, this was not the case in our data. The identification of this pattern at pretest and posttest ranged from 43% to 88% correct, and it was not different from the identification accuracy of the 2nd-syllable accented pattern. The identification of both accented patterns improved with exposure and training.

The confusion between the 1st-syllable and 2nd-syllable accented patterns may result from a F0 peak delay in initially accented CVCV words in Tokyo Japanese (Ishihara, 2006; Sugito, 1982a). Although, peak alignment was not systematically examined in the training stimuli of this experiment, peak delay did occur in the materials (e.g., Appendix C2, Appendix D3). Thus, English listeners who heavily relied on the

acoustic cue of the F0 maximum in order to identify the pattern (the first language bias summarized in Chapter VII), could have mistaken 1st-syllable accented words for 2nd-syllable accented words. The pretest responses of low-scoring listeners to the 1st-syllable accented stimuli conformed to this error pattern (Control Low: 34%; Training Low: 30%). High-scoring listeners showed this error pattern to a much smaller degree (Control High: 14%; Training High: 15%). Thus, only the data of low-scoring listeners seem to agree with the summary in Ayusawa that at the beginning stage non-native listeners find it difficult to identify the location of the pitch accent on the first syllable (1999, p. 10). Higher-scoring listeners may have better sensitivity to the fine-grained F0 peak location.

The unaccented pattern is arguably the most difficult new category for English listeners, because in English every lexical word has prominence, reinforced by a pitch accent in the focus position (Chapter II, section 2.2). The target words in the training materials were likely to be in focus since they were varied (new) information in each sentential context. The analyses of listeners' responses confirmed that the unaccented pattern posited the greatest challenge for English listeners. First, the identification of this pattern was the worst before training (i.e., at pretest) in the high-scoring listeners. Second, it improved the least as compared to the accented patterns at posttest. Third, the exposure to this pattern generalized the least at gen 2 test. Finally, the identification of unaccented pattern was affected the most by sentential context.

This finding of a relative difficulty of the unaccented pattern is inconsistent with the results of the studies that utilized the Tokyo Japanese Accent Listening Test (Chapter II, section 2.6.2). It has been reported in those studies that the unaccented pattern was the easiest to identify in words of different length. This inconsistency may be attributed to

methodological issues, such as the nature of the task, which may lead to listeners' response bias, as reviewed in detail in the introduction. Possible individual and task-induced biases in listeners' responses need to be explicitly discussed.

In Experiment 5, the unaccented pattern tended to be misperceived as the 2nd-syllable accented. This indicated that native English listeners ignored the acoustic cue of the F0 fall after the F0 peak, crucial for distinction of thee two patterns (Chapter II, section 2.4.1; see also the results of the perception experiments summarized in Chapter VII). In both of these patterns, the F0 peak tends to be aligned with the 2nd syllable of a word. Thus, their F0 contours are somewhat similar to English 2nd-syllable stressed pattern. In terms of the Perceptual Assimilation Model, poor discrimination of Japanese unaccented and 2nd-syllable accented patterns may be explained by their Category Goodness assimilation to the corresponding English pattern (Best & Tyler, 2007, p. 23). Perhaps, the 2nd-syllable accented pattern was perceived as a better exemplar, and the unaccented pattern was perceived as a poor exemplar of the English category. The confusion pattern analysis supports this interpretation.

There are several proposed explanations for why English listeners do not utilize the F0 fall in categorization of Japanese pitch patterns. Some of these explanations are related to first language attentional biases. A relative inattentiveness of English listeners to F0 change was discussed elsewhere in the dissertation (Chapter II, section 2.6.3.2; Chapter VII). Another bias would be a suggestion that in English stress prominence involves *increase* in the F0, whereas in Japanese tone prominence involves F0 *decrease* (Nishinuma et al., 1996). Yet another bias would be resulting from a cross-linguistic mismatch of the domain and degree of F0 decrease such as the lexical falling tone in

sentence non-final positions confusing English listeners who are used to the falling boundary tone in sentence-final positions of declarative intonational pattern in English (Broselow, Hurtig & Ringen, 1987). The other type of explanation is articulatory motivated. It may be the case that a falling F0 movement is implemented faster than a rising F0 movement (Ohala, 1978, p. 30; Xu & Wang, 2001, p. 322). Therefore, it is more difficult to attend to F0 fall than F0 rise.

8.4.3. Individual Differences Among Adult Learners

Because the averaged group performance is the most common way to summarize the data and describe trends in populations, individual differences among participants are often overlooked. Persistently, however, a large individual variability in listeners with similar language background has been observed in speech perception and training studies. For example, in work by Iverson and Evans, the range of correct identification of English vowels was 30-83% for native Spanish listeners and 42-89% for native German listeners (2009, p. 868). In the training study by Wang et al., the range of Mandarin tone identification scores was 62-89% in trainees and 42-85% in controls (1999, p. 3655). In a study on perception of linguistic and non-linguistic pitch variation by Bent et al. (2006), naïve English listeners identified Mandarin tones with the accuracy ranging from 26% to 88% correct. In a study on perception of Japanese pitch patterns, Nishinuma et al. (1996) had to divide their 54 American English listeners, all of whom had learned Japanese for two years, in three groups due to individual variability in test performance. The average identification accuracy in these three groups was 42%, 59% and 73%.

In line with these findings, the results of the current experiment showed that the English native listeners varied significantly in their identification accuracy of Japanese

pitch patterns. The differences emerged from the very onset of the study (i.e., pretest), when, to the author's surprise, some English listeners were able to achieve identification accuracy comparable with native Japanese listeners. This excellent performance could not be explained by the history of learning Japanese or other tone language (similar summary of several studies was reported in Ayusawa, 1999, p. 9). The participants were subsequently divided in high-scoring and low-scoring groups, in which the mean pretest score (pooled across trainees and controls) was 75% for high- and 46% for low-scoring participants. In all four tests of Experiment 5, the quantitative difference between the accuracy scores of the two groups was maintained.

Furthermore, the difference between these groups was also qualitative, at least, in some tests. Namely, at pretest, the group difference appeared to be associated with the specific patterns of accent-type confusion. The high-scoring participants tended to erroneously identify the unaccented pattern as 2nd-syllable accented pattern, whereas the low-scoring participants confused all three patterns among each other. Note, that the unaccented and 2nd-syllable accented patterns are mainly distinguished by the acoustic cue of the F0 fall. This observation may be related to the result reported by Chandrasekaran, Sampath and Wong (2010), who found that before and after training, higher-scoring learners attended more to pitch direction of Mandarin tones than lower-scoring learners. These findings suggest that listeners differ in their individual selective attention to the acoustic cues related to F0 change. Why these individual attentional biases are formed and how they interact with first language attentional biases (such as reported in Gandour and Harshman, 1978) is an interesting topic for future research.

Individual attentional biases may account for the large variability not only in non-native listeners, but also in the perception of native listeners who are asked to identify categories in the prosodic domain. This hypothesis is supported by the results reported in Experiment 2, where both native and non-native listeners of Tokyo Japanese exhibited a great variability in using the F0 fall and F0 peak location cues for categorization of pitch patterns. In the current Experiment 5, the overall accuracy of high-scoring English listeners was similar to Japanese native listeners. The fact that the Japanese native listeners did not score at ceiling may be surprising, however, similar results have been reported before. For example, the average accuracy of native speakers in the Tokyo Japanese Accent Listening Test was reported to be 85% (Ayusawa, 1999), which was far from being at ceiling. In a study on a different tone language, Cantonese, Khouw and Ciocca (2007) found that native Cantonese listeners had 58%-70% perception accuracy in identification of level tones, and 75%-84% accuracy in identification of contour tones. Such findings point toward individual attentional biases in native speakers that introduce variability in perception patterns of seemingly homogeneous populations.

To conclude, apart from native language experience, other factors seem to influence listeners' perception of prosodic categories. Traditionally explored subject variables such as previous experience with pitch in one's first and second languages (including tone languages), as well as experience with music and musical instruments may not correlate with pre-existing differences in adult listeners, as reflected in pretest scores. On the other hand, these previous experiences correlate with the degree of individual improvement in training. In this experiment, trainees who self-reported to be

more attentive during the tests and who had experience of studying more languages and playing more musical instruments had the largest gain in their scores after training.

The question of pre-existing individual attentional biases is important for a better understanding of second language speech perception by adult learners. When facilitative of learning of a particular second language, these biases may be referred to as talent or language skills, as in Strange and Shafer (2008):

> Most studies of adult L2 learners' perception of non-native phonetic contrasts report a considerable range in the performance of individual participants that cannot be easily correlated with other subject variables listed here; these individual differences have often been labeled talent (182).

The holistic notion of a talent to languages is somewhat unsatisfactory, because it does not allow us to tease apart subject variables, which are particularly influential for individual performance in language tasks. For example, subject variables unexplored in the current experiment, such as working memory, phonological memory, attention span, and attentional orientation, could be good predictors of individual differences. The challenge of research on individual differences in speech perception lies in careful manipulation of such variables.

