THE ROLE OF LEXICAL CONTRAST IN THE PERCEPTION

OF INTONATIONAL PROMINENCE IN JAPANESE

A Dissertation Presented

by

TAKAHITO SHINYA

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

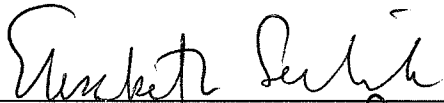February 2009

Department of Linguistics

THE ROLE OF LEXICAL CONTRAST IN THE PERCEPTION

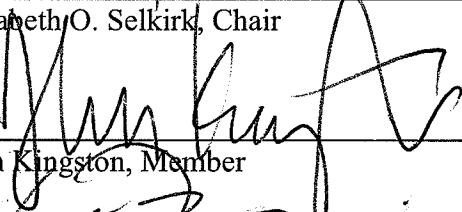OF INTONATIONAL PROMINENCE IN JAPANESE

A Dissertation Presented

by

Takahito Shinya

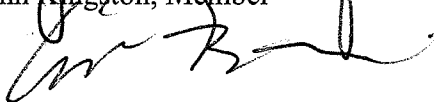Approved as to style and content by:
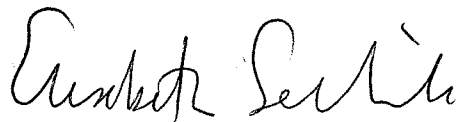
_____

Elisabeth O. Selkirk, Chair

_____

John Kingston, Member

_____

Lyn Frazier, Member

_____

Charles E. Clifton, Member

_____

Elisabeth O. Selkirk, Department Head
Department of Linguistics

ACKNOWLEDGMENTS

As is always the case with any dissertation work, this dissertation would not have been possible without tremendous help from lots of people. First and foremost, I would like to express my most sincere and profound gratitude to the committee members. I cannot find right words to express sufficiently how grateful I am to Lisa Selkirk, the chair of the dissertation committee. She has guided me through literally every aspect of the dissertation project. Her precise and insightful comments always have given me real intellectual excitements and have shown me new ideas and paths that I should take to proceed. Her wholehearted encouragements on the non-academic side have also been of great help. I particularly thank her for her words of encouragement when one of my sons had a serious health problem, which led this work to have a long gestation period before it was out into the world. Thanks so much for everything, Lisa. John Kingston is the best phonetics teacher I have ever had. His critical comments on the experimental side of my research have been particularly invaluable and have opened the door to a new line of investigations. Lyn Frazier has provided lots of critical and important advice on both theoretical and experimental issues from the psycholinguistic point of view. Chuck Clifton has generously spent hours with me discussing the details of the experiments and the proposals.

The UMass community has given me exceptionally-fun academic and non-academic experiences. I thank the faculty members of the Linguistics Department for their teaching and their enthusiasm on linguistics: Kyle Johnson, Angelika Kratzer, John

ABSTRACT

THE ROLE OF LEXICAL CONTRAST IN THE PERCEPTION

OF INTONATIONAL PROMINENCE IN JAPANESE

FEBRUARY 2009

TAKAHITO SHINYA, B.A., KANAGAWA UNIVERSITY

M.A., SOPHIA UNIVERSITY

PH.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Elisabeth O. Selkirk

In this dissertation, I examine the effects of lexical accent on the perception of intonational prominence in Japanese. I look at how an F0 accent peak is perceived relative to another flanking F0 peak in the same utterance with respect to perceived intonational prominence. Through four experiments, I show that the lexical prosodic structure plays a significant role in the perception of intonational prominence.

I first show that two distinct perceptual processes are at play in the perception of relative perceived prominence in Japanese: *accentual boost normalization* and *downstep normalization*. Accentual boost normalization normalizes the accentual boost of an accented word. In this process, the extra F0 boost assigned by a lexical accent does not count as part of the F0 peak's excursion that contributes to the perceived prominence of the F0 peak. I demonstrate that when an accented word and an unaccented word are perceived as having the same prominence, the accented word has a higher F0 peak value than the unaccented word does.

Downstep normalization compensates for the production effect of downstep, a pitch range compression phenomenon after a lexical accent. Experiments show that for an F0 peak to be perceived as having equivalent prominence to a preceding F0 peak, the second peak is always lower in F0 when the first word is accented than when it is unaccented. This suggests the existence of a perceptual process that normalizes the effect of downstep.

I then examine the nature of accentual boost normalization and downstep normalization and show that they refer to two distinct types of lexical accent properties when they are applied. One is the phonetic F0 contour shape that is characteristic of accented words. The other is the phonological lexical accent information that is uniquely specified for accented words. The experimental results show that the perceptual effects of the normalization processes are seen when only the phonological lexical accent information of a word is present with its F0 contour shape being ambiguous as well as when the same word is acoustically manipulated into different F0 contour shapes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

xvi

CHAPTER 1

INTRODUCTION

1.1   Introduction

This dissertation explores the interaction of two distinct phonological systems in Japanese: lexical accent and intonation. Lexical accent is a word-level system that mainly serves to distinguish words. Intonation is a sentence-level system that plays many roles including signaling prosodic and syntactic structures, prominence relations and sentence type. This dissertation focuses on intonational prominence and looks at how it interacts with lexical accent from the viewpoint of speech perception. The fact that is important to this study is that in Japanese lexical accent and intonational prominence are realized in the same acoustic dimension: fundamental frequency (F0). How these two phonological systems affect with one another is the main theme of this dissertation.

Lexical accent in principle has a strong impact on the perception of intonational prominence, but its specific effects have so far been completely unexplored. Research on the perceptual assessment of intonational prominence can roughly be categorized into two types. One is a line of research that focuses on 'intonation languages' such as Dutch and English and examines the effects of sentence accent on prominence in perception (Gussenhoven, Repp, Rietveld, Rump & Terken 1997, Gussenhoven & Rietveld 1988, 1998, Pierrehumbert 1979, Rietveld & Gussenhoven 1985, Terken 1991, 1994). Such studies typically manipulate the heights or positions of the F0 peaks in an utterance and examine the effects of those manipulations on the perception of the peaks' prominence. The other kind of research

1

investigates the effects of 'focus structure' on the perceived prominence of a word in an utterance (Vanio & Järvikivi 2006) or on the processing of the sentence contained in the utterance (Cutler 1976, Cutler & Fodor 1979, Cutler & Foss 1977, Cutler & Darwin 1981). Vanio & Järvikivi (2006), for example, showed that a word is perceived to be more intonationally prominent in a syntactically focused position than in an unfocused position.

Given these two kinds of research mentioned above, neither examines the effects of lexical accent on perceived prominence. One reason for this might be that researchers have concentrated on Dutch and English, which are typologically classified as 'intonation languages.' These languages use pitch accents only at the level of sentence and hence show no lexical contrasts in the use of F0.

Not only in intonation languages but also in pitch-accent languages, accent has a strong influence on intonational patterns. In Japanese, typologically a pitch-accent language (Lehiste 1970, Pike 1948), the F0 pattern of a word differs considerably depending on when the word is lexically accented or unaccented: an accented word shows a higher F0 peak than an unaccented word showing a sharp F0 fall following the accentual peak. An unaccented word does not show either of these F0 patterns. I call this phenomenon *accentual boost*. Moreover, an accented word affects the F0 pattern of the *following* word such that it is realized as lowered with respect to the preceding word. This is a phenomenon called *downstep*. An unaccented word does not cause downstep on the following word. (See §1.3 as well as the references on Japanese prosody such as Haraguchi 1977, 1991, Pierrehumbert & Beckman 1988, Poser 1984, Kubozono 1993 and Sugahara 2003 for background information).

Accentual boost and downstep both potentially interact with perceived prominence. It is

generally the case that, when everything else is the same, the higher the F0 peak of a word the more perceived prominence it yields (see Chapter 2 for detailed discussions of the factors that can influence perceived prominence). Given this general relationship between F0 peak height and perceived prominence, it might be expected that an accented word would be perceived with greater prominence than an unaccented word, because the former would have a higher F0 peak than the latter. It may as well be expected that a downstepped word occurring after an accented word is perceived less prominent than a non-downstepped word occurring after an unaccented word, because a downstepped word would show a lower F0 peak than a non-downstepped word.

In this study I show that neither of these perceptual effects is the case. Instead I demonstrate that the interaction between lexical accent and perceived prominence is not that simple and cannot be explained by the direct relationship between F0 peak and perceived prominence at the acoustic level.

The following sections of this chapter are organized as follows. In the following section I present the specific questions that I ask in this study with the brief answers to them (§1.2). Following that, I give the reader some background information on Japanese prosody (§1.3). I conclude this chapter with an overview of the following chapters (§1.4).

1.2   Research questions

In this dissertation I ask two questions. The first concerns the possibility of perceptual processes that compensate for lexical accent and downstep: those that normalize the high F0 peak of a word that is lexically accented and the low F0 peak of a word that is preceded by a lexically accented word.

Despite the fact that accented words have higher F0 peaks than unaccented words in Japanese, there is no research that shows that accented words are more prominent than unaccented words. Accent in this language is lexical and naturally its main function is to show lexical contrast. Indeed, it has been pointed out in the literature that Japanese lexical accent shares properties with sentence accent in languages like English with respect to several phonological properties. For instance, no more than one pitch accent can occur in a word (McCawley 1968), pitch accent plays a role in Minor Phrase (or Accentual Phrase) formation (Selkirk & Tateishi 1988), and it serves the base for the rhythm of the language (Kubozono 1993). Based on these shared properties, lexically accent words are arguably more prominent than unaccented words. Crucially, however, there is no direct evidence that shows that. In this dissertation, I will provide evidence against the idea of 'more prominent' accented words.

Likewise, no evidence has been reported that downstepped words are less prominent than non-downstepped words. As briefly mentioned in the previous section, the trigger of downstep in Japanese is lexical accent and it is the only trigger of the phenomenon. The F0 peak of a word is reduced precisely because it is preceded by an accented word. If downstep signaled reduction in prominence, it would be very hard to understand why a word preceded by an accented word is less prominent than a word preceded by an unaccented word by virtue of being preceded by an accented word. A natural expectation is that downstep does not seem to be relevant to any linguistic prominence relations.

Given the possibility that lexical accent and downstep both do not play a major role in marking prominence, one can come up with the idea that a word's high F0 peak given by lexical accent and a word's F0 peak reduced by downstep are compensated when

perceptually assessing them in terms of prominence. In this study I show that this is the case. Based on a series of perceptual experiments, I propose two perceptual compensation processes: *accentual boost normalization* and *downstep normalization*. Accentual boost normalization is a process where perceived prominence is assessed in a way that the extra high F0 peak of an accented word is reduced so that, with other things being equal, an accented word has the same perceived prominence as an unaccented word. Downstep normalization compensates for the effect of F0 peak reduction seen in a word that preceded by an accented word. More simply, it perceptually undoes the reduction of a downstepped word and consequently two words are perceived to have the same prominence even if one of them has a significantly lower F0 peak because of it being preceded by an accented word.

The second question asked in the dissertation is about the nature of accentual boost normalization and downstep normalization. The specific question is at what level of lexical accent property the perceptual processes make reference to when they are applied? In speech perception, two distinct types of property are available for the listener to know the accent type of a word. One is the *acoustic property*, namely the F0 contour shape. The listener can use the extra high F0 peak (and the sharp F0 drop that follows) to judge that the word is lexically accented. The other type of property is what I call the *structural property*, which is the information of the accent type stored in the lexicon, specified for each lexical item. With this property of lexical accent, the listener does not have to hear the F0 pattern of a word to know its accent type: all he needs is what the word is. If he can obtain enough segmental information to identify the word, its accent type automatically follows from its structural property. Thus, the precise question I ask is whether accentual boost normalization and downstep normalization are triggered by the acoustic property and/or the structural property.

I give a positive answer to this question. I show that accentual boost normalization and downstep normalization are both triggered by either of the lexical accent properties. More concretely, the results of perceptual experiments show that compensation effects are seen not only when the physical F0 contour is present but also when it is ambiguous or even contradicts to what the segmental content signals.

1.3   Summary

Summing up the discussions given above, I first discussed that the interaction of accent and perceived prominence has not been studied in languages that have lexical accent like Japanese. I pointed out that although lexical accent radically changes the intonation pattern of an utterance, which potentially has a great impact on prominence perception, the effects of lexical accent on perceived prominence have not been subject to serious study. I then presented the two research questions asked in this study. One was whether there are any perceptual compensation processes for accentual boost and downstep. The other was whether the acoustic and/or structural properties trigger the perceptual compensation processes.

1.4   Structure of Japanese prosody

This section outlines the basic facts and theoretical treatment about Japanese prosody (the Tokyo variety). I only give the minimum background that is necessary for the reader to understand the experiments reported in Chapters 3-6. For a more thorough treatment of Japanese prosody, see Haraguchi (1977, 1991), Kubozono (1993), McCawley (1968), Pierrehumert & Beckman (1988), Poser (1984), and Sugahara (2003).

1.4.1 Accent H*+L, phrase accent H, and boundary tones

Japanese exploits F0 to distinguish lexical items and is often called a pitch-accent language (Lehiste 1970, Pike 1948) precisely because of this function. Every lexical item in Japanese is either lexically *accented* or *unaccented*. An accented word has a lexical accent that is associated to one of the moras in the word while an unaccented word does not have such a lexical accent. Because of the lexical accent, an accented word is realized acoustically as having a higher F0 peak and a sharper F0 fall than an unaccented word. The F0 contours in Figure 1.1a and Figure 1.1b show this difference graphically.


```
L%   H        H*+L      L%              L%   H                      L%
ku   bi       ka'    za   ri            hi   ya   ke   do    me
          'necklace'                            'sunscreen'
          a. accented                           b. unaccented
```
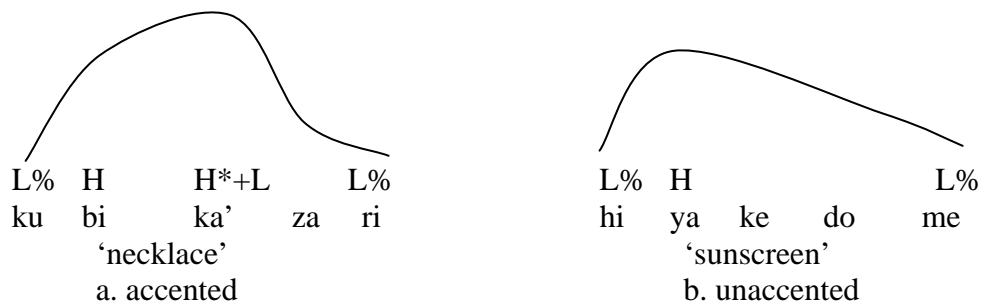
Figure 1.1 F0 contours of an accented and an unaccented word. (a) accented word (*kubika'zari* 'necklace'); (b) unaccented (*hiyakedome* 'sunscreen'). Lexical accent is indicated by an apostrophe.

In the accented word *kubika'zari* 'necklace' in Figure 1.1a, the lexical accent falls on the third mora (indicated by an apostrophe) whereas the unaccented word *hiyakedome* 'sunscreen' in Figure 1.1b lacks a lexical accent. In Tokyo Japanese, any word shows an F0 rise from the first to the second moras unless there is a lexical accent on the first mora, in which case the word starts with an F0 fall instead of a rise.

The F0 contour patterns in accented and unaccented words described above can be characterized by a sequence of *tones* (Pierrehumbert & Beckman 1988). There are two kinds:

lexical tone and *peripheral tones*. Lexical accent is represented by the bitonal accent H*+L, which is considered as a property of each lexical item. A H*+L tone is shown at the third mora in Figure 1.1a becuase *kubika'zari* is accented with the lexical accent on the third mora whereas the word in Figure 1.1b *hiyakedome* is not represented with such a H*+L tone because this word is unaccented.

There are two subclasses of peripheral tones: *phrasal tone* and *boundary tones*. There is only one phrasal tone assumed in Tokyo Japanese, phrasal H tone. The high-pitched point seen in the second mora of an accented and an unaccented word is captured by positing a phrasal tone H. The F0 contours in Figure 1.1 both show a peak at the second mora regardless of the word's lexical accent type. This peak is represented by a phrasal H tone. The phrasal H tone is considered to appear at the beginning of the prosodic constituent called the *Minor Phrase* (*MiP*), also as known as the *Accentual Phrase* (Pierrehumbert & Beckman 1988) (see below for prosodic constituency).[1]

Two types of boundary tones are assumed in Japanese, namely L% and H%.[2] A H% tone typically appears at the end of an utterance in an interrogative sentence, which accounts for the rise observed at that position. The treatment of the L% tone is more complex. The low-pitched point that appears at the beginning of a MiP is explained by positing a L% tone. For instance, consider the noun phrase *tomodachi-no hiyakedome* '(my) friend's sunscreen' (both words are unaccented) in (1). A typical rendition of this noun phrase would contain two

---

[1] The phrasal H tone disappears (or at least cannot be separated from the accent H*), when the lexical accent falls on the second mora of a word. For example, the accented word *ani'yome* 'sister in law' has its lexical accent on the second mora *ni*, the mora on which the phrasal H would appear if the lexical accent fell on another mora. Since the lexical accent of this word is on the second mora, there is only a single F0 peak of the lexical accent observed in an F0 curve of this word.

[2] Recent studies on spontaneous speech suggest that there are more types of boundary tones in Japanese (Maeda & Venditti 1998, Maekawa, Kikuchi, Igarashi & Venditti 2002, Venditti, Maeda & van Santen 1998). For example, Maekawa et al. (2002) assume complex boundary tones that could appear at the MiP-final position such as LH%, HL% and HLH%.

MiPs with an F0 dip after *–no*, indicating the boundary between them, as shown in Figure 1.2.

In this F0 contour, $L\%_2$ could be interpreted as a boundary tone of MiP_2 since it is linked to

the first mora of the second word marking the left edge of MiP_2. According to Pierrehumbert

& Beckman (1988), however, $L\%_2$ is a boundary tone for MiP_1 rather than MiP_2. The reason

for this is that the L% tone would inherit the effects of *downstep* (a pitch range-lowering

phenomenon in which a word is realized lower when it is preceded by an accented word then

when by an unaccented word. See the next section). That is, Pierrehumbert & Beckman

showed that the L% tone between two MiPs is realized lower when the preceding phrase is

accented than when it was unaccented. Based on this observation, the L% tone is interpreted

as a property of the preceding MiP rather than the following MiP.

(1) Tomodachi   -no   hiyakedome
     friend   Genitive   sunscreen
                  case marker
     '(my) friend's sunscreen'

$[L\%_1 H_1$ $L\%_2]$ $[$ H_2 $L\%_3]$

[to   mo   da   chi   -no]$_{MiP1}$ [hi ya   ke   do   me]$_{MiP2}$

Figure 1.2 F0 curve of the noun phrase *tomodachi-no hiyakedome* '(my) friend's sunscreen'.
Both words are lexically unaccented.

A L% tone also appears at the beginning of an utterance, which explains the

utterance-initial low pitch as illustrated as $L\%_1$ in Figure 1.2. Since an utterance-initial MiP is

not preceded by any other MiP, the low point cannot be attributed to the boundary tone of the

preceding MiP as seen in the case of $L\%_2$. L%1 is considered to be inserted by a

phonological rule and linked to the first mora of the initial word (Pierrehumbert & Beckman 1988).

It is necessary to point out that there is another way of rendering an unaccented word. Compare the contours Figure 1.1b and Figure 1.3.
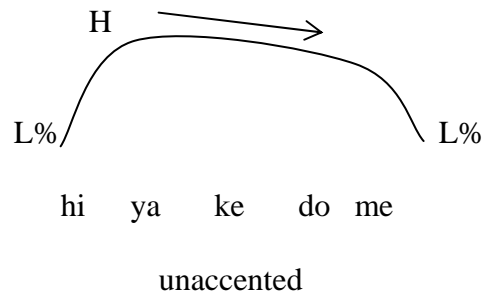
Figure 1.3 A variation of F0 contours of an unaccented word. The phrasal H tone "spreads" to the final mora.

Sugahara (2003) pointed out that there are two types of speakers with respect to the phonetic realization of the H phrasal tone in an unaccented word. One type of speaker shows a "linear interpolation" between the H phrasal tone to the following L% tone that is associated to the beginning of the next word (Figure 1.1b). The speakers who participated in Pierrehumbert & Beckman's (1988) study also exhibited this pattern. However, Sugahara showed that there is another type of Japanese speaker who shows a different realization pattern for the H phrasal tone in an unaccented word. These speakers "spread the H phrasal tone" from the second mora to the final mora of the unaccented word (Figure 1.3). This difference will become important when we discuss the effect of F0 plateaux on perceived prominence in Chapter 3.

1.4.2 Downstep

Lexical accent in Japanese influences not only the F0 pattern of the word that has the accent

but also that of the following word. Specifically, in a sequence of two or more words, a word is realized lower when it is preceded by an accented word than by an unaccented word. This phenomenon is called *downstep* or *catathesis* (Poser 1984, Pierrehumbert & Beckman 1988). Example F0 curves are given in Figure 1.4.

Downstepped   Not downstepped

a. accented   accented   b. unaccented   accented
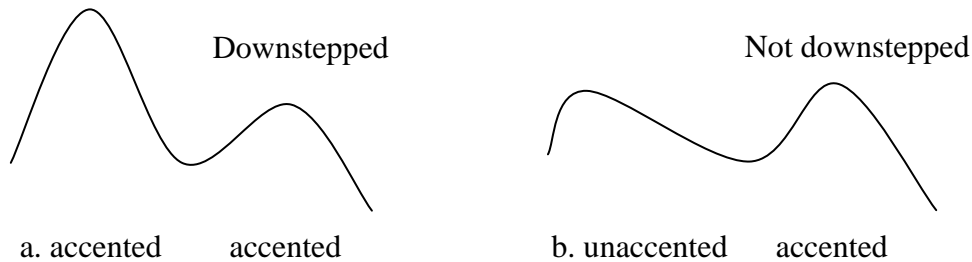
Figure 1.4 An illustration of downstep.

Pierrehumbert & Beckman (1988) claim that downstep is confined within the prosodic constituent called the *Major Phrase* (*MaP*), also as known as *Intermediate Phrase* (Pierrehumbert & Beckman 1988). MaP is the prosodic constituent that is one-level larger than MiP. According to Pierrehumbert & Beckman, downstep is only observed between two words are within the same MaP. They argue that when a word in an utterance is given a contrastive focus (Rooth 1992), it does not undergo downstep even if the preceding word is accented. Instead, the focused word is realized with an expanded pitch range, which yields a very high, non-downstepped F0 peak. Pierrehumbert & Beckman explain that the focused word is not downstepped because the contrastive focus introduces a MaP boundary at its left edge, and because the focused word is not within the same MaP as the preceding word, downstep is not observed.

However, this view has been called into question by Kubozono's (2005) experimental results, based on which he argues that downstep persists across MaPs. Kubozono tested

whether focus blocks downstep using the *wh*-word *nani* 'what'. He took advantage of the property that a *wh*-word is inherently focused. He found that the F0 peak of the *wh*-word was consistently lower after an accented word than after an unaccented word, which suggests that focus does not block downstep.

Based on Kubozono's (2005) findings suggest that MaP is not the domain of downstep in Japanese, leaving the question of which prosodic constituent is the domain of downstep for the future investigation.[3] We can only say at this point that the domain of this tonal phenomenon, if it has one, must be some prosodic constituent larger than MaP.

If MaP is not the domain of downstep, what tonal phenomenon would be characterized by this prosodic constituent? One possibility is the F0 *upstep* phenomenon seen at the left edge of a focused word. Though Kubozono's (2005) results suggest that focus may not introduce a MaP, still a significant F0 rise is always observed at the left edge of a focused word. An important additional observation is that this F0 rise is consistently greater than the rise typically observed at the left edge of an unfocused word; in that case a steep F0 downtrend from the preceding phrase is observed between the two MiPs. This suggests that the upstep observed at the left edge of a focused word marks a prosodic boundary that is larger than MiP, and the most straightforward candidate is MaP.

1.4.3   Prosodic structure

In this study, I assume the prosodic constituent hierarchy in Figure 1.5 (Kawahara & Shinya 2005, Selkirk 2005, and Sugahara 2003 among others). The prosodic constituents that are

---

[3] Kubozono's (2005) interpretation of the results is different from the one presumed here. He argues that since downstep is not blocked by focus, the boundary appearing at the left edge of a focused word is not a MaP-level boundary, based on the assumption that MaP is defined as the domain of downstep. He proposes a structure that involves recursive prosodic constituents, following Ladd (1986).

relevant here are MiP and MaP. MiP, also as known as 'Accentual Phrase' by Pierrehumbert & Beckman's (1988) work, roughly corresponds to a word plus a particle and is considered as the domain of an F0-rise phenomenon called *initial lowering*, in which F0 rises from the first to the second moras of a word signaling an MiP boundary.[4] With respect to tonal specification, MiP is characterized by the presence of a L% boundary tone followed by a phrasal H tone. The prosodic constituent above MiP is MaP, also known as 'Intermediate Phrase' (Pierrehumbert & Beckman 1988). As has been pointed out above, MaP is often claimed to be the domain of downstep because it has been considered that downstep is cancelled at the left edge of a MaP. However, this claim has been called into question by Kubozono (2005). MaP may be the domain of some other tonal phenomenon, and one such possibility is the upstep observed at the left edge of a focused word and a syntactic XP.

Utterance
|
Intonational Phrase
|
Major Phrase (MaP) (a.k.a. Intermediate Phrase)
|
Minor Phrase (MiP) (a.k.a. Accentual Phrase)
|
Prosodic Word
|
Foot
|
Syllable
|
Mora
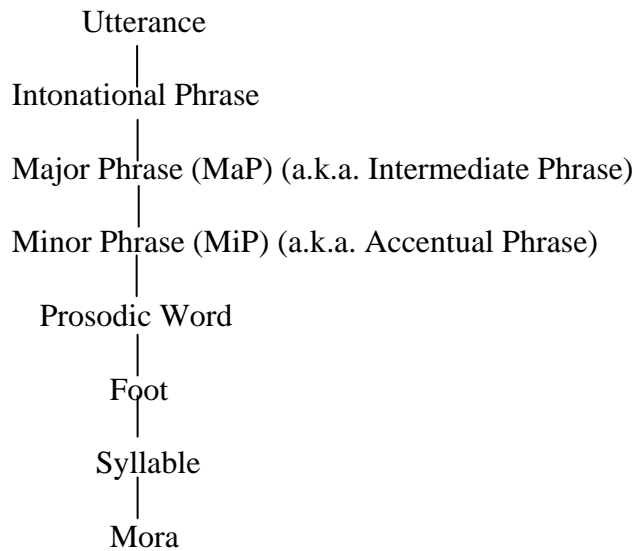
Figure 1.5 Prosodic hierarchy assumed in the dissertation.

---

[4] The term 'initial lowering' reflects the view that F0 shows a dip at the between two MiPs. However, it is possible to characterize the same phenomenon in a different way, which I am adopting here: F0 shows a rise at the beginning of a MiP. This 'initial rising' and initial lowering are thus different characterizations of the same phenomenon.

1.5 Overview of the dissertation

The remainder of the dissertation is organized as follows. Chapter 2 reviews previous research that studied factors that affect perceived prominence. This review covers virtually all of the factors that influence the perception of intonational prominence including linguistic and paralinguistic ones.

In Chapter 3, I report on Experiment 1, which provides the basis for the experiments that follow. The experiment examines how lexical accent affects perceived prominence using naturally-rendered utterances, which gives a basis for the later experiments in which acoustically manipulated stimuli are used. The use of natural utterances means that listeners can use both the acoustic and the structural properties of lexical accent to identify the accent types of the words in an utterance. Based on the results of Experiment 1, I propose a perceptual process that normalizes accentual boost: accentual boost normalization.

Chapter 4 describes Experiment 2, which examines the validity of the accentual boost normalization account by testing it together with another account. This alternative account is based on Knight's (2003) finding that an F0 contour with an F0 plateau induces more perceived prominence than an F0 contour without one. I show that the results of Experiment 2 lend support to the accentual boost normalization account.

In Chapter 5, where I report on Experiment 3, I show that the perceptual compensations seen in Experiment 1 are also found with only the structural property of lexical accent. In Experiment 3 uses acoustically manipulated stimuli in which the F0 contours are ambiguous between the accented and the unaccented types and hence listeners cannot identify the accent types from the acoustic property of the words in the stimuli. Yet, they can still readily recognize the accent types since they do so the words from their segments. That is, they can

judge the accent type of the words from their segmental contents. The results of the experiment show that the structural property plays a crucial role in the perception of prominence.

Chapter 6 focuses on the flip side of what was examined in Chapter 5: the effect of the acoustic property of lexical accent on prominence perception. Experiment 4 is reported, which tests whether or not accentual boost normalization and downstep normalization are observed based only on the acoustic property of lexical accent. To test this, stimuli are created such that the acoustic property conflicts with the structural property. Experiment 4 shows that accentual boost normalization and downstep normalization are both triggered by the F0 contour pattern that is characteristic of an accented word, regardless of whether the words' structural property marks that it is accented or unaccented.

Finally in Chapter 7, I summarize the main findings from the four experiments and conclude the dissertation with discussions of theoretical implications that the experimental results have and of the issues that remain to be investigated.

CHAPTER 2

FACTORS AFFECTING THE PERCEPTION

OF INTONATIONAL PROMINENCE

In this chapter, I review different factors that influence the perceived prominence of intonational peaks. They are sorted into three classes with respect to whether they reflect (1) phonetic/acoustic properties, (2) linguistic properties, and (3) other properties. There are five phonetic factors: excursion size (§2.1), position in utterance (§2.2), rate of declining baseline (§2.3), dependency of P2 on P1 height (§2.4), F0 plateaux (§2.5). One linguistic factor, focus structure (§2.6) and two other factors, experimental instructions (pitch height or prominence) (§2.7) and voice gender (§2.8) are discussed.

2.1 Excursion size

The first and best known factor that affects perceived prominence is *excursion size* of an F0 peak – how much F0 rise or fall the F0 peak exhibits (Gussenhoven et al. 1997, Gussenhoven & Rietveld 1988, 1998, 't Hart 1981, Ladd, Verhoeven & Jacobs 1994, Pierrehumbert 1979, Rietveld & Gussenhoven 1985, Terken 1991, 1994, Terken & Hermes 2000). The idea is very simple. As shown in Figure 2.1, an F0 peak with larger F0 movement (Figure 2.1a) evokes greater perceived prominence than another F0 peak with smaller F0 movement (Figure 2.1b). All of the cited studies experimentally showed a positive correlation between excursion size and perceived prominence or pitch height such that larger F0 movement yields greater perceived prominence (Gussenhoven et al. 1997, Gussenhoven & Rietveld 1988,

16

1998, Ladd, Verhoeven & Jacobs 1994, Knight 2003, Rietveld & Gussenhoven 1985, Terken 1991, 1994) or pitch height ('t Hart 1981, Knight 2003, Pierrehumbert 1979, Terken 1991, 1994)[1].

a   b
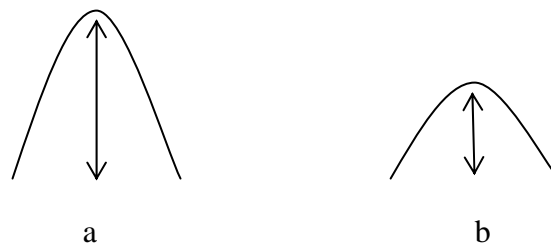
Figure 2.1 Two F0 peaks with different excursion sizes. The contour in (a) evokes more perceived prominence than the one in (b).

2.2   Position of F0 peak in utterance: the declination effect

It is also known that the location of an F0 peak in an utterance affects perceived prominence (Gussenhoven et al. 1997, Gussenhoven & Rietveld 1988, Pierrehumbert 1979, Terken 1991, 1994). Pierrehumbert (1979) showed that when perceiving the relative height of two F0 peaks in an utterance, the second peak (P2) has to be lower than the first peak (P1) in order for it to sound as high as P1. In one of her experiments, Pierrehumbert used reiterative speech, namely the repetition of a syllable (e.g. *mamamama…*), that mimicked the real utterance *the baker made bagels*. This pseudo-utterance contained two F0 peaks, which corresponded to the first syllables of *baker* and *bagels*. Pierrehumbert moved the P2 value up and down in small increments, with two different F0 values of P1: high and low. She asked her English listeners to indicate whether the first or the second stressed syllable sounded higher in pitch.[2]

---

[1] As will be discussed in the next section, Knight (2003) and Terken (1991, 1994) employed two different instructions in the tasks used in their experiments: one based on pitch height and the other based on perceived prominence. For this reason, I included these studies for both perceived pitch height and prominence.

[2] Here I am precise about how Pierrehumbert exactly asked her listeners to do, because it has been reported that

She found that for the high P1 value, when the two syllables were perceived to have the same perceived pitch height, P2 was much lower in F0 than P1. For the low P1 value, however, P2 was slightly higher than P1 when the pitch heights of two syllables were perceived to be the same.

Based on these results, Pierrehumbert (1979) proposed a perceptual process which normalizes *declination* – the gradual decline of F0 over the course of an utterance (Cohen, Collier & 't Hart 1982, Ladd 1984). Specifically, she assumed that listeners know that an F0 peak which occurs later in an utterance is realized lower than an earlier peak due to F0 declination and compensate for it in perception. As for the difference between the high and low P1 conditions, Pierrehumbert suggested that listeners expect more declination in wide pitch range utterances than in narrow pitch range utterances. When P1 is low, the declination rate that listeners expect may be flat (or even positive), but when P1 is high the expected declination rate may become well negative so that P2 is lower than P1 in F0 when the two peaks are perceived to have the equal pitch height. The relation between the pitch range (= peak height) of P1 and its effect on P2 is discussed more in §2.4.[3]

Hereafter I mainly use the term '(perceived) prominence' rather than '(perceived) pitch height' except where the distinction between prominence and pitch height is relevant (particularly in §2.7). The reason for this is that I want to avoid unnecessary confusion of the terminology and that most of the studies that are subsequent to Pierrehumbert (1979) focus on prominence rather than pitch height.

---

listeners response differently depending on which they are asked to judge between pitch height and prominence (Terken 1991, 1994). Pierrehumbert asked her listeners to judge pitch height, not prominence. I discuss the instruction issue in §2.7.

[3] Studies that carried out similar experiments showed that it is only when P1 is scaled with a quite low pitch range that the F0 value of P2 can be higher than P1 when P1 and P2 have the same perceived pitch height or perceived prominence (Gussenhoven et al. 1997, Terken 1991). Under most pitch ranges P2 is lower than P1 when they are perceived to be equal in pitch height/prominence.

2.3   Rate of declining baseline

Research after Pierrehumbert (1979) intensely examined the effects of declination on perceived prominence. The question was how to incorporate declination into theories of prominence assessment, which has its basis on excursion size. More specifically, researchers have focused on lines fitted to F0 peaks or valleys in an utterance contour like the one shown in Figure 2.2, which are called the *topline* and the *baseline*. Of the two lines, the baseline has especially drawn researchers' attention ('t Hart, Collier & Cohen 1990).
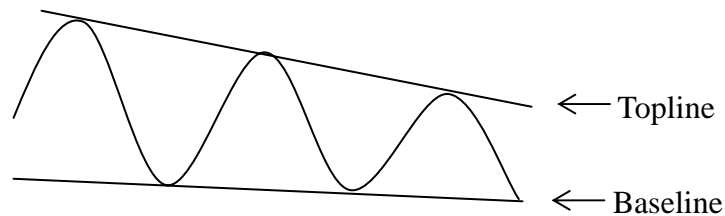
Figure 2.2 The topline and baseline of an utterance contour.

Effects of baseline are more opaque than those of topline. However, we can expect different slopes of baseline to lead to different excursion sizes. As seen in Figure 2.3, Terken (1991) hypothesized that the excursion sizes of adjacent F0 peaks should differ if the slope of the baselines differs. Note that the toplines of the two contours in Figure 2.3 are identical. However, because the baseline is steeper in Figure 2.3b than Figure 2.3a, the excursion size of P1 is smaller than that of P2 in Figure 2.3b while the relation is reversed in Figure 2.3a. Therefore, it was predicted that the perceived prominence of P2 relative to that of P1 would be greater in the contour in Figure 2.3b than in the one in Figure 2.3a.