CHAPTER IX

GENERAL DISCUSSION AND CONCLUSIONS

9.1. Second Language Perception of Prosodic Contrasts

The results reported in Chapters 3-8 indicate that cross-linguistic perception of word-level Japanese pitch patterns can be accounted for by 1) general auditory processes; 2) tasks and materials that listeners are engaged with; 3) first language attentional biases with respect to acoustic cues; and 4) individual attentional biases. Specifically, the same-different discrimination task in Experiment 4 demonstrated both native and non-native listeners were sensitive to the F0 fall, and this auditory sensitivity did not vary by location in the word. The three categorization tasks conducted in Experiments 1-3 showed that slight differences in experimental design and instructions influenced the responses of listeners. A first language attentional bias with regard to the F0 fall and its interaction with the F0 peak location was evident from all perception experiments. Individual biases analyzed in Experiments 3 and 4 showed a large variability in attention to the relevant acoustic cues in both native and non-native listeners. These biases also surfaced in the results of Experiment 5 as a dramatic performance difference between the high-scoring and low-scoring non-native listeners. The effects of these four major factors on cross-linguistic perception of a prosodic contrast, such as Japanese pitch patterns, are discussed below.

9.1.1. Auditory Representations of F0 Contours

The representation of the F0 change in the auditory system may affect perception of this acoustic property as a cue to prosodic categories (the idea similar to Holt & Lotto, 2006). Auditory biases are expected to be similar across languages. They may be implied

from comparable performance of native and non-native listeners in perception tasks that encourage psychoacoustic mode of stimulus processing (e.g., AX discrimination with relatively short ISI). Experiment 5, in which such task was used, showed that both Japanese and English listeners were sensitive to a particular kind of F0 change, that is, F0 fall. This sensitivity was similar at different temporal points of a three-syllable nonword. Although the sensitivity to F0 fall was heightened in native listeners (85% accuracy, $d' = 2.53$) as compared to non-native listeners (72% accuracy, $d' = 1.43$), it seems reasonable to hypothesize that F0 fall has a robust auditory representation in both listener groups.

F0 fall is related to universal phonetic processes such as declination. F0 decrease over a course of an utterance can be observed in many, if not all world languages, although the cause of this common pattern is under dispute (Ohala, 1978, pp. 31-32). F0 decrease may be phonologized as an F0 fall cue to lexical tones, pitch accents, and boundary tones (Table 2.1, high-low tone sequences). Some evidence suggests that falling and rising tones differ in their typical F0 range and in the rate of F0 change (see review in Ohala, 1978, pp. 30-31). These articulatory differences may lead to different auditory representations of the falling and rising F0 contours. If so, the differences in auditory processing of these contour types should be observed across languages.

It appears that listeners may differ in discrimination and identification accuracy for falling versus rising (or high plateau) F0 contours. In Experiment 5 in this dissertation, the accented Japanese patterns with F0 fall had relatively higher identification accuracy than the unaccented pattern without F0 fall, a result common for both native and non-native listeners. Masaki, Takasawa and Arai (2001) also found that native Japanese listeners were more sensitive to falling than rising synthesized F0

contours. This conclusion was drawn from categorical perception of falling F0 movement, which was consistently perceived as the accented pattern, versus non-categorical perception of rising F0 movement, which was accompanied by inconsistent responses of listeners.

Interestingly, in languages like Mandarin where categorical perception of F0 fall is also required to distinguish between level and falling lexical tones, the falling tone 4 has been reported to have the highest accuracy in both native and non-native listeners, which may be interpreted as a response bias (Lee et. al., 2009). The rising tone 2, on the other hand, has been reported to have the lowest perception accuracy (Broselow et al., 1987; Lee et. al., 2009; Wang, 1999). Further cross-linguistic studies are needed to confirm that rising and falling F0 contours may have different auditory representations, which would introduce similar across listeners response biases to speech and non-speech F0 variations.

9.1.2. Task Effects on Listeners' Performance

Task effects on perception of sound categories have been described in many studies, and their overview is provided by Strange and Shafer (2008, pp. 160-168). Methodological variables may influence native and non-native listeners differently, which is a challenge for the tasks targeting psychoacoustic mode of stimuli processing. For instance, Francis and Ciocca (2003) reported that, although similar sensitivity to the differences between Cantonese tones was observed in native and non-native listeners, the groups performance differed with respect to the stimuli presentation order. For native listeners, mean sensitivity across nine pairs of Cantonese tones was twice as large when in tone pairs low F0 stimulus was followed by high F0 stimulus than vise versa. Non-

native listeners were not affected by the order of stimulus presentation. The authors interpreted these results as language-specific effect of the stimulus presentation order, observed in addition to psychophysical effect. One of the lessons here is that all possible task variables should be carefully considered for encouraging auditory versus phonological mode of stimuli processing.

In the current dissertation, the following task effects were observed. First, the responses of non-native listeners were more comparable with the responses of native listeners in discrimination than in categorization. This result is consistent with previous findings of relatively high auditory sensitivity of non-native listeners to acoustic properties in less language-like tasks such as discrimination (Bent, 2005; Eda, 2004; Halle et al. 2004; Miyawaki et al., 1975; Sakamoto, 2008; Wayland & Guion, 2003).

Second, a comparison of the results in the three categorization tasks showed that slight variations in the experimental design influenced the responses of non-native listeners to a larger degree than the responses of native listeners (Table 7.1; Figures 7.1-7.2). Native Japanese listeners consistently attended to the F0 fall and its interaction with the F0 peak location for categorization of F0 contours. Naïve English listeners were more sensitive to the F0 fall when they had a No Prominence answer option in addition to the Syllable 1 Prominent and Syllable 2 Prominent (Experiment 3). No such sensitivity was observed in the two-alternative forced choice tasks (Experiments 1-2), perhaps because they encouraged listeners to map Japanese prominence patterns on English prominence patterns, and thus, to disregard the F0 fall cue. Furthermore, English listeners' sensitivity to the interaction between F0 fall and F0 peak location was increased when the magnitude of the fall was manipulated in eight steps via the height of the starting point of

the fall (Experiment 1), but not when it was manipulated in just three steps via the height of the fall endpoint (Experiments 2-3). Note that the magnitude of the fall was comparable in all experiments. Thus, the number of categories that listeners had to sort the stimuli into, and the manipulation details of parameters affected categorization responses.

Third, the results regarding identification accuracy of the unaccented pattern as compared to the accented patterns differed in Experiment 5 in this dissertation and in previous studies that adopted the methodology of the Tokyo Japanese Accent Listening Test. In Experiment 5, the unaccented pattern was the most challenging for listeners, whereas in the representative work by Nishinuma, Arai and Ayusawa (1996) the unaccented pattern was the easiest. This result inconsistency may be explained by differences in the tasks. In Experiment 5, listeners were asked to identify a pattern in disyllabic words followed by particles, whereas in the Tokyo Japanese Accent Listening Test listeners were asked to identify presence of the pitch accent and mark its location in words of varied length. I argued that the latter methodology introduced a response bias towards the unaccented pattern, which would lead to higher identification accuracy of this pattern as judged by percentage of correct responses.

Lastly, sentential contexts of the target words and speakers' voices or enunciation of the target prosodic contrast also influenced non-native listeners' responses (Experiment 5). These effects are discussed in sections 8.3.2 and 8.3.3.

In addition to the task variables described above, variables specific to research on prosody as compared to research on segments need to be considered. Such variables include prosodic groupings that serve as domains for the investigated prosodic

phenomena, the cumulative nature of many prosodic categories, and metalinguistic (non-) awareness of listeners with regard to prosody. The strategy adopted in this study was to assume that any word uttered in isolation constituted an utterance, and that the category of prominence was more feasible for cross-linguistic comparisons than the categories of accentual patterns and stress patterns.

9.1.3. Native Language Effect on Perception of Prosody

Native language experience alters the perception of sounds by mature speakers of the language. This effect has been observed in many studies on perception of segments: Korean stops (Francis & Nusbaum, 2002); Hindi obstruents [t̪ʰ-ʈʰ] (Guion & Pederson, 2007); French vowels (Gottfried & Beddor, 1988); English vowels (Iverson & Evans, 2009); English /ɹ/-/l/ (Bradlow et al., 1997, 1999; Iverson et al., 2003, 2005; Lively et al, 1993; Logan et al, 1991; Miyawaki et al., 1975;), English /t/-/d/ (Flege, 1989), English /ð/-/θ/ (Jamieson & Morosan, 1986, 1989), among many other segmental contrasts. A similar effect has been reported for perception of lexical tones: in Thai (Wayland & Guion, 2003); in Thai and Yoruba (Gandour & Harshman, 1978); in Cantonese (Francis & Ciocca, 2003; Francis et al., 2008); and in Mandarin (Guion & Pederson, 2007; Wang et al., 1999). Japanese lexical pitch patterns are also subject to first language attentional biases as Experiments 1-4 in this dissertation have demonstrated, and as has been shown in previous studies with different methodologies (Nishinuma et al., 1996; Sakamoto, 2008; Shport; 2008).

9.1.3.1. Categorical perception in prosodic domain. First language attentional biases are often investigated in the paradigm of categorical perception, which includes both identification and discrimination tasks. An identification task is used to show a sharp

category boundary. A discrimination task is used to examine whether the boundary corresponds to a discrimination peak, which suggests that listeners ignore phonetic variation within the categories, and tune in to phonetic variation between the categories. Although not without dispute, consonants are thought to be perceived in a more categorical manner than vowels. Ongoing work on perception of tones suggests that native speakers of tone languages perceive both speech and non-speech F0 variations categorically, whereas non-native speakers show categorical perception only for non-speech (Xu et al, 2006). Furthermore, level tones may be perceived in a more categorical fashion than contour tones by native speakers (Francis et al., 2003).