P1 P2
Excursion size: P1 > P2

P1 P2
Excursion size: P1 < P2

a

b
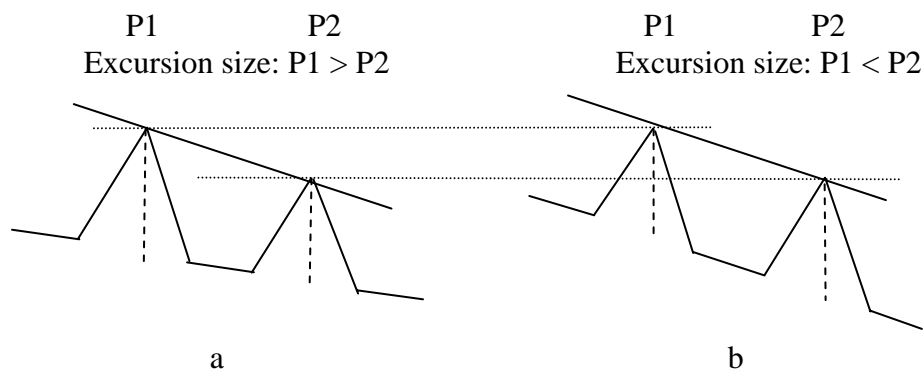
Figure 2.3 Different relations between two adjacent F0 peaks with respect to excursion size. Redrawn based on Terken (1991), p.1769.

Terken (1991) tested this hypothesis with Dutch listeners using reiterative speech. His stimuli consisted of a set of utterances that contained two F0 peaks with different baseline slopes, just like the contours shown in Figure 2.3. The listeners were asked to adjust the frequency of P2 so that it was judged to have the same pitch height or prominence as P1.[4] The results supported the prediction: Terken found that when the baseline was steeper P2's F0 value was much smaller than P1's F0 value for the two peaks to be judged equally prominent. Terken's experiment made clear that the assessment of perceived prominence needs information about the slope of the baseline, which suggests that listeners perceptually normalize the rate of declination.

Based on Terken's (1991, 1994) findings, Gussenhoven et al. (1997) examined further the effects of the properties of the baseline on the perceived prominence of F0 peaks in an utterance contour. Using real Dutch speech materials, they showed that the declination effect is observed in an utterance with only one F0 peak: in one of the experiments reported in their paper, a single peak with a fixed F0 was perceived to be more prominent if it appeared later in the utterance. Gussenhoven et al. (1997) interpreted this result as indicating that the

---

[4] Recall that Terken employed two experimental instructions: the pitch height instruction and the prominence instruction.

perceived prominence of an F0 peak is not computed based on adjacent F0 peaks in the same utterance.

In other experiments, Gussenhoven et al. (1997) showed that changes of the onset frequency of an utterance affect the perceived prominence of an F0 peak in the utterance but changes of the offset frequency of an utterance do not. They obtained these findings using two kinds of stimuli shown in Figure 2.4.

a   b
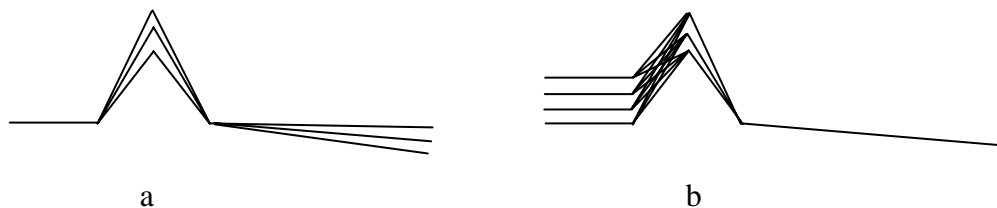
Figure 2.4 Schematic F0 contours used in Gussenhoven et al.'s (1997) Experiments III (a) and IV and V (b). Redrawn based on Gussenhoven et al. (1997), p.3016 and p.3018.

In Figure 2.4a, the height of the utterance-final frequency is varied independently from the peak height of the F0 peak, and in Figure 2.4b, the height of the utterance-initial frequency is varied independently from the height of the following peak. The listeners' task was to rate the degree of the speaker's emphasis on the stressed word between 0 (least emphasis) and 100 (most emphasis). Gussenhoven et al. found no effect of the offset frequency on the perceived prominence of the F0 peak, but there was a significant effect of the onset frequency height such that raising it reduced the perceived prominence of the F0 peak. Also, this onset effect was only observed when the duration of the onset was long enough (over 400 ms).

One of the important theoretical implications of these results is that the perceived prominence of an F0 peak is not assessed based on the excursion size that is measured as the distance between the peak height and the "observable" baseline. Rather, it is assessed on the

basis of the distance between the peak height and an "abstract baseline" that is computed by the listener. The results of Gussenhoven et al.'s (1997) "offset" experiment can be interpreted such that the observable baseline does not provide the basis for perceived prominence assessment. In the experiment a considerable portion of the observable baseline was changed when the offset height was varied up and down. Despite the manipulations, there was no effect on the perceived prominence of the F0 peak. Conversely in Gussenhoven et al.'s (1997) "onset" experiments, a large part of the observable baseline was left intact even when the onset frequency was changed, but the results showed a significant effect on the perceived prominence of the F0 peak. Both of those findings strongly suggest that the actually observed baseline is not the reference with respect to which the perceived prominence of F0 peaks is computed. Gussenhoven et al. argue that the baseline based on which perceived prominence is assessed is abstract and not directly obtained from the F0 contour.

2.4   Dependency of P2 on P1 height

The next factor that plays a role in the interpretation of intonational prominence is related to the finding by Pierrehumbert (1979) about the difference in the size of the declination compensation effect between her low P1 and high P1 conditions. She found that when P1 and P2 were perceived to have the same pitch height, the difference between the two peaks was greater when P1 was in a high pitch range than when it was in a low pitch range. This effect has been replicated in subsequent experiments in which relative perceived prominence of F0 peaks was judged (Gussenhoven & Rietveld 1988, Gussenhoven et al. 1997, Ladd, Verhoeven & Jacobs 1994, Terken 1991, 1994).

Figure 2.5 graphically illustrates this perceptual effect. Typically, when two peaks in an

utterance sound equal in perceived prominence, P2 is lower than P1, just as Pierrehumbert (1979) and other studies showed. This is shown in the arrow in Figure 2.5a. When P1 is raised P2 needs to be raised too, but crucially, P2 does not have to be raised by the same amount as P1. As a result, the difference between P1 and P2 is greater when P1 is higher, as shown in Figure 2.5b.[5]
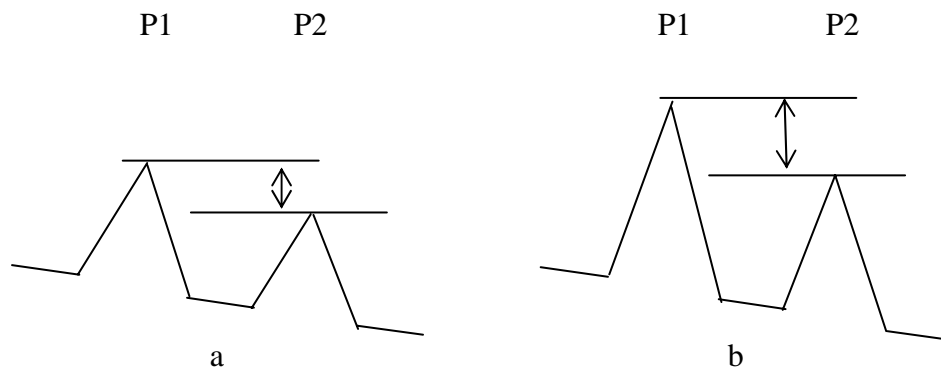
Figure 2.5 Illustration of the baseline adjustment effect.

Ladd, Verhoeven & Jacobs (1994) observed essentially the same effect as the one illustrated in Figure 2.5 through a set of perceptual experiments with English listeners. Their finding was that with a fixed P2 height, an increase of P1 height induced more perceived prominence on P2. However, there is an interesting complication in Ladd et al.'s experiments that is worth noting. This effect of perceptual dependence was seen only when the height of P2 is moderate and at the same time the listeners were phonetically untrained. Ladd et al. (1994) tested different P2 values. They observed the perceptual dependence effect when P2 was 140 Hz and their listeners were phonetically trained. They did not observe the effect

---

[5] Recall that in Pierrehumbert's (1979) experiment P2 was actually slightly higher than P1 when P1 was low. This can be considered as an extreme case of this effect. If the F0 difference between the two peaks is a function of P1 height when they have the same perceived prominence, it is possible that P2 is very close or higher when P1 is very low.

when the listeners were untrained or when P2 was a higher 160 Hz. They argue that when F0 peaks are within a normal non-emphatic pitch range, listeners treat the pitch range of the two-peak utterance as a whole, and their prominence are assessed on the basis of the single pitch range in which they are scaled. However, when F0 peaks are higher than a normal pitch range, they are interpreted as being scaled in different pitch ranges and their prominence is individually assessed based on each pitch range (see Ladd 1994 for a phonological account that is consistent with this view and Hayes 1994 for a critique of that account).

Gussenhoven et al. (1997) proposed a model that accounts for the perceptual dependence between two successive F0 peaks. The basic idea goes along with that of Pierrehumbert's (1979), which says that greater pitch range induces more declination on the listener's side. Gussenhoven et al. claim that the slope of the abstract baseline becomes greater as P1 increases. I call the perceptual effect *the baseline adjustment (BA) effect*.

2.5   F0 plateaux

The discussions we have had so far only consider F0 peaks with a steep rise followed by a steep fall, making a single turning point. However, Knight (2003) demonstrated that flat stretches of F0 contour, or F0 plateaux, also play a role in the interpretation of intonational prominence. Specifically, she showed that the perceived prominence of a word relative to the preceding word was greater when the word had a plateau contour than when it had a peaked contour. Consequently, the difference between the height of P1 and that of P2 was greater in the contour with an F0 plateau in P2 than in the contour without one. This is shown in Figure 2.6. P2 in Figure 2.6b is perceived to have the same prominence as P1, just like in the utterance shown in Figure 2.6a, but P2 in Figure 2.6b is lower than P2 in Figure 2.6a because

the former peak is an F0 plateau while the latter peak is not.

Equal prominence   Equal prominence

P1   P2   P1   P2

a   b

Figure 2.6 Schematic F0 contours showing that a contour with a plateau has more perceived prominence than a contour without it. From Knight (2003), p.163, redrawn with a slight modification by the author.

In a different experiment, Knight (2003) tested the effects of duration of F0 plateaux. The experiment had three conditions: peaked contour, 50 ms plateau, and 100 ms plateau. She found a significant difference in perceived prominence between the peaked contour condition and the other two plateau conditions, but no difference between the two plateau conditions.

Knight (2003) argued that the plateau effect can be explained in terms of the sluggishness of auditory detection of pitch. That is, in a peaked contour, the F0 stays on its peak value only momentarily, which may make it difficult for the auditory system to phase lock to the very highest frequency. The duration is too short for the auditory system to detect the peak frequency. In a plateau contour, on the other hand, it is much easier for the auditory system to detect the information of the F0 peak because the highest value is sustained long enough. Knight explained that there was no durational effect between the two plateau conditions because both had a plateau that was long enough for the auditory system to phase lock to its peak frequency.

2.6   Focus structure

Focus structure is one of the most well-known linguistic factors that affect perceived prominence. In many languages a focused word, whether it is a presentational focus or a contrastive focus (Kiss 1998, Belletti 2004, Katz & Selkirk 2005), is spoken with an accent, by virtue of which a greater degree of stress than the other words in the utterance is placed on the word. As a result, the stressed syllable of the focused word becomes longer in duration, higher in F0, and greater in its peak amplitude (Lehiste 1970). Thus, we can say that a focused word is more prominent than an unfocused word.

The nature of accented words was extensively studied with psycholinguistic methods. One of the primary findings in the research is that accented words are processed more easily than non-accented words (Shields, McHugh & Martine 1974, Cutler 1976, Cutler & Fodor 1979, Cutler & Foss 1977, Cutler & Darwin 1981, Hornby 1972). One explanation of this effect is that accented words have greater acoustic clarity than non-accented words. However, Cutler (1976) showed that that is not all that is going on. She showed that a word is processed more easily when it is in a focused position than when it is not, even when there are no acoustic cues that tell the word is focused. Cutler recorded two kinds of sentences, one in which the target word is in a focused position and one in which it is in an unfocused position. An example of each kind of the sentences is shown in (1a) and (1b):

(1)   a. She managed to remove the _dirt_ from the rug, but not the berry stains.
       b. She managed to remove the *dirt* from the <u>rug</u>, but not from their clothes.
       c. She managed to remove the dirt from the rug.

The sentence in (1a) can be interpreted as the target word *dirt* (indicated by italic) being in a focused position (indicated by underline) while it is impossible to have such an interpretation

26

for the target word in the sentence in (1b); rather the focus is on *rug*, not on *dirt*. Cutler spliced out of the target words from (1a) and (1b) and replaced them with the same target word that was taken from a separately recorded control sentence, shown in (1c), in which no focus occurred in *dirt*. This manipulation made the acoustic properties of the target word identical between (1a) and (1b). Using a phoneme-monitoring task (Connie & Titone 1996 for a review of the method), Cutler tested whether initial phoneme /d/ of the target word is detected faster in (1a) or (1b). Results showed that the target phoneme was detected significantly faster in (1a) than in (1b). Crucially, this difference cannot be explained by the acoustics of the target words since they were acoustically identical. Cutler suggested that the contextual prosodic difference before the target word between the two sentences is the source of this effect. She proposed that listeners search for sentence accent based on the prosodic information that precedes the target word.

Cutler & Fodor (1979) further showed that the same kind of advantage to focus in phoneme monitoring is seen when a target word is directed to be focused only by the preceding question sentence. In their experiment, sentences such as the one in (2a), recorded with the neutral intonation pattern, were presented to listeners after either of the two types of question sentences as in (2b) and (2c). When (2c) precedes (2a), (2a) can be said to have the focus structure in which *blue* is focused. However, when (2b) precedes (2a), the same sentence can be said to have the focus structure in which *corner*, not *blue*, is focused. Under these circumstances, Cutler & Fodor showed that the initial phoneme /b/ of *blue* in (2a) was detected significantly faster when it was preceded by (2c) than by (2b). Also, the initial phoneme /k/ of the word *corner* was detected faster after (2b) than after (2c).

(2) a. The man on the corner was wearing the *blue* hat.
     b. Which man was wearing the hat?
     c. What hat was the man wearing?

An important implication of these results is that Cutler's (1976) proposal that listeners actively search for words which will be accented can be reinterpreted from the perspective of a search for focus. In Cutler & Fodor's experiment the sentences were acoustically constant with neutral intonation contours, which indicates that listeners cannot use contextual prosodic cues to predict an upcoming accent. Rather, as Cutler & Fodor suggested, it is appropriate to consider that they use semantic information to search for focus in an utterance. Also, Cutler & Fodor's results suggest that the processing of focus does not crucially depend on any acoustic properties.

Cutler & Darwin (1981) extended Cutler's (1976) experiment and obtained the evidence that supports Cutler & Fodor's (1979) proposal. They showed two points. First, they demonstrated that when F0 contour is monotonized the predicted accent effect was unchanged. Second, they showed that this accent effect is not influenced by the manipulation of the duration of closure for the target stop phoneme. These findings suggest that the search for focus is at least partly independent of the acoustic properties of the utterance including the target word.

The psycholinguistic studies just reviewed above have made a significant contribution to out understanding of the processing of focus. However, they all deal with focus with respect to processing, not with respect to prominence. What insights would they have into perceived prominence?

Here is one consideration. The reviewed work on the processing of focus showed that a focused word in an utterance is processed more easily than an unfocused word because

28

listener's attention is directed to searching for focus in sentence comprehension. The work also showed that focus is not crucially dependant on its acoustic manifestation in the sense that a focused word does not have to have a higher F0, longer duration, and greater intensity than an unfocused word for it to be interpreted as focused. Based on these findings, we may consider that a focused word is processed more easily than an unfocused word because it is more prominent by virtue of the abstract property of focus, which directs listeners' attention to the focused constituent.

Vanio & Järvikivi (2006) has recently proved the idea that the abstract property of focus contributes to perceived prominence. They took advantage of the effect of word order in Finnish on the interpretation of a phrase as contrastively focused. In Finnish, the sentence *Menemme laivalla Lemille* "We go by boat to Lemi" has the canonical word order, and can be used as an answer to the question "What will you do tomorrow?". When the last two words are switched, as in *Menemme Lemille laivalla*, the sentence now expresses *laivalla* "by boat" as contrastively focused; thus it can be translated into "It is by boat we are going to Lemi" as an answer to the question "By what means are you going to Lemi?"

Vanio & Järvikivi (2006) first conducted a perceptual experiment that examined how prominence is assessed between two F0 peaks in an utterance. They used the sentence *Menemme laivalla Lemille*, with the canonical word order. They then carried out another experiment with *laivalla* and *Lemille* reversed in their order. In both of the experiments, the F0 peak heights of *laivalla* and *Lemille* as well as the baseline declination were varied in five levels using the F0 synthesis model developed by Fujisaki and his colleagues (Fujisaki & Hirose 1984). The only difference between the two experiments was in the word order between *laivalla* and *Lemille*. The listeners in the experiments were asked to indicate which

of the last two words is given the main stress of the utterance by choosing one of the alternatives from (1) on the word *Lemille*, (2) on the word *laivalla*, or (3) either of them (the sentences had two F0 peaks (or pitch accents) on the first syllables of the second and third words). The results of the two experiments showed that when everything else was equal, the third word always had more perceived prominence in the sentence with the focus word order than the one with the canonical word order. That is, when the two F0 peaks in the utterance are perceived as equal in prominence, the second F0 peak height must be lower when it was in the syntactically focused position than when it was not. This suggests that listeners expect more prominence in the syntactically-focused position than in the syntactically-unfocused position and compensate for it by perceptually reducing the prominence of the word in the syntactically-focused position. The results also indicate that not only characteristics of F0 but more abstract linguistic structure affects the interpretation of intonational prominence.

To sum up, focus, as a linguistic factor, contributes significantly to perceived prominence. It affects perceived prominence not only with its acoustic manifestation as higher F0, greater duration and higher intensity but also with its grammatical property whereby listeners' attention is directed to items in which they expect it to occur.

2.7   Pitch height or perceived prominence?

Two 'other' factors influencing the interpretation of intonational prominence are discussed below. The first one is the type of experimental instructions. Terken (1991) found that the way the perceptual relation between two F0 peaks is judged depends on the type of instructions for the task. He compared two different instructions in his experiment: one in which his Dutch listeners were asked to judge relative pitch height between two F0 peaks in

reiterant speech (*mamama…*) and the other in which they were asked to judge relative prominence. In addition to the perceptual effect for declination that was described above, Terken found that the F0 difference between P1 and P2 was greater with the prominence instruction than with the pitch height instruction.

Terken suggested several possible accounts of this effect, which are based on the assumption that listeners are aware of the "linguistic" structure of the utterance in assessing prominence but they are not in assessing pitch height. For example, one of his accounts of this instruction effect is a perceptual normalization for *final lowering* – an extra F0 lowering at the end of an utterance (cf. Arvaniti & Godjevac 2003, Liberman & Pierrehumbert 1984). According to this account, listeners perceptually undo the extra lowering applied to the final F0 peak, and as a consequence, when two F0 peaks are perceived to have the same prominence, the physical height of P2 with respect to that of P1 is lower when P2 is in the final position than when it is in a non-final position. Crucially, this perceptual mechanism is assumed to be at play only when listeners assess prominence, not pitch height. However, Terken (1994) ruled out this account based on a similar experiment using three F0 peaks instead of two. He found the same instruction effect between the first and the second F0 peaks: the difference between them was greater in the prominence instruction than in the pitch height instruction when they were judged to have the same pitch height or prominence. Since the second peak was not in the utterance-final position, Terken rejected the final lowering account. Instead, he referred to an alternative possibility that concerns listeners' expectation about the application of downstep.

Terken's instruction effect itself is interesting since it implies that all that matters is listener's strategy rather than materials that they actually hear. However, his proposal needs

to be interpreted with caution. The underlying assumption of Terken's proposal seems to be the well-known distinction between non-speech and speech (cf. Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967). That is, Terken's assumption that instructions would matter itself rests on the more general assumption that listeners will respond differently when asked to make a linguistic judgment such as relative prominence, than a nonlinguistic one, such as relative pitch height. However, it is not clear why listeners could be aware of the linguistic structure of an utterance under the prominence instruction but not under the pitch height instruction. Moreover, Knight (2003) recently carried out a similar perceptual experiment with real speech adopting the pitch height and the prominence instructions. She found no instruction effects, which also suggests that we have to be careful in interpreting Terken's proposal.

2.8   Voice gender

The second non-linguistic factor affecting perceived prominence is the gender of the speaker voice, which was nicely demonstrated by Gussenhoven & Rietveld (1998). Using utterances produced by a Dutch female speaker, Gussenhoven & Rietveld resynthesized them into two versions of single-F0-peak contours, one that sounded as though it were produced by a female speaker and the other by a male speaker, by changing the first, second and fifth formants in the original speech. Thus, the artificially created male speaker utterances had a pitch range that was considerably high for a male speaker. Native Dutch listeners were asked to rate the degree of emphasis of the F0 peak. Gussenhoven & Rietveld found that listeners judged F0 peaks that had the formant properties appropriate for a male voice to be more prominent than those that had the formant properties appropriate for a female voice. This

result suggests that listeners compensate for voice gender based on the knowledge of the gender difference in pitch range. Listeners know that the neutral pitch range of a male voice is lower and narrower than that of a female voice. With the assumption that the stimulus utterance was produced by a male speaker, its perceived pitch level was higher than his neutral pitch level. Naturally, the F0 peak is interpreted as emphatic under this greater pitch level. As a result, female speech that is made to sound like male speech is perceived to have more perceived prominence than non-manipulated female speech.

2.9   Chapter summary

This chapter provided a thorough review of the past studies that investigated factors affecting the perception of intonational prominence. We saw five acoustic factors (excursion size, position in utterance, rate of baseline, dependency of P2 on P1 height, and F0 plateaux), one linguistic factor (focus structure), and two non-linguistic factors (experimental instructions and voice gender). It is rather surprising that only one factor has been seriously studied as a linguistic factor affecting perceived prominence. This situation ensures the necessity of the current study, which explores the effect of another linguistic factor on perceived prominence, namely lexical accent.

CHAPTER 3

EXPERIMENT 1: PERCEIVED PROMINENCE WITH

NATURALLY-RENDERED STIMULI

3.1   Introduction

This chapter presents Experiment 1, which builds a basis for the experiments that follow. The experiment tests the effects of lexical accent on perceived prominence in Japanese using 'naturally-rendered' stimuli. Natural speech contains both types of the properties about lexical accent briefly mentioned in the previous chapter: the acoustic and the structural properties. We look at the effects of these different types of lexical accent properties in the later experiments, but such further examination makes sense only if we find any perceptual effect of lexical accent on prominence with naturally-rendered speech. Experiment 1 tests the possibility of such effect.

In production, accent type in Japanese has two different sorts of influence on F0 contours. First, an accented word is realized with a higher F0 peak (hence larger excursion size) than an unaccented word, which is characterized as *accentual boost*. Second, within the same Major Phrase (MaP), an accented word triggers the downstep on the following word (Poser 1984, Pierrehumbert & Beckman 1988).

Experiment 1 tests the possibility of perceptual effects that compensate form accentual boost and downstep. The primary reason for suspecting such perceptual effects can be found in the way that accentual contrast is expressed in Japanese: it is mainly expressed by the

presence or absence of lexical accent.[1] For example, *de'nki* 'electricity' and *denki* 'biography' form a minimal pair by the presence of the former word's lexical accent, which the latter word lacks. If we consider this way of showing lexical contrast in Japanese in terms of perceived prominence, it is unlikely that lexical accent gives additional perceived prominence. If lexical accent gave prominence to words, we would have a strange situation where the more accented words a phrase contains the more prominence the listener perceive when hearing it. There is no purpose for the intonation system to do that.

A similar argument as the one above is applied to downstep. A word is downstepped when it is preceded by an accented word in the same MaP; it is not downstepped when it is preceded by an unaccented. If a downstepped word was perceptually less prominent than a non-downstepped word, it would be less prominent precisely because the former is preceded by an accented word whereas the latter by an unaccented word. This is again a strange situation. There is no good reason why the downstepped word has to have less perceived prominence in such an environment.

If there are perceptual normalization processes for accentual boost and downstep, what sort of processes could they be? The most straightforward conceptualization would be to consider that they are processes that undo the production effects. That is, the compensation for accentual boost would nullify the difference in F0 peak height between an accented word and an unaccented word by undoing the extra F0 boost that is added to the accented word in production. The compensation for downstep would nullify the differences in F0 peak height between a downstepped word and a non-downstepped word by undoing the F0 lowering

---

[1] Lexical contrast is also expressed by the position that the accent appears in a word. For example, the initially-accented *ha'shi* 'chopstick' contrasts with the finally-accented *hashi'* 'brdge'. However, the number of words that show this way of lexical contrast is much smaller than that of words showing contrast with the presence or absence of lexical accent.

applied to the downstepped word in production. I call these two normalization processes *accentual boost normalization* and *downstep normalization*, respectively.

The effects of accentual boost normalization and downstep normalization are expected to be observed when comparing sequences of two F0 peaks (or words) with various combinations of accent types in terms of relative perceived prominence.

First, if there is any perceptual normalization for accentual boost on perceived prominence, we should observe it by comparing sequence of *accented-accented* (*aa*) words with sequence of *accented-unaccented* (*au*) words. Specifically, comparing the difference between the first F0 peak (P1) and the second F0 peak (P2) when P1 and P2 are perceived to be equal in prominence allows us to see the accentual boost effect. What we expect to see is that the difference between P1 and P2 in F0 is much smaller in *aa* than in *au* when P1 and P2 have the same perceived prominence. This is so because listeners may use the knowledge that an accented word is realized higher than an unaccented word and compensate for it by perceptually undoing the extra F0 peak height brought by the accentual boost. Since the first word is accented in both *aa* and *au*, this compensation effect has to be attributed to the accent type difference in the second word. When P1 and P2 are perceived to have the same prominence, P2 needs to be higher in *aa* than in *au* because accentual boost normalization reduces to P2's perceived prominence. Consequently, P1-P2 difference becomes smaller in *aa* than in *au*.

Second, if we compare the relative perceived prominence of *aa* with that of *uu*, we can assess the downstep effect. Downstep is expected to appear in *aa* but not in *uu* because the first word is accented only in *aa*. The second words also differ in their accent types between the two conditions. Despite the potential concern that the accent type of the second word

36

should be identical across the conditions, what is crucial in this comparison is that the accent type of P1 and P2 should not differ *within* the condition. As long as this is the case, the effect of accentual boost normalization that is hypothesized to occur in the accented words in *aa* cannot emerge or be measurable since accentual boost normalization should occur in both of them, canceling out each other's effects. Therefore, it is appropriate to do the comparison between *aa* and *uu* in terms of the degree of P1-P2 difference in F0 when P1 and P2 are equal in perceived prominence. If downstep compensation is at play between *aa* and *uu*, P1-P2 difference should be smaller in *aa* than in *uu*. Listeners are expected to compensate for the downstep appearing in *aa* such that the perceived prominence of P2 is greater in this condition than the one estimated by the physical excursion size.

3.2   How to tease apart different compensation effects

Experiment 1 employs all of the accent type combinations (*aa, au, ua,* and *uu*). We take this tack because these four accent-type conditions make it possible to distinguish the different perceptual compensation effects. Here, we consider two kinds: the accentual boost effect and the downstep effect.[2]

We should note that if any perceptual normalization effect is found, it is very likely that more than one type of normalization is involved. For example, the *au* condition would in principle show the effects of accentual boost normalization and the downstep normalization. The accentual boost effect is expected on P1 because it is accented. The downstep normalization effect is also expected on P2 since it is preceded by the accented P1. Clearly, we need a way to tease apart these different "components" of perceptual normalization effect.

---

[2] Note that the declination effect is also expected. However, this effect is always expected in any utterance and we do not discuss it here.

One way to accomplish this task is to compare pairs of conditions where they differ only in one component in normalization effect and compare their effect sizes. Consider Table 3.1 below. This table shows which perceptual normalization effect is involved in the four accent-type conditions. The short thick horizontal lines represent the F0 peak values when P1 and P2 have the same perceived prominence. The symbols A and D denote the accentual boost effect and the downstep effect, respectively. Each symbol in the table has an arrow to its left, either upward (↑) or downward (↓). They represent how the normalization process affects the physical value of the F0 peak. An upward arrow, seen with the accentual boost effect indicates that the normalization process has the effect of increasing the F0 value of the accent peak. A downward arrow indicates that the normalization process has an effect such that it lowers the value of the affected F0 peak.

Accent combination	a	a	a	u	u	a	u	u
Perceptual effect	↑ A	↑ A / ↓ D	↑A	↓ D	—	↑ A	—	—

Table 3.1 Different components for perceptual normalization in four accent conditions. *A* indicates accented and *u* unaccented. The thick vertical lines represent F0 peak values of the words contained in the utterances. A and D stand the accentual boost effect and the downstep effect, respectively. The arrows represent how the normalization process affects the physical value of the F0 peak.

Let us consider Table 3.1 closely. First, the accentual boost effect (A) is at play in the *aa, au* and *ua* conditions. However, its effect can only be observed in the *au* and *ua* conditions, namely in the conditions where the accent types of the two F0 peaks differ from one another.

The *aa* condition involves this perceptual process since both of the F0 peaks are accented. Even so, however, we can not observe any perceptual effect in this condition because accentual boost normalization is expected to occur in *both* of the accent peaks and the accentual boost effects on P1 and P2 will cancel each other. In other words, the accentual boost effect exists but is invisible in the *aa* condition.

Turning to the downstep effect (D), assuming that the two F0 peaks are in the same MaP, downstep is expected to appear only in the conditions where P1 is accented, namely, *aa* and *au*. The downstep effect is expected to occur based on listeners' knowledge of P2 being realized lower than P1 when P1 is accented. Therefore, the P2 value should be much lower than the P1 value, whereas P2 would not necessarily be lower than P1 when P1 is unaccented.

Note that P2 in the *aa* condition has two arrows showing opposite directions. The accentual boost effect is expected on this accent peak because it is accented. Due to the accentual boost effect, the P2 value should be higher than the P2 values in the other conditions in which P2 is unaccented. However, in Table 3.1 the P2 value in the *aa* condition is not very high: in fact it is at the same level as P2 in *uu*. That is because the downstep effect is also seen in the *aa* condition and its effect cancels out the accentual boost effect that is also at work on P2.

When we take into consideration the two kinds of perceptual normalization effects discussed above (the accentual boost effect and the downstep effect), we have specific predictions about the magnitude of the overall normalization effect among the four conditions being considered here. Specifically, it is expected that the four conditions are ordered *au > aa > uu > ua* (where ">" indicates "greater in P1-P2 difference).

(1)   *au > aa > uu > ua*
        where ">" indicates "greater in P1-P2 difference"

As shown in Table 3.1, the *au* condition is predicted to show the greatest P1-P2 difference. The accentual boost normalization is expected to yield the effect that the F0 value of P1 is higher when P1 is accented than it is unaccented. At the same time, the downstep normalization would affect the F0 value of P2 such that it is lower when it is preceded by an accented P1 than when it is preceded by an unaccented P1. Consequently, the difference between P1-P2 is greatest among the four conditions. The *aa* condition is like the *au* condition in that accentual boost normalization and downstep normalization are both involved. However, as pointed out above, the accentual boost effect is invisible in this condition. We could only see the downstep effect, which decreases the value of P2. There is not much to comment on the *uu* condition. Since both accent peaks are unaccented, no major perceptual normalizations are expected. The F0 values of P1 and P2 should not be at the same level. Finally, The *ua* condition is predicted to show a negative P1-P2 value due to the accentual boost effect on P2. Note that only *au* and *aa* are the conditions where P1-P2 value is positive.

The hierarchy in (1) allows us to tease apart the different perceptual normalization effects. We can compare the entire compensation effect of *au* with that of *aa* to see the pure accentual boost effect. This is possible because these two conditions share the same component effects except for A of P2 in *aa*. Therefore, the pure accentual boost effect should in principle be observed by taking the difference between the P1-P2 difference in *au* and the P1-P2 difference in *aa*. Similarly, the pure downstep effect can be seen by comparing *aa* with *uu*. The *uu* condition is considered to involve no perceptual compensation. Note that

although the *aa* condition has the accentual boost effect and the downstep effect, the accentual boost effect is invisible, as discussed above.

There is an important assumption concerning the predictions in (1). The different perceptual normalization effects are assumed to be additive. That is, there is no interaction assumed among the different types of effects when they co-occur. We are not considering potential effects that would be caused by one effect combined with another effect. If the relationship is not additive, then the magnitude of the compensation effect will vary in a much more complex way when more than one effect surfaces, depending on which effects are involved, and we would not be able to see the pure effect of the accentual boost effect or the downstep effect. Therefore, such a situation will invalidate the predictions in (1). However, if our experimental results support the predictions, it can be safely concluded that the different perceptual effects are additive.

Note that we need to take Gussenhoven, Repp, Rietveld, Rump and Terken's (1997) baseline adjustment (BA) effect into account as well. The BA effect is described such that P1-P2 difference becomes greater as P1 increases. This effect is expected to occur in all of the other conditions since it only depends on the excursion size increase of P1. It might be the case that the degree of the BA effect (rate of increase in P1-P2) varies depending on the condition. However, our current consideration of the issue does not make any particular predictions regarding the BA effect.

The results of Experiment 1, which are reported below, provide solid evidence that there is a perceptual normalization effect that is purely attributed to the accentual boost effect and to the downstep effect. Moreover, the results confirm the predicted hierarchy of the magnitude of perceptual compensation in (1).

3.3 Methods

3.3.1 Stimuli

The author recorded sentences shown in (2) based on which the experimental stimuli were created. The experimental sentences were four sentences each of which consisted of three words, where the first two words were varied in accent type so that we had all possible combinations of accent type: *aa, au, uu,* and *ua*. The third word was kept unaccented. The syntactic structure of the sentences was $[[N_1\text{-Gen } N_2\text{-Nom}]_{NP} [V]_{VP}]_S$,[3] where the first two words form the subject NP followed by an intransitive verb. This structure was assumed to be mapped into the prosodic structure where there is no MaP boundary between the first and the second words and the entire sentence is in a single MaP because there is no syntactic XP boundary between the first and the second words (Selkirk and Tateishi 1991). A typical rendition of a sentence with this structure shows an initial lowering between the first and second words thus a MiP boundary, but does not show one between the second and final words, which means that the last two words are grouped into a single MiP. The first and the second words were four-moras long with lexical accent on the second mora if they were accented.

(2) a. *aa*
   Ina'mori-no   an'iyome-ga   inai.
   Inamori-Gen   brother-in-law-Nom not found
   Inamori's bother-in-law is not found.

   b. *au*
   Ina'mori-no   omiyage-ga   kieta.
   Inamori-Gen   souvenir-Nom   disappeared
   Inamori's souvenir disappeared.

---

[3] Gen=Genitive, Nom=Nominative

c. *ua*
    Inamura-no   an'iyome-ga   inai.
    Inamura-Gen   brother-in-law-Nom not found
    Inamura's bother-in-law is not found.

d. *uu*
    Inamura-no   omiyage-ga   kieta.
    Inamura-Gen   souvenir-Nom   disappeared
    Inamura's souvenir disappeared.

The recorded sentences were stylized and resynthesized using the "pitch synchronous overlap add" (PSOLA) method supplied with the *Praat* program (Boesma & Weenink, 1992-2006). The F0 peak height (the H* tones for the accented words; the phrasal H tones for the unaccented ones) was varied, in 4 steps for the first word and in 7 steps in the second word. The step size is 1 semitone with the base frequency of 100 Hz, so it was approximately 10 Hz. The exact F0 peak values used in the experiment are shown in Table 3.2. The other tonal values were fixed. These fixed values are shown with the actual F0 contours used in the experiment in Figure 3.1.

The two peak heights were not orthogonal in that the lowest/highest P2 value became higher as P1 gets higher. The values in brackets in Tables 3.2c and 3.2d are ones for the "F0 shoulder" in the unaccented words: the phrasal H tone spreading to the right edge of the word (cf. Sugahara 2003). The symbol "×" in Table 3.2d indicates that these stimuli were not used for the reasons that P2 value was lower than the preceding L% and, therefore, they sounded highly unnatural. The overall number of stimuli was 109 (4 P1 heights × 7 P2 heights × 4 accent-type conditions − 3 unused stimuli = 109).

aa

P1	P2 1	2	3	4	5	6	7	8	9	10
	133	140	149	158	167	177	188	199	210	223
1 194								▓	▓	▓
2 206	▓								▓	▓
3 218	▓	▓								▓
4 231	▓	▓	▓							

a

au

P1	P2 1	2	3	4	5	6	7	8	9	10
	121	128	136	144	153	162	171	186	192	204
	109	115	122	129	137	145	154	163	177	183
1 194								▓	▓	▓
2 206	▓								▓	▓
3 218	▓	▓								▓
4 231	▓	▓	▓							

b

ua

P1	P2 1	2	3	4	5	6	7	8	9	10
	137	145	153	162	172	182	193	205	217	230
1 168 (158)								▓	▓	▓
2 178 (167)	▓								▓	▓
3 189 (177)	▓	▓								▓
4 200 (188)	▓	▓	▓							

c

uu

P1	P2 1	2	3	4	5	6	7	8	9	10
	114	120	127	135	143	152	161	170	180	191
	(102)	(109)	(115)	(122)	(129)	(139)	(149)	(159)	(169)	(172)
1 168 (158)	✗	✗						▓	▓	▓
2 178 (167)	▓	✗							▓	▓
3 189 (177)	▓	▓								▓
4 200 (188)	▓	▓	▓							

d

Table 3.2 F0 values of the peak of the first (P1) and the second (P2) words. The shaded cells indicate the combinations of P1 and P2 values that were not used. The values in brackets are the values of the "F0 shoulder" of the unaccented word *Inamura* (P1) and *omiyage* (P2): phrasal H tone at the right edge of the unaccented words. The cells with "✗" indicates the combinations of P1 and P2 values that sounded so unnatural that they were not used in the experiment.
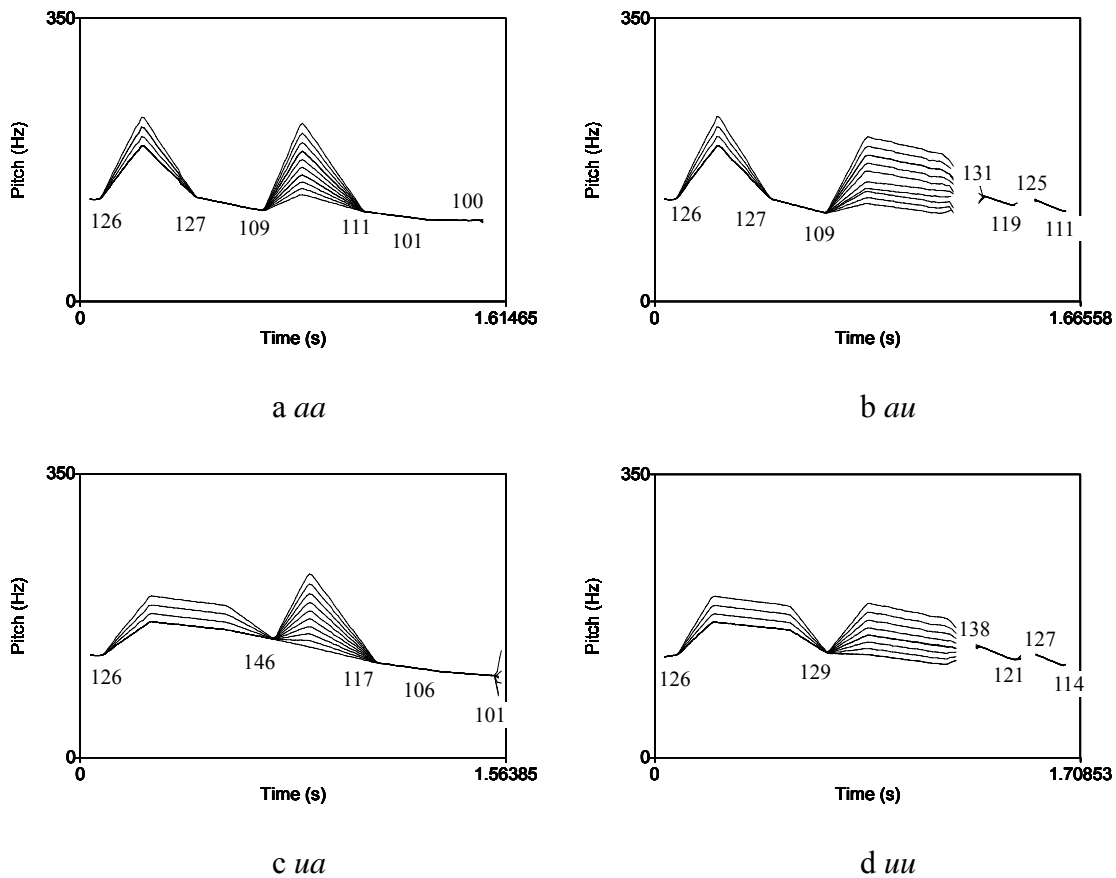
Figure 3.1 The F0 contours used in Experiment 1 with the fixed F0 values (Hz). The varied peak values are shown in Table 3.2.

3.3.2 Listeners

Nineteen native speakers of Japanese were recruited in Amherst, Massachusetts, USA. Mostly they were students at the University of Massachusetts Amherst and their spouses or people who lived in the area working for branches of Japanese companies. Their ages ranged early twenties to early forties. The mean was 30.5. They were originally from various different areas in Japan. Dialectal difference could not be controlled since there was a relatively small number of native speakers of Japanese in the area who were available for the experiment.

3.3.3   Procedure

The listeners each sat on a computer terminal and listened to the 109 stimulus sentences randomly 5 times through headphones. They were asked to judge which of the first two words were given more emphasis or importance by the speaker. They were also asked to respond as quickly as possible. The Multiple Forced Choice (MFC) in the *Praat* program was used to present the stimuli. The listeners had five 109-stimulus blocks of trials with an optional short break after each block of trials and a forced break after the third. The listeners heard the same stimuli in each block of trials, in a different order. In total 545 stimuli were presented. The listeners had a practice session to be familiarized with the stimuli where they heard two repetitions of 16 of the experimental stimuli. These stimuli consisted of the two endpoint P1 stimuli combined with the two endpoint P2 stimuli for each other four accent type conditions. The session lasted about 80 minutes including the instructions, the practice, the blocks of trials, and the breaks.

3.3.4   Analysis

The response rates for P2 (P2 sounding more emphatic than P1) were calculated for each listener. Then, using probit analysis (Finney 1971), the 50 % crossover point of the response rates expressed as a function of P2 height was obtained for each P1 height and for each listener. This point is assumed to represent the P2 value that is equivalent to P1 in its perceived prominence.

3.4   Results

3.4.1   Effects of accentual boost

The two graphs in Figure 3.2 plot the mean P2 F0 values obtained by probit analysis that have the same perceived prominence as P1, as a function of P1. Figures 3.2a and 3.2b show the data where P1 is accented and unaccented, respectively. The straight line in each graph represents the *y=x* function. Since it tells us the points where P1 and P2 have the same physical value, it serves as the reference line when we examine the normalization effects seen in the experimental conditions.

In Figure 3.2a, one can immediately notice that all of the mean P2 values reside well below the reference line, indicating that there is some normalization process involved. If the P1 and P2 values were identical when they sounded equal, these mean values would be right on the reference line. Moreover, the P2 mean values are much greater in the *aa* condition than in the *au* condition, which can be attributed to the difference in the accent type of the second word between the two conditions. In addition, the listeners' response was orderly in the sense that the P2 mean values increase as P1 does.

A 2×4 repeated-measures ANOVA with accent type and P1 height as the independent variables revealed that the P2 mean values were significantly lower in *au* than *aa* ($F(1,18)=68.57$, $p<0.001$) and that the increase of those values as a function of P1 was also significant ($F(3,54)=52.13$, $p<0.001$). No significant interaction between accent type and P1 height was found ($F(3,54)=2.248$, $p=0.093$).

The *uu* and *ua* conditions in Figure 3.2b show patterns similar to the other two conditions. Most of the mean P2 values are below the *y=x* function, just like the data in Figure 3.2a. However, the mean values are much closer to the reference line in Figure 3.2b,

which suggests that the sizes of the normalization effects, shown as the P1-P2 difference in F0 here, are smaller in *uu* and *ua* than in *aa* and *au*.
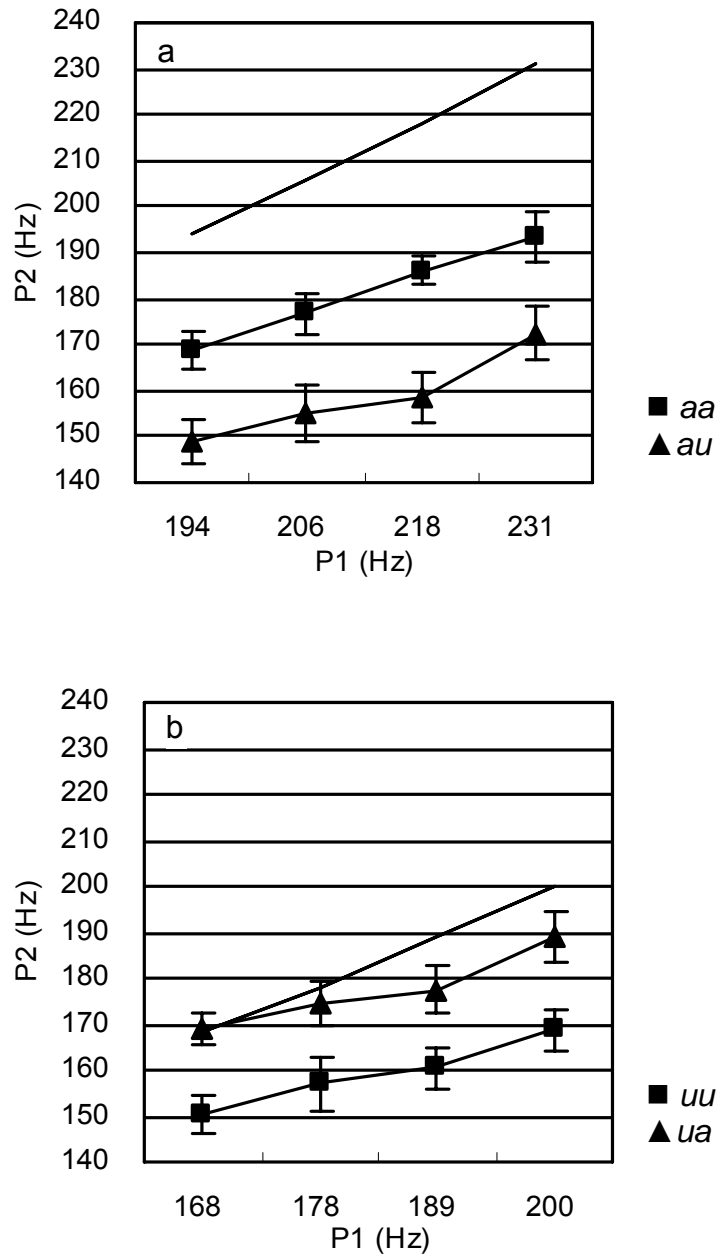
Figure 3.2 Mean values of P2 giving the same prominence as P1 as a function of P1 height for different accent-type conditions: (a) *aa* vs. *au*; (b) *ua* vs. *uu*. The straight line in each graph represents the *y=x* function. The error bars are 95 % confidence intervals.

One more noticeable feature in Figure 3.2b is the rate of increase of P2 mean values from low to high P1 values; it is smaller than *y=x*. In other words, the slopes formed by P2 mean values are less steep than *y=x*. This is readily observed in the *ua* function. This indicates that the difference between P1 and P2 gets greater as the height of P1 increases, which is the BA effect reported by Gussenhoven et al. (1997).

An ANOVA analysis confirmed the above observations. The same analysis as we did on the data in Figure 3.2a was carried out. There were significant differences for accent type ($F(1,18)=69.92$, $p<0.001$) and for P1 height ($F(3,54)=30.51$, $p<0.001$). The interaction was not significant ($F(3,54)=0.638$, $p=0.594$).

3.4.2   Effects of downstep

Figure 3.3 represents the same data as those in Figure 3.2 above, but here the data are shown with respect to the difference between the reference line and the mean P2 values (= P1-P2 difference). The figure contains all of the four accent-type conditions. This is possible because the values on the *y*-axis are not the values of P2 but the differences between P1 and P2, with the step size being identical among the conditions.
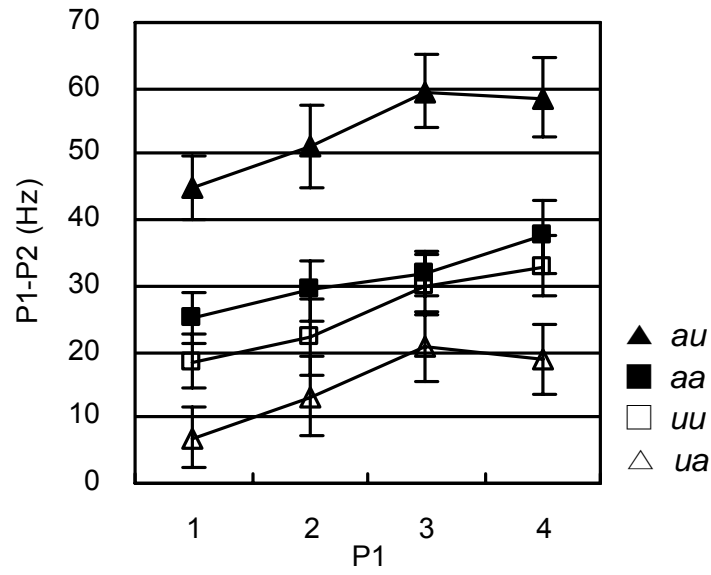
Figure 3.3 Mean *P1-P2* difference values for different accent-type conditions (*aa, au, ua,* and *uu*). The error bars represent 95 % confidence intervals.

As we have already observed in Figures 3.2a and 3.2b, *au* and *uu* show greater normalization effect than *aa* and *ua*, respectively. The mean P1-P2 values are greater in *au* than in *aa*, and are greater in *uu* than in *ua*. These results indicate that P2 is lower in *au* than *aa* and in *uu* than *ua* when P1 and P2 are assessed equivalent in perceived prominence. This shows that the perceptual normalization for accentual boost is at play on P2. Moreover, as discussed in 3.2, these differences can be considered the pure effect of accentual boost.

The downstep effect is also observed. *Aa* exhibits greater P1-P2 differences than *uu*, though the size of the differences between *aa* and *uu* are much smaller than the differences between *au* and *aa*. However, if this difference between *aa* and *uu* is statistically reliable, this can be considered the pure effect of the perceptual normalization for downstep. Since *aa* and *uu* do not differ from one another in accent type, the difference between the two conditions cannot be explained by the accentual boost effect. Presumably, accentual boost normalization

50

occurs in both peaks in *aa*, which makes its effect unobservable, and no accentual boost normalization should be at play in *uu*. This difference should only be explained by perceptual normalization for downstep.

Moreover, all of the mean P1-P2 values are positive in *uu*, which is indicative of the presence of the declination effect. Finally, the mean values are also positive in the *ua* condition, suggesting that the declination effect is greater than the accent effect in its magnitude. The overall order we obtained is *au > aa > uu > ua*, which is exactly what we predicted in (1) above.

We can also see that the P1-P2 values become greater as P1 increases; namely, the BA effect. We predicted that the BA effect would be seen in all of the accent type conditions, which are borne out.

To confirm these observations, a 4×4 repeated-measures ANOVA was carried out with accent type and P1 as the independent variables. There was a significant difference both for accent type ($F(3,54)=60.64$, $p<0.0001$) and for P1 ($F(3,54)=28.32$, $p<0.0001$). The interaction was not significant. Since we already established that there is a significant difference between *au* and *aa* and between *uu* and *ua*, the only concern here is whether or not *aa* and *uu* differ with each other significantly. The difference between *aa* and *uu* was small but significant ($F(1,18)=4.411$, $p=0.050$). Thus, the hierarchy *au > aa > uu > ua* has been statistically confirmed.

The non-significant result of the interaction suggests that the different components of perceptual normalization effects, the accentual boost effect and the downstep effect, are indeed additive. That is, perceptual normalization will be larger if two effects in the same direction are involved than if only one effect is involved.

51

As for the BA effects seen in all of the conditions, a regression analysis revealed that the slopes were significantly greater than zero ($F(1,18)=51.39$, $p<0.0001$), which confirmed the presence of the effect.

Note that the difference between *au* and *aa*, the one that is attributed to the accentual boost effect, is much greater than the difference between *aa* and *uu*. Shown in Figure 3.4 are the difference values between the two pairs of conditions. The triangles show the differences between *au* and *aa* in different P1 heights. We consider them to reflect the magnitudes of the accentual boost effect. The other pair is *aa* and *uu*, which shows the magnitude of the downstep effect as represented by the squares. As is clear from the figure, the mean difference values between *au* and *aa* (triangles) are much greater than those between *aa* and *uu* (squares). This suggests that though both the accentual boost effect and the downstep effect are induced by lexical accent, the size of the accentual boost effect is much greater than the size of the downstep effect.

This observation was statistically supported by a 2×4 repeated-measures ANOVA comparing the differences between *au* and *aa* with the ones between *aa* and *uu*, accent type and P1 as the independent variables. A significant result was found for accent type ($F(1,18)=16.07$, $p=0.001$). Neither P1 nor the interaction was significant (P1: $F(3,54)=0.89$, $p=0.450$), interaction: $F(3,54)=1.91$, $p=0.139$).
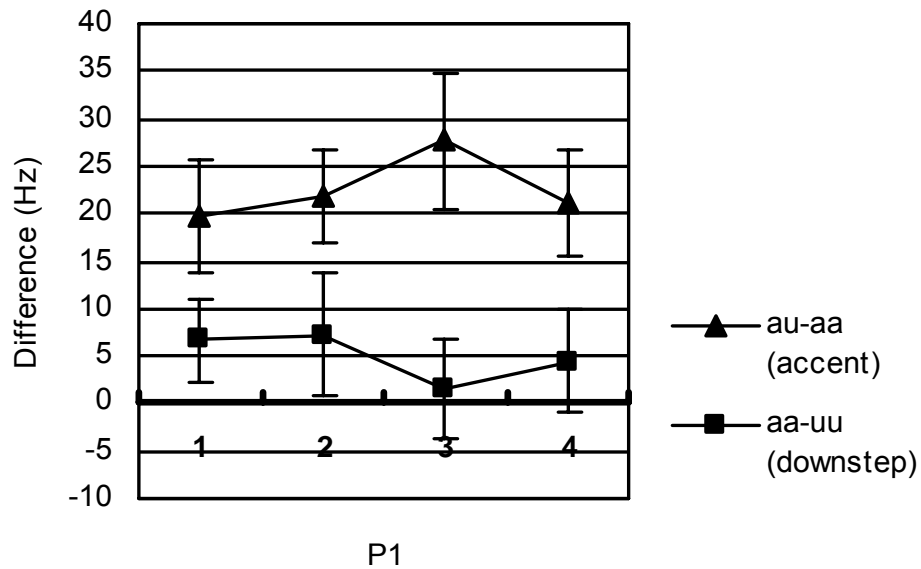
Figure 3.4 Mean difference values between *au* and *aa* (triangles) and between *aa* and *uu* (squares). The numbers 1 through 4 on the *x* axis indicate different p1 values from low to high. The actual values are not shown since the P1 values used differ between the condition where P1 is accented and the condition where it is unaccented. The error bars represent 95 % confidence intervals.

3.5   Discussion

3.5.1   Summary of the experimental results

Experiment 1 tried to give an answer to the question of whether perceptual normalizations for accentual boost and downstep play a role in the perception of intonational prominence. The answer is positive to both kinds of effect. The results have uncovered two perceptual normalization processes that have not been reported in the literature: accentual boost normalization and downstep normalization. Accentual boost normalization is triggered by the presence of a lexical accent (= accented word) and perceptually undoes the effect of accentual boost in the word that is brought by production. Downstep normalization is also triggered by the presence of a lexical accent, but the effect appears not on the word itself but

on the following word. The process undoes the production effect of F0 lowering. The extent of the effect differed such that accentual boost normalization was greater than downstep normalization in its magnitude.

The results of Experiment 1 have also revealed a certain relationship between the two different perceptual compensation effects being considered here, namely the accentual boost effect and the downstep effect. The results have shown that those effects are additive. This finding contributes to our understanding of intonational prominence interpretation in such a way that when considering the perception of prominence we can assume a fairly simple picture where different factors affecting perceived prominence are at play independently without interacting with each other.

In the remainder of this section, I would like to do two things. First, I describe the mechanism of accentual boost normalization and downstep normalization, the proposed perceptual processes that explain the results of Experiment 1. I then briefly discuss the theoretical implications of the experimental results.

3.5.2 Mechanism of accentual boost normalization and downstep normalization

I propose that accentual boost normalization and downstep normalization both have the same mechanism: undo the production effect on F0 in perception. In production, the F0 peak of an accented word is higher than that of an unaccented word by accentual boost. In perception, the portion of the F0 height added by accentual boost is perceptually removed by accentual boost normalization.

Let us see how this process explains the results of Experiment 1. The experiment has shown that the prediction in (1), repeated in (3), is correct.

(3)   *au > aa > uu > ua*
         where ">" indicates "greater in P1-P2 difference"

In this hierarchy, *au > aa* and *uu > ua* can be accounted for by accentual boost normalization.

Let us look at *au > aa* first. The account is visually shown in Figure 3.5.
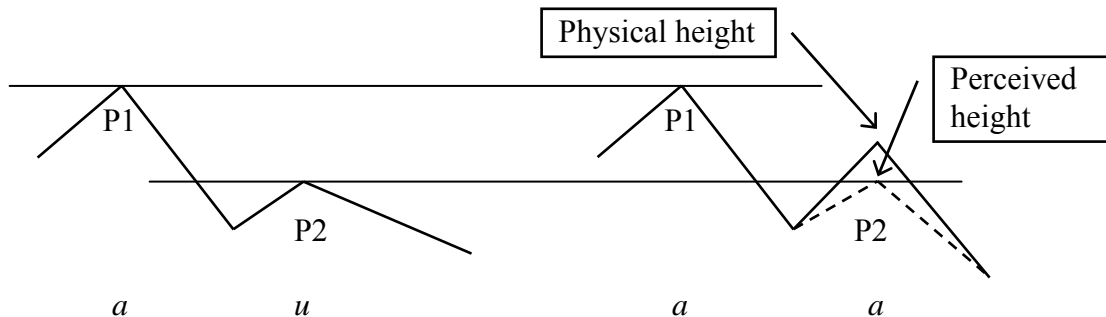
Figure 3.5 Explanation of the greater P1-P2 difference in *au* than *aa* by accentual boost normalization.

This figure schematically shows the two F0 contours for *au* and *aa* when P1 and P2 have the same perceived prominence in each of the contours. The relevant experimental result is that the F0 difference between P1 and P2 was greater in *au* than in *aa* when P1 and P2 were perceived to have equal prominence. What needs be explained here is how P1 and P2 could have the same perceived prominence in *au* and in *aa* despite the fact that the P1-P2 difference values are different between the two conditions.

In Figure 3.5, P2 in *aa* is accented so accentual boost normalization is at play. The perceived peak height of P2 is lower than the physical peak height because the portion of the P2 peak height added by accentual boost in production is taken away by accentual boost normalization in perception. As a result, the peak height difference between P1 and P2 are perceptually identical even though they are physically different. Note that accentual boost normalization should be seen in P1 in both *au* and *aa*. But this effect is not shown in Figure

3.5 because P1 is accented in *both* of the conditions, which makes the effect invisible.

The same account is possible for *uu* > *ua*, as shown in Figure 3.6. P1-P2 difference was greater in *uu* than in *ua* in Experiment 1 because the accentual boost applied in the P2 in *ua* is perceptually undone by virtue of accentual boost normalization whereby its perceived peak height is lower than its physical peak height. Consequently, P1-P2 difference is the same between *uu* and *ua* after accentual boost normalization is applied.
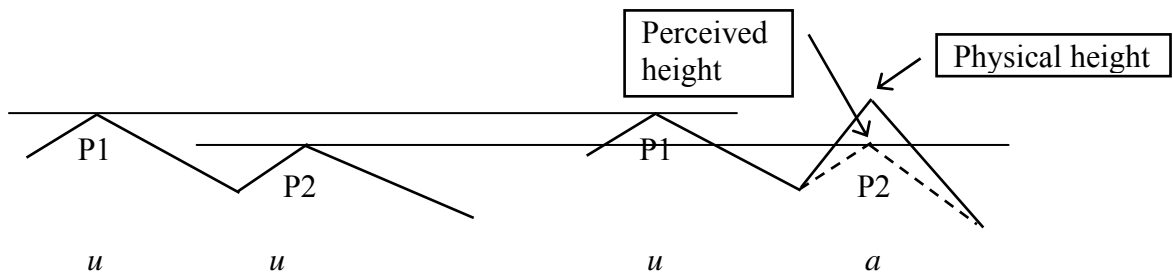
Figure 3.6 Explanation of the greater P1-P2 difference in *uu* than *ua* by accentual boost normalization.

The process of downstep normalization is shown in Figure 3.7. The relevant result is that P1-P2 difference is greater in *aa* than in *uu*. Note that to account for this result we cannot rely on accentual boost normalization. In *uu,* this perceptual process is irrelevant since both P1 and P2 are unaccented. In *aa,* where both peaks are accented, the effect of accentual boost normalization cannot be seen even though the process is at work in this sequence. A different perceptual normalization is necessary to explain the greater P1-P2 difference in *aa* than in *uu,* and that, I propose, is downstep normalization.
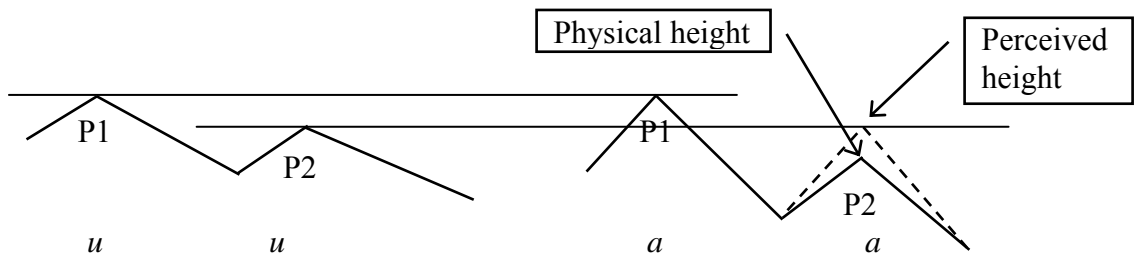
Figure 3.7 Explanation of the greater P1-P2 difference in *uu* than *aa* by downstep normalization.

Here is how downstep normalization works. In downstep normalization, the perceived peak height is higher than the physical peak height. In Figure 3.7, downstep normalization is at work on P2 only in *aa* since P1 is accented in *aa* while it is unaccented in *uu*. The P2 in *aa* undergoes downstep normalization whereby its peak height is boosted in perception. This boost makes P1 and P2 have the equal perceived prominence even though the physical value of P2 is much lower with respect to the preceding P1 in *aa* when we compare *uu* with *aa*. Note that the direction of the downstep normalization effect is opposite to that of the accentual normalization effect: accentual boost reduces perceived peak height whereas downstep normalization boosts it. However, the two perceptual processes share the mechanism of peak height adjustment: they both undo production effect.

3.5.3 Implications

Let us now move on to the theoretical implications. I want to make two points. First, the discovery of the accentual boost effect and the downstep effect clearly argue against the idea that F0 excursion size is one of the most dominant indicators of the prominence of F0 peaks. Many experiments reported in the literature have shown that how much F0 movement an F0 peak has directly affects its perceived prominence (Gussenhoven & Rietveld 1988, Rietveld

& Gussenhoven 1985, Terken 1991, 1994). However, Experiment 1 has shown that the situation is not that straightforward. The experimental results suggest that for a given P1 height, an unaccented P2 may have a much smaller excursion size than an accented P2 when the two peaks are perceived to be equal in prominence, due to accentual boost normalization. A downstepped P2 may also have the same perceived prominence as a non-downstepped P2 when its F0 excursion size is much smaller for the downstepped peak than for the non-downstepped peak. Without taking additional factors such as accentual boost normalization and downstep normalization into account, we cannot fully understand the interpretation of intonational prominence.

Second, the results of Experiment 1 also suggest that theories of the perception of intonational prominence need to take language-specific effects into account. The two perceptual normalization processes proposed on the basis of the experiment are both closely related to lexical accent, which is specific to Japanese. In production, lexical accent affects intonational patterns in ways that are specific to the language. In perception, as we found, lexical accent also affects the perceived prominence of intonational peaks in ways that are specific to Japanese. Theories of prominence perception have to account for this interaction between lexical and intonational features with respect to perceived prominence.

3.6   Chapter summary

In this chapter I tried to establish that the assessment of perceived prominence involves normalization processes that are triggered by lexical accent in Japanese. The stimuli were naturally spoken in the sense that both the acoustic and structural properties of lexical accent were present. I tried to tease apart different perceptual normalization processes. Besides the

declination effect (Pierrehumbert 1979) and the BA effect (Gussenhoven et al. 1997), the experimental results suggested that two novel normalization processes are at work: accentual boost normalization and downstep normalization. Accentual boost normalization perceptually undoes the accentual boost of an accented word so that the perceived prominence of a word is smaller than the prominence that the word would have if it is computed solely based on the F0 excursion size. Downstep normalization perceptually undoes the effect of the post-accent F0 lowering in production. Furthermore, it was found that the magnitude of accentual boost normalization is significantly greater than that of downstep normalization.

The most important theoretical implications that the findings of Experiment 1 is that theories of prominence interpretation have to take the effects of take lexical, and hence language-specific, accent features on perceived prominence into account. Current theories on the perception of intonational prominence only consider phenomena at the intonation level and do not discuss interactions between lexical accent and intonational perceived prominence. Experiment 1 indicates that a whole new set of language-specific perceptual effects needs to be explored in tone languages and pitch accent languages.

CHAPTER 4

EXPERIMENT 2: F0 PLATEAUX AND PERCEIVED PROMINENCE

4.1   Introduction

4.1.1   F0 plateaux in unaccented words

Based on the results of Experiment 1, I proposed two perceptual normalization processes: accentual boost normalization and downstep normalization. This chapter focuses on accentual boost normalization. We test another possible account that equally explains the results of Experiment 1. The account is based on Knight's (2003) finding that an F0 peak evokes more perceived prominence when it has a long stretch of its peak or *a plateau* than when it does not. An unaccented word in Japanese with spreading of the H phrasal tone has an F0 contour shape that is close to a plateau, which indicates that if the peak height is identical, an unaccented word might have more perceived prominence than an accented word. In relation to the results of Experiment 1, Knight's account predicts that in the conditions where the accent types of P1 and P2 are different, namely *au* and *ua*, the height of the unaccented peak is always lower in F0 than that of the accented peak when P1 and P2 have the same equal prominence.[1]

Sugahara (2003) reported that in Japanese there are two distinct types of speakers with respect to the phonetic realization of the H phrasal tone in unaccented words. One type of

---

[1] Knight's account does not say anything about the sequences where P1 and P2 have the same accent types. That is, even if an unaccented evoked more perceived prominence than an accented word, its effect could not be observed if both P1 and P2 are both unaccented.

speaker shows a "linear interpolation" between the H phrasal tone and the following L% tone. The speakers who participated in Pierrehumbert & Beckman's (1988) study exhibited this pattern. However, Sugahara showed that there is another type of Japanese speaker who shows a different realization pattern for the H phrasal tone in unaccented words. This type of speaker "spreads" the H phrasal tone from the second mora to the final mora of the word, showing a stretch of the F0 peak that is sustained at more or less the same level. This is the pattern that traditional scholars of Japanese tonology assume (Haraguchi 1977, Poser 1984) in the analyses based on their native speakers' intuition. Thus, the F0 contour has a high F0 plateau for the most of the word. The F0 contours for "linear interpolation" and "spreading" are schematically shown in Figure 4.1.

a. Linear interpolation   b. Spreading
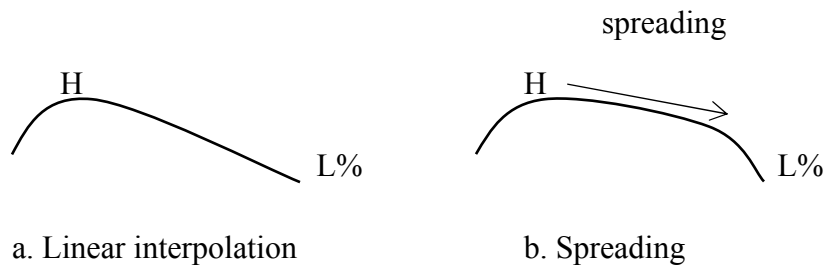
Figure 4.1 Two different realization patterns of an unaccented word.

This phonetic realizational difference may be a sociophonetic difference based on speakers' age (Sugahara 2003). Among the five native Japanese speakers in Sugahara's work, only one showed the linear interpolation pattern and she was in her late fifties at the time of recording, and the other speakers were in their late twenties or early thirties. If generation were the sole factor involved, then the difference between linear interpolation and spreading

61

might only serve as an acoustic cue that signals the generation that the speaker belongs to.

The question I want to ask is whether the F0 plateaux in unaccented words contribute to perceived prominence. Knight (2003) showed that in English an F0 contour with a plateau yields more perceived prominence than an F0 contour with its highest point lasting only momentarily, namely a peaked contour. She accounts for this effect in terms of a very general principle in auditory perception. That is, a peaked contour causes difficulty for the auditory system to "phase lock" to the frequency of the highest F0 point because of the extreme brevity of its duration, whereas a plateau contour does not have this problem because the highest F0 point lasts long enough for the auditory system to phase lock the frequency (see the following section more on this point).

If this low-level neurophysiological account is correct, the effect of F0 plateaux on perceived prominence should be seen in any languages including Japanese. It is predicted that an unaccented word has more perceived prominence than an accented word if they have the same F0 peak height. Experiment 2, which is presented in this chapter, tests this prediction. Before proceeding, we review Knight's (2003) experiments in more detail.

4.1.2   Knight (2003): F0 plateaux and perceived prominence

Knight (2003) tested the effect of sustained F0 peaks on perceived prominence in English. Based on several perceptual experiments, she showed that F0 contours with F0 plateaux are perceived to be more prominent than peaked contours without F0 plateaux. In her Experiment 3, Knight had her listeners hear the utterance *Anna came with Manny* in which

the F0 of *Manny* was varied with respect to its height and shape, as shown in Figure 4.2. The

F0 height and the overall duration were fixed for both of the words. The F0 contour shape for

Anna was kept constant.

Anna   Manny

Figure 4.2 F0 contours used in Knight's (2003) Experiment 3 (p.164).

As illustrated in *Manny* in the figure, Knight (2003) created a number of pairs of F0

contours in which each pair contained contours which had a sharp peak (represented by a

solid line) and a plateau of 100 ms with the same height (represented by a dotted line). Her

listeners randomly listened to each of the sentences and were asked to answer which of the

two words was higher in pitch or more prominent.[2] Knight found that *Manny* was

consistently judged higher or more prominent when the contour had an F0 plateau than when

it did not.

In another experiment (Experiment 2), Knight (2003) tested the effects of the duration

of F0 plateaux. She cut out *came with Manny* from *Anna came with Manny* and manipulated

---

[2] Knight (2003) in fact divided her listeners into two groups and gave different instructions to them. One group
   of listeners were asked to judge relative "height" in pitch, and the other were asked to judge relative
   "prominence", following Terken (1991). However, she did not find a significant difference between the two
   different instruction groups, and, therefore, I ignored these instructional conditions here.

the F0 height and the F0 contour shape of *Manny*, as shown in Figure 4.3. For each of the different F0 peak heights she prepared three different contour shapes: one with a peaked contour (solid line), one with a 50ms plateau (broken line), and one with a 100ms plateau (dotted line). She paired up different stimuli, as well as the identical stimuli for control and carried out a discrimination test. She asked her listeners (there were only 6) which of the two extracted phrases (*Manny*) is higher in pitch. Knight found that the F0 plateau stimuli were perceived to be higher than the peaked stimuli when they had the same peak height. That is, there were significant differences in perceived pitch between the peaked contours and the plateau contours of 50ms or 100ms. However, she found no effect of plateau duration: there was no difference in perceived pitch between the two types of plateau contours with different durations.
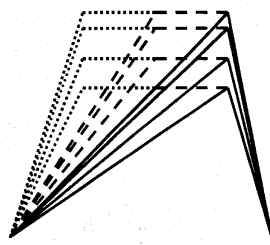
Figure 4.3 F0 contours used in Knight's (2003) Experiment 2 (p.153).

Knight (2003) argued that these effects can be explained in terms of general neurophysiological mechanisms at the level of auditory detection of pitch. To know how her explanation works, it is necessary to review the physiology of hearing.