To my knowledge, categorical perception of Japanese lexical pitch patterns has not been tested in this classic paradigm. In discrimination and categorization tasks in this dissertation, at least two acoustic parameters were varied – F0 fall and F0 peak location. The results suggest, however, that native speakers' perception of F0 contours was more categorical than non-native speakers' perception, in the sense that category boundaries were better defined along the two acoustic dimensions. As Figures 7.1 and 7.2 show, in both 2AFC and 3AFC tasks the stimulus space was categorized by native Japanese listeners more consistently than by English listeners. This tendency was especially clear in the 3AFC task, in which English listeners failed to discern three prosodic patterns and gave even more inconsistent answers than in the 2AFC task. This result supports the proposal in Hallé, Chang and Best (2004) that English listeners may perceive lexical tones as uncategorized speech or non-speech categories. However, such findings do not speak to the classic idea of categorical perception along one acoustic dimension such as VOT, F3, or duration.

9.1.3.2. Learned attentional bias toward specific acoustic information. The most investigated bias in perception of lexical tones by native listeners of non-tone languages is insufficient attention to F0 change. Since the seminal work of Gandour and Harshman (1978) showing that Thai and Yoruba listeners attached relative importance to the direction of the F0 change and its slope in their perception of Mandarin lexical tones, whereas English listeners did not, other studies reported similar findings for a variety of languages. Experiments 1-3 in this dissertation showed that this result holds for Japanese lexical pitch patterns as well.

The specific case of F0 change investigated in our experiments was the falling F0 movement, which is the primary acoustic correlate in production and the primary acoustic cue in perception of the lexical pitch accent H*+L in Japanese. The results showed that native Japanese listeners better attended to this acoustic parameter than native English listeners in the categorization tasks, and had a higher auditory sensitivity to the F0 fall in the discrimination task. Thus, with regard to F0 fall, native listeners outperformed non-native listeners in both language-like categorization and auditory discrimination.

The perception experiments reported here also have shown that the most meaningful description of the first language attentional bias for Japanese pitch accent patterns is based not just on one acoustic dimension of the F0 fall, but on a combination of two dimensions, namely the F0 fall and the F0 peak location. The perceptual integration of these cues was inferred from significant interactions between the manipulated parameters in the categorization experiments. This integration was likely to be language-specific, because no interaction between the F0 fall and the F0 peak location

was observed in the discrimination experiment, suggesting equal sensitivity to the fall across a word in a less language-like task.

Cross-linguistic differences in perceptual integration of acoustic dimensions are well known from studies on segmental and prosodic contrasts that involve several acoustic correlates. Examples include co-varying duration and spectrum in French /o/-/ɔ/ (Gottfried & Beddor, 1988), English /iː/-/ɪ/ (Morrison, 2005), and Dutch /aː/-/ɑ/ (Escudero et al., 2009); co-varying F2 and F3 in English /ɹ/-/l/ (Iverson et al., 2003); co-varying F0 height and direction in stops (Francis & Nusbaum, 2002); co-varying direction of F0 change, its slope and length in lexical tones (Gandour & Harshman, 1978); and co-varying F0 peak location and F0 fall in this study.

Previous studies have convincingly demonstrated that perceptual integration of the acoustic cues to a given contrast does not simply follow from acoustic properties of the sounds pertaining to the contrast. Such perceptual integration, or the lack of it, is language-specific. For instance, in perception of the native contrast /kot/-/kɔt/, French listeners show no evidence of the integration between spectral and temporal information, even though both spectrum and duration are physical characteristics of this vowel contrast (Gottfried & Beddor, 1988). English listeners show perceptual integration of this acoustic information, resembling another native contrast /iː/-/ɪ/. In the same vein, in perception of word-level pitch patterns, Japanese listeners show evidence of perceptual integration between F0 fall and F0 peak information, whereas English listeners do not. Thus, the presence of this perceptual integration constitutes one of the first language attentional biases.

9.1.3.3. Between-category similarity as an explanation of attentional biases.

Cross-linguistic perception of acoustic properties is often assessed in terms of phonetic and phonemic similarity between native and non-native sound patterns. It has been proposed that the more dissimilar is a non-native sound to the closest native sound, the easier it is to learn (Aoyama et al., 2004; Best & Tyler, 2007; Flege, 1995). In perception of segments, such cross-language comparisons are easier to make, because segments belong to the same level in the phonological structures of languages. In perception of prosody, a prediction based on cross-linguistic similarity is more difficult to test, because the cues are often used in both native and non-native languages at different levels of the prosodic hierarchy. For instance, in English, the variation in F0 has linguistic significance at the level of the phrase. In Japanese, or other tone languages with contrastive lexical pitch patterns, the variation in F0 is linguistically important at both the word and phrase levels.

For cross-linguistic research in the prosodic domain, the first hypothesis is that phonetic experience with F0 variations is important regardless of the prosodic grouping where such variations occur. Theories of segment perception make a similar prediction that phonetic familiarity with acoustical properties of a non-native sound may facilitate learning of that sound. Under such a hypothesis, the unaccented pitch pattern in Japanese is perhaps the most dissimilar, and the accented patterns are the most similar to English intonational contours. The alternative hypothesis is that the phonological prosodic grouping, in which a particular F0 variation occurs, matters for cross-linguistic perception more than phonetic experience with that F0 variation. This prediction is not the same as the effect of phonemic experience predicted by theories of segment

perception, because prosodic groupings may belong to different levels of the prosodic hierarchy. Under such a hypothesis, the notion of similarity becomes impractical for the cross-linguistic perception of lexical pitch patterns. These two hypotheses need to be tested for the prosodic domain, as they were for segments (e.g., Polka, 1992; Werker & Tees, 1984).

As far as the interpretation of the results is concerned, whether similar F0 variations such as rises and falls are cross-linguistically comparable or not is typically a judgment call of a researcher. When such comparisons are made between English and tone languages, the experience of English listeners with boundary tones (e.g., F0 falls in statements and F0 rises in yes-no questions, or lists) and universal phonetic processes such as declination (i.e., F0 decrease across an utterance) are often appealed to.

For example, Broselow, Hurtig and Ringen (1987) found that the falling tone 4 in Mandarin is the easiest to identify in isolation, but it returns the poorest identification accuracy rate by English native listeners when it occurs in non-final sentence position. The authors propose that sentence-final tone 4 is similar to a familiar declarative intonation pattern in English, hence, better identification accuracy. Shen (1989), Wang et al. (1999), and So and Best (2010) also found that the Mandarin tone 4 is one of the most challenging for English learners, which may be attributed to the perceptual interference between the non-native lexical tone and native boundary tones. So and Best further suggested that lexical tones perceived as uncategorized speech categories might assimilate to English intonation patterns, emotion intonation categories, or even rhythmic patterns. In a similar vein, Hallé et al. (2004) proposed that the Mandarin tone 4 could suggest a forbidding intonation to a French listener, and the rising tone – a stunned

intonation (p. 418). Although these prosodic properties cannot be directly mapped on each other due to the difference in the domain – lexical in Mandarin and phrasal in French, – the similarities are still alluded to in the discussion of findings.

Interestingly, another common finding in the well researched topic of Mandarin lexical tones is the confusion of tone 4 and tone 1, which have similar high pitch level at the beginning and differ along the dimension of F0 fall, namely, the sharp fall in the former tone but not in the latter. Such findings suggest that phonetic similarity between non-native prosodic categories is another major factor in listeners' judgments. A similar result was obtained in Experiment 5 in this dissertation, where the patterns similar in their global F0 contours tended to be confused the most, that is the 2nd-syllable accented and unaccented patterns. To further investigate the effect of this similarity, acoustic cues to the patterns need to be compared in terms of their distributional properties.

9.1.3.4. Within-category characteristics as an explanation of attentional biases. Holt and Lotto (2006) investigated two distributional properties of acoustic cues that might affect cue weighting, namely, cue informativeness and variance. Informativeness was defined as relative increase in accuracy over chance performance that listeners would demonstrate by using a particular cue. The researchers found that within-category distribution variances led to a substantial change in cue weighting in perception of the category, whereas the central tendencies of the distributions (with concomitant changes in informativeness) did not.

This finding would be important to test for Japanese pitch patterns, as a follow up of the training Experiment 5, where multiple words were produced in different sentential contexts by several speakers as training and testing materials. A detailed analysis of the

acoustic correlates of lexical pitch patterns in these materials was not conducted, keeping the chapter focus on learning. However, in addition to the similarity explanation between the patterns of prominence in Japanese and English, uneven learning of the 1st-syllable accented, 2nd-syllable accented and unaccented categories may be also explained by the distributional properties of the acoustic properties: rise to the lexical H* tone or phrasal H- tone, the location of F0 maximum, the value of F0 maximum, and the magnitude of the subsequent F0 fall. This investigation may be especially revealing because the native listeners in the Experiment 5 did not reach the accuracy ceiling and showed uneven accuracy of responses for the three patterns, similar to non-native listeners.

9.1.4. Individual Attentional Biases in Perception of Pitch Patterns

Apart from the first language attentional biases, individual attentional biases to particular acoustic cues of the pitch-accent contrast became apparent over the course of this dissertation project. The individual differences among listeners were analyzed in the categorization Experiment 2, discrimination Experiment 4, and training Experiment 5. The results of the Experiment 2 suggested that individual variability in attention to the F0 fall and F0 peak location was large not only in non-native but also in native listeners (Figures 4.4 and 4.5). Individual biases exhibited group patterns in native but not non-native listeners. Only a quarter of the native Japanese listeners were heavily relying on the F0 fall cue in the categorization task, and half of the listeners relied on the integrated perception of the F0 fall and its location in the word.

These individual differences manifested themselves in the discrimination task as overlapping distributions of the *d′* scores. This means that some non-native listeners were as sensitive to the F0 fall as native listeners. The sensitivity in the native group, although

higher than in non-native group, was not at ceiling (consistent with Eda, 2004). Neither was the identification of real word pitch patterns in the training experiment.

The most unexpected result of the training experiment was that at pretest the high-scoring English listeners had identification accuracy similar to the native listeners, and the low-scoring English listeners were close to the chance level. In contrast to the degree of training-induced learning, these pre-existing individual differences did not correlate with experiences of learning second languages, including tone languages or Japanese, musical training, or with other subject variables. It seems implausible that the high-scoring listeners (mostly naïve to Japanese) could have instantaneously tuned in to the relevant acoustic cues, a hypothesis supported by different confusion patterns of these listeners from native Japanese listeners (Table 8.5). Thus, it appears that either they had relatively high auditory sensitivity to F0 variations, or they were able to consistently apply their native language criteria to categorization of non-native patterns. The low-scoring English listeners, on the other hand, did not use their auditory sensitivity to the F0 fall, or their experience with F0 variations in English intonational patterns in perception of the novel prosodic contrast.

Another subject variable that correlated with success of learning was self-reported attention during the task. This general attentiveness is different from implicit selective attention as used in terms *first language attentional biases* or *individual attentional biases*. It is also different from explicit orientation of attention, which has been previously investigated by manipulating the task instructions in training (Guion & Pederson, 2007) and testing (Werker & Tees, 1984). It would be interesting to examine

the relationship between explicit attentional orientation and learning of a contrast in the prosodic domain, as compared to segments.

To conclude, striking individual differences in perception of Japanese lexical pitch patterns that were described in this dissertation are consistent with previous findings on perception of segments (Bradlow et al., 1997; Golestani & Zatorre, 2009; Iverson & Evans, 2009; Polka, 1992) and prosodic contrasts (Bent et al., 2006; Chandrasekaran et al., 2010; Golestani & Zatorre, 2009; Nishinuma et al., 1996; Wang et al, 1999). The current state of research on individual attentional biases in perception and learning is not conclusive with regard to their causes. They are different from first language attentional biases, and they may be related to other personal experiences (e.g., musical training) as well as to general cognitive skills (e.g., working memory).

9.2. Implications for Second Language Learning of Japanese Pitch Patterns

The findings in this dissertation relevant to applied learning of Japanese lexical pitch patterns are described in the rest of this chapter. First and most encouraging finding is that learners can improve in their perception of the Japanese pitch-accent contrast even after an hour of training. Learning of pitch patterns is asymmetric in that some patterns are learned better than the others. The unaccented pattern is particularly challenging for students whose native language is English. English listeners may or may not learn to perceive the unaccented pattern with great accuracy, however, this pattern is the most difficult to identify for native Japanese speakers as well. Therefore, it is unreasonable to expect perfect accuracy from successful learners in tasks such as identification of the unaccented pattern among several alternatives, or in tests such as the Tokyo Japanese

Accent Listening Test. Based on the results of high-scoring listeners in Experiment 5, even if the unaccented pattern is never acquired, the acquisition of the accented patterns may be sufficient for relatively accurate identification of pitch-accent patterns.

The finding that the unaccented pattern was the most poorly identified does not mean that the difference between the accented and unaccented patterns cannot be perceived by beginning learners of Japanese. They may be able to better attend to the distinction in certain contexts, in materials produced by certain speakers, or in certain tasks. These facilitative task variables are reviewed in three following sections.

9.2.1. Varying Contexts

In the materials used in Experiment 5, the unaccented pattern in a word was identified the best when it was followed by another unaccented word (Appendices D2 and E2). In the contexts like *Watáshi wa __ ga kakemas* and *__ ga hatsuon shiyasui*, the target word formed a long accentual phrase with the following word, where a phrasal high tone on the target word and the boundary tone at the end of the sentence were separated from each other by a larger amount of segmental material, and by a F0 plateau contour. This longer plateau, uninterrupted by other accents or tones, was likely to have a facilitating effect on identification. It is interesting to note that the context preceding a target pitch pattern does not seem to affect listeners' perceptual judgments (Kitahara, 2001, for Japanese), or its effect is considerably weaker than that of the following context (Francis et al., 2006, for Cantonese).

Controlling following contexts may be beneficial at the beginning stages of learning, when the perceptual fading technique is used for training (Jamieson & Morosan, 1986, 1989). With this technique, learners are first trained on materials that are

maximally contrastive in their acoustical properties. Then, the contrastiveness is gradually reduced approximating naturally occurring utterances. For the contrast between the unaccented and accented patterns, the contexts such as described above are likely to be maximally contrastive. For the contrast between the 1st- and 2nd- syllable accented patterns, the contexts such as __ *ga kakemas* and __ *ga hatsuon shiyasui* may be maximally contrastive (Appendix D1 and Appendix E2, respectively). In these contexts, in addition to the word-level cues, the phrase initial F0 rise for the 2nd- but not 1st-syllable accented word may facilitate perception.

9.2.2. Varying Speakers

Pitch-accent contrast is enunciated more clearly in the speech of some speakers than others, and this perceived clearness does not seem to be related to the gender (and pitch range) of the speaker. At the end of the training Experiment 5, the participants reported that the pitch pattern contrast as produced by female KC was more difficult to hear compared to male CH and female AK. Although the averaged F0 contours of the target words in Appendix C do not clearly suggest a basis for these impressions, relative intelligibility of the speakers seemed to correlate with the listeners' scores in the tests recorded by those speakers. Therefore, perceived difficulty of a speaker's voice should be considered for the training purposes. These individual differences in native speakers' productions, or synthesized multi-voice stimuli, may be used for the benefit of learners adopting the same perceptual fading technique described above.

9.2.3. Varying Tasks

The training task itself may affect listeners' perception and learning. Experiment 3 suggested that offering three answer choices (i.e., first syllable prominent, second

syllable prominent, no prominence) for categorization of the pitch patterns in two syllable words could, in fact, have facilitated the performance of non-native listeners. At first glance, *no prominence* response did not seem to make sense for English listeners, in whose native language every lexical word has prominence on a particular syllable. Surprisingly, however, in the presence of this answer option English listeners tended to pay some attention to the F0 fall, which is a necessary condition for pitch pattern learning. Furthermore, when in addition to F0 fall and F0 peak location, the height of the F0 peak was manipulated in Experiment 1, English listeners paid some attention to the interaction between the first two parameters, which was typical for the native listeners' perception.

These findings suggest that introducing non-native listeners to more complex materials (i.e., three manipulated parameters in Experiment 1 rather than two in Experiments 2-3), and introducing the listeners to more demanding tasks (three-alternative forced choice in Experiments 3 and 5 rather than two-alternative forced-choice in Experiments 1-2) may facilitate learning. Although this may seem to language teachers as counterintuitive advice, one should keep in mind that, in the process of the first language acquisition, a similar situation takes place. Namely, children are exposed to sound patterns that occur in different contexts, spoken by different speakers, and requiring different responses. The richness of the sound environment facilitates rather than interferes with learning. Variable classroom materials, instructors and tasks will help learners of Japanese to develop sensitivity toward lexical and intonational pitch patterns of the language.

9.2.4. Explicit Instructions and Differences Between Learners

Lastly, teachers of Japanese as a foreign language need to provide explicit instructions about the lexical pitch patterns in Japanese, because mere exposure to pitch patterns is not likely to lead to learning of this prosodic contrast. Even though lexical prosody is not typically taught in second language Japanese classrooms (Shport, 2008), learning of lexical pitch patterns may serve several goals. It will enable second language speakers to distinguish between (or at least be aware of) homophones such as like *hana* 'nose' and *haná* 'flower', which constitute about 13% of short words in the Japanese lexicon (Kitahara, 2001). It may help learners to better recognize prosodic groupings in the language such as accentual phrases, and thus result in a more effective speech parsing. Perceptual learning may generalize to more accurate production of pitch patterns, which would lead to better intelligibility and less effortful processing of learners' speech by native speakers of Japanese. In sum, learning of prosody can contribute to more effective communication.

Teachers may want to orient learners' attention toward the F0 fall cue of the pitch-accent contrast, and its realization in words with initial, medial, and final accented syllables. Large individual variability in students' perception of this particular acoustic cue has to be expected.

In general, individual variability in learning of Japanese lexical prosody is not necessarily related to the length of students' exposure to Japanese in classroom or immersion settings. It may not even be related to the length of targeted training on Japanese prosody. There are pre-existing differences in individuals' perception, and this fact has to be taken into account for learning assessment. Students with richer linguistic

and musical experiences are likely to make a faster progress in learning of the pitch-accent contrast, and, perhaps, second-language prosody in general. In addition, a larger effort is required for learners who are less sensitive to this prosodic contrast than others as determined during initial pretest or pre-training assessment. For low-scoring listeners, learning of this contrast may involve consistent application of English intonational F0 cues to the word domain, or it may involve a gradual development of Japanese-like cue weighting. Which of these strategies is less time consuming, while leading to robust categorization, is an open empirical question. For teaching Japanese as a second language that prioritizes communicative competence either strategy should be encouraged.