There have been two different theories of pitch perception.[3] They are known as the "place" theory and the "temporal" theory. The place theory assumes that acoustic signals undergo some kind of spectral analysis in the inner ear: different frequencies excite different places along the basilar membrane and therefore neurons with different characteristic frequencies. Although it is an established fact that different places of the basilar membrane are excited by different frequencies of the input signal, it is generally considered that the place mechanism is less useful for complex tones, since with a complex tone, the produced pattern of excitation along the basilar membrane does not show a single maximum.

The temporal theory suggests that the pitch of acoustic signals is related to the pattern of the neural impulses evoked by the signals in the time domain. The basilar membrane vibrates in response to an acoustic signal, whose temporal patterning is also transmitted to the hair cells. The excitation of the neural spikes tends to be "phase locked" or synchronized to the signal waveform. As a result, the intervals between successive spikes approximate integer multiples of the period of the waveform. The auditory system is considered to be able to compute the pitch of the signal waveform from this temporal pattern seen in the excitation of the auditory nerves. However, the temporal theory also has a drawback: it does not work for signals with very high frequencies like those above 5 kHz: phase locking is found not to occur for signals of such high frequencies.

In auditory perception, both the place mechanism and the temporal mechanism are considered to be utilized. However, Knight (2003) focuses on this temporal mechanism to

---

[3] See Moore (2003) and Pickles (1988) for a review.

account for her experimental results that the plateau F0 contours evoked more perceived prominence than the peaked contours and that there was no effect of plateau duration.

Her account is based on the sluggishness of auditory detection of pitch. It is known that pitch discrimination becomes worse as the frequency of the stimulus changes more rapidly (e.g. Sek & Moore 1999). Thus, Knight considered that in a peaked contour, the F0 only stays on its peak value only momentarily, which would make it difficult for the auditory system to phase lock to the very highest frequency. In a plateau contour, on the other hand, because the highest value is sustained long enough, it would be much easier for the auditory system to detect the information of the F0 peak. Therefore, the perceived height of a peaked contour is lower than that of a plateau contour when the two contours have the same F0 peak height.

As for the lack of durational effect of F0 plateau, Knight argues that a plateau of 50ms may be long enough for phase locking to occur and for the precise frequency information to be detected. A 100ms plateau would be redundantly long for the auditory system to obtain the information of the maximum F0 value.

Since Knight's (2003) account is based on the neurophysiology of the auditory system, the effect of F0 plateau should be universal. Thus, it is predicted that in Japanese unaccented words with plateaux contours like the one in Figure 4.1b will evoke more perceived prominence than those with linear interpolation contours as the one in Figure 4.1a. Though the plateau pattern does not show a real "plateau" with a flat F0 pattern as in Knight's stimuli, it is still much more plateau-like than the linear interpolation contour.

Based on the discussion so far, it seems possible to quantify the relationship between F0 plateaux and perceived prominence in terms of a negative correlation between F0 slope and perceived prominence. The hypothesis is that as the slope of an F0 contour decreases the perceived prominence increases. I call this the *slope-prominence hypothesis*. This is shown in the Figure 4.4 below. Experiment 2 tests this hypothesis by manipulating the F0 slope of an unaccented word, just like shown figure.
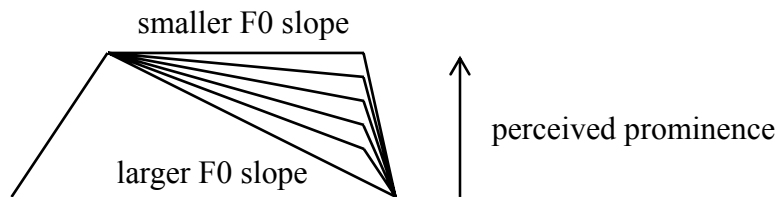
Figure 4.4 Illustration of the slope-prominence hypothesis.

4.2   Methods

4.2.1   Stimuli

There were two stimulus sentences in this experiment, which are shown in (1). Of the four possible combinations of the two accent types for a sentence with two major F0 peaks, the combinations with different accent types (*au* and *ua*) are chosen. The sentences were recorded several times by the author in a sound attenuated booth. The mean values of the tones that constituted the F0 contour were used as the basis of the stimulus sentences, that is, the mean values were used for tones whose F0 values were fixed. One of the experimental sentences (1a) had an unaccented noun followed by an accented one (*au* sequence), and the

other (1b) the other way round (*ua* sequence). The verbs were both intransitive and unaccented.

(1) a. [*ua*] Inamura-no   ani'yome-ga   inai.
              Inamura-Gen   sister-in-law-Nom   not fond
              'Inamura's sister-in-law is not found.'

     b. [*au*] Ina'mori-no   omiyage-ga   ureta.
              Inamori-Gen   souvenir-Nom   sell-past
              'Inamori's souvenir sold'

Like Experiment 1, the experimental sentences were naturally uttered in the sense that their phonological lexical accent patterns were consistent with their phonetic F0 contour shapes.

However, the F0 contour shape of the unaccented nouns (the first word in 1a and the second word in 1b) was acoustically manipulated and resynthesized, with the "pitch synchronous overlap add" (PSOLA) method using *Praat*. More specifically, what was manipulated was the F0 slope from the H phrasal tone that is associated to the second mora of the unaccented words to the following L% tone. This is illustrated in Figure 4.5.

(2) **Plateau**   (3) **Shoulder**

H

1/2

(1) **Linear interpolation**

L%

Turning point: Onset of the vowel
of the case particle /no/ or /ga/
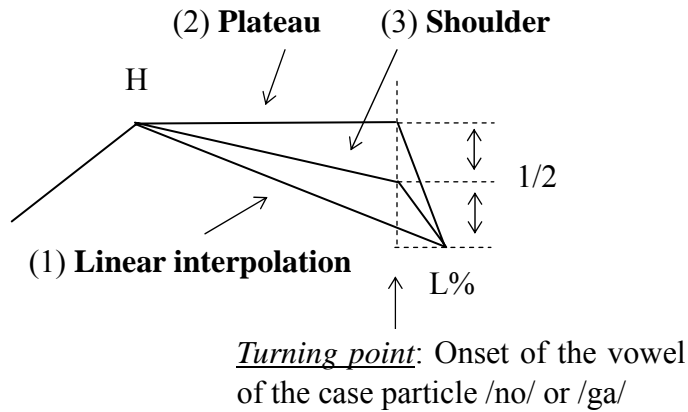
Figure 4.5 Manipulation of F0 contour shape of the unaccented words in the experimental sentences in (1).

The unaccented words had three different F0 contour shapes. The first contour shape was the "linear interpolation" between the H phrasal tone to the L% tone at the end. This is the F0 pattern that Pierrehumbert & Beckman (1988) observed in their data of unaccented words. The second F0 pattern contained a complete F0 "plateau". In this contour, the F0 height was kept constant from the H phrasal tone to the onset of the vowel /o/ or /a/ of the case marker /no/ (genitive) or /ga/ (nominative), the point I call the "turning point", and then it fell into the L% tone. The third F0 contour shape was an intermediate contour between the linear interpolation and the plateau contours. The F0 started from the same H phrasal tone but the F0 height was physically 1/2 lower than the plateau contour at the turning point. This contour is called the "shoulder" contour. In terms of F0 slope, the linear interpolation contour was the steepest, the plateau contour was the shallowest, and the shoulder contour was intermediate.

In addition to F0 contour shape, F0 peak height was also manipulated. This was done

for both of the accented and unaccented words. The overall F0 contour shape was kept the same. For the accented words, only the peak F0 value that corresponds to the H*+L tone was varied. For the unaccented words, the values of the H phrasal tone and the turning point were covaried so that the relation between those two points was kept constant. Regardless of accent type, the F0 peak height was varied in 4 steps with the step size being 1 semitone (st) (with the base frequency of 100Hz) for the first word, and in 6 steps with the step size being 1.4 st for the second word. The actual F0 values of the peak tones (H phrasal tone and the H* accented tone) and those of the turning point are shown in Table 4.1. The other fixed F0 values can be found in Figure 4.6, with the F0 contours actually used in the experiment. Overall, the number of stimuli was 144 (2 sentence types $\times$ 3 F0 contour shapes $\times$ 4 P1 heights $\times$ 6 P2 heights).

ua				P2						
P1	Turning point			111	120	130	141	153	166	180
H	LI	Shoulder	Plateau							
142	116	129	142							▨
150	118	134	150							▨
159	120	140	159	▨						
168	123	146	168	▨						

a

au	P2											
	112			122			132			143		
	Turning point			Turning point			Turning point			Turning point		
P1	LI	S	P	LI	S	P	LI	S	P	LI	S	P
H*	111	110	112	114	118	122	118	125	132	122	134	143
168												
178												
189	▨	▨	▨									
200	▨	▨	▨									

(LI = linear interpolation; S = shoulder; P = plateau)

b

Continued from above table for *au*

au	P2								
	155			168			182		
	Turning point			Turning point			Turning point		
P1	LI	S	P	LI	S	P	LI	S	P
H*	126	141	155	131	150	168	136	159	182
168							▨	▨	▨
178							▨	▨	▨
189									
200									

b

Table 4.1 F0 values of the H phrasal tone and the H* accent tone used in Experiment 2: (a) *ua* sequence, (b) *au* sequence. The cells that are shaded indicate that they were not used.

a. ua   b. au
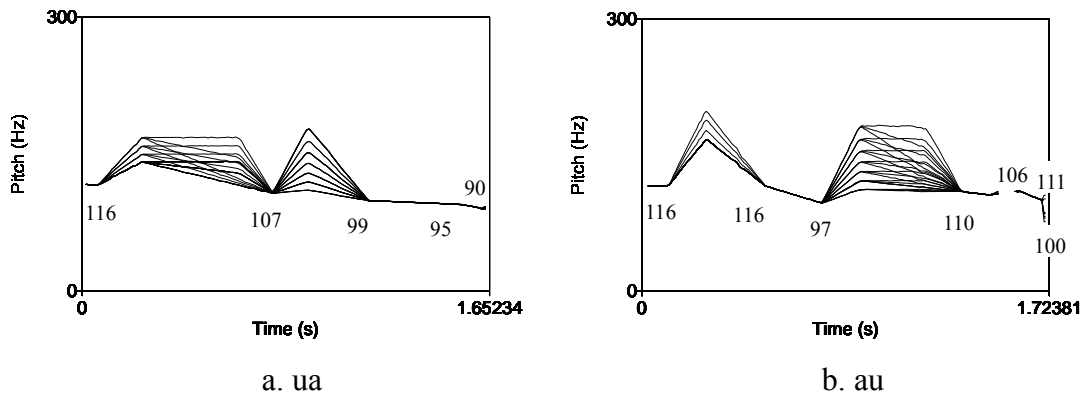
Figure 4.6 F0 contours used in Experiment 2.

4.2.2 Listeners

Twenty-nine native speakers of Japanese were recruited at Sophia University, Tokyo University of Agriculture and Technology, and Waseda University (mean age 22.8). All institutions were in Tokyo, Japan. Most of them were undergraduate students. None had any kind of hearing disorders. Of the 29 listeners, 23 were originally from Tokyo and its neighboring areas (Chiba, Kanagawa, and Saitama). The remaining 6 were from various areas (Fukui, Fukushima, Hyogo, Ibaraki, and Nigata). All listeners were paid for participation.

4.2.3 Procedure and analysis

The basic experimental procedure of Experiment 2 was identical to that of Experiment 1. Listeners were tested in groups of two to five people. They each sat at a computer terminal facing a display. The experimental stimuli described in §4.2.1 were randomly presented to

72

the listeners, using the Multiple Forced Choice (MFC) program equipped with *Praat*. The listeners' task was identical to the one adopted in Experiment 1: to judge which of the two words (first or second) was given more emphasis by the speaker as quickly as possible. The stimulus presentation involved repetitions, but the number of repetitions differed depending on the stimulus. For each of the four P1 heights, the highest and the lowest P2 values were presented with only 2 repetitions, the second highest and the second lowest P2 values with four repetitions, and the remaining two intermediate P2 heights with 6 repetitions. This manipulation was made because it was expected that the judgment would be relatively easy when the F0 difference between P1 and P2 is very large. Also, this made the size of the experiment reasonable. Overall, there were 576 tokens of stimuli were presented to the listeners.

The stimuli were divided into 7 blocks of trials each of which contained 72 stimuli. The listeners were instructed to take a short break after every block of trials if they wanted. They had a practice session to be familiarized with the stimuli where they heard two repetitions of 16 of the experimental stimuli. The session lasted about 80 minutes including the instructions, the practice, the blocks of trials, and the breaks.

As for the analysis, the same procedure as the one adopted in Experiment 1 was used. That is, using probit analysis the response rates for P2 (P2 sounding more emphatic than P1) obtained for each listener were submitted to calculate the 50% crossover point of the response rates, the point where P1 and P2 would sound the same in perceived prominence.

4.2.4   Predictions

Recall that the slope-prominence hypothesis, which is tested here, says that there is a negative correlation between the slope (downtrend) of an F0 contour and its perceived prominence: smaller F0 slopes evoke more perceived prominence than larger F0 slopes. Based on this hypothesis, it is predicted that a plateau contour will have greater perceived prominence than a shoulder contour, which in turn will have greater perceived prominence than a linear interpolation contour.

Let us consider these predictions in more detail using schematic graphs. The two graphs in Figure 4.7 represent the P2 values as a function of P1 that has the same perceived prominence as P2. Let us first consider Figure 4.7a the *ua* sequence. In this sequence the manipulation of the F0 contour was made in the first word (note that contour shape is manipulated for only unaccented words) and thus it is predicted that as unaccented P1 changes from the linear interpolation through the shoulder to the plateau contours with a constant peak height, the F0 value of accented P2 will need to be increased in order for it to have the same perceived prominence as P1. This should be so because it is hypothesized that for a given peak height the plateau contour has more perceived prominence than the shoulder contour, which in turn has more perceived prominence than the linear interpolation contour. Therefore, in the ua sequence the prediction is plateau > shoulder > linear interpolation, where ">" means "more prominent than".

The order is reversed in the *au* sequence. In Figure 4.7b, the plateau contour is predicted to show the smallest P2 values, the linear interpolation contour the greatest values, and the shoulder contour the intermediate values. This is because the unaccented word, whose F0 contour shape is manipulated, now appears in the second word rather than in the

74

first word. Thus, if we assume that perceived prominence increases as F0 slope becomes smaller, it is predicted that for a given P1 height, the P2 F0 value that sounds equal to P1 in prominence will have to be decreased from the linear interpolation through the shoulder to the plateau contours.
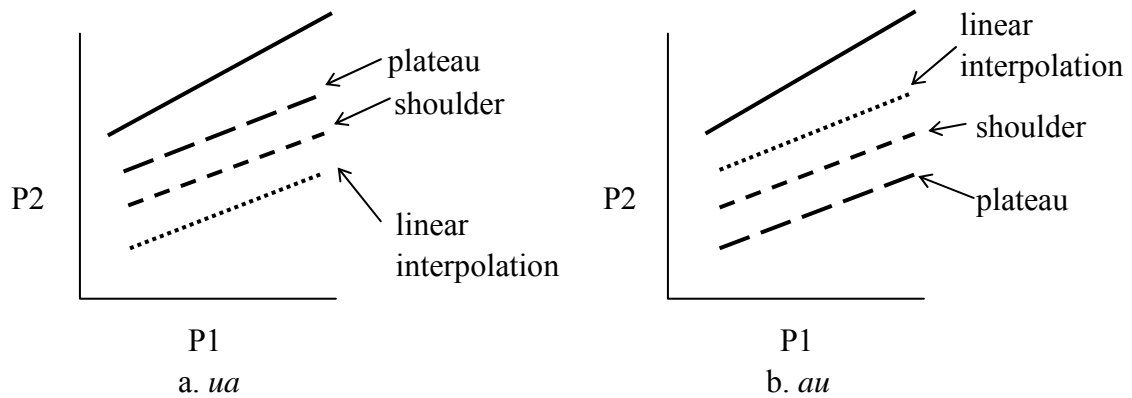
Figure 4.7 Predictions on the effects of F0 slope of the unaccented words. P2 values are represented against P1 values when P1 and P2 have the same perceived prominence. (a) *ua* sequence, (b) *au* sequence.

4.3 Results

In Figure 4.8 the P2 values obtained by the probit analysis are plotted. These P2 values represent those that have the same perceived prominence as P1, which are represented on the *x*-axis.
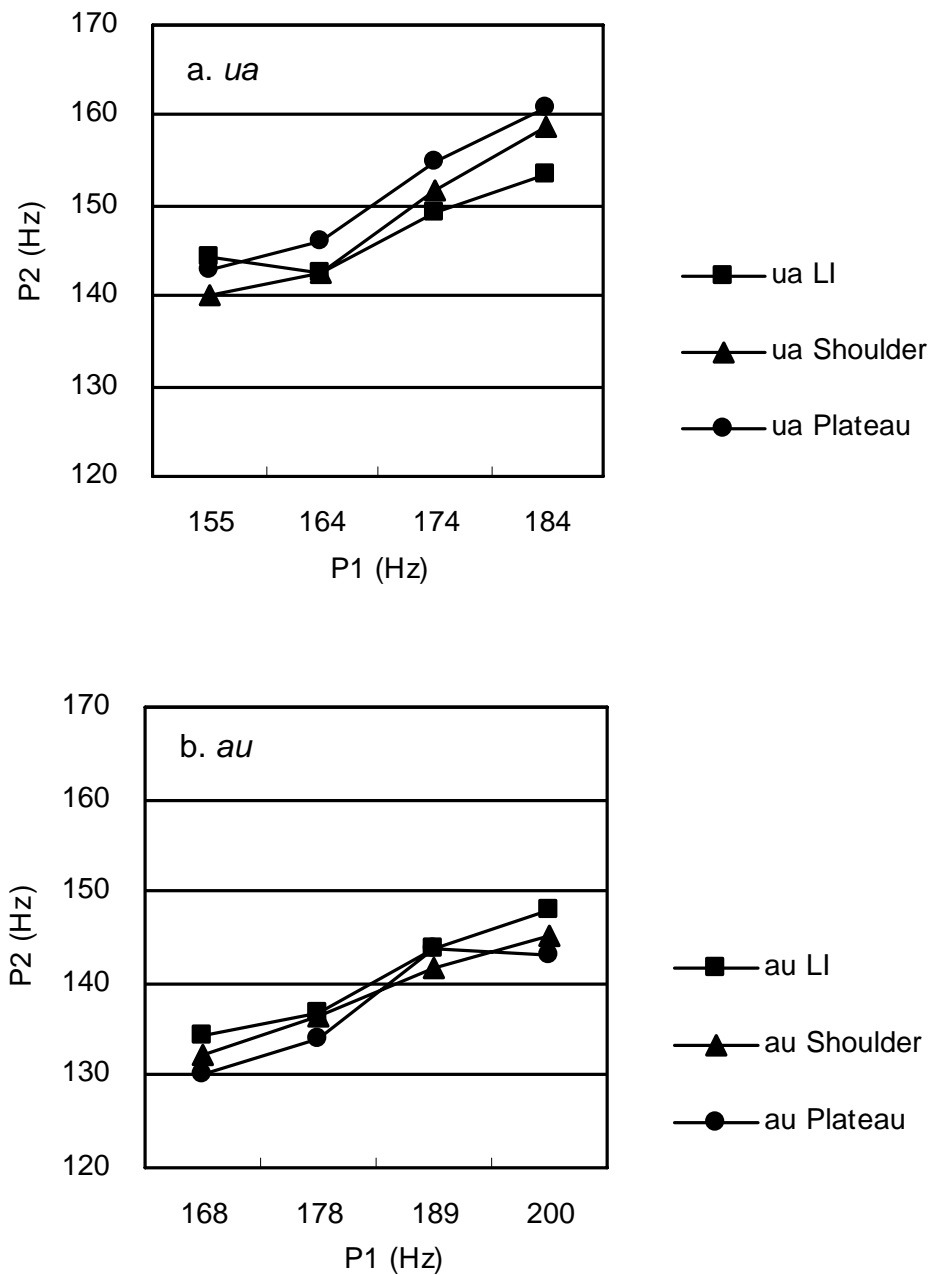
Figure 4.8 Mean values of P2 giving the same prominence as P1 as a function of P1 height among the linear interpolation (LI), the shoulder, and the plateau conditions: (a) *ua* sequence, (b) *au* sequence.

The overall patterns observed among the three different contour shape conditions confirm the slope hypothesis. In Figure 4.8a, the plateau condition (circles) yields the greatest P2 values of the three contour shape conditions, the linear interpolation condition (squares) the smallest P2 values, and the shoulder condition (triangles) the intermediate P2 values, although the differences are small, on the order of 5Hz. These patterns suggest that for a given unaccented P1 value, the following accented P2, which has the equivalent perceived prominence, needs to be higher in F0 when P1 has an increasingly smaller slope. This is a finding that supports the slope-prominence hypothesis.

In Figure 4.8b, which represents the results of the *au* sequence, the basic patterns are the same as those observed in Figure 4.8a, though again the differences are small. The order of the effect is reversed due to the location of the unaccented word, as explained in the previous section. P2 shows the greatest values in the linear interpolation condition (squares), intermediate values in the shoulder condition (triangles), and the smallest values in the plateau condition (circles). These results can again be explained by the slope-prominence hypothesis. When an accented P1 and an unaccented P2 have the same perceived prominence, P2 needs to be lower in the plateau condition than the other two conditions because its F0 slope is the smaller (it is zero) than the slopes of the contours in the other conditions. Thus it evokes the most perceived prominence of the three F0 contour shape conditions. P2 is higher in the shoulder condition than in the linear interpolation condition because the F0 slope is smaller in the former condition than in the latter condition.

However, the general pattern is less orderly in the *au* sequence than in the *ua* sequence,

and moreover the differences among the F0 contour shape conditions are smaller in the *au* sequence than in the *ua* sequence.

To statistically assess our results, the obtained data were submitted to a repeated-measures ANOVA analysis. For each of the two different sequences, an ANOVA was done with F0 contour shape and P1 as the independent variables. For both of the sequences there was a significant main effect for both F0 contour shape and P1, and the interaction between the two factors, as shown in Table 4.2. Post-hoc pair-wise comparisons showed that in the *ua* sequence the P2 value were significantly higher in the plateau condition than the shoulder condition ($F(1,28)=24.255$, $p<0.001$) and the linear interpolation condition ($F(1,28)=20.649$, $p<0.001$), but the shoulder condition did not differ significantly from the linear interpolation condition ($F(1,28)=2.456$, $p=0.128$). In the *au* sequence, on the other hand, the linear interpolation condition showed P2 values that were significantly higher than the shoulder condition ($F(1,28)=8.689$, $p=0.006$) and the plateau condition ($F(1,28)=25.033$, $p<0.001$), but there was no difference between the shoulder and the plateau conditions ($F(1,28)=4.429$, $p<0.044$, note for the significance-level correction for multiple comparisons ($\alpha = 0.05/6 \approx 0.0083$)).

ua sequence	
F0 contour shape	$F(2,56)=17.524, \ p<0.001$
P1	$F(3,84)=104.915, \ p<0.001$
F0 contour shape * P1	$F(6,168)=6.710, \ p<0.001$
au sequence	
F0 contour shape	$F(1,56)=13.041, \ p<0.001$
P1	$F(3,84)=67.295, \ p<0.001$
F0 contour shape * P1	$F(6,168)=3.159, \ p=0.006$

Table 4.2 Results of the ANOVAs.

A closer examination of the data suggest that the significant interactions between F0 contour shape and P1 were due to the relatively high P2 mean value at P1=155Hz in the shoulder condition in the *ua* sequence and the relatively high P2 mean value at P1=189Hz in the plateau condition in the *ua* sequence. To confirm these observations, separate repeated-measures ANOVAs were conducted without the P1=155Hz condition in the *ua* sequence and the P1=189Hz condition in the *au* sequence. The results showed that the significant interactions were gone from both of the sequences (*ua*: $F(4,112)=2.319, p<0.061$, *au*: $F(4,112)=0.913, p=0.459$).

In sum, the general patterns observed in Experiment 2 are consistent with the slope-prominence hypothesis, which claims that the F0 contour slope of a word is negatively correlated with perceived prominence of the word: the perceived prominence increase as the F0 slope gets smaller.

4.4   Discussion

4.4.1   The slope-prominence account cannot supplant the accentual boost account

The important finding of Experiment 2 is that the F0 downtrend slope of an unaccented word in Japanese affects perceived prominence: the smaller the F0 slope is the greater the perceived prominence the word evokes. This F0 slope effect on perceived prominence is consistent with the findings of Knight's (2003) for English: English native speakers perceive F0 contours with plateaux to be higher (and therefore more prominent) than peaked F0 contours, if everything else is equal. Since F0 contours with plateaux show smaller F0 slopes than those without them, the F0 slope effect found in Experiment 2 is compatible with Knight's result. The spreading of the H phrasal tone in an unaccented word contributes to the perceived prominence of the word by evoking more prominence than an unaccented word with a linear interpolation contour. Knight's universal neurophysiological mechanism for the perception of prominence seems to be at play in Japanese as well as in English.

The question that should be asked is whether it is possible to account for the accentual boost effect found in Experiment 1 with the neurophysiological mechanism without recourse to accentual boost normalization? I argue that it is not.

Accentual boost normalization assumes a perceptual compensation for the accentual boost of an accented word such that its extra F0 boost is factored out when the perceived prominence of the word is assessed. As a result, when an *au* sequence is compared with an *aa* sequence, the P1-P2 difference in F0 is greater in *au* than in *aa*. However, the same effect can in principle be explained if we assume the F0 slope effect found in Experiment 2: since an

unaccented word evokes more perceived prominence than an accented word due to its smaller F0 slope, the F0 peak height is lower for the unaccented word than for the accented word when they have the same perceived prominence. Therefore, between *au* and *aa*, P2 needs to be higher in F0 *aa* than *au* in order for it to be perceived as having the same prominence as P1.

However, the F0 slope effect observed in Experiment 2 is quite small. As we saw in Figure 4.8, the size of the effect is about 5 Hz or less between the linear interpolation contours and the shoulder contours and between the shoulder contours and the plateau contours. The maximum difference is about 10Hz at best between the linear interpolation and the plateau contours. If we look at the results of Experiment 1 represented in Figure 3.2a, reproduced below as Figure 4.9, the difference in the effect size is obvious: the differences in mean P2 values between the *aa* and the *au* sequences in Experiment 1 are much greater than those seen in the F0 contours with different shapes in Experiment 2. The differences in Figure 4.9 are 20Hz or more. Therefore, it is impossible to explain these large differences in terms of the neurophysiological mechanism that Knight (2003) proposed based on the declining slope of an F0 contour.
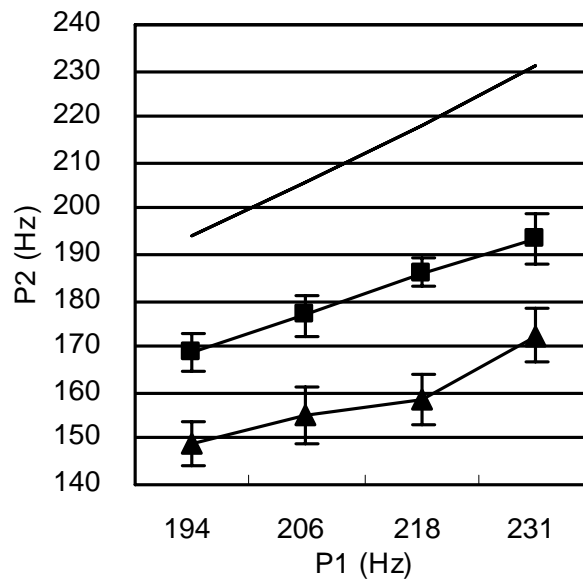
Figure 4.9 (= Figure 3.2) Mean values of P2 giving the same prominence as P1 as a function of P1 height for *aa* and *au*. The straight line in each graph represents the *y=x* function. The error bars are 95 % confidence intervals.

4.4.2 Comparison with Knight (2003)

The conclusion reached in the previous section was based on the slope-prominence hypothesis, which was developed from the work by Knight (2003) in which she carried out experiments that are similar to the one presented in this chapter. In this section, I compare Knight's experiments with the current one to see to what extent they are directly comparable. I point out similarities and differences seen between the two studies. I clarify several points points that we need to be careful about when considering the relation between F0 slope and perceived prominence in future work.

Let us start with similarities. Figure 4.10 represents part of the data from Knight's Experiment 3 (see §4.1.2 for the methodology). This graph shows the results of the

prominence judgment about the stimulus sentence *Anna came with Manny*. The *y*-axis represents the mean response rates for *Manny* being judged as more prominent than *Anna*. The *x*-axis shows different F0 peak values of *Manny*. Because the way Knight's data are represented in this graph is different from the way ours in Figure 4.8, we need to estimate the P1-P2 difference when *Anna* and *Manny* have he same perceived prominence. To do this we first need to find the F0 peak height of *Manny* that corresponds to the 50% mean response rate, the point where *Anna* and *Manny* have the same perceived prominence for each of the peaked and the plateau contours. Those values – approximately 193 Hz and 205 Hz for the plateau contour and the peaked contour, respectively – are indicated by the broken lines in Figure 4.10. The F0 difference between the two contour types, 12 Hz (205 Hz – 193 Hz, indicated by the thick arrow), is the mean difference in the peak value of *Manny* when *Anna* and *Manny* have he same perceived prominence in each contour type.

This difference found in Knight's experiment and the differences between the plateau and the linear interpolation conditions in Experiment 2 are in the same ballpark. As pointed out above, the size of the F0 slope effect found in Experiment 2 is approximately between 5 to 10 Hz between the plateau and the linear interpolation conditions. It is fair to say that the effect size of Knight's experiment is not distinctively larger than that of Experiment 2. At least it seems much smaller than the size of the differences that were found in Experiment 1 (represented in Figure 4.9), based on which I proposed accentual boost normalization.
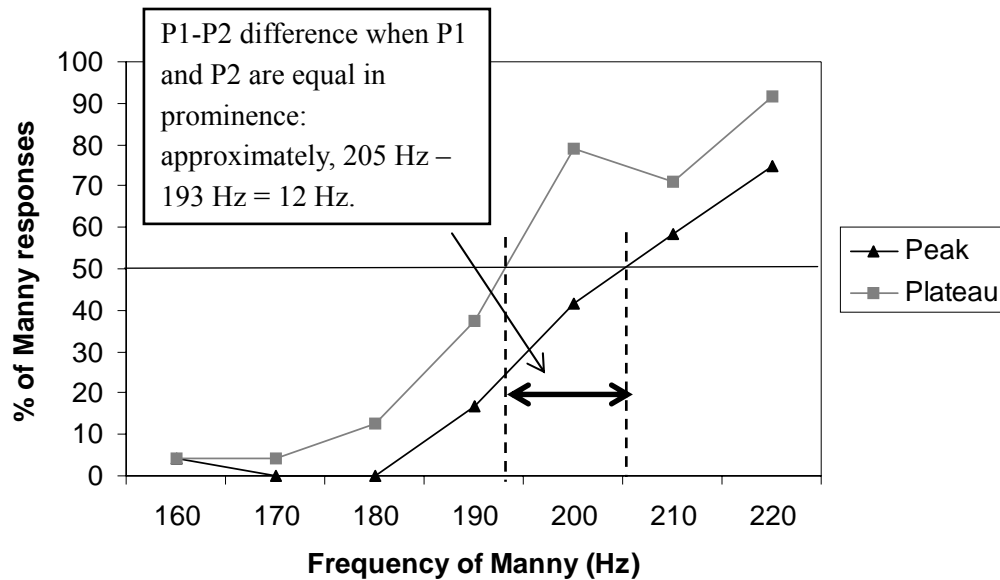
Figure 4.10 Effects of F0 plateaux found in one of Knight's Experiment 3. The data are for prominence judgment task. From Knight's (2003), Figure 5.12, p.166.

Let us move on to the differences. One difference between Knight's (2003) study and the current study lies in the difference between rising and falling slopes. Knight manipulated the slope of F0 rise instead of that of F0 fall. In her F0 contours, the slope of the onset rise varies depending on whether the contour has a single peak or a plateau, with the F0 declining slope being fixed. In Experiment 2, on the other hand, the slope of the offset of the F0 contour was manipulated with the slope of the onset rise in tact. Based on this stimulus difference between the two experiments, we might not know whether the English effect comes from the sharper F0 rise at the beginning of the plateau contour or from its smaller F0 declining slope.

Though the difference between rising slope and falling slope is acoustically obvious, we should notice the fact that Knight (2003) found no difference in perceived height between

the shorter plateau and the longer plateau contours observed in her Experiment 2 (Figure 4.3). In Figure 4.3, the contours with different plateau durations have different onset slopes. Knight obtained the categorical effect of difference in perceived prominence only between the contours with plateaux and those without plateaux, and no difference between the conditions that differed with respect to plateau duration. This suggests that F0 rising slope does not matter to perceived prominence. Rather the right criterion seems to be whether the contour has a plateau or not.

However, still, the lack of plateau duration effect does not mean that the results of Knight's experiments and those of the current experiment are directly comparable. Even if the true characterization was whether the contour has a plateau or not, we still could not compare Knight's experiments and the current experiment, because there is still a difference between them. The difference is in the position of the F0 peak occurring in a word differs from each other, as shown in Figure 4.11.

On the one hand, Knight (2003) found a significant difference in perceived prominence between a plateau contour and a peaked contour with its peak located toward the *end* of the word (between Figure 4.11a and Figure 4.11b). On the other hand, there was a significant difference between a plateau contour and a peaked contour with its peak located toward the *beginning* of the word (Figure 4.11c and Figure 4.11d). It is not entirely clear how this F0 peak alignment difference affects perceived prominence.

Knight's (2003) experiments
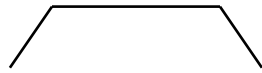
a. Plateau contour   b. Peaked contour

Experiment 2
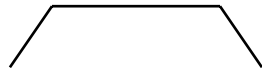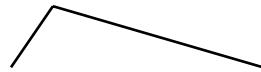
c. Plateau contour   d. Peaked (linear interpolation) contour

Figure 4.11 Differences between the F0 contours used in Knight's (2003) experiments (a and b) and those used in the current experiment (c and d).

In fact, there is a piece of evidence that suggests that the perceived prominence of an F0 contour differs depending on whether the F0 contour has an early or delayed peak. Ladd & Morton (1997) showed that an F0 contour with a delayed peak alignment evokes more perceived prominence than an F0 contour with no delayed peak alignment. Also, Gussenhoven (2004) makes the same suggestion, based on the physical requirement that a higher F0 peak takes longer to reach than a lower one with rate of F0 rise is constant (pp.90-91). If listeners have knowledge of this relationship between F0 peak height and peak alignment, it is conceivable that an F0 contour with a delayed peak is perceived to be more prominent than a contour without one. Given this relation between peak delay and perceived prominence, it is predicted that Knight's (2003) peaked contour (Figure 4.11b) is more perceptually prominent than the peaked contour of the current experiment (Figure 4.11d). I leave the testing of the prediction for future research.

The comparison presented above between Knight's experiments and Experiment 2

have revealed that the two studies have more differences than similarities, which indicates a direct comparison of the experimental results between is not possible. What is certain, however, is that as long as the slope-prominence hypothesis is concerned, it cannot adequately account for the results of Experiment 2.

4.5   Chapter summary

Prompted by Knight's (2003) experimental results, this chapter tested the slope-prominence hypothesis, which claims that as the declining slope of an F0 contour decreases the perceived prominence increases. The results of Experiment 2 supported the hypothesis. The experiment revealed that unaccented words with F0 plateaux show more perceived prominence than those with F0 shoulders, which in turn yielded more perceived prominence than unaccented words with linear interpolation. These results are generally in accordance with the results that Knight (2003) obtained, which is that F0 plateaux evokes more perceived prominence in English. However, the size of the F0 slope effect found in Experiment 2 was quite small, based on which I argue that this effect cannot be an alternative account for accentual boost normalization.

CHAPTER 5

EXPERIMENT 3: STRUCTURAL PROPERTY OF LEXICAL ACCENT AND

PERCEIVED PROMINENCE

5.1   Two sources of accent type identification

Experiments 1 and 2 used experimental sentences containing words that were lexically

accented or unaccented and that, at the same time, acoustically exhibited the properties that

are characteristic of each accent type. In other words, the listeners of the experiments could

use two different kinds of information to identify whether the experimental words were

accented or unaccented.

The first type of information is the lexical accent status specified for each word in the

Japanese lexicon (MaCawley 1969, Haraguchi 1977, Vance 1987). I call this information on

lexical accent type the *underlying* property of lexical accent.