APPENDIX A

TRAINEES: DEMOGRAPHIC, ACADEMIC AND LANGUAGE BACKGROUND

Group	Listener	Pretest %	Age	Sex	Academic experience Major	GPA	Foreign language experience Languages studied	Months abroad
Training Low	220	34	22	M	economics	3.4	Spanish	
	223	39	21	F	linguistics	3.9	Spanish, ASL, Hebrew	
	219	39	19	F	pre-nursing	3.4	Spanish, ASL	
	216	43	22	F	journalism	3.5	Spanish	
	211	45	19	F	family & human services	3.1	Spanish	
	218	49	20	F	political science	3.2	ASL	
	215	49	21	F	biology	3.1	Spanish	6
	234	53	20	F	linguistics, political science	3.8	German, Arabic	1
	212	56	19	M	human physiology	3.4	Spanish	
	217	56	24	M	human physiology	3	Spanish	
Training High	235	61	19	F	linguistics, Spanish	3.8	Spanish	
	221	61	20	M	product design	3.6	Japanese, Spanish	
	236	69	22	M	linguistics	3.7	Spanish, Sahaptin	
	214	71	23	F	linguistics, Spanish	2.9	Spanish, German	
	224	71	23	F	linguistics	3.5	Japanese, Spanish, French	6
	213	74	19	F	linguistics, English	3.5	Spanish	
	230	77	19	M	sociology	2.4	Spanish, German	12
	222	77	22	M	journalism	3.1	Spanish	
	226	77	23	F	linguistics	3.8	German, French, Mandarin	
	228	87	27	M	linguistics	3.4	Japanese, Mandarin, Spanish	7
	229	91	21	F	linguistics	3.9	French	4

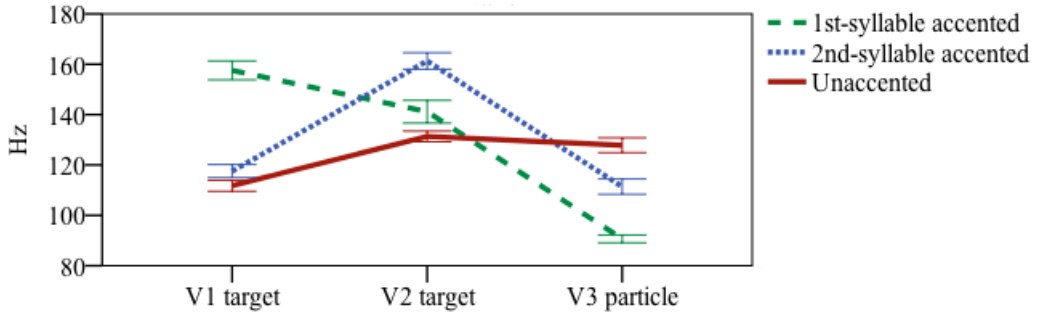
APPENDIX B

CONTROLS: DEMOGRAPHIC, ACADEMIC AND LANGUAGE BACKGROUND

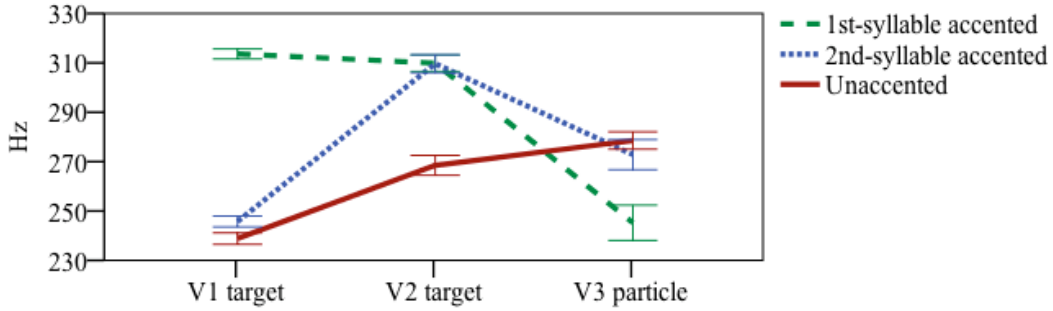
Group	Listener	Pretest %	Age	Sex	Academic experience Major	GPA	Foreign language experience Languages studied	Months abroad
Control Low	117	27	19	F	linguistics	4	Spanish, Arabic	
	130	41	19	M	history	3.4	none	
	136	44	21	F	journalism	3.2	Spanish	1
	131	44	20	M	undeclared	2.7	Spanish	
	137	45	22	F	linguistics	3.4	French, Danish, Russian	12
	126	45	22	F	linguistics	3.6	Spanish, French, Arabic	
	120	47	20	M	human physiology	3.3	French, German, Danish	
	122	50	21	F	linguistics	3.6	Spanish, German, Portuguese	
	133	51	19	F	undeclared	2.7	Spanish, French, German	
	113	56	19	M	linguistics	4	Japanese, Russian	
Control High	128	69	22	F	art (painting)	2.8	Spanish, Italian	
	135	70	18	F	linguistics	4.2	Spanish	11
	132	71	21	F	linguistics, Spanish	3.2	Spanish	4
	116	73	25	F	linguistics	4.2	French, Swahili, Malagasy	6
	119	74	24	F	linguistics	3.5	Japanese, Spanish, French	
	129	76	20	M	public policy	3.1	Spanish	
	125	78	23	M	linguistics	3.6	German, French	
	127	79	20	M	linguistics	3.4	Spanish, Korean	
	118	81	22	M	linguistics	2.9	French, Russian, Korean	
	123	90	21	M	linguistics	3.8	Spanish, Arabic	

APPENDIX C

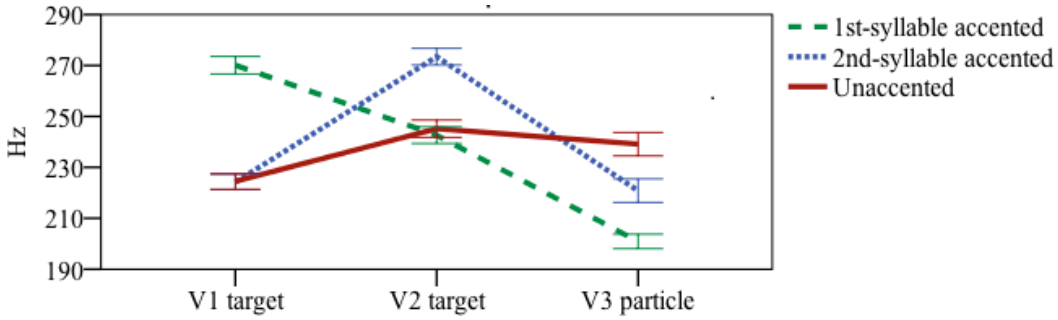
LEXICALLY CONTRASTIVE PITCH PATTERNS IN TEST MATERIALS



C1. Speaker CH, male (pretest, posttest)



C2. Speaker AK, female (training, gen1 test)