The second type of information that the listeners could exploit for accent type

identification in Experiments 1 and 2 is the specific F0 contour pattern that each of the accent

types shows, to which I refer as the *acoustic* property of lexical accent. In Japanese an

accented word is characterized by a set of F0 properties that are distinct from those of an

unaccented word: an accent word shows an F0 higher peak and a sharper and bigger F0 fall

than an unaccented word (Pierrehumbert & Beckman 1988, Kubozono 1993, Sugahara 2003).

These acoustic F0 differences are robust in perception too. There is ample evidence that in

perception the acoustic property of lexical accent is used to distinguish one accent type from

the other (Sugito 1982, Hasegawa & Hata 1992, 1995, Kitahara 2001). In particular, through

a series of perceptual experiments on accent identification, Kitahara (2001) argues that the

most reliable F0 property for accent-type identification is the degree of F0 fall: a word with a large F0 fall is likely to be perceived to be accented and one a small F0 fall is likely to be perceived to be unaccented.

The issue we have to consider then is how the underlying and surface properties of lexical accent are related to accentual boost normalization and downstep normalization. Identification of lexical accent type is crucial to these normalization processes because only accented words trigger them. The specific question we should ask then is which of the two properties does so. In this chapter, we test whether the structural property of lexical accent plays a role in the perceptual assessment of prominence.

5.2 Aims of Experiment 3

Experiment 3 reported below consists of two sub-experiments. In the first sub-experiment, called Experiment 3a, an accent-type identification experiment was carried out. In this sub-experiment, listeners heard short sentences with three words (N1-N2-V) and two F0 peaks corresponding to N1 and N2. N1 and N2 were varied in their accent types. N1 was acoustically manipulated such that continua from accented to unaccented were created. Listeners judged the accent type of the first word by accent-type matching: they matched the accent type of the word that they heard with that of one of the words on the computer screen. In Experiment 3b, listeners performed relative prominence judgments with the same stimulus sentences used in Experiment 3a.

Experiment 3a is a necessary experiment for three reasons. The first reason is practical. It provides a basis for Experiment 3b. In Experiment 3b, it is essential to know in advance what F0 contours are ambiguous between the accented and unaccented patterns in N1, since

our interest lies in how prominence assessment is affected when the accent type of a word is not acoustically signaled; that is, in the circumstance in which the acoustic property of lexical accent is unavailable and listeners can only use the structural property of lexical accent.

The second reason is to examine whether the so called "lexicality effect" or the "Ganong effect" (Ganong 1980) is seen at the lexical accent level in Japanese. The lexicality effect refers to the influence of existing lexical items on phoneme processing such that when hearing an ambiguous segment in a carrier string of segments, the percept is biased toward the segment that constitutes a real word with the carrier string. For example, a segment which is acoustically ambiguous between [k] and [g] is perceived as [k] before –*iss* (*kiss*) but as [g] before –*ift* (*gift*). This means that the very same stimulus invokes different percepts depending on the environment under which it appears. The listener's percept is biased by his knowledge about what string of segments constitutes a word and what string of segments does not. Experiment 3a tests this effect in terms of Japanese lexical accent.

The lexicality effect has been investigated overwhelmingly on phoneme processing (see Pitt & Samuel 1993 for a comprehensive literature review on phonemic processing). Very little work has been carried out on the processing of prosodic features. The only study that came to my attention is Connine, Clifton & Cutler (1987). They showed that word stress information is relevant to the lexicality effect in English. Their experiments used two pairs of words in which the VOT (voice onset time) of the initial stop consonant ([t] or [d]) is varied into continua such that only one member of the pair forms a real word. One of the pairs was *diGRESS* and *tiGRESS*, where only *diGRESS* is a real word, and the other pair was *DIgress* and *TIgress*, where only *TIgress* is a real word (stressed syllables are denoted by capitals). Connine et al. observed that when the word-initial stop had an ambiguous VOT value, it was

90

perceived to be [d] in the context of *–iGRESS* while it was perceived to be [t] in the context of *–Igress*. These results showed that the lexicality effect is seen at the prosodic level suggesting that word stress information in English plays a role in lexical processing.

No experimental work of the kind that Connine et al. (1987) did has been done in other languages. This is surprising, because many languages are known to use prosodic features to distinguish lexical items. Tone languages and pitch-accent languages are good examples. Experiment 3a is an attempt to see whether the lexicality effect is observed in Japanese, a pitch-accent language.

The third reason that Experiment 3a is necessary is that it tests a hypothesis formulated on the basis of the claim entertained by Kubozono and his colleagues (Kubozono 1996, Kubozono & Fukui 2004, Kubozono & Ogawa 2004). They argue that the accent types of Japanese words are not entirely determined by the lexical specification in the lexicon but are largely determined by independent phonological factors such as syllable length and syllable structure. More specifically, through examining the accent pattern of loanwords and alphabetic acronyms, they showed that the accent type tends to be unaccented when a word is (1) four-moras long, (2) has a word-final sequence of light syllables, and (3) does not contain epenthetic vowels in word-final position. For example, *o.ru.gan.*[1] 'choir organ' and *FM (e.fu.e.mu.)* 'Frequency Modulation', in which both are four-moras long and end with successive light syllables, are unaccented words, but *e'n.jin.* 'engine' and *su.ko'c.chi.* 'scotch' are accented words since both contain a heavy syllable within the final two syllables (both of the two syllables in *e'n.jin.* and the penult in *su.ko'c.chi.*). Kubozono & Ogawa (2004) contend that even nonwords that meet the above three conditions such as /batabina/ and /berugiwa/ follow this pattern and are pronounced as unaccented by most of native

---

[1] Periods indicate syllable boundaries.

speakers of Japanese, presumably because those nonwords can be considered as extremely unfamiliar loanwords.

Experiment 3a tests this "unaccented law" hypothesis. To this end, this sub-experiment employs a nonword for N1 in one of the conditions. This nonword is created by minimally changing the segmental strings of the real words in the other conditions, which made the nonword conditions and the real word conditions comparable to one another.

Experiment 3b is an extension of Experiment 1. The basic design is similar to Experiment 1 in that listeners hear sentences consisting of three words with different accent types and were asked to judge relative prominence between the first and the second words. Experiment 3b uses a subset of the sentences used in Experiment 3a. What is novel in Experiment 3b was that N1 involved continua from acoustically accented to unaccented words so that listeners heard N1 whose accent type was ambiguous. The interest of this sub-experiment was to see whether prominence judgments are affected by the structural property of lexical accent without the acoustic property in such ambiguous cases.

5.3 Experiment 3a: Accent type identification

5.3.1 Methods

5.3.1.1 Stimuli

The four sentences used in Experiment 1, repeated in (1), plus two new sentences in which the first nouns were nonwords were rerecorded by the author. The new sentences are shown in (2). All sentences had three words: noun 1 (N1), noun 2 (N2) and intransitive verb (V). In the sentences in (1), the accent types of N1 and N2 were lexically varied such that we had all of the four combinations of accented/unaccented N1 and N2. The nonword used for N1 in the

nonword sentences was *inamire*, which was chosen to make its phonological form

comparable to the real words used in N1: *Ina'mori* and *Inamura*.

(1) a. *aa*
      Ina'mori-no   an'iyome-ga   inai.
      Inamori-Gen   sister-in-law-Nom   not found
      Inamori's sister-in-law is not found.

    b. *au*
      Ina'mori-no   omiyage-ga   kieta.
      Inamori-Gen   souvenir-Nom   disappeared
      Inamori's souvenir disappeared.

    c. *ua*
      Inamura-no   an'iyome-ga   inai.
      Inamura-Gen   sister-in-law-Nom   not found
      Inamura's sister-in-law is not found.

    d. *uu*
      Inamura-no   omiyage-ga   kieta.
      Inamura-Gen   souvenir-Nom   disappeared
      Inamura's souvenir disappeared.

(2) a. *na* ("*n*" indicating "nonword")
      Inamire-no   ani'yome-ga   kieta.
      nonword -Gen   sister-in-law-Nom   disappeared
      Inamire's sister-in-law is not found.

    b. *nu*
      Inamire-no   omiyage-ga   kieta.
      nonword -Gen   souvenir-Nom   disappeared
      Inamire's souvenir disappeared.

The recording was carried out in a sound attenuated booth. The sentences in (1) and (2)

were read seven times. The tonal values of the stimulus sentences were determined based on

the average values. After recording, they were stylized with *Praat* such that they had

schematized F0 contours expressed only by the tones constituting the contours. The

algorithm used was the PSOLA method.

After stylization, the F0 properties of N1 were manipulated. For each N1 of the six experimental sentences, a 5-step continuum was created such that N1 showed canonical F0 properties for a 4-mora accented word at one endpoint and canonical properties for a 4-mora unaccented word at the other endpoint. The mean F0 values obtained from the seven naturally produced tokens were used for the canonical tonal values. The intermediate stimuli were prepared by changing the F0 height and alignment of N1's peak and the following two valleys (as denoted by V1 and V2 in Figure 5.1). The peak values (P1) and the values of V1 and V2 are shown in Table 5.1. It turned out that peak height and its alignment was systematically related between the accented and the unaccented contours in such a way that the peak height was higher and aligned later in the accented contour than in the unaccented one. Therefore, these two properties were covaried. The step size was 3.4 Hz for F0 height and 3 ms for alignment. A systematic but reversed relationship between F0 height and alignment was seen in V1. The V1 height was lower and aligned earlier in the accented contour than in the unaccented one. The two properties were also covaried with the step size of 5.8 Hz for F0 height and 10 ms for alignment. The actual F0 contours used in the experiment are shown in Figure 5.1, together with the fixed values that were used in Experiment 3a.

The F0 properties of N2 were fixed in Experiment 3a. For the accented N2, the peak height was 148 Hz. For the unaccented N2, there were two F0 peaks that correspond to the phrasal H tone and the "shoulder" as a consequence of the phrasal H spreading. Their values were 148 Hz and 137 Hz. Actually, N2 was varied with respect to its F0 peak height in 5 steps but only the middle values were used in Experiment 3a (all of the 6 F0 heights were used in Experiment 3b). The actual F0 contours of the stimulus sentences are shown in

Figure 5.1. Note that both phonological and phonetic information is available to listeners for N2.

Overall, there were 6 different stimuli of N1 for each of the 6 experimental sentences given in (1) and (2), yielding 36 stimuli.

N1	Step						
	unaccented 1	2	3	4	5	accented 6	
P1	179	182	186	189	193	196	
V1	122	127.8	133.6	139.4	145.2	151	
V2	106	108.6	112.2	113.8	116.4	119	(Hz)

Table 5.1 Height of N1 peak and the following valleys (V1 and V2).
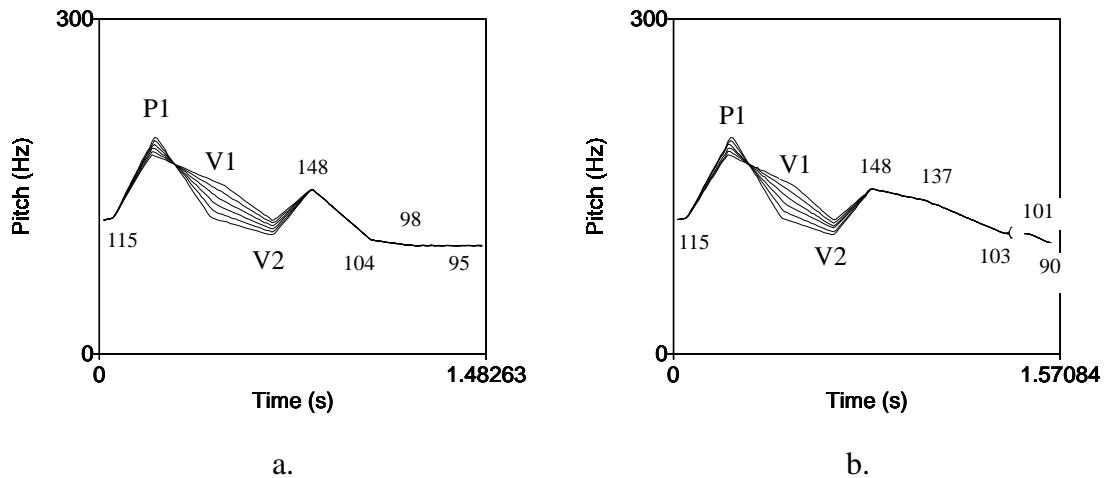
Figure 5.1 F0 contours of the stimuli used in the experiment: N2 shows F0 properties for an accented word in (a) and for an unaccented word in (b).

5.3.1.2 Listeners

The listeners of Experiment 3a was 24 undergraduate and graduate students recruited at Sophia University in Tokyo. All were native speakers of Japanese and were from the Tokyo dialect areas such as Chiba, Saitama, Tokyo, and Yokohama, except 3 of them. The areas which those 3 listeners were originally from were Aichi (about 350km west of Tokyo),

Hokkaido, and Kagawa (western Japan). All listeners were paid for their participation.

Unfortunately, one listener's data were lost due to a mistreatment in the data analysis. The data from the remaining 23 listeners were used for analysis.

5.3.1.3 Procedure

Listeners were tested in groups of two or three people at a sound-attenuated room. They each sat at a computer terminal facing a display. To have them identify accent patterns of words and not the words themselves, the following procedure was adopted. On the screen, two "matching words" whose accent types differ from each other were displayed: *kuda'mono* "fruit" and *kuchibiru* "lips". The former word is lexically accented and the latter unaccented. Listeners heard the stimulus sentences and were asked to focus on the first word in each sentence, the portion in which F0 was manipulated into a continuum. Their task was to match up the accent pattern of the first word that they heard with that of either of the two matching words. They indicated which of the two words' accent patterns matches the accent pattern of the word they heard by clicking on one of the matching words. Also, they were instructed to do the task as quickly as they could. The Multiple Forced Choice (MFC) of *Praat* was used to present the stimuli.

To be familiarized with the stimuli, the listeners first had a practice session where they heard the two endpoint stimuli for each of the 6 continua (*aa, au, ua, uu, na,* and *nu*) three times. Thus, they heard 36 stimuli in the practice session (2 endpoints × 6 continua × 3 repetitions).

The test session followed the practice session. Since it was considered that the endpoint stimuli would be unambiguous and it would be easier to identity their accent types

than the non-endpoint ones, the listeners heard fewer repetitions for those stimuli than the non-endpoint stimuli. More specifically, the endpoint stimuli were presented to the listeners 3 times while the non-endpoint ones were presented 6 times. Overall, the listeners heard 180 stimulus sentences: [(2 endpoint stimuli × 3) + (4 non-endpoint stimuli × 6)] × 6 conditions. The entire experiment took about 25 minutes.

Accent-type differences are normally well apparent to native speakers of Japanese since there are a certain number of minimal pairs that can only be accentually distinguished such as *ka'mi* "God" vs. *kami* "hair" and *a'me* "rain" vs. *ame* "candy". Most of the listeners were able to perform the task described above quite well. Some of them had difficulty in doing so at first, but they got used to the task after hearing the practice stimuli.

5.3.2 Results

The data were analyzed with respect to the response rates in which N1 was judged to be accented. The results are shown in the two graphs in Figures 5.2, which plot the response rates as a function of the P1 values.[2]

The overall pattern observed in Figure 5.2 is that in both graphs listeners' responses change from 'unaccented' to 'accented' as the unaccented endpoint stimulus (low P1) to the accented endpoint stimulus (high P1), which indicates that the manipulated F0 contour properties properly signaled the acoustic cues to identify the accent type of N1. This supports the validity of the stimuli for Experiment 3b.

---

[2] Each stimulus is represented by its P1 values only for representation's sake. Note that there are other differences in other F0 properties among the stimuli such as the height and alignment of V1 and V2 (See Figure 5.1).

N2 = accented
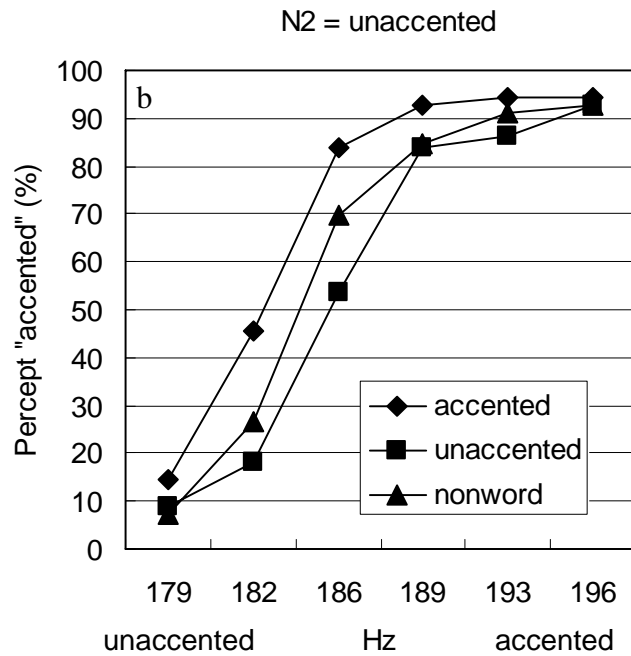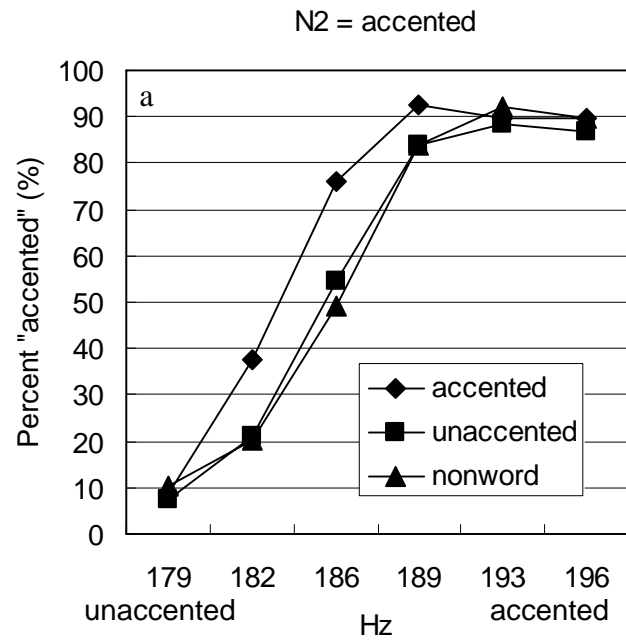
N2 = unaccented

Figure 5.2 Mean response rates for N1 being judged as "accented" as a function of the continuum steps of the stimuli, represented by the peak F0 values of N1 (P1). On the *x*-axis, the lowest value represents an unambiguous unaccented pattern and the highest value an unambiguous accented pattern. (a) N2 = accented, (b) N2 = unaccented.

A comparison of the accented N1 condition (diamonds) with the unaccented N1 condition (squares) in Figures 5.2a and 5.2b reveals that the lexicality effect is observed regardless of the accent type of N2. In both of the graphs the response rates are greater in the accented N1 condition than in the unaccented N1 condition when the acoustic properties of N1 do not strongly signal the accent type, which is most clearly seen when P1 is 182, 186 and 189 Hz. As a result of this effect, the category boundary (50% response rate) is found earlier when N1 is lexically accented than when it is lexically unaccented. This suggests that when the acoustics does not strongly cue the accent type of a word, listeners rely on the word's structural property of lexical accent to determine its accent type.

Another observation concerns the conditions in which N1 is a nonword (the triangles in Figures 5.2a and 5.2b). What is observed in these conditions is a perceptual bias toward the unaccented type. In Figure 5.2a, the mean response rates for the nonword condition pattern almost identical to the unaccented condition. In Figure 5.2b, however, the pattern slightly differs from Figure 5.2a in that the degree of the bias toward the unaccented type is somewhat weaker. That is, when N1 is a nonword, listeners are biased toward the unaccented pattern more when N2 was accented than when it was unaccented.

A three-way repeated-measures ANOVA with the accent type of N1 (AccN1), that of N2 (AccN2), and stimulus step (referred to by P1 height) as the independent variables was carried out. Since the lexicality effect was conspicuous in P1 = 182 Hz, 186 Hz, and 189 Hz, only the data obtained from those three steps were included in the analysis. The results were shown in Table 5.2 below:

Factor	F value	P value
AccN1	$F(2,44) = 18.267$	$p < 0.001$*
AccN2	$F(2,44) = 74.479$	$p < 0.001$*
P1 height	$F(1,22) = 4.747$	$p = 0.040$*
AccN1*AccN2	$F(2,44) = 3.437$	$p = 0.041$*
AccN1*P1 height	$F(4,88) = 3.465$	$p = 0.011$*
AccN2*P1 height	$F(4,88) = 2.008$	$p = 0.146$
AccN1*AccN2*P1 height	$F(4,88) = 1.926$	$p = 0.113$

Table 5.2 Results of ANOVA for the results of Experiment 3a (accent identification). A significant effect is indicated by an asterisk.

As shown, the effect of AccN1 was significant showing that the response rate is significantly higher when N1 is lexically accented than when it is lexically unaccented. This supports the presence of the lexicality effect caused by the lexical accent status of N1. The significant result of P1 height suggests that the listeners judged word accent status appropriately such that they judged the stimuli that were made to sound accented as accented ant those that were made to sound unaccented as unaccented.

The significant effect for AccN2 shows the overall response rates are higher when N1 is unaccented than when it is accented. However, it seemed that this was particularly caused by the relatively high response rates in the nonword N1-unaccented N2 condition (triangles in Figure 5.2b). The evidence comes from a separate repeated-measures ANOVA carried out excluding this condition: there was no significant effect on AccN2 ($F(1,22)=1.113$, $p=0.303$).

There were two significant interactions: AccN1*AccN2 and AccN1*P1 height. The first significant interaction suggests the relation among the three N1 conditions is different between when N2 is accented and when it is unaccented. What seems to be mainly responsible is that the weaker bias toward the unaccented pattern observed in the nonword N1-unaccented N2 condition. The repeated-measures ANOVA excluding the nonword N1 conditions, which was mentioned in the previous paragraph, showed a raised p-value for the

interaction between AccN1 and AccN2 but still a marginally significant effect ($F(1,22)=4.119$, $p=0.055$). This result implies that there are some other effects contributing to the interaction, the uncovering of which I leave for future research.

The second interaction, AccN1*P1 height, suggests that the relation among the three accented N1 conditions differs among different P1 heights, which can readily be seen in Figure 5.2. An examination of the mean response rates at P1=182, 186 and 189 Hz – the values used in the statistical analysis – tells us that in both graphs in Figure 5.2 the difference between the accented N1 condition and the other two conditions is relatively large when P1 is 182 Hz and 186 Hz (setting aside the exceptionally high mean value for the nonword N1 condition at P1=186 in Figure 5.2b). However, at P1=189 Hz the difference is much smaller. These patterns are not surprising: they just show that the stimulus is more ambiguous with respect to lexical accent type at P1=182 Hz and 186 Hz than at P1=189 Hz.

5.3.3   Discussion

An important finding of Experiment 3a is the significant difference between the accented N1 and the unaccented N1 conditions in Figures 5.2a and 5.2b. What is observed is that upon hearing an acoustically ambiguous word, the listeners' response is more likely to be "accented" when the word is lexically accented than when it is lexically unaccented, and conversely their response is more likely to be "unaccented" when the word is lexically unaccented than when it is lexically accented. This pattern can be considered to be the lexicality effect (Ganong 1980), in which acoustically ambiguous stimuli are perceived such that they conform to lexical forms in the lexicon of the listener's language. The current experimental results suggest the presence of the lexicality effect at the prosodic level.

Another significant pattern found in Experiment 3a is the overall bias toward the unaccented type in the nonword N1 conditions. In Figure 5.2 the nonword N1 conditions show the same pattern as the unaccented N1 conditions with much fewer accented percepts than in the accented conditions, especially when the stimuli were acoustically ambiguous (P1=182, 186, 189 Hz). This pattern is in accordance with Kubozono and his colleagues' "unaccented law" (Kubozono 1996, Kubozono & Fukui 2004, Kubozono & Ogawa 2004). The unaccented law says that the accent type of a loanword (and of even a nonword, with the assumption that such an item can be regarded as an extremely unfamiliar loanword) can largely be determined by its syllable length and structure: specifically, there is a strong tendency that the accent type is unaccented when the loanword is four-moras long, ends with a sequence of light syllables and does not contain epenthetic vowels.

The perceptual bias toward the unaccented pattern observed in Experiment 3a can be considered to be evidence for the unaccented law. The nonword used in the experiment *inamire* is four-moras long, ends with a sequence of light syllables and does not contain any epenthetic vowels, which satisfies the conditions in which the default unaccented pattern emerges. Thus, we can explain the obtained bias in such a way that the listeners perceived the nonword *inamire* as an extremely infrequent loanword and assigned the unaccented type based on its syllable structure and mora length when it was acoustically ambiguous.

The observed perceptual bias does not just serve as evidence for the validity of the unaccented law. It suggests that the law is psychologically real. In their studies mentioned above, Kubozono and his colleagues reached the unaccented law as the emergent pattern by examining existing loanwords (and alphabetic acronyms) in a dictionary. The bias toward the unaccented type found in Experiment 3a indicates that the unaccented law is a perceptual

process that is actively processed in real time by the listeners.

However, we should be careful about one point. The explanation of the perceptual bias holds by the unaccented law holds only if we assume that the listeners categorized *inamire* as a loanword, since the syllable structure conditions for the unaccented law are based on the accent patterns of loanwords. Studies show that the Japanese lexicon consists of several different "lexical strata" and words in a stratum show different phonological restrictions from words in another stratum, such as voiced geminates being only seen in the foreign stratum (Amano & Moreton 1999, Gelbart 2005, Ito and Mester 1995, 1999, Moreton 2002). We do not know which stratum our listeners perceived the nonword as belonging to. The nonword does not have any phonological cues that signal that it belongs to any particular single lexical stratum. In that sense, it might be perceived as an extremely rare loanword. But it was a nonword created by minimally changing the real person mamas *Inamori* and *Inamire*, which themselves are native (Yamato) words. Thus it is also equally possible that it is perceived as a Yamato word. The relationship between nonwords and the lexical strata in Japanese is an interesting topic for future investigation.

One could think of another possible account of the observed perceptual bias. The account says that when listeners hear a nonword, they are biased toward the accent pattern that is most dominant in the Japanese lexicon. Based on the Japanese corpus developed by Amano & Kondo (1999), approximately 60 % of the four-mora words are unaccented (Kitahara 2001). Given that listeners assigned this most commonly seen accented type to the nonword that does not have any structural property of lexical accent, the perceptual bias found in the experiment has a natural explanation. Further study is warranted to determine which of the two accounts given above (the unaccented law account and the dominant accent

type account) is on the right track.

5.3.4   Summary

In Experiment 3a, we have found the lexicality effect on accent identification. It has been shown that when the acoustics does not provide strong cues for the accent type of a word, the lexical specification of the word, which is available through the segmental string of the word, is used to identify its accent type. In other words, when the acoustic property of lexical accent is not available listeners make use of the structural property to determine the accent type of a word. Moreover, it has been shown that when listeners heard a nonword, the accent type that they assigned to it was biased toward the unaccented pattern. Two possibilities that can account for this perceptual bias have been suggested, but further research is needed to properly understand this effect.

The next sub-experiment in Experiment 3 (Experiment 3b) focuses on the structural property of lexical accent. Specifically, it examines how the structural property of lexical accent is used to assess the perceived prominence of an F0 peak. The crucial question is whether lexical accent is perceptually normalized with only the structural property of lexical accent when no acoustic property is available.

5.4   Experiment 3b: Prominence judgment

5.4.1   Methods

5.4.1.1 Stimuli

The stimuli used in Experiment 3b were basically identical to those used in Experiment 3a. However, there were two differences. One was that the nonword stimuli were excluded

simply to keep the size of the experiment manageable. That is, only the sentences in (1) with different combinations of accent types between N1 and N2 were used. The sentences are reproduced in (3) below for each of reference.

The other difference was that the F0 heights of N2 were varied to indicate 6 different heights, as shown in Figure 5.3 below. N2 was varied only in its F0 peak height (P2) when it was accented and in its F0 peak height (P2) and the following F0 peak (P3) when it was unaccented, with its overall F0 contour shape kept constant. The step size was 14 Hz. The actual F0 values used for the N2 height in Experiment 3b are shown in Table 5.3. The P2 values were identical between the accented and the unaccented N2 conditions. One might object that we would need a range that was shifted to a bit higher range for the accented word condition because we know that accented words are realized higher than unaccented words. However, this manipulation actually did not pose a problem since the purpose of the manipulation was to obtain the point where N1 and N2 are judged to have the same perceptual prominence. Thus, as long as we obtain a pattern in which the responses for N2 that are judged to be more prominent than N1 are an orderly function of P2 height, there should be no problem with the same F0 range between the accented and unaccented conditions. As seen in the results section, we did obtain such orderly functions in the experiment.

(3) = (1)
      a. *aa*
      Ina'mori-no   an'iyome-ga   inai.
      Inamori-Gen   sister-in-law-Nom   not found
      Inamori's sister-in-law is not found.

    b. *au*
      Ina'mori-no   omiyage-ga   kieta.
      Inamori-Gen   souvenir-Nom   disappeared
      Inamori's souvenir disappeared.

c. *ua*

Inamura-no	an'iyome-ga	inai.
Inamura-Gen	sister-in-law-Nom	not found

Inamura's sister-in-law is not found.

d. *uu*

Inamura-no	omiyage-ga	kieta.
Inamura-Gen	souvenir-Nom	disappeared

Inamura's souvenir disappeared.

The 6 different F0 heights for N2 were combined with the 6 N1 stimuli, yielding 36 stimuli per continuum. The overall number of stimuli was 144 (36 × 4 sentences).

N2	Step					
	1	2	3	4	5	6
P2 (*a* or *u*)	120	134	148	162	176	190
P3 (for *u* only)	109	123	137	151	165	179	(Hz)

Table 5.3 Heights of N2 peak (P2) and the F0 shoulder (P3).
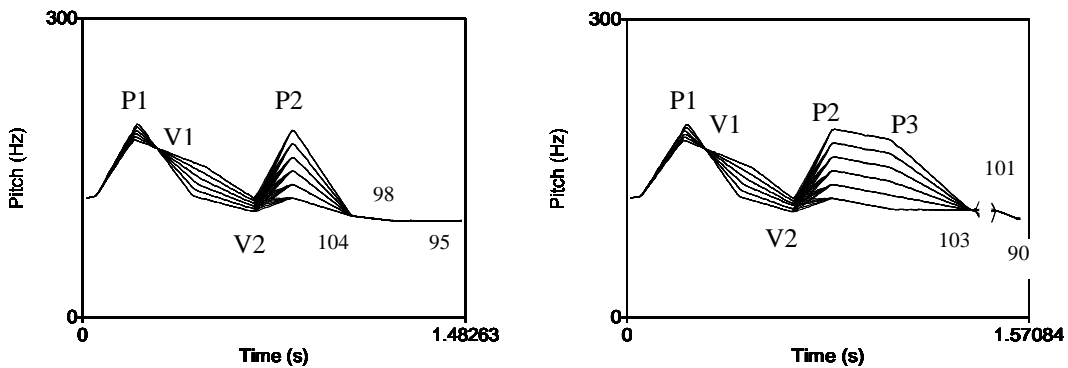
Figure 5.3 F0 contours of the stimuli used in Experiment 3b: N2 shows F0 properties for an accented word in (a) and for an unaccented word in (b).

5.4.1.2 Listeners

The listeners for Experiment 3b were the same as in Experiment 3a. Note, however, that the

data from all of the 24 listeners were used for the analysis including the listener whose data were mistakenly lost in Experiment 3a.

5.4.1.3 Procedure

Experiment 3b was run following Experiment 3a. The stimulus sentences in (3), with continua on N1 and different peak heights on N2, were randomly presented to the listeners. The listeners' task was identical to Experiments 1 and 2: to judge relative perceived prominence between N1 and N2. More specifically, they were asked to judge which of N1 and N2 was given more emphasis by the speaker.

The stimulus presentation involved repetitions and the number of repetitions depended on the stimuli. Specifically, the lowest and the highest N2 values were presented only 3 times while the other 4 intermediate values were presented 6 times. This was simply because it was expected that listeners would reliably respond that N1 was more emphatic than N2 when N2 was at the lowest value and that N2 was more emphatic than N1 when N2 was at the highest. For each of the N2 heights, the 6 N1 stimuli were combined. Since there were 4 accent-type conditions (*aa, au, ua,* and *uu*), the total number of the stimuli presented was 720 ([2 endpoint N2's × 3 repetitions + 4 intermediate N2's × 6 repetitions] × 6 N1 × 4 accent-type conditions). Experiment 3b took approximately 70 minutes including the instructions, a practice session and breaks.

5.4.1.4 Analysis

Just as Experiments 1 and 2, the response rates for N2 (N2 sounding more emphatic than N1) were obtained for each listener. Then, the 50% crossover point of the response rates was

calculated for each P1 height and for each listener, by probit analysis.

5.4.1.5 Predictions

Recall that in Experiment 1 we found perceptual effects for accentual boost and downstep. The accentual boost effect was found between *aa* and *au*: the difference between P1 and P2 was greater in *au* than in *aa*. The downstep effect was found between *aa* and *uu*: the difference between P1 and P2 was greater in *aa* than in *uu*. Based on these findings, I proposed two perceptual normalization processes: accentual boost normalization and downstep normalization.

If the structural property of lexical accent plays a role in these perceptual normalization processes, we should see their effects between *aa* in (3a) and *ua* in (3c) and between *au* in (3b) and *uu* in (3d). Note that the N1s in each of these pairs have exactly the same F0 contour properties. They only differ from each other with respect to segmental content. The lexically accented N1 in *aa* (*Ina'mori*) and the lexically unaccented N1 in *ua* (*Inamura*) are identical with respect to F0 pattern. The F0 properties of these two N1s were manipulated in the same way with the segmental content intact. The same manipulation was done for *au* and *uu*. Therefore, the only information that the listeners can rely on in order to identify the accent type of N1 in a sequence is the lexically-specified accent information, namely the structural property of lexical accent.

The specific predictions are illustrated in Figure 5.4, using the same type of graph that we have been using in describing experimental results.
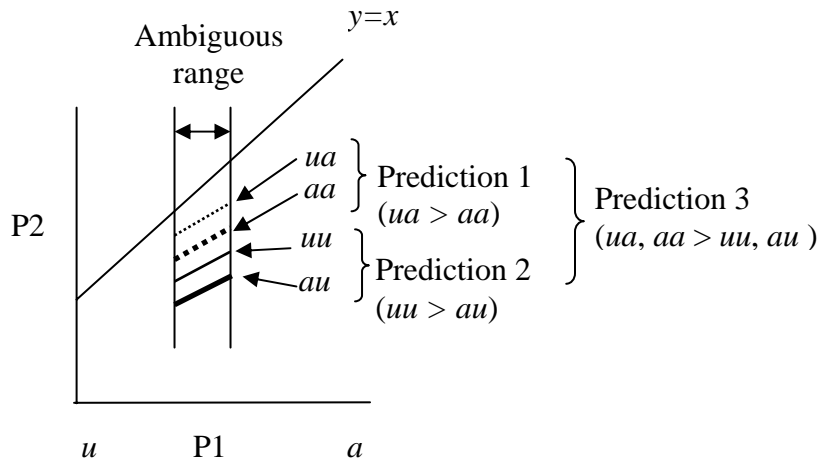
Figure 5.4 Predictions made when the effects of the structural property of lexical accent is assumed.

There are three specific predictions. The first prediction is that for a given height of P1, P2 is higher in F0 for *ua* than for *aa* when P1 and P2 are perceived to have the same prominence. In terms of P1-P2 difference, it is predicted to be greater in *aa* than in *ua*.[3] This effect is expected for two reasons. First, accentual boost normalization is applied to only the second word in *ua* while this perceptual process is applied in both of the words in *aa*, which causes their accentual boost effects to be unobservable. Therefore, it should pattern as if the accentual boost effect was seen only in the second word of *ua*, and P2 is expected to be higher in *ua* than in *aa* for a given P1 value when P1 and P2 have the same prominence. Second, downstep normalization is expected in the second word of *aa* but not in *ua*, which will bring a lower P2 value in *aa* than in *ua* for a given P1 value in the circumstance in which P1 and P2 are perceived to be equally prominent. Again, P1-P2 difference will be greater in *aa* than in *ua*. So accentual boost normalization and downstep normalization are predicted to

---

[3] Note that the "P1-P2 difference" means the subtraction of P2 from P1 and not the absolute difference between P1 and P2. Thus, if the difference was negative in *ua* while it was positive in *aa* (which might well be the case) with the absolute difference being greater in *ua*, I would still say P1-P2 difference is greater in *aa* than in *ua*.

give a combined effect such that P1-P2 difference is greater in *aa* than in *ua*.