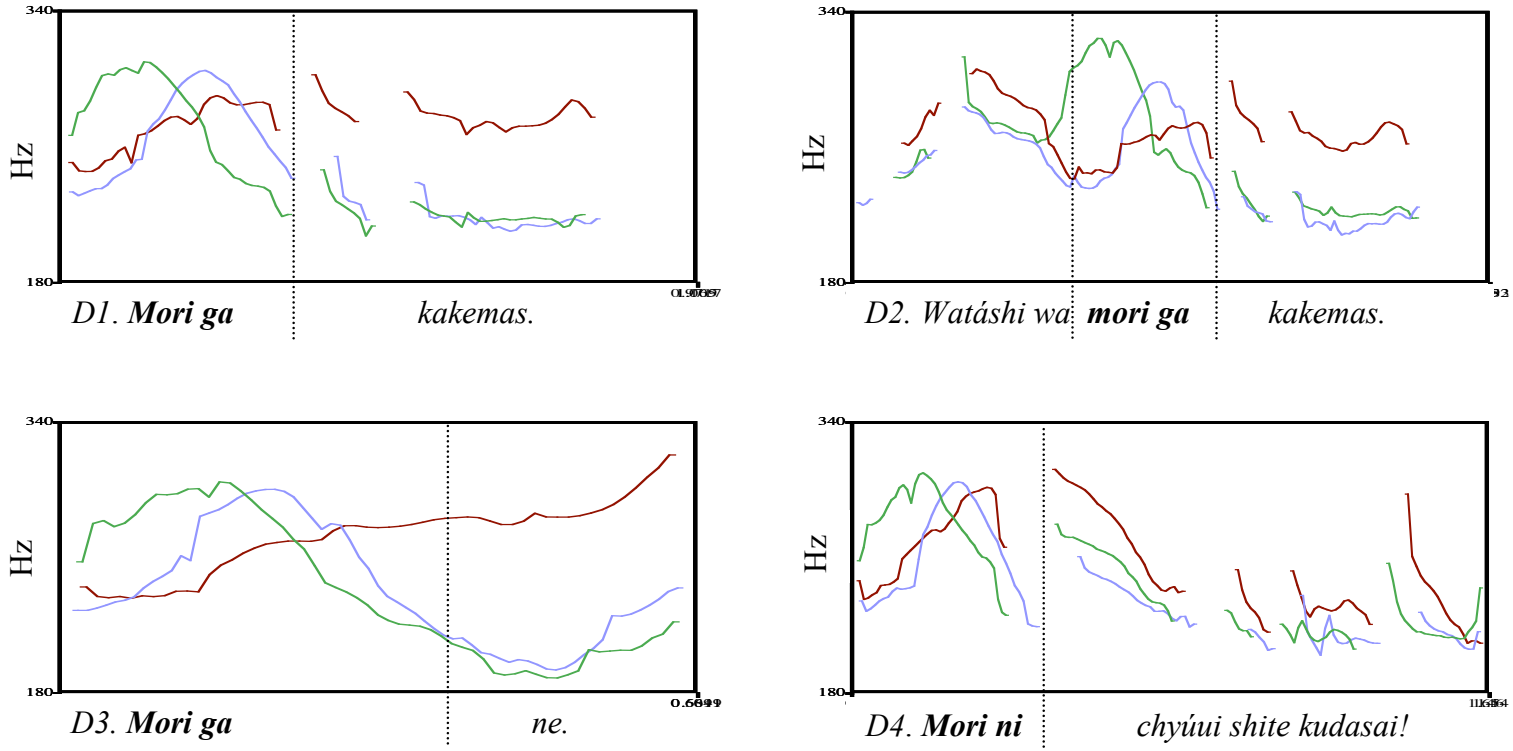


C3. Speaker KC, female (gen2 test)

Note. F0 was measured in Praat (Boersma & Weenink, 2010) at the vowel midpoints of disyllabic target words (Table 8.2) followed by one-syllable particle or copula. For each measurement point, F0 was averaged across all words of the same accentual type.

APPENDIX D

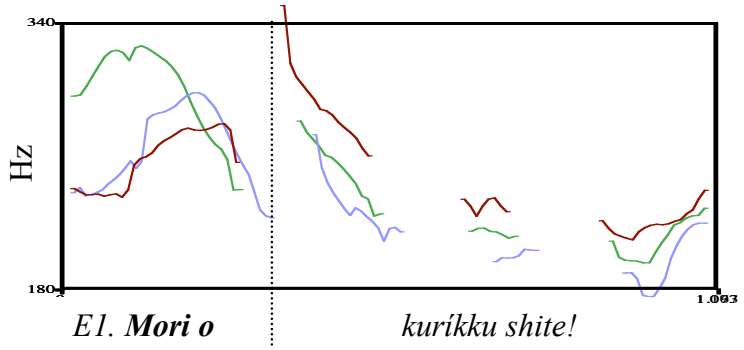
EXAMPLES OF INTONATION CONTOURS: SPEAKER AK, TRAINING MATERIALS

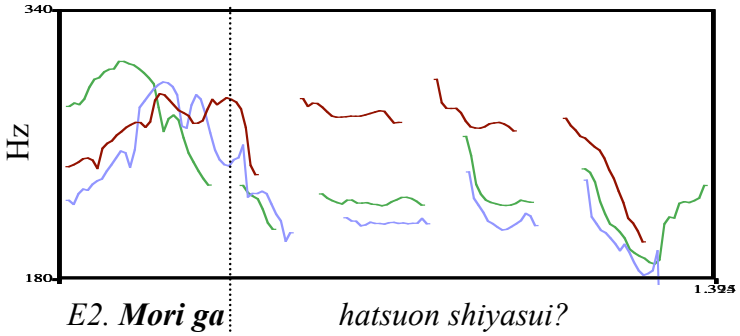


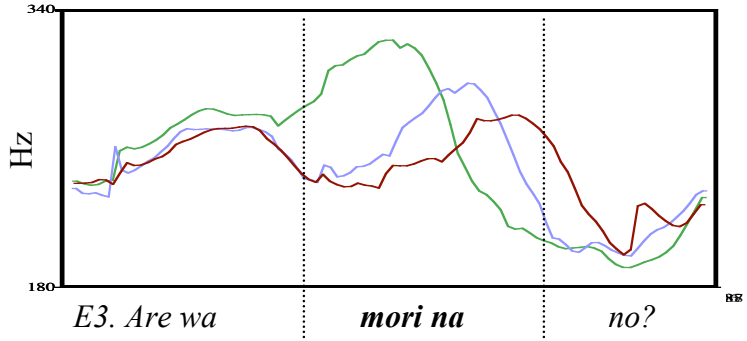
Note. F0 contours in green: sentences with the 1st-syllable accented word *móri*; F0 contours in blue: sentences with the 2nd-syllable accented word *morí*; F0 contours in red: sentences with the unaccented word *mori*. Vertical lines show the approximate boundaries of the target word ***mori* with the following particle or copula**.

APPENDIX E

EXAMPLES OF INTONATION CONTOURS: SPEAKER AK, GEN 1 TEST MATERIALS







Note. F0 contours in green: sentences with the 1st-syllable accented word *móri*; F0 contours in blue: sentences with the 2nd-syllable accented word *morí*; F0 contours in red: sentences with the unaccented word *mori*. Vertical lines show the approximate boundaries of the target word ***mori* with the following particle or copula**.

REFERENCES CITED

Aoyama, K., Flege, J. E., Guion, S. G., Yamada, T., & Akahane-Yamada, R. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /r/ and /l/. *Journal of Phonetics*, *23*, 233–250.

Arai, M. (1997). The results of a longitudinal survey on the percpetion of the Tokyo accent: American learners living in Kyoto. *Spoken Japanese language education: Looking to the 21st century* (pp. 73-79). Tokyo: The National Language Research Institute.

Atterer, M., & Ladd, D. R. (2004). On the phonetics and phonology of "segmental anchoring" of F0: Evidence from German. *Journal of Phonetics, 32*, 177-197.

Ayusawa, T. (1999). Interlanguage studies: Japanese learners' sound systems. *Journal of the Phonetic Society of Japan, 3*(3), 4-12.

Ayusawa, T. (2003). Acquisition of Japanese accent and intonation by foreign learners. *Journal of the Phonetic Society of Japan, 7*(2), 47-58.

Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412.

Beckman, M. E. (1986). *Stress and non-stress accent*. Dordrecht: Foris.

Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In S.-A. Jun (Ed.), *Prosodic typology: The phonology and intonation of phrasing* (pp. 9-54). New York: Oxford University Press.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. In *Phonology yearbook: Vol. 3*. (pp. 255-309). Great Britain: Cambridge University Press.

Behne, D., Wang, Y., Rø, M. H., Hoff, A.-K., Knutsen, H. A., & Schmidt, M. (2006). Effects of musical experience on linguistic pitch perception training [Abstract]. *Journal of the Acoustical Society of America, 120*(5), 3168.

Bent, T. (2005). *Perception and production of non-native prosodic categories.* Unpublished doctoral dissertation. Northwestern Univsersity, Evanston.

Bent, T., Bradlow, A. R., & Wright, B. A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology: Human Perception and Performance, 32*(1), 97-103.

Best, C. T. (1995). A Direct Realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-204). Timonium, MD: York Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception. In O.-S. Bohn & M. Munro (Eds.), *Language experience in second language speech learning* (pp. 13-34). Amsterdam: John Benjamins.

Boersma, P., & Weenink, D. (2010). *Praat: doing phonetics by computer* (Version 5.1) [Software]. Available from http://www.fon.hum.uva.nl/praat/

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems: II. Effect of inequality of variances and of correlation of errors in the two-way classification. *Annals of Mathematical Statistics, 25*, 484-498.

Bradlow, A. R. (2008). Training non-native language sound patterns: Lessons from training Japanese adults on the English /ɹ/-/l/ contrast. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 287-308). Amsterdam: John Benjamins.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning. *Journal of the Acoustical Society of America, 101*(1), 2299-2310.

Broselow, E., Hurtig, R. R., & Ringen, C. (1987). The perception of second language prosody. In G. Ioup & S. H. Weinberger (Eds.), *Inter-language phonology: The acquisition of second language sound system* (pp. 350-361). Cambridge: Newbury House Publishers.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007). Mismatch negativity to pitch contours is influenced by language experience. *Brain Research, 1128*, 148-156.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2009). Relative influence of musical and linguistic experience on early cortical processing of pitch contours. *Brain and Language, 108*, 1-9.

Chandrasekaran, B., Sampath, P. D., & Wong, P. C. M. (2010). Individual variability in cue-weighting and lexical tone learning. *Journal of the Acoustical Society of America, 128*(1), 456-465.

Cutler, A., & Otake, T. (1999). Pitch accent in spoken-word recognition in Japanese. *Journal of the Acoustical Society of America, 105*(3), 1877-1888.

Dilley, L. C., Ladd, D. R., & Schepman, A. (2005). Alignment of L and H in bitonal pitch accents: Testing two hypotheses. *Journal of Phonetics*, *33*, 115-119.

Dupoux, E., Pallier, C., Sebastián-Gallés, N., & Mehler, J. (1997). A destressing "deafness" in French? *Journal of Memory and Language, 36*, 406-421.

Dupoux, E., Sebastián-Gallés, N., Navarette, E., & Peperkamp, S. (2008). Persistent stress "deafness": The case of French learners of Spanish. *Cognition, 106*, 682-706.

Eda, S. (2004). *Processing of intonational patterns in Japanese: Implications for Japanese as a foreign language.* Unpublished doctoral dissertation. Ohio State University, Columbus.