The second prediction concerns *au* and *uu*. It is predicted that for a given P1 height, P2 will be higher in *uu* than in *au* for it to have the same perceived prominence as P1. Here again, accentual boost normalization and downstep normalization are both considered to be responsible for this effect. First, accentual boost normalization is expected only in the first word of *au*, whereby P1 has to be higher in F0 in *au* than in *uu* when P1 and P2 are perceived to have the same prominence. From the perspective of P1-P2 difference, the difference value should be greater in *au* than in *uu*. Second, downstep normalization is expected in *au* but not in *uu*, thus for a given P1, P2 is lower in *au* than in *uu* when P1 and P2 have the same perceived prominence. Like accentual boost effect, the downstep effect is predicted to give a greater P1-P2 difference in *au* than in *uu*. Just like in the first prediction, the effects of accentual boost normalization and downstep normalization are predicted to show a combined effect on P1-P2 difference such that it is greater in *au* than in *uu*.

The third prediction is that P2 will be higher in the sequences in which the second word is accented (*aa* and *ua*) than when it is unaccented (*au* and *uu*). Recall that in the experimental sentences the second words differ from one another with respect to the structural as well as the acoustic properties of lexical accent. Thus, the way the second words differ between *aa* and *ua* on the one hand, and *au* and *uu* on the other, is the same as the way the experimental sentences in Experiment 1 did, which suggests a robust normalization effect. The effect that is expected to be seen is the accentual boost effect: for a given P1 height, P2 has to be higher in *xa* (where *x* could be *a* or *u*) than in *xu* when P1 and P2 are perceived to be equally prominent. In other words, greater P1-P2 difference is expected in *xu* than in *xa* when P1 and P2 are perceived to have the same prominence.

We also have to consider another set of predictions, based on the assumption that the structural property of lexical accent does not play a role in prominence perception. In other words, this second set of predictions is based on the hypothesis that only the acoustic property of lexical accent is relevant. This hypothesis makes the three predictions shown in Figure 5.5. They are (1) there will be no difference between *ua* and *aa*; (2) there will be no difference between *au* and *uu*; (3) there will be a difference between *xa* and *xu* such that for a given P1 height, P2 is higher in *xa* than in *xu* when P1 and P2 are perceived to be equally prominent. The differences between Figure 5.5 and Figure 5.4 are clear: the two sets of predictions differ from one another with respect to Predictions 1 and 2 but share Prediction 3.
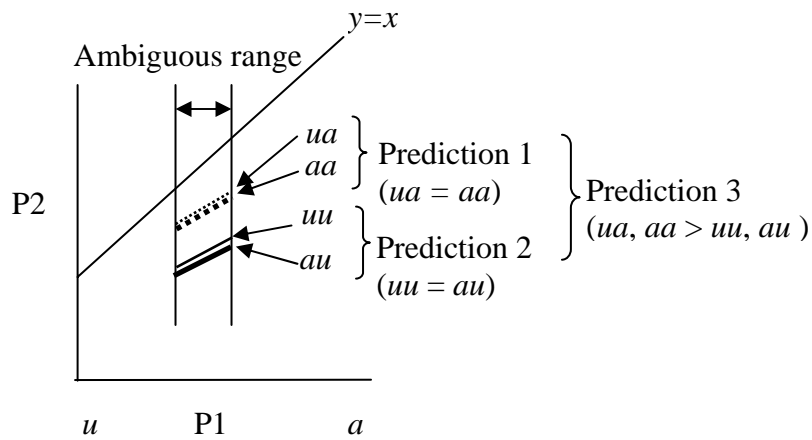
Figure 5.5 Predictions made when the effects of the structural property of lexical accent is not assumed.

5.4.2   Results

The results of Experiment 3b are represented in Figure 5.6. This graph shows P2 values that have the same perceived prominence as P1. The 5-step N1 continuum from the unaccented to the accented patterns is represented on the *x*-axis.
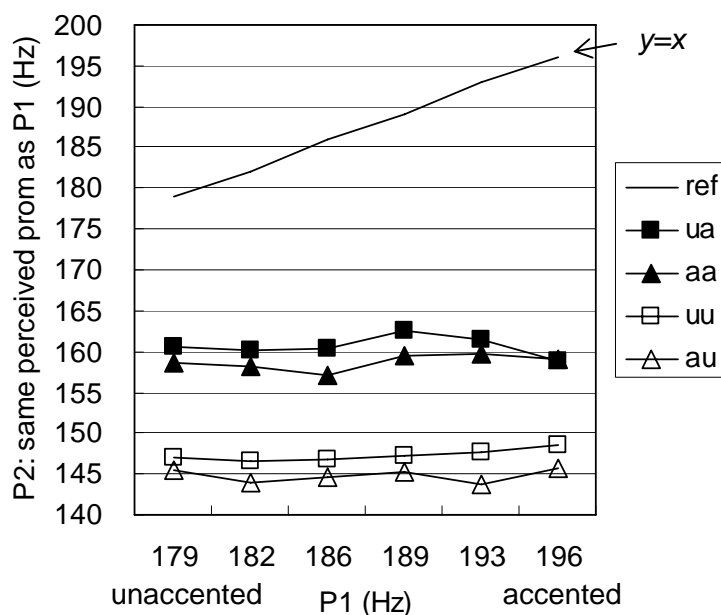
Figure 5.6 Mean values of P2 giving the same prominence as P1 as a function of P1 height for different accent-type conditions.

One can readily see that all of the mean P2 values are well below the reference line *y=x*, showing that for P1 and P2 to be equal in perceived prominence, P2 has to be much lower in F0 than P1. In other words, this is a replication of the effect of perceptual compensation for F0 declination (Pierrehumbert 1979).

In addition to the F0 declination effect, we can make two observations. First, compare the cases in which only the lexical accent status of N1 is different, namely *ua* vs. *aa* and *uu* vs. *au*. The mean P2 values are consistently lower in *aa* than in *ua* and in *au* than in *uu*. In other words, *aa* and *au* need larger P1-P2 differences than *ua* and *uu*, respectively, when P1 and P2 are perceived to be equal in prominence (*ax > ux*, where *x* is *a* or *u*). These results confirm Predictions 1 and 2 in Figure 5.4. Recall that the first words in *ua* and *aa* are identical to one another in their F0 contour shape. So are the first words in *ua* and *uu*. Thus, the observed differences in P2 height should be explained in terms of the difference in their

lexical accent status. The structural property of lexical accent plays a role in the perception of intonational prominence.

The second observation concerns the large differences between the cases where N2 is accented and those where it is unaccented (*xa* vs. *xu*). Across the all N1 stimuli, the *ua* and the *aa* conditions, in which N2 is accented, show remarkably higher mean P2 values than the *uu* and the *au* conditions, in which N2 is unaccented. To put it in other words, a greater P1-P2 difference is observed when N2 is unaccented than when it is accented. This second pattern is what we expected in Prediction 3 in Figure 5.4 (and in Prediction 3 in Figure 5.5). The second words were manipulated only for F0 peak height. Their F0 contour shapes were kept intact. Both the structural and the acoustic properties of lexical accent were present in them for the listeners to identify their accent types. We expected a robust perceptual effect and that is confirmed in the experiment.

Although both of the two observed patterns are consistent with the idea of accentual boost normalization (and downstep normalization for the first pattern), they differ from one another in their effect size. The size of the effect is much greater when the accent type of N2 differs (*ux* vs. *ax* or filled vs. unfilled symbols) than when the accent type of N1 differs (*xa* vs. *xu* or squares vs. triangles). Given that only the structural property of lexical accent is available on N1 but both the structural and the acoustic properties of lexical accent are available on N2, this bigger difference may be seen as the intrinsically greater influence of the acoustic property of lexical accent.

The statistical analysis supports our observations. A three-way repeated-measures ANOVA was performed with accent type of N1 (AccN1), that of N2 (AccN2), and stimulus step as the independent variables. In Experiment 3a, we found that the stimuli that were

ambiguous in terms of accent type were the three intermediate stimuli in the N1 continua, i.e. ones at P1 = 182 Hz, 186 Hz, and 189 Hz. Thus, only those three stimuli were used in the ANOVA analysis. The results were summarized in Table 5.4. Only the main effects of accent types of N1 and N2 were significant. That is, when P1 and P2 are equal in perceived prominence, the mean P2 values were significantly lower when N1 was accented than when it was unaccented (*aa* and *au* are lower than *ua* and *uu*) and were significantly higher when N2 was accented than when it was unaccented (*ua* and *aa* are higher than *uu* and *au*). From these results, we can reasonably conclude that (1) the structural property of lexical accent plays a role in perception of intonational prominence and that (2) the effect is greater when the structural and the acoustic properties of lexical accent are expressed together.

Factor	F value	P value
AccN1	$F(1,23) = 5.584$	$p = 0.027^*$
AccN2	$F(1,23) = 70.204$	$p < 0.001^*$
P1 height	$F(2,46) = 1.640$	$p = 0.205$
AccN1*AccN2	$F(1,23) = 0.00001$	$p = 0.997$
AccN1*P1 height	$F(2,46) = 0.268$	$p = 0.766$
AccN2*P1 height	$F(2,46) = 0.801$	$p = 0.455$
AccN1*AccN2*P1 height	$F(2,46) = 0.594$	$p = 0.556$

Table 5.4 Results of ANOVA for the results of Experiment 3b. Significant results are indicated by the asterisks.

5.4.3   Discussion

In this section, I have two tasks. The first one is to show how the patterns observed in Experiment 3b can be explained in terms of accentual boost normalization and downstep normalization. The second task is to discuss the theoretical implications of the experimental findings.

5.4.3.1 Explanations of the observed patterns

What has been found in Experiment 3b can be summarized as the following two patterns. First, P1-P2 difference is greater when the first word is accented than when it is unaccented (*ax > ux*). This pattern is observed when the accented type of the first word is signaled only by its structural property of lexical accent. Second, P1-P2 difference is greater when the second word is unaccented than when it is accented (*xu > xa*). Unlike the first pattern, this difference is observed when the structural and the acoustic properties of lexical accent are both present, and the effect is quite robust in the sense that the effect size is much larger in the second pattern than in the first one.

The first pattern (*ax > ux*) can be explained as the combined effect of accentual boost normalization and downstep normalization, as shown in Figure 5.7 using the same type of graphical technique as the one employed for the explanations of the results of Experiment 1 (Chapter 3).
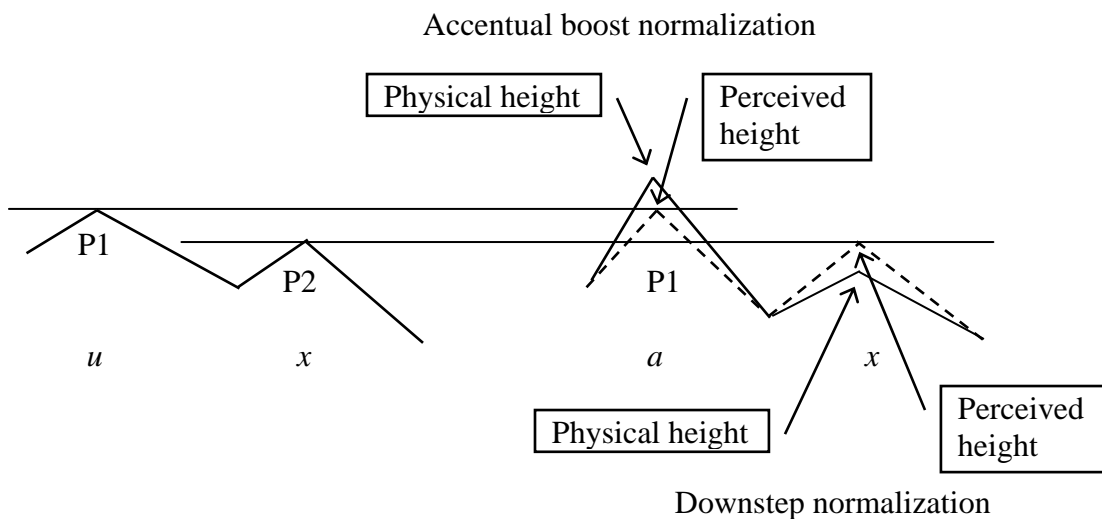
Figure 5.7 Explanation of the greater P1-P2 difference in F0 in *ax* than in *ux* by accentual boost normalization and downstep normalization.

Given accentual boost normalization, the perceived height of P1 in *ax* is lower than its physical height due to the accentual-boost-cancelling mechanism of the perceptual process. The first word of *ux* is unaccented, and no perceptual normalization is at work there. As a result P1-P2 difference is greater in *ax* than in *ux*. Moreover, physical P1-P2 difference gets even greater due to downstep normalization on the second word of *ax*. The first word of *ax* is accented, which triggers downstep normalization on the following word. The production effect of downstep is perceptually undone on the second word of ax such that the perceived P2 height is higher than the physical P2 height. This perceptual adjustment makes physical P1-P2 difference greater in *ax* than in *ux*.

Let us now turn to the explanation of the second pattern: greater P1-P2 difference in the sequences where the second word is unaccented than those where it is accented (*xu* > *xa*). This pattern can properly be explained by accentual boost normalization,[4] as graphically shown in Figure 5.8.
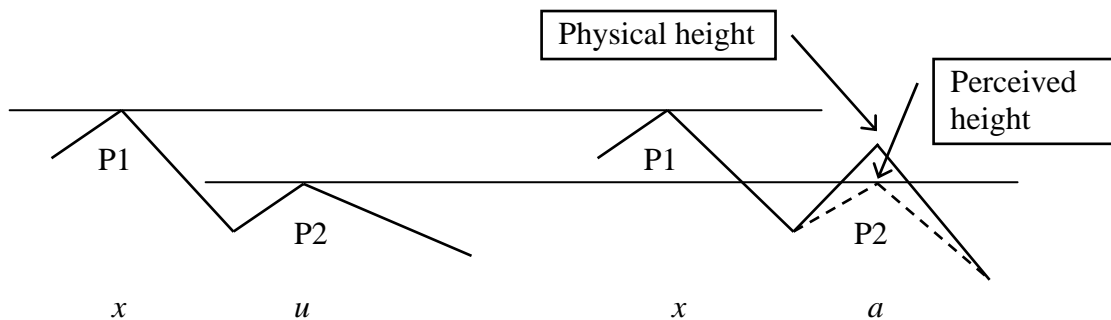
Figure 5.8 Explanation of the greater P1-P2 difference in F0 in *xu* than *xa* by accentual boost normalization.

---

[4] One may want to point out that downstep normalization is involved in the *aa-au* pair because the first word is accented in both sequences. However, the perceptual effect for downstep normalization can not be reflected on P1-P2 difference precisely because the first words are *both* accented. It is considered that downstep normalization is applied to both of the second words because of which no difference in relative perceived prominence can be observed between *aa* and *au*. It should be concluded that only accentual boost normalization should be responsible for the second pattern.

What is shown in this figure is that P1-P2 difference is greater in *xu* than *xa* because the accented second word of *xa* triggers accentual boost normalization whereby its F0 peak height is higher than its perceived height. Thus, the smaller P1-P2 difference in *xa* is equivalent to the larger P1-P2 difference in *xu* when P1 and P2 are equal in perceived prominence in each sequence.

5.4.3.2 Implications

This section discusses two issues. The first pertains to the importance of lexical accent in the perception of intonational prominence. The finding that the structural property of lexical accent plays a role in the perception of intonational prominence lends further support for the claim I made in Chapter 3: models of the perception of intonational prominence have to take lexical accent into account. The crucial point is that the greater P1-P2 differences found in *ax* than in *ux* are not caused by the difference in F0 properties between the two first words but by the difference in lexical accent status between them. The two first words are identical to one another with respect to F0 contour shape. They only differ with respect to the accent specification as part of the lexical information of each word, which can be accessed by identifying the word itself through the segmental content. Thus, the F0 properties can never be responsible for the greater P1-P2 differences in *ax* than in *ux*. The result of *ax* being greater than *ux* in P1-P2 difference can be taken as evidence that listeners' knowledge of phonological lexical accent status affects perceived prominence of intonational peaks.

Note that the Japanese lexicon plays a decisive role here. The result mentioned above would not be observed if the listeners were not native speakers of Japanese who have the Japanese lexicon. It is the knowledge on accent that is specific to Japanese that causes that

result. This is solid evidence that lexical accent needs to be included as an important factor that influences perceived prominence in prominence perception models.

As the second issue I want to take up the nature of accentual boost normalization and downstep normalization. Specifically, I want to discuss the level at which they operate in the processing of speech signal. The question is whether the two perceptual processes are phonological or phonetic.[5]

The experimental finding that accentual boost normalization and downstep normalization can both be informed by the structural property of lexical accent suggests that these normalization processes operate with reference to the phonological representation where the information of lexical accent status is available. This means that the perceptual processes are phonological (rather than phonetic).

While this claim conforms to the conception of lexical accent as phonological entity, it runs counter to the characterization of downstep by Pierrehumbert & Beckman (1988). For these authors, downstep is a phonologically-conditioned phonetic implementation rule. In their model, the phonetic implementation exploits the difference in tonal type between accented and unaccented words, namely lexical accent H* and phrasal H. A lexical accent H* triggers a phonetic implementation rule on pitch range scaling whereby the pitch range is compressed for the tones that follow the lexical accent H*, but a phrasal H does not trigger the implementation rule and as a result no pitch range compression occurs. This difference in the application of the implementation rule accounts for the different patterns between downstepped and non-downstepped F0 peaks. Thus, in Pierrehumbert & Beckman's (1988)

---

[5] Whether a linguistic phenomenon is phonological or phonetic is a difficult issue because the answer crucially depends on what is considered phonetic or phonological. Here I take one of the most common views: Phonology deals with categorical and symbolic entities while phonetics deal with gradient and non-symbolic entities (Keating 1988, Pierrehumbert 1990, Cohn 1993 among others). For a recent review, see Kingston (2007).

model, the distinction between a downstepped and a non-downstepped F0 tone is not represented in the phonological representation: it is a job of phonetics.

Pierrehumbert & Beckman's (1988) model of downstep is based on speech production. Under the reasonable assumption that the production and the perception of downstep are governed by the same system, the most reasonable modeling of downstep normalization would be to give the process the same characterization as Pierrehumbert & Beckman's production-based downstep. That is, downstep normalization is conceptualized as the process that operates at the phonetic level but makes reference to the information at the phonological representation. This way, downstep normalization can be the perceptual counterpart of the downstep in production. The phonetic implementation rule for downstep and downstep normalization both do not manipulate symbols but only refer to symbolic tonal strings in the phonological representation. I will elaborate this suggestion in Chapter 7.

5.5   Chapter summary

This chapter showed that the structural property of lexical accent is crucially relevant to the perceived prominence of intonational peaks. Experiment 3a demonstrated the lexicality effect in the identification of lexical accent in Japanese. When the acoustic signal does not have the strong cues for a certain accent type, listeners' response is biased toward the phonological accent type that is specified for the word in the lexicon. Experiment 3a also used nonword stimuli. Results showed a bias toward the unaccented pattern rather than the accented one. Several possible accounts were discussed, including the claim made by Kubozono and his coworkers (Kubozono 1996, Kubozono & Fukui 2004, Kubozono & Ogawa 2004) that a large part of the accent patterns of loanwords are not determined by the lexical specifications

but largely by independent phonological factors such as syllable length and syllable structure.

Experiment 3b tested whether listeners' phonological knowledge of lexical accent status (= the structural property of lexical accent) affects the assessment of perceived prominence. The most important finding is that when P1 and P2 are equally prominent, P2 is significantly lower in F0 when N1 is accented than when it is unaccented. This finding is obtained crucially when N1's accent type is differentiated by its lexical accent status alone. This finding is important because it indicates that phonological knowledge influences the perception of intonational prominence.

CHAPTER 6

EXPERIMENT 4: ACOUSTIC PROPERTY OF LEXICAL ACCENT AND PERCEIVED PROMINENCE

6.1   Introduction

In the previous chapter, we considered the interaction between lexical accent and perceived prominence with respect to the structural property of lexical accent, namely the phonological lexical accent status of words specified in the lexicon. Experiment 3 showed that the proposed perceptual normalization processes – accentual boost normalization and downstep normalization – play a role in the interpretation of perceived prominence. This chapter examines the flip side of Experiment 3. We explore the interaction between prominence perception and the *acoustic* property of lexical accent, namely the phonetic F0 properties associated with lexical accent.

We have not yet investigated the effects of the acoustic property of lexical accent on prominence assessment. In Experiment 1, we used naturally-produced utterances in which the structural and the acoustic properties of lexical accent were both present in the stimuli. In Experiment 3, we used F0 contours that contained the structural property of lexical accent but did not contain the acoustic property in order to study the pure effects of the structural property on perceived prominence. What we need to do next is obvious: we need to study the pure effects of the acoustic property of lexical accent on perceived prominence. Experiment 4, which is reported in this chapter, attempts to fulfill this purpose.

There are two specific research questions asked in Experiment 4: (1) Is accentual boost

normalization triggered by the acoustic property of lexical accent (henceforth phonetically-driven accentual boost normalization)? and (2) Is downstep normalization triggered by the acoustic property of lexical accent (henceforth phonetically-driven downstep normalization)? Of these, the second question is the particular interest of this chapter.

6.2 Theoretical status of downstep in Japanese

The question of whether there is phonetically-driven downstep normalization is interesting when we consider it in terms of the theoretical status of downstep in Japanese. Downstep in Japanese is often characterized as a phonologically-conditioned phonetic implementation rule: a MiP is realized lower when it is preceded by an accented MiP than by an unaccented MiP. Pierrehumbert & Beckman (1988) claimed that this lowering of the second MiP is achieved by the reduction of the pitch range after the first MiP, and as a result, all subsequent tones are scaled in the reduced pitch range. In their characterization, downstep in Japanese is phonologically-conditioned because its occurrence is crucially dependent on the presence of a lexical accent in the preceding phrase, but it operates at the phonetic level because it manipulates the scaling of tones in a gradient tonal space.

However, as we have been discussing in this dissertation, Japanese lexical accent has two facets, its acoustic and the structural properties. It has never been clearly said which property of lexical accent downstep refers to in the traditional conception of downstep in Japanese. This is not surprising, however, since in naturally-produced utterances these two properties are always compatible with each other, which in turn suggests that this issue needs to be scrutinized from the viewpoint of perception.

In previous theoretical work on downstep in Japanese, Sugahara's (2003) work is

exceptional. She proposed that Japanese downstep is phonetic in that it is triggered by the physically realized F0 value of a tone and not by the phonological lexical accent property of a word. She also claimed that what downstep refers to is not the accent H* tone but the following trailing low tone, i.e. +L.

Her evidence for these claims comes from a finding in one of her production experiments. In the experiment, she showed that the magnitude of the effect of downstep on a tone following an accent H* tone diminishes as more tones intervene between the tone and the preceding accent H* tone. This is graphically illustrated in Figure 6.1.
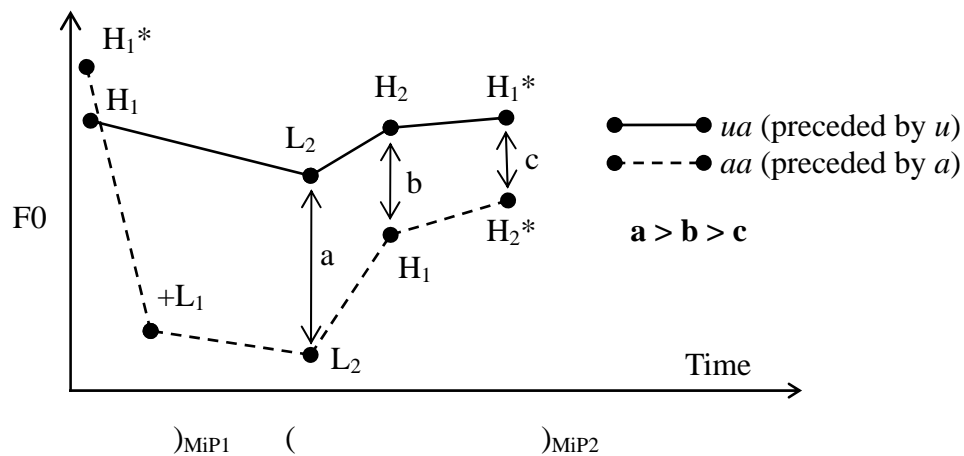
Figure 6.1 Diminution of the effect of downstep across tones. Sugahara's (2003) Figure 4.9 (p.113) is redrawn with slight modifications by the author.

The two F0 contours shown in the figure represent the sequences of an unaccented MiP followed by an accented MiP (*ua*, solid line) and of an accented MiP followed by an accented MiP (*aa*, broken line). The second MiP in the *aa* sequence undergoes downstep and all of the tones in the MiP are realized lower than the tones in the second MiP in the *ua* sequence, in which no downstep occurs on the second MiP because it is preceded by an unaccented MiP.

123

What Sugahara (2003) demonstrated in her experiment is that the differences between the temporally-comparable tones in the two F0 contours, namely those indicated by (a), (b) and (c) in the figure, diminish from (a) to (c).

Note that this result cannot be accounted for by Pierrehumbert & Beckman's (1988) pitch-range reduction model. In their model, all tones in a post-accent MiP are realized in a uniformly-reduced pitch range in such a way that they are reduced by the same proportion so that the scaling relation among them are identical between the original and the reduced pitch ranges. Therefore, the diminution of the downstep effect such as the one in Figure 6.1 is not expected in this model of downstep.

Based on this result, Sugahara proposes that downstep in Japanese is not based on a pitch range reduction triggered by the lexical accent H* but the phonetic value of a tone is computed relative to the value of the previous tone. She claims that downstep is a consequence of the second member of the accent H*+L tone (= +L) being scaled realized lower than a L% edge tone. Because the +L tone is realized lower that the L% tone, the value of the following tone is also lower than the tone following the L% tone (see Sugahara 2003 for the exact algorithm).

Note that Sugahara's (2003) claim has two parts. One is that downstep in Japanese is a phonetic phenomenon that refers to the physically-realized F0 value of some tone. The other is that the relevant tone whose value has an effect on what follows is the trailing part of the accent tone +L, not H*. This chapter only examines the first claim.

In Experiment 4, the effects of the *acoustic* property of lexical accent on downstep normalization and on accentual boost normalization are examined, with a particular focus on downstep normalization. Together with this principal purpose, the effect of the *structural*

property of lexical accent on downstep normalization is also investigated. To these ends, Experiment 4 uses stimuli whose acoustic and structural properties of lexical accent are varied independently of one another. Consequently, the F0 contour information and the lexical accent status information are compatible with each other in some conditions but conflict with each other in others. Thus, it is possible to examine whether there is any perceptual normalization effect between conditions which only differ from one another with respect to F0 contour shape or lexical accent status. In the following section, I describe the basic design of Experiment 4 and the ideas behind it.

6.3   The basic idea of Experiment 4

When one wants to study two different factors that are involved in a single phenomenon, the usual approach is to control one of the factors so as to be able to examine the other. The phenomenon that we are facing is lexical accent in Japanese, and the two factors are the acoustic and the structural properties of lexical accent. In the last chapter, by using ambiguous stimuli (or more precisely continua that contain ambiguous stimuli), we controlled the acoustic property of lexical accent in order to study the effects of the structural property of lexical accent on the assessment of perceived prominence. In this chapter, we want to study the flip side: the effects of the acoustic property of lexical accent on perceive prominence. To achieve this goal we need to control the structural property of lexical accent.

The most straightforward tack one could take is probably to use nonwords, since nonwords, by virtue of their own nature, do not have any lexical accent status. However, this approach has a problem. It is known that when one hears a word-like string of segments, some existing words are activated on the basis of segmental similarity, even if the string does

125

not match any of those existent words in the language. Thus, it would be difficult to see the effects that are brought out only by the properties of F0 contour shape. Lexical activation needs to be properly controlled.

To overcome this difficulty, Experiment 4 takes the following approach. It employs real words with naturally-produced F0 contours. However, in some conditions the F0 contour is manipulated such that its structural property of lexical accent is inconsistent with its acoustic property. For example, in the experiment, we have a sequence of two words which are naturally-produced as accented-accented (*aa*). In this sequence, the acoustic and the structural properties of lexical accent are matched with each other. Let us indicate this sequence as $[aa]_{aa}$, where $[aa]$ represents the acoustic properties and $[\quad]_{aa}$ the structural properties. We also have a separate sequence of *aa* using the same words as $[aa]_{aa}$, but in this sequence the words are accented only phonologically: the F0 contour shape is unaccented for both of them. The acoustic and the structural property of lexical accent conflict with each other in this sequence (assuming that there are no real words that correspond to combination of the segments and the accent type). This sequence can be denoted as $[uu]_{aa}$. Thus, $[aa]_{aa}$ and $[uu]_{aa}$ only differ from one another with respect to the acoustic property of lexical accent. We compare these two sequences in the experiment, which allows us to see the effects of the acoustic property of lexical accent on downstep normalization.[1] Of course we also have the complementary sequence pair of sequences in the experiment: $[aa]_{uu}$ and $[uu]_{uu}$.

An important point here is that lexical activation is properly controlled in these comparisons. The words in $[uu]_{aa}$ can still be perceived as those that their segmental strings indicate, because the experimental words were selected such that there are no existing words

---

[1] Recall that *aa* and *uu* were compared to see the pure effect of downstep normalization in Experiment 1. It is important to keep the accent type identical within a sequence because the effect of accentual boost normalization would emerge if adjacent words had different accent types, which causes a confounding effect.

126

that match their segmental strings and the manipulated phonetic properties at the same time. That is, it is possible that listeners activate the intended lexical items when their phonetic properties are inconsistent with their lexical entries.[2] This means that we can study the effects of the acoustic property of lexical accent on prominence assessment with controlled lexical activation. For example, the word *oma'wari* 'police officer' is a phonologically accented word and is pronounced with a lexical accent on the second mora. However, it is possible to identify this word when it is pronounced as if it was an unaccented word, because there is no real word in Japanese that has the segmental string of *omawari* and is lexically unaccented.

6.4 Methods

6.4.1 Stimuli

The five sentences shown in (1) were recorded several times by the author in a sound-attenuated booth. The stimulus sentences were based on the mean tonal values of those recorded utterances. The sentences all had the syntactic structure of [N(oun)1 N(oun)2 V]. The two N1s (*Yama'nashi* and *Yamanote*) and the two N2s (*oma'wari* and *omamori*) all were four-moras long and the two verbs (*yameta* and *kieta*) were three moras long. The two words for N1 *Yama'nashi* and *Yamanote* are lexically accented and unaccented, respectively, but their phonetic F0 properties were independently controlled at the recording of the sentences. That is, even though *Yama'nashi* is lexically accented and *Yamanote* is lexically unaccented, recording was made such that *Yama'nashi* was pronounced as if it was a lexically unaccented word and *Yamanote* was pronounced as if it was a lexically accented word depending on the

---

[2] Listeners may have an impression that the intended words are misaccented and that the speaker speaks some dialect that is different from theirs, since accent patterns are one of the most conspicuous features that distinguish regional dialects in Japanese.

experimental conditions. Similarly, *oma'wari* and *omamori* are lexically accented and unaccented respectively, and their phonetic properties were different depending on the condition.

(1) a. $[aa]_{aa}$ Yama'nashi-no   oma'wari-ga   yameta.
                     Yamanashi-Gen   police officer-Nom   quit-past
                     'The police officer from Yamanashi quit.'

     b. $[uu]_{aa}$ Yama'nashi-no   oma'wari-ga   yameta.

     c. $[aa]_{uu}$ Yamanote-no   omamori-ga   kieta.
                     Yamanote-Gen   charm-Nom   quit-past
                     'The charm of Yamanashi disappeared.'

     d. $[uu]_{uu}$ Yamanote-no   omamori-ga   kieta.

     e. $[au]_{uu}$ Yamanote-no   omamori-ga   kieta.

     (An underlined word indicates that it is phonetically accented; an apostrophe indicates that the word is lexically accented.)

In sentence (1a), N1 and N2 both had the phonetic F0 contours of an accented word whereas in (1b) N1 and N2 had those of an unaccented word, keeping the segmental string identical to the words in (1a). In other words, the acoustic properties of N1 and N2 were consistent with their structural properties in (1a) but were not in (1b); in (1b) these two types of information conflicted with each other.

Let us turn to the sentences in (1c)-(1e). The sentence in (1d) was a natural rendition of the words *Yamanote* and *oma'wari*, so they did not have any conflict between their acoustic and the structural properties of lexical accent. Just like the sentence in (1b), the acoustic properties were manipulated in (1c) and (1e). In (1c), N1 and N2 were both phonetically accented but phonologically unaccented. In (1e), only N1 had the phonetic F0 property that conflict with its phonological property (phonetically *a* but phonologically *u*). Both kinds of

properties were matched in N2 in (1e).

The recorded sentences were stylized and resynthesized in terms of F0 peak height, using the PSOLA algorithm in the *Praat* program. The basic procedure was identical to that adopted in the earlier experiments. The F0 peak height (the H* tones for the accented words; the phrasal H tones for the unaccented ones) was varied, in 5 steps for N1 (first peak, P1) and in 6 steps for N2 (second peak, P2). The step size was 1.0 st for P1 and 1.4 st for P2 with the base frequency of 100 Hz. The exact F0 peak values used in the experiment are shown in Table 6.1. The other tonal values were fixed (see Appendix for those values). The actual F0 contours used in the experiment are shown in Figure 6.2. Overall, the number of stimuli was 150 (5 P1 heights × 6 P2 heights × 5 sentence types).

Table a — [aa]_aa & [aa]_uu

[aa]$_{aa}$ & [aa]$_{uu}$	P2							
P1	102	111	120	130	141	153	166	180
155							░	░
164							░	░
174	░							░
184	░	░						
195	░	░						

a

Table b — [uu]_uu & [uu]_aa

[uu]$_{uu}$ & [uu]$_{aa}$	P2							
P1	105 (97)	113 (105)	123 (114)	133 (124)	145 (134)	157 (145)	170 (158)	184 (171)
133 (121)						░	░	░
141 (128)							░	░
149 (136)	░							░
158 (144)	░	░						
167 (153)	░	░						

b

Table c — [au]_uu

[au]$_{uu}$	P2							
P1	105 (97)	113 (105)	123 (114)	133 (124)	145 (134)	157 (145)	170 (158)	184 (171)
155						░	░	░
164							░	░
174							░	░
184								░
195	░							

c

Table 6.1 F0 values of the peak of N1 (P1) and N2 (P2) in Experiment 4. The cells that are shaded indicate that they were not used. The values in brackets in (b) and (c) are the ones for the "F0 shoulder", i.e. the spread phrasal H tone at the right edge of the unaccented words.

a

b

c
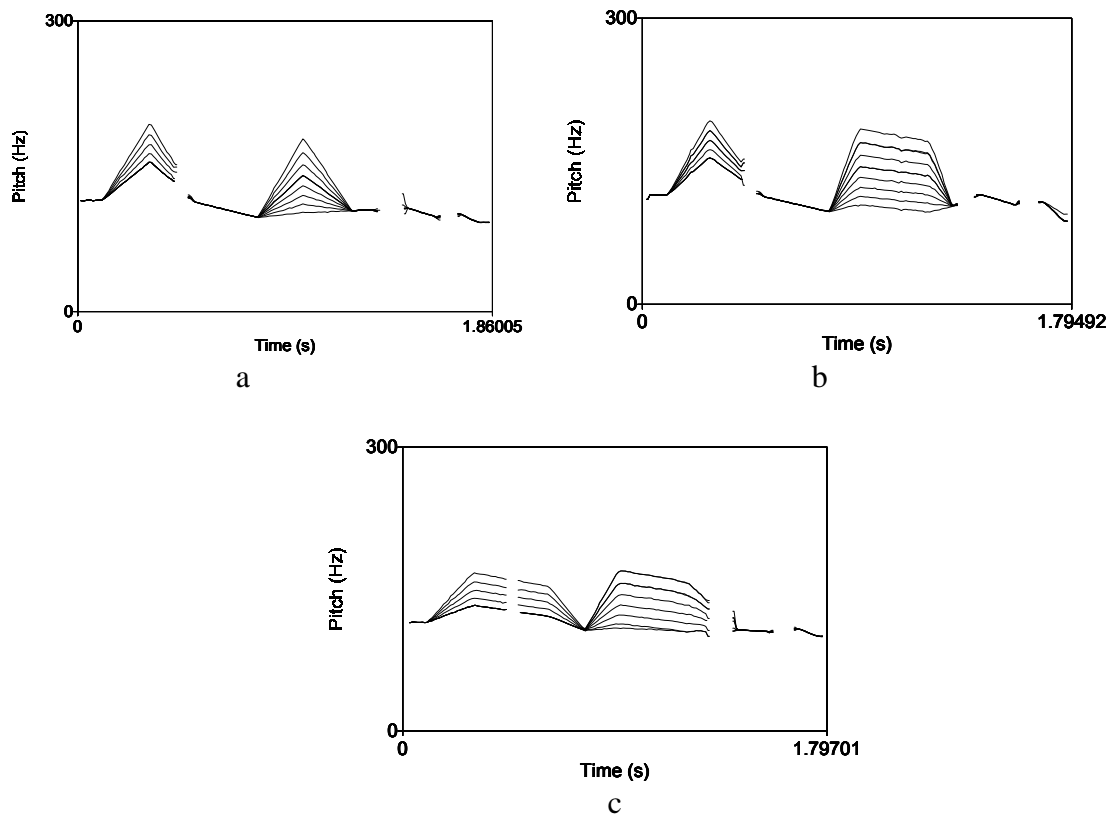
Figure 6.2 The F0 contours used in Experiment 4: (a) $[aa]_{aa}$ and $[aa]_{uu}$, (b) $[au]_{uu}$, (c) $[uu]_{aa}$ and $[uu]_{uu}$.