Escudero, P., Benders, T., & Lipski, S. C. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish learners. *Journal of Phonetics, 37*, 452-465.

Flege, J. E. (1989). Chinese subjects' perception of the word-final English /t/-/d/ contrast: Performance before and after training. *Journal of the Acoustical Society of America, 86*, 1684-1697.

Flege, J. E. (1995a). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Timonium, MD: York Press.

Flege, J. E. (1995b). Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics, 16*, 425-442.

Francis, A. L., Baldwin, K., & Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception and Psychophysics*, *62*, 1668-1680.

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics, 36*, 268-294.

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance, 28*(2), 349-366.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America, 27*, 765-768.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech, 1*, 126-152.

Fujisaki, H., & Sudo, H. (1971). Nihongo tango akusento no kihon shūhasū patan to sono sēsēkikō no moderu [A model for the generation of fundamental frequency contours of Japanese word accent]. *Nihon Onkyō Gakkai Shi* [Journal of the Acoustical Society of Japan]*, 27*, 445-444.

Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics, 11*, 149-175.

Gandour, J. T., & Harshman, R. A. (1978). Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Language and Speech, 21*, 1-33.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel-hierarchical models.* Cambridge: Cambridge University Press.

Golestani, N., & Zatorre, R. J. (2009). Individual differences in the acquisition of second language phonology. *Brain and Language, 109*, 55-67.

Gösta, B. (1977). *Swedish word accents in sentence prospective.* Lund: Gleerup.

Gottfried, T., & Beddor, P. (1988). Perception of temporal and spectral information in French vowels. *Language and Speech, 31*, 57-75.

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O.-S. Bohn & M. Munro (Eds.), *Language experience in second language speech learning* (pp. 57-77). Amsterdam: John Benjamins.

Hallé, P. A., Chang, Y.-C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics, 32*, 395-421.

Haraguchi, S. (1999). Accent. In N. Tsujimura (Ed.), *The handbook of Japanese linguistics* (pp. 1-30). Oxford: Blackwell Publishers.

Hasegawa, Y., & Hata, K. (1992). Fundamental frequency as an acoustic cue to accent perception. *Language and Speech, 35*(1, 2), 87-98.

Hattori, S. (1960). Akusento so ni tsuite [On accent]. *Gengogaku no hōhō* [Methods in linguistics]. Tokyo: Iwanami Shoten.

Hayes, B. (1995). *Metrical stress theory: Principles and case studies.* University of Chicago Press.

Hirata, E., Ayusawa, T., Nakagawa, C., & Odaka, K. (1997). Longitudinal study of the effects of prosodic education as seen in the results of the Tokyo Japanese Accent Listening Test. *Spoken Japanese language education: Looking to the 21st century* (pp. 61-72). Tokyo: The National Language Research Institute.

Hirata, Y. (1999). *Acquisition of Japaneser rhythm and pitch accent by English native speakers.* Unpublished doctoral dissertation. University of Chicago, Chicago.

Hirata, Y. (2003). Learning to form new L2 phonetic categories in sentence contexts. *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, Spain.

Hirata, Y. (2004). Training native English speakers to perceive Japanese length contrasts in word versus sentence contexts. *Journal of the Acoustical Society of America, 116*(4), 2384-2394.

Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Infant Behavior and Development, 32*, 262-274.

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical Society of America, 119*(5), 3059-3071.

Homma, Y. (1973). An acoustic study of Japanese vowels. *The Study of Sounds, 16*, 347-368.

Huss, V. (1978). English word stress in postnuclear position. *Phonetica, 35*, 86-105.

Imada, S. (1989). *Pronunciation*. Tokyo: The Japan Foundation Japanese Language Institute.

Ishihara, T. (2003). A phonological effect on tonal alignment in Tokyo Japanese. *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, Spain.

Ishihara, T. (2006). *Tonal alignment in Tokyo Japanese.* Unpublished doctoral dissertation. University of Edinburgh, Edinburgh.

Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *Journal of the Acoustical Society of America, 126*(2), 866-877.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *Journal of the Acoustical Society of America, 118*(5), 3267-3278.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., et al. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87*, B47-B57.

Jamieson, D. G., & Morosan, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /ð/-/θ/ contrast by francophones. *Perception and Psychophysics, 40*(4), 205-215.

Jamieson, D. G., & Morosan, D. E. (1989). Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques. *Canadian Journal of Psychology, 43*(1), 88-96.

Jun, S.-A. (2005). Prosodic typology. In S.-A. Jun (Ed.), *Prosodic typology: The phonology and intonation of phrasing* (pp. 430-458). New York: Oxford University Press.

Kaan, E., Wayland, R., Bao, M., & Barkley, C. M. (2007). Effects of native language and training on lexical tone perception: An event-related potential study. *Brain Research, 1148*, 113-122.

Kawakami, S. (1957). Jun akusento ni tsuite [About accent]. *Kokugo Kenkyū* [Research on Japanese Language]*, 7*, 44-60.

Kawakami, S. (1961). On the relationship between word-tone and phrase-tone in Japanese. *The Study of Sounds, 9*, 169-177.

Kawakami, S. (1995). Akusento genshō to akusento kannen [Accent as a phenomenon and as a concept]. In *Nihongo akusento ronshū* [Collection of essays on Japanese accent] (pp. 23-29). Tokyo: Iwanami Syoten. (Original work published 1966)

Khouw, E., & Ciocca, V. (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics, 35*, 104-117.

Kitahara, M. (2001). *Category structure and function of pitch accent in Tokyo Japanese.* Unpublished doctoral dissertation. Indiana University, Bloomington.

Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *Journal of the Acoustical Society of America, 118*(2), 1038-1054.

Kochanski, G., & Orphanidou, C. (2008). What marks the beat of speech? *Journal of the Acoustical Society of America, 123*(5), 2780-2791.

Komatsu, H. (1989). Nihongo no rekishi: akusento [History of Japanese: Accent]. In T. Kamei, R. Kono & E. Chino (Eds.), *Gengogaku daijiten* [Linguistic dictionary] (pp. 1651-1653). Tokyo: Sanseido.

Kubozono, H. (1993). *The organization of Japanese prosody.* Tokyo: Kuroshio.

Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the Perceptual Magnet Effect. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121-154). Timonium, MD: York Press.

Ladd, D. R., & Schepman, A. (2003). "Sagging transitions" between high pitch accents in English: Experimental evidence. *Journal of Phonetics, 31*, 81-112.

Leather, J. (1990). Perceptual and productive learning of Chinese lexical tone by Dutch and English speaker. *Proceedings of the 1990 Amsterdam Symposium on the Acquisition of Second-Language Speech*. University of Amsterdam.

Lee, C.-Y., Tao, L., & Bond, Z. S. (2009). Speaker variability and context in the identification of fragmented Mandarin tones by native and non-native listeners. *Journal of Phonetics, 37*, 1-15.

Lehiste, I., & Peterson, G. E. (1959). Vowel amplitude and phonemic stress in American English. *Journal of the Acoustical Society of America, 31*, 428-435.

Lehiste, I., & Peterson, G. E. (1961). Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America, 33*(4), 419-425.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America, 94*, 1242-1255.

Lively, S. E., Pisoni, D. E., Yamada, R. A., Tokhura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America, 96*, 2076-2087.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America, 89*, 874-886.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). New Jersey: Lawrence Erlbaum Associates.

Masaki, A., Takasawa, M., & Arai, T. (2001). Akzentwahrnehmung von Japanern bei technisch kontrollirten F0-Konturen [Accent perception by Japanese listeners in manipulated F0 Contours]. *Sophia Linguistica: Working Papers in Linguistics, 48*, 213-224.

Masuda-Katsuse, I. (2006). Contribution of pitch-accent information to Japanese spoken-word recognition. *Acoustical Science and Technology, 27*(2), 97-103.

McCandliss, B. C., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the [r]-[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cogntitive, Affective, and Behavioral Neuroscience, 2*(2), 89-108.

Miyata, K. (1927). Atarashī akusento kan to akusento hyōki hō [A new view on accent and accent marking]. *The Study of Sounds, 1*, 18-22.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics, 18*, 331-340.

Mo, Y. (2008). *Acoustic correlates of prosodic prominence for naïve listeners of American English*. Paper presented at the 34th Meeting of the Berkeley Linguistic Society, Berkeley, CA.

Mo, Y. (2009). *F0 max and formants (F1, F2) as perceptual cues for naïve listeners' prominence perception*. Paper presented at the 35th Annual Meeting of the Berkeley Linguistic Society, Berkeley, CA.

Morrison, G. S. (2005). An appropriate metric for cue weighting in L2 speech perception: Response to Escudero and Boersma (2004). *Studies in Second Language Acquisition, 27*, 597-606.

Morrison, G. S. (2007). Logistic regression modelling for first- and second-language perception data. In M. J. Solé, P. Prieto & J. Mascaró (Eds.), *Segmental and prosodic issues in Romance phonology* (pp. 219-236). Amsterdam: John Benjamins.

Neustupný, J. V. (1980). Nihongo no akusento wa kōtē akusento ka? [Is the Japanese accent a pitch accent?]. In M. Tokugawa (Ed.), *Akusento* [*Accent*] (pp. 230-239). Tokyo: Yuseido. (Original work published 1966)

NHK (1998). *Nihongo hatsuon akusento jiten* [Pronunciation and accent dictionary of Japanese]. Tokyo: Japan Broadcast Publishing Co.