6.4.2 Listeners

Thirty undergraduate and graduate students at 4 different universities in Tokyo (Sophia University, Tokyo University of Agriculture and Technology, and Waseda University) participated in Experiment 4 (mean age 25.0). The majority of the listeners (21) were from Tokyo and its neighboring areas. The remaining 9 were from various other areas (Fukuoka, Gifu, Hokkaido, Hyogo, Ibaraki, Kagawa, Mie, Nagano, and Niigata). All listeners were paid for their participation.

6.4.3   Procedure and analysis

The general procedure of Experiment 4 was identical to that of the earlier experiments. The experimental sessions took place in a sound-attenuated room or a quiet classroom at the universities mentioned in the previous section. Listeners were tested in groups of two to five people. They each sat at a computer terminal facing a display. The stimulus sentences were randomly presented to the listeners, using the MFC in the *Praat* program. The listeners' task was identical to that adopted in the earlier experiments: to judge which of the two words (N1 or N2) was given more emphasis by the speaker.

Stimuli were presented more than once, but the number of repetitions differed depending on the stimulus: for each P1 value, the highest and the lowest P2 values had 2 repetitions, the second highest and the second lowest P2 values 4 repetitions, the remaining intermediate P2 heights had 6 repetitions. In total, there were 600 different tokens in the experiment.

The stimuli were divided into 8 blocks of trials, each of which contained 75 stimuli. The listeners were instructed to take a short break after every block of trials if they wanted. They had a practice session to be familiarized with the stimuli. The session lasted about 80 minutes including the instructions, the practice, the blocks of trials, and the breaks.

The procedure for the probit analysis was also identical to the earlier experiments. The response rates for P2 (P2 sounding more emphatic than P1) obtained for each listener were submitted to calculate the 50 % crossover point of the response rates.

6.4.4   Predictions

We tested two kinds of effects of the acoustic property of lexical accent on perceived

prominence: accentual boost normalization and downstep normalization. To study the effects of accentual boost normalization we can compare the two experimental conditions $[aa]_{uu}$ and $[au]_{uu}$. Because the difference between them is only in their phonetic F0 contours, this comparison allows us to see if there is any effect of the acoustic property of lexical accent on prominence perception. Recall that in Experiment 1 we compared the *aa* and the *au* sequences, where their phonetic and phonological information were matched (following the notation here, they would be $[aa]_{aa}$ and $[au]_{au}$) in order to see the effect of accentual boost normalization. We had the result that for a given P1 F0 height P2 needs to be lower in $[au]_{au}$ than in $[aa]_{aa}$ in order to perceive P1 and P2 as sounding equal in prominence. In other words, P1-P2 difference was greater in $[au]_{au}$ than in $[aa]_{aa}$. If the acoustic property of lexical accent plays a role in accentual boost normalization, the same kind of difference should be observed, as illustrated in the schematic graph in Figure 6.3, which represents the P2 values when they have the same perceived prominence as P1. This graph is the same type of graph we had in the earlier experiments. On the other hand, if the process of accentual boost normalization is triggered only by the structural property of lexical accent, there should be no difference observed between $[au]_{au}$ and $[aa]_{aa}$ in the value of P2 when the two peaks are perceived to be equally prominent.
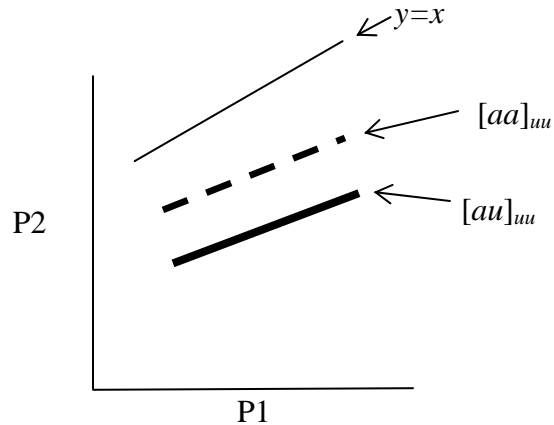
Figure 6.3 Prediction made by the phonetically-based accentual boost normalization. The graph represents P2 values when P1 and P2 have the same perceived prominence between $[aa]_{uu}$ and $[au]_{uu}$.

Next, let us consider phonetic effects of downstep normalization. If this perceptual process is triggered by the acoustic property of lexical accent, we should see differences between the conditions in which the words are phonetically *aa* ($[aa]_{aa}$ and $[aa]_{uu}$) and between the conditions in which the words are phonetically *uu* ($[uu]_{aa}$ and $[uu]_{uu}$). Recall that when assessing whether there is any perceptual normalization for downstep, it is necessary to compare the patterns of perceived prominence between *aa* with *uu*. Accentual boost normalization is not capable of accounting for any difference that may be observed in this pair of sequences since the two words in each pair share the lexical accent type. When the two words in a sequence are both accented, the accentual boost effects on them cancel out each other. Consequently, no difference is expected from accentual boost normalization between *aa* and *uu*. The downstep effect would be found if differences between P1 and P2 are larger in *aa* than in *uu*. This is because in downstep normalization the production effect of downstep on P2 is perceptually undone after an accented P1 and is lower in F0 when it is

preceded by an accented P1 than by an unaccented P1. Consequently, P1-P2 F0 difference is greater in *aa* than in *uu* when P1 and P2 are equal in perceived prominence. Therefore, if there is downstep normalization that is based on phonetic F0 contour information, [*aa*]*aa* should show a greater P1-P2 F0 difference than [*uu*]*aa* and [*aa*]*uu* than [*uu*]*uu* because, regardless of the structural property of lexical accent for these sequences, [*aa*]*aa* and [*aa*]*uu* are phonetically accented and [*uu*]*aa* and [*uu*]*uu* are phonetically unaccented.

The predictions depicted above are illustrated visually in Figure 6.4. Note that the these schematic graphs differ from the one in Figure 6.3 in that it represents the difference value between P1 and P2 (P2-P2) on the *y* axis rather than the P2 value. We could not directly compare [*aa*] and [*uu*] for the P2 value that has the same perceived prominence because P1 height differs between [*aa*] and [*uu*]. Taking the difference between P1 and P2 heights makes the P1 height irrelevant and makes it possible to compare [*aa*] and [*uu*].

Figure 6.4 Predictions on the phonetic effects of lexical accent on downstep normalization. P1-P2 values are represented against P1 values when P1 and P2 have the same perceived prominence.

The current experimental setup also allows us to test the phonological effects of lexical accent on prominence perception. The relevant comparisons are the ones of [*aa*]*aa* with [*aa*]*uu*

and of $[uu]_{aa}$ with $[uu]_{uu}$. The two members in each pair share phonetic F0 patterns but differ in phonological lexical accent status. Thus, if downstep normalization is crucially informed by the structural property of lexical accent, we should observe a difference between $[aa]_{aa}$ and $[aa]_{uu}$ and between $[uu]_{aa}$ and $[uu]_{uu}$. The way the two sequences differ in each pair will be such that the phonologically *aa* $([\ ]_{aa})$ condition yields larger P1-P2 differences than the phonologically *uu* $([\ ]_{uu})$ condition in each comparison. It is probably good to recall again here that the *aa-uu* comparison is necessary to see the pure downstep normalization effect as discussed in Experiment 1. If, on the other hand, the structural property of lexical accent does not play a role for downstep normalization, we should see no difference between $[aa]_{aa}$ and $[aa]_{uu}$ and between $[uu]_{aa}$ and $[uu]_{uu}$.

The predictions described in the previous paragraph are shown in the schematic graph in Figure 6.5 together with the predictions that were seen above about the phonetic effects of lexical accent on downstep normalization. As seen in the figure, if the predictions about the phonetic effects are right, we will observe substantial differences between $[aa]_{aa}$ and $[uu]_{aa}$ and between $[aa]_{uu}$ and $[uu]_{uu}$ in such a way that the phonetically $[aa]$ conditions show greater P1-P2 differences than the phonetically $[uu]$ conditions. On the other hand, if there is an effect of phonological lexical accent status on the perception of prominence, then we should see differences between the conditions whose difference is only in phonological lexical accent status: between $[aa]_{aa}$ and $[aa]_{uu}$ and between $[uu]_{aa}$ and $[uu]_{uu}$.
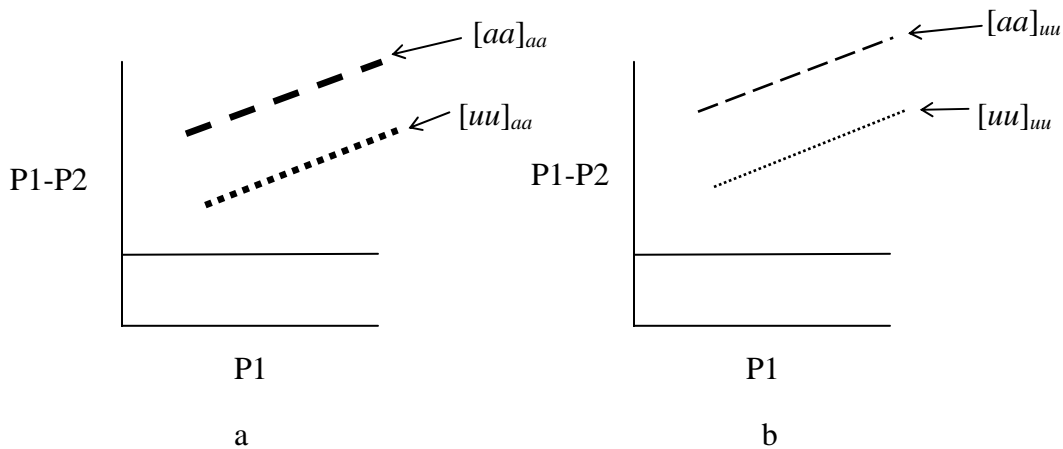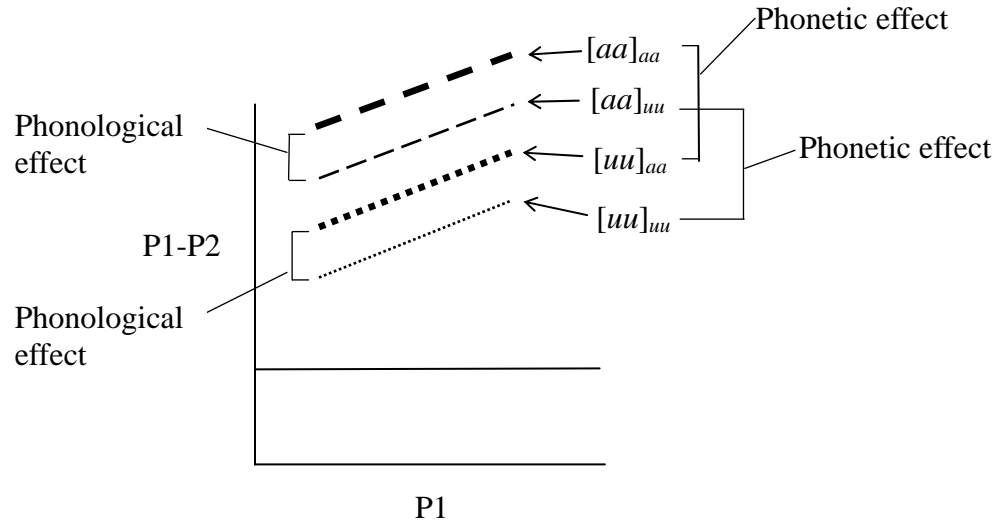
Figure 6.5 Predictions of the phonological and phonetic effects of lexical accent downstep normalization. P1-P2 values are represented against P1 values when P1 and P2 have the same perceived prominence.

6.5 Results

6.5.1 Phonetic effects on accentual boost normalization

Figure 6.6 plots the P2 values that sound equal in prominence to P1 for different P1 values in $[aa]_{uu}$ and $[au]_{uu}$. It corresponds to Figure 6.3, which shows the prediction about $[aa]_{uu}$ and $[au]_{uu}$.

Figure 6.6 shows that our prediction about the $[aa]_{uu}$–$[au]_{uu}$ comparison is borne out. We can see that all of the P2 values for $[au]_{uu}$ are lower than those for $[aa]_{uu}$. This is the same pattern as the one that was obtained in Experiment 1, which used naturally-rendered stimuli with no conflict between the acoustic and the structural properties of lexical accent. This result supports the idea that phonetically-driven accentual boost normalization can be triggered by F0 information only, without making reference to phonological lexical accent information.

137
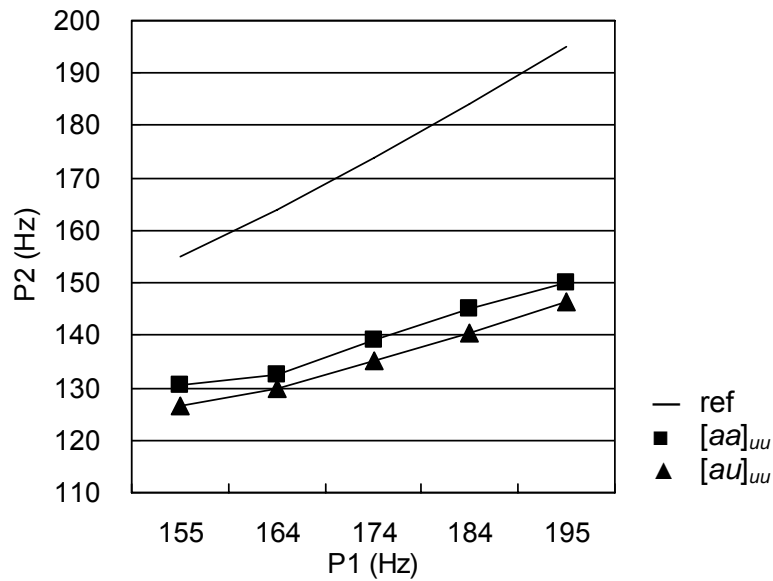
Figure 6.6 Mean values of P2 giving the same prominence as P1 as a function of P1 height between $[aa]_{uu}$ and $[au]_{uu}$. "Ref" is a shorthand for "reference", which represents the $y = x$ function.

However, the differences in P2 value observed between $[aa]_{uu}$ and $[au]_{uu}$ is relatively small. We need to check to see whether they are statistically reliable. A two-way repeated-measures ANOVA was conducted with sentence type ($[aa]_{uu}$ and $[au]_{uu}$) and P1 height (5 heights) as the independent variables. It revealed that there was a statistically significant main effect for both of sentence type and P1 height, and there was no interaction between sentence type and P1 height (Sentence type: $F(1, 29)=10.314$, $p=0.003$, P1 height: $F(4,116)=145.293$, $p<0.001$, Interaction: F(4,116)=0.510, $p=0.728$).

6.5.2 Phonetic effects on downstep normalization

In addition to the effect of the acoustic property of lexical accent on accentual boost normalization, this experiment also tested the effect of the acoustic property of lexical accent on downstep normalization. Figure 6.7 represents the differences in F0 between P1 and P2

138

(P1-P2) when the two peaks were perceived to have equivalent prominence for the two pairs of phonetically-identical but phonologically-different sequences of words ($[aa]_{aa}$–$[aa]_{uu}$ and $[uu]_{aa}$–$[uu]_{uu}$). This graph corresponds to Figure 6.5, where the predictions of the experiment on the phonetic and the phonological effects on downstep normalization were illustrated (note that the *y*-axis is now the P1-P2 values, which is different from Figures 6.3 and 6.6 where it is the P2 values).
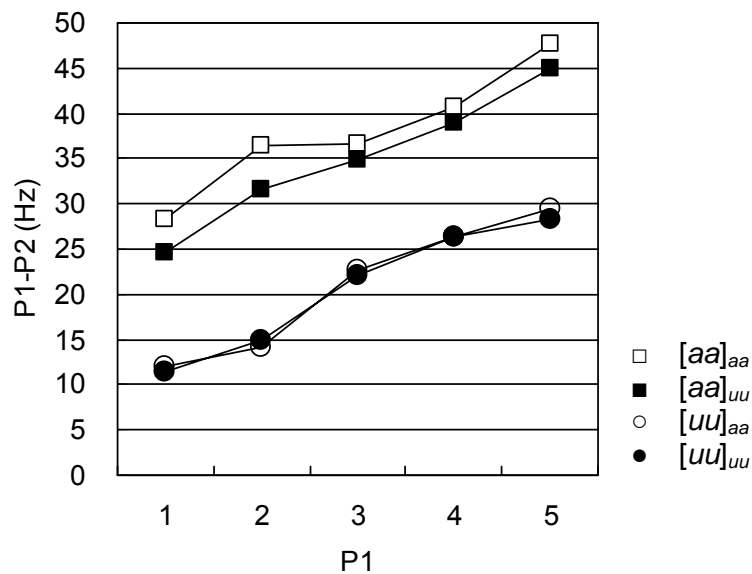
Figure 6.7 Mean P1-P2 values Mean *P1-P2* differences against P1 values for $[aa]_{aa}$, $[aa]_{uu}$, $[uu]_{aa}$, and $[uu]_{uu}$.

In the figure, the P1-P2 values are substantially greater in the two phonetically [*aa*] conditions ($[aa]_{aa}$ and $[aa]_{uu}$, filled and unfilled squares) than in the two phonetically [*uu*] conditions ($[uu]_{aa}$ and $[uu]_{uu}$, filled and unfilled circles). This difference provides evidence that the phonetic F0 properties of words play a role in downstep normalization. It suggests that the phonetically [*aa*] sequences trigger downstep normalization due to their phonetically accented pattern of the first words while the phonetically [*uu*] sequences do not involve such a perceptual process because they only contain words that are phonetically unaccented.

139

Consequently, when P1 and P2 are perceived to have the same perceived prominence, the P1-P2 difference is greater in $[aa]_{aa}$ and $[aa]_{uu}$ than in $[uu]_{aa}$.

6.5.3 Phonological effects on downstep normalization

We can also see a small but clear phonological effect of downstep normalization between the two phonetically *aa* conditions: $[aa]_{aa}$ (unfilled squares) exhibits greater P1-P2 values than $[aa]_{uu}$ (filled squares). Since these two sequences are phonetically identical, these differences must be attributed to the difference in the lexical accent status: the words in $[aa]_{aa}$ are phonologically accented whereas those in $[aa]_{uu}$ are phonologically unaccented. Given this phonological difference, it can be understood that the larger P1-P2 values in $[aa]_{aa}$ than $[aa]_{uu}$ are consistent with the pattern that indicates downstep normalization (*aa* > *uu* in P1-P2). This result replicates what we obtained in Experiment 3 in the sense that not only the acoustic property of lexical accent but also the structural property play a role in downstep normalization.

However, in spite of the interpretation we reached above, we observe another, rather puzzling, pattern which seems to contradict the pattern found between $[aa]_{aa}$ and $[aa]_{uu}$. In Figure 6.7, there is virtually no difference between the two phonetically *uu* conditions ($[uu]_{aa}$ and $[uu]_{uu}$, filled and unfilled circles). This leads us to have the question of why the effect of phonological lexical accent emerges only when the words were phonetically accented and not when they were phonetically unaccented. I discuss this asymmetry in the discussion section.

We observe yet another interesting asymmetry. The size of the normalization effect for downstep is much greater for the phonetically-based effect than for the phonologically-based effect. As seen in Figure 6.7, the size of the differences between $[aa]_{aa}$ (unfilled squares) and

[*uu*]*aa* (unfilled circles) is far greater than the size of the differences between [*aa*]*aa* (unfilled squares) and [*aa*]*uu* (filled squares). This suggests that the phonetic F0 contour influences much more than the phonological lexical accent information in downstep normalization and that downstep normalization is largely (but not entirely) based on the phonetic F0 contour properties. This in turn implies that when the effect of downstep normalization is seen in a sequence with no conflict between the acoustic and the structural properties of lexical accent, the phonetically-based compensation is responsible for most of its effect. In relation to Sugahara's (2003) proposal on downstep, the results of Experiment 4 lends support for her proposal because downstep normalization is largely phonetic but provides evidence against it because phonological accent status also plays a role in the perceptual process. Again, this point is further discussed in the discussion section below.

To statistically assess the patterns described above, a three-way repeated-measures ANOVA was performed with the phonological lexical accent type ([ ]*aa* and [ ]*uu*), the phonetic F0 pattern ([*aa*] and [*uu*]), and P1 height (5 heights) as the independent variables. The results are summarized in Table 6.2:

Phonetic F0 pattern	$F(1,29)=360.911, \ p<0.001*$
Phonological lexical accent	$F(1,29)=9.799, \ p=0.004*$
P1	$F(4,116)=176.186, p<0.001*$
Phonetic F0 pattern * Phonological lexical accent	$F(1,29)=5.605, \ p=0.025*$
Phonetic F0 pattern * P1	$F(4,116)=71.469, \ p<0.001*$
Phonological lexical accent * P1	$F(4,116)=0.667, \ p=0.616$
Phonetic F0 pattern * Phonological lexical accent * P1	$F(4,116)=1.620 \ p=0.174$

Table 6.2 Results of ANOVA. Significant result is indicated by an asterisk.

There was a main effect for all of the three the independent variables. The significant effects of phonetic F0 pattern and phonological lexical accent suggest that the phonetic F0 contour

properties as well as the phonological lexical accent status play an important role in the compensation process for downstep. As for the interaction, the significant interaction between phonetic F0 pattern and phonological lexical accent indicates that the two phonological conditions ([ ]$_{aa}$ and [ ]$_{uu}$) differ between the two phonetic condition. In other words, it reflects the asymmetry in which the phonological difference between [ ]$_{aa}$ and [ ]$_{uu}$ was substantial the two words involved were phonetically *aa* but was gone when they were phonetically *uu*. This asymmetry was statistically confirmed by a separate repeated-measures ANOVA with only [*aa*]$_{aa}$ and [*aa*]$_{uu}$ and a yet separate ANOVA with [*uu*]$_{aa}$ and [*uu*]$_{aa}$. They showed that the difference was significant only between [*aa*]$_{aa}$ and [*uu*]$_{aa}$ but not between [*uu*]$_{aa}$ and [*uu*]$_{aa}$ ([*aa*]$_{aa}$–[*uu*]$_{aa}$: $F(1,29)=13.489$, $p=0.001$, [*uu*]$_{aa}$–[*uu*]$_{aa}$: $F(1,29)=0.192$, $p=0.665$).[3] The other significant interaction, the one between Phonetic F0 pattern and P1, shows that the size of the phonetic differences between the two phonetically [*aa*] conditions and the two phonetically [*uu*] conditions (the large separation between the squares and the circles) differs depending on P1 height. A further examination revealed that the difference was larger at the second and the fifth P1 heights than at the other heights.[4]

Summarizing the results, we have found that both the acoustic and the structural properties of lexical accent are used to trigger downstep normalization. However, two asymmetries are observed. One is that the size of the phonetic effect was much greater then that of the phonological effect. This result suggests that downstep normalization is largely based on the phonetic F0 properties of words. The other asymmetry is that the phonological

---

[3] These separate ANOVAs were two-way ANOVAs, with phonological lexical accent and P1 height as the independent variables. However, only the results for phonological lexical accent are reported here, as those for the other variable are not of our interest here.

[4] There does not seem to be a reasonable explanation for these differences, since they do not pattern with P1 heights in a principled way. My best guess is that these differences are accidental.

effect was seen only when the target words were phonetically accented.

We also found that there is accentual boost normalization that is based purely on the acoustic properties of lexical accent. This phonetic effect was robust but its size was much smaller than the phonetic effect of downstep normalization.

6.6   Discussion

6.6.1   Nature of downstep normalization

Experiment 4 revealed two types of phonetic effects of lexical accent on the perception of intonational prominence in Japanese. It showed that accentual boost normalization and downstep normalization can both be triggered by phonetic F0 contour information alone. Moreover, the experiment tested the effect of the structural property of lexical accent about downstep normalization. It demonstrated that the structural property plays a significant role in downstep normalization. One reasonable conclusion drawn from these results about downstep normalization is that the acoustic and the structural properties of lexical accent are both important to the perceptual normalization process.

However, Experiment 4 revealed that the effect of downstep normalization reflects the word's acoustic property of lexical accent far more than its structural property. The differences between P1 and P2 were greater in *aa* than in *uu* not only when the words' accent types are signaled phonetically ([*aa*] > [*uu*]) but also when they are signaled phonologically ([   ]$_{aa}$ > [   ]$_{uu}$), which itself can be interpreted as the evidence for the phonological and the phonetic effects on downstep normalization. However, the effect size was much larger for the phonetic effect than for the phonological effect.

These findings have two implications for the analysis of downstep in Japanese by

143

Sugahara (2003). Based on her production experiment, she proposes a model in which downstep is characterized as an entirely phonetic process. However, Experiment 4 showed that the phonological lexical accent status also affects the perceptual normalization for downstep, which is incompatible with her purely phonetic view if we assume that the production and the perception of downstep are governed by the same system. The experimental results suggest that at least in perception listeners do make reference to phonological lexical accent status in assessing the relative perceived prominence of words.

At the same time, Experiment 4 showed a result that supports Sugahara's model of downstep. It showed an asymmetry in which the phonetic effect of downstep normalization is far greater than its phonological effect. This result indicates that the effect of downstep normalization can largely be explained by the physical F0 properties of the experimental words. In Figure 6.7, the differences between phonetically [*aa*] and [*uu*] conditions (squares and circles) are quite large across all of the P1 heights whereas the differences between phonologically [   ]*aa* and [   ]*uu* conditions (filled and unfilled symbols) are solid but much smaller than between the phonetic conditions. (I discuss the absence of differences between [*uu*]*aa* and [*uu*]*uu* later.) Therefore, if we assume that a single mechanism is responsible for both of the production and the perception of downstep in Japanese, the appropriate characterization should be that it is both the phonetic and phonological accent properties that are responsible for downstep in Japanese, but the phonetic property plays a greater role.

6.6.2   On the small phonetic effect of accentual boost normalization in [*aa*]*uu* vs. [*au*]*uu*

Experiment 4 showed the phonetic effect of accentual boost normalization but the effect size was rather small. It is usually the case that phonetic or bottom-up effects take precedence

over top-down or phonological effects in speech perception and that the former type of effects is greater and more robust than the latter type of effects. For example, Experiment 3a, in which we looked at the lexicality effect in accent type identification, revealed a robust phonetic effect: listeners' responses were invariably "accented" regardless of the word's structural property of lexical accent when its F0 contour showed an accented pattern and "unaccented" when the F0 contour shape showed an unaccented pattern, whether the word itself was *Ina'mori* (inherently accented) or *Inamura* (inherently unaccented).

Given this phonetic dominance, the small phonetic effect of accentual boost normalization is atypical. By way of the exploration of the reason for it, let us compare this result in Experiment 4 with the relevant result found in Experiment 1. In Experiment 1, we found a robust effect of accentual boost normalization between *aa* and *au* using naturally-rendered stimuli where the acoustic and the structural properties of lexical accent are matched with one another. The P1-P2 differences in the *aa* and the *au* condition in Experiment 1 are shown in Figure 6.8a. We can see that *aa* shows much greater P1-P2 differences than *au* by approximately 20 to 30 Hz.

With this result of Experiment 1 on mind, let us look at the P1-P2 differences observed between $[aa]_{uu}$ and $[au]_{uu}$ in Experiment 4, which is shown in Figure 6.8b. What we notice is that the mean P1-P2 differences for the $[aa]_{uu}$ sequence in Experiment 4 are quite comparable with the *aa* sequence in Experiment 1, exhibiting the values between 25 Hz and 40 or 45 Hz. However, a comparison between the $[au]_{uu}$ sequence and the *au* sequence reveals that the former shows much greater P1-P2 values than the latter. This suggests that the $[au]_{uu}$ sequence may be the source of the small phonetic effect of accentual boost normalization.
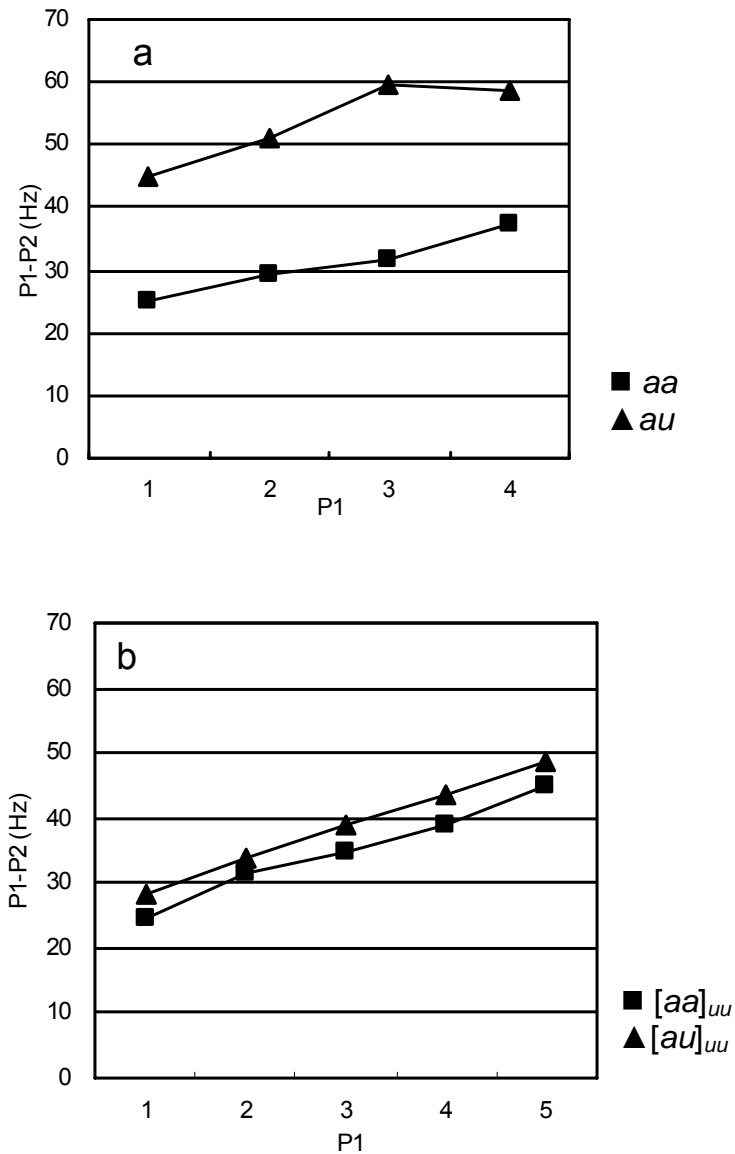
Figure 6.8 Comparison of the mean P1-P2 differences observed between *aa* and *au* in Experiment 4 with those in Experiment 1. (a): Experiment 1, (b) Experiment 4.

At this point, I cannot spot how exactly the [*au*]$_{uu}$ sequence brought the small phonetic effect. However, I want to point out one difference between [*au*]$_{uu}$ and the other experimental sequences that contain words whose acoustic and structural properties of lexical accent are compatible with one another, which can be the starting point for further research. Among the sequences used in Experiment 4, [*aa*]$_{uu}$, [*uu*]$_{aa}$ and [*au*]$_{uu}$ involve mismatch between the

acoustic and the structural properties of lexical accent. But such mismatch is seen both of the two words uniformly in $[aa]_{uu}$ and $[uu]_{aa}$ while it is seen only in one of the two words in $[au]_{uu}$. Note that the robust phonetic effect of downstep normalization was found in Experiment 4 between sequences that involve mismatch between the acoustic and the structural properties of lexical accent, namely between $[aa]_{aa}$ and $[uu]_{aa}$ and between $[aa]_{uu}$ and $[uu]_{uu}$. So the mismatch of accent information itself is not crucial for the small phonetic effect of accentual normalization. Instead, I consider that the mixture of match and mismatch within a sequence may have introduced some confounding factor to Experiment 4. Further investigation is necessary on this point.

6.6.3 Asymmetry between $[aa]_{aa} - [aa]_{uu}$ and $[uu]_{aa} - [uu]_{uu}$

Another interesting asymmetry found in Experiment 4 is that no difference was found in P1-P2 value between $[uu]_{aa} - [uu]_{uu}$ while a small but statistically reliable differences was found between $[aa]_{aa} - [aa]_{uu}$. These two comparisons were intended to test the effect of the structural property of lexical accent on the perceptual normalization for downstep. The way the two members differed from one another in each pair was identical between the two comparisons: one member was phonologically specified as *aa* and the other member as *uu*. The question we have to ask is why the effect of the structural property of lexical accent was only observed between the pair in which the words were phonetically accented ($[aa]_{aa}-[aa]_{uu}$) and not between the one in which they were phonetically unaccented ($[uu]_{aa}-[uu]_{uu}$).

I suggest one possible account, which is based on the naturalness or unmarkedness of the historical change of lexical accent. A well-known historical change of lexical accent in

Japanese is *deaccenting* (cf. Kobayashi 2002). In this process, words that are lexically accented lose their accents over time. For example, the word *denwa* 'telephone' used to be pronounced as an accented word with a pitch accent on the first mora (*de'nwa*), but currently no Tokyo Japanese speakers pronounce the word that way: it is invariably unaccented with no variation. Also, we can find many words that are in the middle of their transitions: for example, *ongaku* 'music' has two variants – *o'ngaku* (accented) and *ongaku* (unaccented).

Crucially, the direction of the accent-type change is from accented to unaccented and not the other way round. If it is hypothesized that this direction of the accent change, namely deaccenting is more "natural" than the other direction, the accent-type change from the accented to the unaccented type may have less perceptual impact on listeners' ears and hence more tolerable than the change in the opposite direction. Consequently, if a word which is lexically accented is pronounced as if it was unaccented, they would consider the unaccented pattern to be permissible as a possible variant of the word. Hence, the discrepancy between the acoustic and the structural properties of lexical accent in $[uu]_{aa}$ may not have been perceived as anomalous as those in $[aa]_{uu.}$

6.7   Chapter Summary

This chapter reported on Experiment 4, which tested the effects of the acoustic property of lexical accent on accentual boost normalization and downstep normalization. The experiment has shown a phonetic effect on both of the perceptual processes. Accentual boost normalization and downstep normalization can both be triggered by purely phonetic cues.

Experiment 4 has also revealed an important relationship between the phonetic and the phonological effects of lexical accent on downstep normalization. It has shown that both the

acoustic and the structural properties of lexical accent play a role but the former is considerably larger than the latter. That is, the effect of downstep normalization can largely be accounted for by the phonetic F0 contour properties of words.

Based on these findings, I argued against Sugahara's (2003) model on Japanese downstep because it cannot capture the phonological effect obtained in the experiment. However, at the same time, her view is in accordance with the results in that to a large extent the effect of downstep normalization can be accounted for by the phonetic F0 contour shape.

Experiment 4 has shown two other results. One is that the phonetic effect of accentual boost normalization is relatively small, which is a puzzle given that phonetic effects are usually greater and more robust than phonological effects in speech perception. Clearly further research is needed on this point. The other result is the asymmetry in which there was a small but statistically reliable phonological effect of downstep between $[aa]_{aa}$ and $[aa]_{uu}$ but there was no effect between and $[uu]_{aa}$ and $[uu]_{uu}$. I suggested one possible account of this asymmetry based on the deaccenting process, a diachronic process where accented words lose their accents.