Nishinuma, Y. (1997). Perception of Japanese accent by American and French learners: First language intervention. *Spoken Japanese language education: Looking to the 21st century* (pp. 5-12). Tokyo: The National Language Research Institute.

Nishinuma, Y., Arai, M., & Ayusawa, T. (1996). Perception of tonal accent by Americans learning Japanese. *Proceedings of the 4th International Conference on Spoken Language Processing*, Philadelphia, PA.

Nishinuma, Y., di Cristo, A., & Espesser, R. (1983). Loudness as a function of vowel duration in CV syllables. *Speech Communication*, *2*, 167-169.

Ohala, J. J. (1978). Production of tone. In V. A. Fromkin (Ed.), *Tone: A linguistic survey* (pp. 5-39). New York: Academic Press.

Ortega-Llebaria, M., del Mar Vanrell, M., & Prieto, P. (2010). Catalan speakers' perception of word stress in unaccented contexts. *Journal of the Acoustical Society of America, 127*(1), 462-471.

Ortega-Llebaria, M., Prieto, P., & Vanrell, M. M. (2007). Perceptual evidence for direct acoustic correlates of stress in Spanish. *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America, 32*(6), 693-703.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation.* Bloomington, Indiana: Indiana University Linguistics Club.

Pierrehumbert, J. (2000). Tonal elements and their alignment. In M. Horne (Ed.), *Prosody: Theory and experiment* (pp. 11-36). Dordrecht: Kluwer.

Pierrehumbert, J. B., & Beckman, M. E. (1988). *Japanese tone structure.* Cambridge, MA: MIT Press.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan & M.E. Pollack (Eds.), *Intentions in communication* (pp. 271-311). Cambridge, MA: MIT Press.

Polivanov, E. D. (1976). Tōkyō hōgen ni okeru ongakuteki akusento [Melodic accent in Tokyo dialect]. In S. Murayama (Ed.), *Nihongo kenkyū* [Research on Japanese language] (pp. 18-31). Tokyo: Kobundo. (Original work published 1915)

Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception and Psychophysics, 52*, 37-52.

Poser, W. J. (1984). *The phonetics and phonology of tone and intonation in Japanese.* Unpublished doctoral dissertation. Massachusetts Institute of Technology, Cambridge.

Pruitt, J. S. (1995). Perceptual training on Hindi dental and retroflex consonants by native English and Japanese speakers [Abstract]. *Journal of the Acoustical Society of America, 97*, 3417.

Rietveld, A. C. M., & Gussenhoven, C. (1985). On the relation between pitch excursion size and pitch prominence. *Journal of Phonetics, 13*, 299-308.

Rø, M. H., Behne, D., & Wang, Y. (2006). The effects of musical experience on linguistic pitch perception: A comparison of Norwegian professional singers and instrumentalists [Abstract]. *Journal of the Acoustical Society of America, 120*(5), 3168.

Sakamoto, E. (2008). Second language acquisition of Japanese pitch accent by native speakers of English: An experimental perception study. *Sophia Linguistica: Working Papers in Linguistics, 56*, 139-150.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide.* Pittsburgh: Psychology Software Tools, Inc.

Sekiguchi, T., & Nakajima, Y. (1999). The use of lexical prosody for lexical access of the Japanese language. *Journal of Psycholinguistic Research, 28*(4), 439-454.

Shen, X. S. (1989). Toward a register approach in teaching Mandarin tones. *Journal of Chinese Language Teachers, 24*, 27-47.

Shibata, T. (2005). *Prosody acquisition of Japanese as a second language: View from an integrative perspective.* Unpublished doctoral dissertation. University of Iowa, Iowa City.

Shport, I. A. (2008). Acquisition of Japanese pitch accent by American learners. In P. Heinrich & Y. Sugita (Eds.), *Japanese as foreign language in the age of globalization* (pp. 165-187). München: Iudicium Verlag.

Shport, I. A., & Guion, S. G. (2008). The effect of segmental structure on F0 patterns of words in Tokyo Japanese. *Journal of the Phonetic Society of Japan, 12*(2), 4-16.

Silipo, R., & Greenberg, S. (2000). *Prosodic stress revisited: Reassessing the role of fundamental frequency.* Paper presented at the NIST Speech Transcription Workshop, College Park, MD.

Silverman, K., & Pierrehumbert, J. (1990). The timing of prenuclear high accents in English. In J. Kingston & M. E. Beckman (Eds.), *Between the grammar and physics of speech* (pp. 71-106). Cambridge: Cambridge University Press.

Sluijter, A. M. C., & van Heuven, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America, 100*, 2471-2485.

Sluijter, A. M. C., van Heuven, V. J., & Pacilly, J. A. (1997). Spectral balance as a cue in the perception of linguistic stress. *Journal of the Acoustical Society of America, 101*, 503-513.

So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech, 53*(2), 273-293.

Strange, W. (in press). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics*.

Strange, W., & Shafer, V. L. (2008). Speech perception in second language learners: The re-education of selective perception. In J. G. Hanson Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 193-218). Amsterdam: John Benjamins.

Sugito, M. (1980). Ososagari-kō: Dōtai-sokutei ni yoru nihongo akusento no kenkyū [Thoughts on peak delay: An instrumental study of Japanese accent]. In M. Tokugawa, *Akusento* [Accent] (pp. 65-94). Tokyo: Sanseido. (Original work published 1972)

Sugito, M. (1982a). Kakō onchō to akusento-gata no chikaku [Perception of falling intonation and accentual patterns]. *Nihongo akusento no kenkyū* [Studies on Japanese accent] (pp. 135-146). Tokyo: Sanseido.

Sugito, M. (1982b). Tōkyō akusento ni okeru 'hana' to 'hana' no sēsē to chikaku [The production and perception of 'flower' and 'nose' with Tokyo accent]. *Nihongo akusento no kenkyū* [Studies on Japanese accent] (pp. 182-201). Tokyo: Sanseido.

Terken, J., & Hermes, D. J. (2000). The perception of prosodic prominence. In *Prosody: Theory and Experiment, Studies Presented to Gösta Bruce* (pp. 89-127). Dordrecht: Kluwer Academic.

Turk, A., & Sawusch, J. (1996). The processing of duration and intensity cues to prominence. *Journal of the Acoustical Society of America*, *99*, 3782-3790.

Ueyama, M. (2000). *Prosodic transfer: An acoustic study of L2 English vs. L2 Japanese.* Unpublished doctoral dissertation. University of California, Los Angeles.

Ueyama, M., & Jun, S.-A. (1998). Focus realization in Japanese English and Korean English intonation. In N. Akatsuka, H. Hoji, S. Iwasaki, S. Sohn, & S. Strauss (Eds.), *Japanese/Korean linguistics: Vol. 7*. Stanford, California: CSLI Publications.

Vance, T. J. (1995). Final accent vs. no accent: Utterance-final neutralization in Tokyo Japanese. *Journal of Phonetics, 23*, 487-499.

Vanderslice, R., & Ladefoged, P. (1972). Binary suprasegmental features and transformational word-accentuation rules. *Language, 48*(4), 819-838.

Venditti, J. J. (2005). The J-ToBI model of Japanese Intonation. In S.-A. Jun (Ed.), *Prosodic typology: The phonology and intonation of phrasing* (pp. 172-200). New York: Oxford University Press.

Venditti, J. J. (2006). Prosody in sentence processing. In M. N. R. Mazuka & Y. Shirai (Eds.), *The handbook of East Asian psycholinguistics. Vol. II: Japanese* (pp. 208-217). Cambridge: Cambridge University Press.

Venditti, J. J., Maeda, K., & van Santen, J. P. H. (1998). *Modeling Japanese boundary pitch movements for speech synthesis.* Paper presented at the 3rd ESCA Workshop on Speech Synthesis, Jenolan Caves, Australia.

Venditti, J. J., Maekawa, K., & Beckman, M. E. (2008). Prominence marking in the Japanese intonation system. In S. Miyagawa & M. Saito (Eds.), *The Oxford handbook of Japanese linguistics* (pp. 456-512). New Tork: Oxford University Press.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America, 106*(6), 3649-3658.

Wayland, R., & Guion, S. (2003). Perceptual discrimination of Thai tones by naive and experienced learners of Thai. *Applied Psycholinguistics, 24*, 113-129.

Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning, 54*(4), 681-712.

Wayland, R., Herrera, E., & Kaan, E. (2010). Effects of musical experience and training on pitch contour perception. *Journal of Phonetics, 38*, 654-662.

Wayland, R. P., & Li, B. (2008). Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics, 36*, 250-267.

Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America, 75*, 1866-1878.

Whalen, D. H., & Levitt, A. G. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics, 23*, 349-366.

Xu, Y., & Wang, E. Q. (2001). Pitch targets and their realization: evidence from Mandarin Chinese. *Speech Communication, 33*, 319-337.

Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *Journal of the Acoustical Society of America, 120*, 1063-1074.

Yamada, T., Yamada, R. A., & Strange, W. (1996). Perceptual learning of Japanese mora syllables by native speakers of American English. *Proceedings of the International Conference on Spoken Language Processing*, Yokohama, Japan.

Yoshimitsu, K. (1981). Gaikokujin gakushūsha no akusento [Foreign learners' accent]. *Nihongo Kyōiku* [Japanese Language Pedagogy], *45*, 63-75.