CHAPTER 7

GENERAL DISCUSSION AND CONCLUSION

7.1   Introduction

Through the four perceptual experiments, this dissertation investigated how Japanese lexical accent interacts with the perceived prominence of words in an utterance. The main interest of this study was in two distinctive perceptual normalization processes on the assessment of prominence, namely *accentual boost normalization* and *downstep normalization*. We considered that the former perceptually undoes the inherently higher peaks (or inherently greater F0 excursion size) of accented words by lowering their perceived F0 height, and the latter perceptually undoes downstepped F0 peaks occurring after accented words by raising their perceived F0 height. The four experiments together showed that the two accented-based perceptual normalization processes play an important role in the perception of perceived prominence in Japanese.

In the remainder of this chapter, I first summarize the main findings for each of the four experiments (§7.2). I then propose a model that explains how accentual boost normalization and downstep normalization work in speech perception as a picture that emerges on the basis of the findings (§7.3). Finally, I provide a few suggestions for future research.

7.2   Summary of the findings

Experiment 1 tried to determine whether there is any perceptual normalization effect on perceived prominence that is crucially dependent on lexical accent in Japanese. The stimuli

were naturally-produced short sentences that contained three words. They contained both the structural and the acoustic property of lexical accent, which were available to the listeners. Two main results were obtained from two comparisons: (1) $[au]_{au}$ vs. $[aa]_{aa}$ and (2) $[aa]_{aa}$ vs. $[uu]_{uu}$. In $[au]_{au}$ vs. $[aa]_{aa}$, the P1-P2 differences were greater in $[au]_{au}$ than $[aa]_{aa}$ and in $[aa]_{aa}$ vs. $[uu]_{uu}$ they were greater in $[aa]_{aa}$ than in $[uu]_{uu}$.[1]

These results suggested that accentual boost normalization and downstep normalization both play a role in the assessment of perceived prominence. I characterized these normalization processes as a process that nullifies the corresponding process in production. Specifically, accentual boost normalization perceptually nullifies the extra prominence assigned to an accented word in production. Downstep normalization perceptually nullifies the reduction of prominence that occurs in production to a word after an accented word.

Accentual boost normalization can nicely account for the first experimental result observed between $[au]_{au}$ and $[aa]_{aa}$. The account is based on the fact that accentual boost normalization occurs only in the first word in $[au]_{au}$ but it occurs in both of the words in $[aa]_{aa}$. As a result, accentual boost normalization that appears in the two words cancel each other and no accentual boost effect is observable in $[aa]_{aa}$. On the other hand, the effect of accentual boost normalization is clearly observable in $[au]_{au}$. The effect is observed in such a way that for a given F0 height of P2, P1 is higher when P1 is accented when it is unaccented, because when P1 is accented it needs to be raised against the accentual boost effect. Consequently P1-P2 difference is greater in $[au]_{au}$ than in $[aa]_{aa}$.

Downstep normalization is responsible for the second result observed between $[aa]_{aa}$

---

[1] Note that the same notation as the one used in Experiment 4 is used here, where letters in square brackets ($[xx]$) indicates acoustic the acoustic property of lexical accent   and subscripted letters ([   ]$_{xx}$) the structural property.

and $[uu]_{uu}$. Since the first word is accented, downstep is expected in $[aa]_{aa}$ but not in $[uu]_{uu}$. For a given F0 height of P1, P2 needs to be lower in F0 in $[aa]_{aa}$ than in $[uu]_{uu}$ because downstep normalization is applied only to the second word in $[aa]_{aa}$, which makes it sound more prominent than a word after an unaccented word. As a result, P1-P2 difference is greater in $[aa]_{aa}$ than in $[uu]_{uu}$ when P1 and P2 are perceived to have the same prominence.

Experiment 2 was a follow-up for accentual boost normalization in Experiment 1 in that it tested what I call the "slope-prominence" hypothesis, which can equally account for the results without recourse to the idea of accentual boost normalization. Based on Knight's (2003) finding, we developed the hypothesis that says that the perceived prominence of a word increases as the declining slope of its F0 contour decreases. The prediction made by the slope-prominence hypothesis was that because unaccented words have smaller (shallower) F0 slopes than accented words, they should evoke more perceived prominence than accented words if their peak F0 heights are the same. The theory behind this hypothesis rests on the nature of the auditory system: the auditory system has difficulty in phrase-locking the top frequency when it lasts only momentarily, while it is much easier to do so when the declining slope is small constituting an F0 "shoulder" or "plateau." The results of Experiment 2 support the slope-prominence hypothesis, but the effect size was too small to account for the large effect of accentual boost normalization found in Experiment 1. Experiment 2 showed that unaccented words with F0 plateaux evoke more perceived prominence than those with F0 shoulders, which in turn yielded more perceived prominence than unaccented words with linear interpolation downward. However, I argued that the slope-prominence hypothesis cannot be an alternative account for accentual boost normalization because the size of the F0 slope effect was too small to explain the accentual boost effect found between *aa* and *au* in

Experiment 1.

Experiment 3 examined the role of the structural property of lexical accent for prominence perception. It had two parts: Experiments 3a and 3b. Experiment 3a served as a preliminary experiment to Experiment 3b to check that the experiment stimuli desirably yield ambiguous percepts with respect to lexical accent type. At the same time, it investigated whether the lexicality effect (Ganong 1980) is observed in a prosodic-level phenomenon like lexical accent. It also explored whether Kubozono and his colleagues' "unaccented law" is observed in the actual speech processing (Kubozono 1996, Kubozono & Fukui 2004, Kubozono & Ogawa 2004). This law says that the accent type of a loanword or an infrequent nonword tends to be unaccented if the word is a four-moras long, ends with a sequence of light syllables and does not contain epenthetic vowls. The stimuli were of the same type as those used in Experiment 1. However, the first words of the experimental sentences were manipulated to form acoustic continua from an accented to an unaccented pattern. Listeners identified the lexical accent type of the words. Results demonstrated the lexicality effect in the identification of lexical accent in Japanese. When the acoustic signal was ambiguous and did not have enough cues for the accent type of a word, listeners' response was biased toward what the word's structural property of lexical accent indicated. The results of Experiment 3a also suggested that the unaccented law has psychological reality. When the acoustically-ambiguous word was a nonword which satisfies the conditions to which the unaccented law applies, listeners' response was biased toward the unaccented pattern.

Experiment 3b tested whether listeners' structural property of lexical accent affects the assessment of perceived prominence when no acoustic property of lexical accent is available. The same stimuli as those in Experiment 3a were used. The crucial comparisons were the

153

pairs of sequences in which the only difference was in the lexical accent status of the first word with acoustically-ambiguous F0 shape, $[a/u\ a]_{aa}$ vs. $[a/u\ a]_{ua}$ and $[a/u\ u]_{au}$ vs. $[a/u\ u]_{uu}$ ("a/u" denotes that the word's F0 contour is ambiguous). Results showed that when P1 and P2 had the same perceived prominence, the P1-P2 difference in F0 was greater when the first word was lexically accented than when it was lexically unaccented. More specifically, P1-P2 difference was greater in $[a/u\ a]_{aa}$ than in $[a/u\ a]_{ua}$ and in $[a/u\ u]_{uu}$ than in $[a/u\ u]_{uu}$. These results were obtained crucially when the accent type of the first word was differentiated with respect to phonological lexical accent status alone. We interpreted the results as the phonological effect of accentual boost normalization as well as downstep normalization. Experiment 3b indicated that phonological knowledge influences the perception of intonational prominence.

Experiment 4 mainly tested the effects of the acoustic property of lexical accent on the proposed two perceptual normalization processes. It drew three sets of comparisons: (1) $[aa]_{aa}$ vs. $[au]_{aa}$, (2) $[aa]_{aa}$ vs. $[uu]_{aa}$ and $[aa]_{uu}$ vs. $[uu]_{uu}$, and (3) $[aa]_{aa}$ vs. $[aa]_{uu}$ and $[uu]_{aa}$ vs. $[uu]_{uu}$. The first set of comparisons (which actually consisted of a single comparison) was intended to test the phonetic effect of accentual boost normalization. the only difference between was in the acoustic property of the second word between a and u. $[aa]_{aa}$ differed from $[au]_{aa}$ only in the F0 contour shape of the first word. The experiment showed that the P1-P2 difference was larger in in $[au]_{aa}$ than in $[aa]_{aa}$ when P1 and P2 had the same perceived prominence, exhibiting the same pattern between $[aa]_{aa}$ and $[au]_{au}$ in Experiment 1. The second set of comparisons examined the role of phonetically-driven downstep normalization. In both of the pairs, the phonetically aa ($[aa]$) conditions needed larger P1-P2 differences than the phonetically uu ($[uu]$) conditions when P1 and P2 were perceived to

have the same perceived prominence, which suggests that downstep normalization is triggered by the acoustic property of lexical accent alone. The last set of comparisons was made in order to investigate whether phonologically-driven downstep normalization plays a role in prominence assessment. A disparity was found between $[aa]_{aa}$ vs. $[aa]_{uu}$ and $[uu]_{aa}$ vs. $[uu]_{uu}$: the effect of phonologically-driven downstep normalization was observed only in the pair $[aa]_{aa}$ vs. $[aa]_{uu}$. While a significant P1-P2 difference was found between $[aa]_{aa}$ and $[aa]_{uu}$ such that it was greater in $[aa]_{aa}$ than in $[aa]_{uu}$, no significant P1-P2 difference was detected between $[uu]_{aa}$ vs. $[uu]_{uu}$. I suggested that there might be a naturalness difference in the direction of accentuation change: listeners tolerate a change from an accented to an unaccented pattern more than the other way around because the former direction is in accordance with the historical change that has been seen in lexical accent in Japanese.

The experimental results described above are summarized in Table 7.1. Experiment 2 is excluded because it tested a hypothesis not directly relevant to the question of whether the acoustic and/or the structural properties of lexical accent affect perceived prominence. Rather, it was a follow-up experiment of Experiment 1, as mentioned above. In the table, two cells are collapsed into one for Experiments 1 and 3. In Experiment 1 it is not possible to tease apart the phonetic and the phonological effects for accentual boost normalization and for downstep normalization, and in Experiment 3 the phonological effect of accentual boost normalization and that of downstep normalization cannot be distinguished.[2] As can be seen in Table 7.1, the experiments did not discover the pure effect of the phonological accentual boost normalization, though it is highly likely that an experiment designed to test its effect would find the active role of phonologically-driven accentual boost normalization.

---

[2] Note in Table 7.1 that the order of the phonetic and phonological effects are reversed between accentual boost normalization and downstep normalization in order to express the indistinguishability of between the two effects

155

	Accentual boost normalization		Downstep normalization	
	Phonetic	Phonological	Phonological	Phonetic
Effect found in…	Experiment 1		Experiment 1	
		Experiment 3		
	Experiment 4		Experiment 4	Experiment 4

Table 7.1 Summary of Experiments 1, 3 and 4 in terms of the phonetic and the phonological effects of accentual boost normalization and downstep normalization.

7.3 The picture emerging from the findings

In this section, we consider what kind of picture emerges when we look at accentual boost normalization and downstep normalization in terms of speech processing. I propose that these perceptual normalization processes are phonologically-conditioned phonetic processes in the sense that they refer to the phonological status of lexical accent when they are applied, but the adjustment of perceived F0 height takes place at the phonetic level. The rationale for this comes from the following two points. First, Experiment 3 showed these perceptual normalization processes can make reference to the structural property of lexical accent of a word when no F0 cues signal its accent type. This indicates that the trigger of the perceptual processes is the phonological entity for lexical accent, namely the pitch accent H*. Second, we characterized accentual boost normalization and downstep normalization as non-symbolic processes in which the perceived F0 height of an F0 peak is boosted or reduced depending on the lexical accent information. Based on these two points, the most natural conceptualization of the perceptual normalization processes is that they are phonetically-conditioned phonetic processes.

I suggest that the specific level in speech processing at which accentual boost normalization and downstep normalization refer to lexical accent information corresponds to

the level that generative phonology models typically call the *surface phonological representation* (Chomsky & Halle 1968, Kenstowicz 1994) or that psycholinguistic models of speech perception call the *prelexical level* (Luce & Pisoni 1998, McCleland & Elman 1986, Moreton 2002, Norris 1994, Vitevich 1997, Vitevich & Luce 1998). The surface phonological representation, which is production-based, refers to the level at which discrete symbolic representations are encoded into gradient phonetic properties. The prelexical level is the level where the gradient acoustic signal is transformed into symbolic representations such as segments and features.

I argue that the normalization processes refer to the surface phonological representation and cannot directly refer to the lexical specification in the lexicon (that is, the underlying phonological representation). We found a robust downstep effect between $[aa]_{aa}$ and $[uu]_{aa}$ in Experiment 4, where they shared the words differing only in their phonetic F0 contour shape. Since $[aa]_{aa}$ and $[uu]_{aa}$ showed the same pattern as the one between $[aa]_{aa}$ and $[uu]_{uu}$ in Experiment 1, it is considered that listeners made their judgments based on the same representation. The representation cannot be the underlying phonological representation because the observed pattern between $[aa]_{aa}$ and $[uu]_{aa}$ crucially rests on the phonetic accent information of $[uu]_{aa}$, not on the phonological accent information. It is suggested that the listeners constructed phonological representations that correspond to the words' phonetic contour shape, which are distinguished from their underlying phonological representations.

The production-based work by Pierrehumbert & Beckman (1988) gives support for downstep normalization as a phonologically-conditioned phonetic process because this way of characterizing downstep normalization is consistent with their characterization of downstep. They propose that downstep in Japanese is a phonologically-conditioned phonetic

157

implementation rule such that the application of downstep is determined by the presence or absence of the pitch accent H* but the reduction of the tonal height itself takes place in the tonal space. Thus, downstep normazaliton and Pierrehumbert & Beckman's downstep are both phonologically-conditioned phonetic processes. If we assume that speech production and perception are governed by the same system, the idea of downstep normalization as a phonologically-conditioned phonetic perceptual process is the natural consequence that mirrors the production process of downstep.

The proposed characterization of accentual boost and downstep normalizations is graphically shown in Figure 7.1 below:
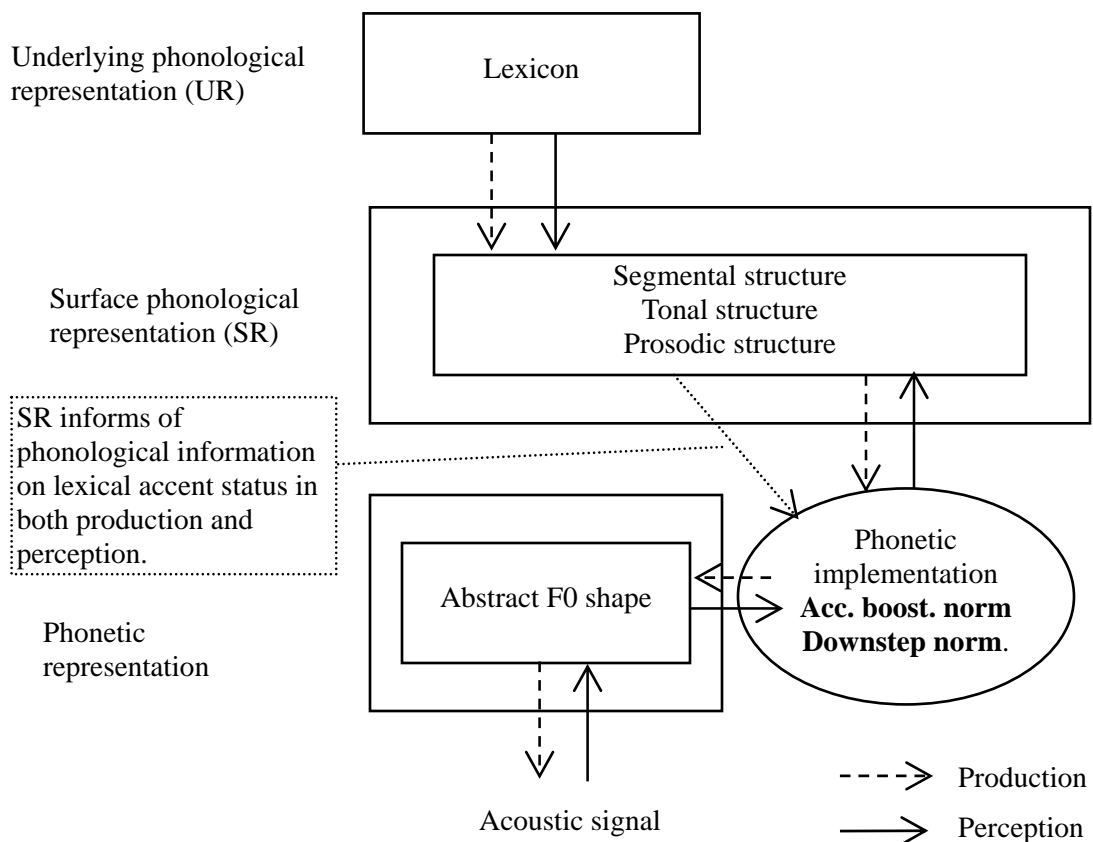
Figure 7.1 Conceptualization of accentual boost normalization and downstep normalization in speech perception.

158

The figure shows a model that captures speech production as well as speech perception. The flow of linguistic information in production and in perception is represented by different kinds of arrows. The surface phonological representation contains segmental, tonal and prosodic structures. As can be seen in the figure, accentual boost normalization and downstep normalization are processes in the phonetic implementation component. When they are triggered, they are informed of lexical accent information by the surface phonological representation. The direction of the arrow expressing the reference of lexical accent information is constantly one way both in production and in perception. This way of making reference to phonological lexical accent information is exactly the same as the downstep as a production process proposed by Pierrehumbert & Beckman (1988) in that it crucially refers to phonological tonal information when it is applied. With the assumption that that production and perception are governed by the same processing system, this would be an ideal situation.

7.4 Future directions

I conclude this dissertation with providing a few directions for future research. One line of research that can be started right away is to test the predictions that the model presented in the previous section makes. One of the predictions that is worth a serious investigation is based on the central proposal of the model: accentual boost normalization and downstep normalization are triggered by making reference to a lexical accent in the surface phonological representation. This proposal implies that the perceptual normalizations are triggered as long as the surface phonological representation is available. We would expect no effect if there was no surface phonological representation to refer to. This kind of situation is

readily conceivable. One such situations is when we hear non-speech like a sine wave that mimics the F0 contour shapes of the stimuli used in our experiments. If we assume that no surface phonological representation is created when haring non-speech stimuli, there should be no effect of accentual boost normalization or downstep normalization.

Another area that needs to be explored is the domain of the influence of relative prominence judgment. In this study, we only considered relative perceived prominence relation between MiPs. There may be perceived prominence relations of different kinds between prosodic constitutes that are larger than MiP. For example, the relative perceived prominence relation between MaPs may be different from that between MiPs because MaP is the level at which the syntactic as well as the focus structures of utterances are reflected (Pierrehumbert & Beckman 1988, Selkirk & Tateishi 1991). How these factors interact with lexical accent at the MaP level in terms of perceived prominence would be an interesting topic.

Downstep might also be relevant at the level of MaP. Kubozono (2005) suggested that the domain of downstep is larger than MiP. He tested whether focus blocks downstep using the *wh*-word *nani* 'what'. A *wh*-word is inherently focused and is normally considered to introduce a MaP boundary (Pierrehumbert & Beckman 1988). He found that the F0 peak of the *wh*-word was consistently lower after an accented word than after an unaccented word, which suggests that focus does not block downstep. Given this result, whether downstep normalization is seen at the MaP level is worth an exploration.

Speech prosody is one of the research fields that interact with phenomena that belong to many different levels. This dissertation examined one of such interplays, the one between lexical accent on the one hand as a property of the grammar and perceived prominence on the

160

other as a property of speech perception. This work can thus be seen as a contribution to the research of the interplay between phonology and speech perception (Hume & Johnson 1999). I hope this study has shed some light on an aspect of such interplay.

BIBLIOGRAPHY

Archangeli, D. & Pulleyblank, D. (1986) The content and structure of phonological representations. Ms. University of Arizona and University of Southern California.

Arvaniti, A. & Godjevac, S. (2003) The origins and scope of final lowering in English and Greek. *Proceedings of the 15th International Congress of Phonetic Sciences*, 1077-1080.

Beckman, M. & Pierrehumbert, J. (1986) Intonational structure in Japanese and English. *Phonology Yearbook, 3*, 255-309.

Boersma, P. & Weenink, D. (1992-2009) *Praat: Doing phonetics by computer* (Version 4.3.02) [Computer program]. Retrieved from http://www.praat.org/

Chomsky, N. & Halle, M. (1968) *The Sound Pattern of English.* Cambridge, MA: The MIT Press.

Cohn, A. (1993) Nasalization in English: phonology or phonetics. *Phonology*, 10, 43-81.

Cohen, A., Collier, R., & 't Hart, J. (1982) Declination: Construct or intrinsic feature of speech pitch? *Phonetica, 39*, 254–273.

Connine, C., Clifton, C.Jr., & Cutler, A. (1987) Effects of Lexical Stress on Phonetic Categorization. *Phonetica, 44*, 133-146.

Cutler, A. (1976) Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception and Psychophysics*, *20*, 55-60.

Cutler, A., & Darwin, C. J. (1981) Phoneme-monitoring reaction time and preceding prosody: effects of stop closure duration and of fundamental frequency. *Perception & Psychophysics, 29*, 217-224.

Cutler, A. & Fodor, J.A. (1979) Semantic focus and sentence comprehension. *Cognition*, *7*, 49–59.

Cutler, A. & Foss, D.J. (1977) On the role of sentence stress on sentence processing. *Language and Speech*, *20*, 1-10.

Cutler, A., Dahan, D. & van Donselaar, D. (1997) Prosody in the comprehension of spoken language: A literature review. *Language and Speech, 40*, 141-201.

Donselaar, W. & Cutler, A. (1997) Exploration of stress information in spoken-word recognition in Dutch. *Journal of Acoustical Society of America, 102*, 3136.

Féry, C. & Truckenbrodt, H. (2005) Sister and tonal scaling. *Studia Linguistica, 59*, 223-243.

Fujisaki, H. & Hirose, K. (1984) Analysis of vice fundamental frequency contours for declarative sentences of Japanese. *Journal of Acoustical Society of Japan*, *5*, 233-242.

Gelbart, B. (2005) *Perception of Foreignness*. Doctoral dissertation, University of Massachusetts Amherst.

Ganong, W.F. (1980) Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception & Performance, 6*, 110-125.

Gussenhoven C. & Rietveld, A.C.M. (1988) Fundamental frequency declination in Dutch: testing three hypotheses. *Journal of Phonetics, 16*, 335-369.

Gussenhoven, C., Repp, B.H., Rietveld, A.C.M., Rump, H.H. & Terken, J. (1997) The perceptual prominence of fundamental frequency peaks. *Journal of the Acoustical Society of America*, *102*, 3009-3021.

Gussenhoven, C. & Rietveld, T. (1998) On the speaker-dependence of the perceived prominence of F0 peaks. *Journal of Phonetics, 26*, 371-380.

Haraguchi, S. (1977) *The Tone Pattern of Japanese: An Autosegmental Theory of Tonology*. Tokyo: Kaitakusya.

Hasegawa, Y. (1999) Pitch accent and vowel devoicing in Japanese. *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, USA.

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

Hasegawa, Y. & Hata, K. (1996) The function of F0-peak delay in Japanese. *Proceedings of Berkeley Linguistic Society, 21*, 141-151.

Hayes, B. (1994) 'Gesture' in prosody: Comments on the paper by Ladd. In P. Keating (ed.) *Phonological Structure and Phonetic Form: Papers in laboratory Phonology III*. Cambridge, UK: Cambridge University Press, 64-75.

Hornby, P.A. (1972) The psychological subject and predicate. *Cognitive Psychology, 3*, 632-642.

Hume, E. & Johnson, K. (eds.) (1999) *The Role of Speech Perception in Phonology*. San Diego, CA: Academic Press.

Inkelas, S. & Leben, W. (1990) Where phonology and phonetics intersect: the case of Hausa intonation. In M.E. Beckman & J. Kingston (eds.) *Papers in Laboratory Phonology I: Between grammar and physics of speech.* Cambridge, UK: Cambridge University Press, 17-34.

Ito, J. & Mester, A. (1995) Japanese Phonology. In J. Goldsmith (ed.) *The Handbook of Phonological Theory.* Malden, MA: Blackwell Publishers, 817-838.

Ito, J. & Mester, A. (1999) The Structure of the Phonological Lexicon. In N. Tsujimura (ed.) *The Handbook of Japanese Linguistics.* Malden, MA: Blackwell Publishers, 62-100.

Kawahara, S. & Shinya, T. (2005) The intonation of gapping and coordination in Japanese: Evidence for Intonational Phrase. *University of Massachusetts Occasional Papers in Linguistics 30: Papers in Prosody*, GLSA, 91-127.

Katz, J. & Selkirk, E.O. (2005) *Two Types of Focus, Two Types of Stress Prominence.* ms. University of Massachusetts Amherst.

Keating, P. (1988) The phonology-phonetic interface. In F.J. Newmeyer (ed.) *Linguistic Theory: Foundations.* Vol.1. Linguistics: The Cambridge survey. Cambridge: Cambridge University Press, 281-302.

Kenstowicz, M. (1994) *Phonology in Generative Grammar.* Cambridge, MA.: Blackwell Publishers.

Kingston, J. (2007) The phonetics-phonology interface. In P. de Lacy (ed.) *The Cambridge handbook of Phonology.* Cambridge: Cambridge University Press, 401-434.

Kitahara, M. (2001) *Category Structure and Function of Pitch Accent in Tokyo Japanese.* Doctoral Dissertation, Indiana University.

Kjelgaard, M.M. & Speer, S.R. (1999) Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language, 40*, 153-194.

Knight, R.A. 2003. *Peaks and plateaux: the production and perception of intonational high targets in English.* Doctoral dissertation, University of Cambridge.

Kobayashi, M. (2002) *A Quantitative Study of accent Change and Variation in Tokyo Japanese: Focusing on adjectives and nouns.* Doctoral dissertation, University of Pennsylvania.

Kubozono, H. (1993) *The Organization of Japanese Prosody.* Tokyo: Kuroshio Publishers.

Kubozono, H. (1996) Syllable and accent in Japanese: Evidence from loanword accentuation. *The Bulletin* 211 (The Phonetic Society of Japan), 71-82.

Kubozono, H. & Ogawa, S. (2004) Emergence of Unaccented Words in Japanese. *Proceedings of Speech prosody 2004*, 33-36.

Kubozono, H. & Fukui, M. (2004) Phonological structure and unaccented nouns in Tokyo and Osaka Japanese. Presented at the 14th Japanese/Korean Linguistics, Tucson, Arizona.

Kubozono, H. (2005) Focus and Intonation in Japanese: Does Focus Trigger Pitch Reset? In S. Ishihara, M. Schmitz, and A. Schwarz (eds.) *Interdisciplinary Studies on Information Structure*.

Ladd, D.R. (1983) Phonological features of intonational peaks. *Language, 59,* 721-759.

Ladd, D.R. (1984) Declination: a review and some hypotheses. *Phonology Yearbook, 1*, 53-74.

Ladd, D.R. (1986) Intonational phrasing: The case of recursive prosodic structure. *Phonology Yearbook, 3*, 311-340.

Ladd, D.R (1988) Declination 'reset' and the hierarchical organization of utterances. *Journal of the Acoustical Society of America, 84,* 530-544.

Ladd, D.R. (1990) Metrical representation of pitch register. In M.E. Beckman & J. Kingston (eds.) *Papers in Laboratory Phonology I: Between grammar and physics of speech.* Cambridge, UK: Cambridge University Press, 35-57.

Ladd, D.R. (1993a) In defense of a metrical theory of intonational downstep. In H. v.d. Hulst & K. Snider (eds.) *The Representation of Tonal Register.* Dordrecht: Foris, 109-132.

Ladd, D.R. (1993b) On the theoretical status of 'The Baseline'' in modeling intonation. *Language and Speech 36*, 435–451.

Ladd, D.R. (1994) Constraints on the gradient variability of pitch range, or, Pitch level 4 lives! In P. Keating (ed.) *Phonological Structure and Phonetic Form: Papers in laboratory Phonology III*. Cambridge, UK: Cambridge University Press, 43-63.

Ladd, D.R. (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.

Ladd, D.R., Verhoeven, J. and Jacobs, K. (1994) Influence of adjacent pitch accents on each other's perceived prominence. *Journal of Phonetics, 22*, 65-85.

Ladd, D.R. & Moreton, R. (1997) The perception of intonational emphasis: Continuous or categorical? *Journal of Phonetics, 25*, 313-342.

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven & N. Warner (dds.), *Laboratory phonology 7*. Berlin: Mouton, 637–676.

Lehiste, I. (1970) *Suprasegmentals.* Cambedige, MA: MIT Press.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*, 431-461.

Liberman, M. & Pierrehumbert, J.B. (1984) Intonational invariance under changes in pitch range and length. In M. Aronoff and R. Oehrle (eds.) *Language Sound and Structure*, Cambridge, MA: MIT Press, 157-233.

Luce, P. & Pisoni, D. (1998) Recognition of spoken words: The neighborhood activation model. *Ear and Hearing*, 19, 1-36.

Maeda, K. & Venditti, J.(1998) Phonetic investigation of boundary pitch movements in Japanese. *Proceedings of the International Conference on Spoken Language Processing*, 631-634.

Maeda, S. (1976) *A Characterization of American English intonation.* Doctoral Dissertation, MIT.

Maekawa, K., Kikuchi, H., Igarashi, Y. &Venditti, J. (2002) X-JToBI: An extended J_ToBI for spontaneous speech. *Proceedings 7th International Congress on Spoken Language Processing*, Denver, USA, 1545-1548.

Marslen-Wilson, W.D. (1987) Function and process in spoken word recognition. *Cognition, 25*, 71–102.

McCleland, J.L. & Elman, J.L. (1986) The TRACE model of speech perception. *Cognitive Psychology*, 18-1-86.

McQueen, J.M., Norris, D. & Cutler, A. (1999) The time course of lexical involvement in phonetic categorization, *Journal of the Acoustical Society of America, 105*, 1398.

McCawley, J.D. (1968) *The Phonological Component of Grammar of Japanese.* The Hague: Mouton.

Moore, B.C.J. (2003) *An Introduction to the Psychology of Hearing.* 5th edition. London: Academic Press.

Moreton, E. (2002) Phonological grammar in speech perception. Doctoral dissertation, University of Massachusetts, Amherst.

Moreton, E. & Amano, S. (1999) Phonotactics in the perception of Japanese vowel length: evidence for long-distance dependencies. Presented at Eurospeech 99, Budapest, Hungary.

Norris, D. (1994) Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52, 189-234.

O'Shaughnessy, D. (1979) Linguistic features in fundamental frequency patterns. *Journal of Phonetics*, 7, 119-145.

Pickles, J.O. (1988) *An Introduction to the Physiology of Hearing.* 2nd edition. London: Academic Press.

Pierrehumbert, J.B. (1979) The perception of fundamental frequency declination. *Journal of the Acoustical Society of America*, *66*, 363-369.

Pierrehumbert, J.B. (1980) *Phonology and Phonetics of English Intonation.* Doctoral dissertation, MIT.

Pierrehumbert, J.B. (1990) Phonological and phonetic representations. *Journal of Phonetics*, 18, 375-394.

Pierrehumbert, J.B. & Beckman, M.E. (1988) *Japanese Tone Structure.* Linguistic Inquiry Monograph 15, Cambridge, MA.: MIT Press.

Pike, K. (1948) *Tone Languages.* Ann Arbor: The University of Michigan Press.

Pitt, M. & Samuel, A.G. (1993) An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human perception and Performance, 19*, 699-725.

Poser, W.J. (1984) *The Phonetics and Phonology of Tone and Intonation in Japanese.* Doctoral Dissertation, MIT.

Price, P.J., Ostendorf, M., Shattuck-Hufnagel, S. & Fong, C. (1991) The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America, 90*, 2956–2970.

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

Selkirk, E.O. (1995) The prosodic structure of function words. In *Papers in Optimality Theory*, ed. Jill Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk, 439-70. Amherst, MA: GLSA Publications. Also in *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, ed. James L. Morgan and Katherine Demuth, 187-214: Lawrence Erlbaum Associates.

Selkirk, E.O. (2000) The interaction of constraints on prosodic phrasing. In Merle Horne (ed.) *Prosody: Theory and Experiment*. Dordrecht: Kluwer Academic Publishing, 231-262.

Selkirk, E.O. (2005) Comments on Intonational Phrasing in English. In S. Frota, M. M. Vigario and J. Freitas (eds.) *Prosodies: Selected papers from the Phonetics and Phonology in Iberia Conference, 2003*. Mouton de Gruyter.

Selkirk, E.O. & Tateishi, K. (1988) Constraints on minor phrase formation in Japanese, *Chicago Linguistic Society, 24*, 316-336.

Selkirk, E.O. & Tateishi, K. (1991) Syntax and downstep in Japanese. In C. Georgopoulos and R. Ishihara (eds.) *Interdisciplinary Approaches to Language: Essays in Honor of S.-Y. Kuroda.* Dordrecht: Kluwer Academic Publishers, 519-543

Shafer, A.J. (1997) *Prosodic Parsing: The Role of Prosody in Sentence Comprehension.* Doctoral dissertation, University of Massachusetts Amherst.

Shafer, A.J., Carlson, C. Clifton, C. Jr. & Frazier, L. (2000) Focus and the interpretation of pitch accent: Disambiguating embedded questions. *Language and Speech, 43*, 75-105.

Shinya, T. (2006) Lexical accent status affects perceived prominence of intonational peaks in Japanese. *Proceedings of Speech Prosody 2006*, 89-92.

Speer, S.R., Kjelgaard, M.M. & Dobroth, K.M. (1996) The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal of Psycholinguistic Research, 25*, 247-268.

Sugahara, M. (2003) *Downtrends and post-FOCUS intonation in Tokyo Japanese.* Doctoral Dissertation, University of Massachusetts Amherst.

Sugito, M. (1982) *Nihongo akusento no kenkyuu* (Studies on Japanese accent). Tokyo: Sanseido.

Terken, J. (1991) Fundamental frequency and perceived prominence of accented syllables. *Journal of the Acoustical Society of America, 89*, 1768-1776.

Terken, J. (1994) Fundamental frequency and perceived prominence of accented syllables. II. Nonfinal accents. *Journal of the Acoustical Society of America, 95*, 3662-3665.

Terken, J. & Hermes, D. (2000) The perception of prosodic prominence. In M. Horne (ed.) *Prosody: Theory and Experiment.* Dordrecht: Kluwer Academic Publishers, 89-127.

't Hart, J. (1981) Differential sensitivity to pitch distance, particularly in speech. *Journal of the Acoustical Society of America, 69*, 811-821.

't Hart, J., Collier, R., & Cohen, A. (1990) *A Perceptual Study of Intonation: An Experimental-Phonetic Approach to Speech Melody.* Cambridge, UK: Cambridge University Press.

Trager, G.L. & Smith, H.L. (1951) *An Outline of English Structure*. Norman, OK: Battenburg Press.

Truckenbrodt, H. (1995) *Phonological Phrases: Their Relation to Syntax, Focus, and Prominence*. Doctoral dissertation, MIT.

Truckenbrodt, H. (1999) On the relation between syntactic phrases and phonological phrase, *Linguistic Inquiry, 30*, 219-255.

Truckenbrodt, H. (2002) Upstep and embedded register levels. *Phonology, 19*, 77-120.

Truckenbrodt, H. & Féry, C. (2004) *More on hierarchical organization and tonal scaling.* Ms. University of Potsdam and University of Tübingen.

van den Berg, R., Gussenhoven, C., & Rietveld, T. (1992) Downstep in Dutch: implications for a model. In G. Docherty & R.D. Ladd (eds.) *Papers in Laboratory Phonology II: Gesture, segment and Prosody*. Cambridge, UK: Cambridge University Press, 335-358.

Vance, T.J. (1987) *An Introduction to Japanese Phonology.* State University of New York Press, Albany.

Vanio, M. & Järvikivi, J. (2006) Tonal features, intensity, and word order in the perception of prominence. *Journal of Phonetics, 34,* 319-342

Vitevich, M.S. (1997) *Phonotactics and Spoken Word Recognition.* Doctoral Dissertation, University at Buffalo.

Vitevich, M.S. & Luce, P.A. (1998) When words complete: Levels of processing I perception of spoken words. *Psychological Science*, 9, 325-329.

Warren, P. (1999) Prosody and language processing. In *Language Processing.* M.J. Pickering Ed, Hove: Psychology Press, 155-188.