DOWNTRENDS AND POST-FOCUS INTONATION IN TOKYO JAPANESE

A Dissertation Presented

by

MARIKO SUGAHARA

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 2003

Department of Linguistics

DOWNTRENDS AND POST-FOCUS INTONATION IN TOKYO JAPANESE

A Dissertation Presented

by

MARIKO SUGAHARA

Approved as to style and content by:

Elisabeth O. Selkirk, Co-Chair

John Kingston, Co-Chair

Joseph Pater, Member

Charles E. Clifton, Jr., Member

Elisabeth O. Selkirk, Department Head
Department of Linguistics

PREFACE

This version of my doctoral dissertation comes with some corrections of typographical errors and minor modifications added to the original dissertation submitted to University of Massachusetts at Amherst in 2003.

ACKNOWLEDGMENTS

Many people supported my dissertation project. First and foremost, I would like to thank the members of my committee: Lisa Selkirk (co-chair), John Kingston (co-chair), Chuck Clifton (member) and Joe Pater (member). Without their academic guidance, helpful feedback to my drafts, constant encouragement and patience, this dissertation would not exist.

Lisa Selkirk's contribution to this dissertation is enormous beyond description. It is Lisa who ignited my interest in prosody. Furthermore, most of the ideas in my thesis were developed in our weekly meetings since 1999. Discussions in those meetings always provided me insight and inspiration. She spent countless hours helping me think globally, organize arguments, and formalize empirical observations. Also, this dissertation was supported in part by National Science Foundation grant BCS0004038 <u>The Reflexes of Focus in Phonology</u> to Lisa. She generously gave me an opportunity to work as a research assistant for the grant project between September 2001 and May 2002. The experience as a research assistant enriched my understanding of Japanese intonation.

John Kingston also played a crucial role in my dissertation. He showed me the importance of scientific and deductive thinking. My thesis improved dramatically after I followed his instruction. Quite a few important claims and proposals made in this thesis originated from his comments on my earlier drafts. In addition, his experience with experimental design and statistical analysis helped me a great deal.

I am grateful to my UMass colleagues for their encouragement, in particular Anna Arregui, Maria Nella Carminati, Angela Carpenter, Ben Gelbart and his wife Ieva, Nancy Hall, Masako Hirotani, Eva Juarros, Kiyomi Kusumoto, Makoto Kadowaki, Ji-Yung Kim, Min-Joo Kim, Meredith Landman, Ania Lubowicz, Marcin Morzycki, Anita Nowak, Junko Shimoyama, Taka Shinya, Jennifer Smith, Steve Parker and his wife Monica, and Pius Tamanji. Discussions with Masako Hirotani and Taka Shinya on Japanese prosody were especially very useful.

Also, I am grateful to department secretaries Kathy Adamczyk and Lynne Ballard. They made my life in the department much easier. Kathy is my music advisor and she introduced me to the local folk music scene. We often visited Iron Horse for our favorite musicians. I really enjoyed the moments that I could be free from thinking of my dissertation. Lynne always listened to me sincerely and encouraged me.

I should add that my housemates, Laura Dahowsky, Liesbet Mandars and Ellen Manning, proofread this acknowledgement page and the following abstract page. Kitchen talk with them was an important part of my life during I was writing this dissertation. I am thankful to them for their help and encouragement.

Finally, I would like to thank my mother Akiko Sugahara, father Masanori Sugahara, grandmother Kiyoko Utsumi, and grandfather Shozo Utsumi for their constant love and support.

ABSTRACT

DOWNTRENDS AND POST-FOCUS INTONATION IN TOKYO JAPANESE

FEBRUARY 2003

MARIKO SUGAHARA, B.A., INTERNATIONAL CHRISTIAN UNIVERSITY

M.A., UNIVERSITY OF TOKYO

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Elisabeth O. Selkirk & Professor John Kingston

This dissertation is concerned with F0 downtrends in Tokyo Japanese: time-dependent declination, post-accent downtrend, i.e. *catathesis*, and post-FOCUS compression of F0 movement.

I investigate in Part I (Chapters 3 and 4) how "local" or "global" those downtrends are. In that part of the thesis, I focus on the time-dependent declination (Chapter 3) and *catathesis* (Chapter 4). Though they have been considered to be global phenomena, I show more local aspects of those downtrends.

The time-dependent declination is usually formalized as a gradually declining slope of the base line unfolding over the whole utterance or across phrases. In Chapter 3, however, I argue for an additional "tone-bound" declination slope which unfolds only between two neighboring tones. This accounts for my observation that F0 of the second tone (T2) gets substantially lower as the duration between two neighboring tones (T1 and T2) increases, while tones that follow T2 are barely affected by the duration change.

The post-accent downtrend, i.e. *catathesis*, has been formalized as tonal space lowering. In Chapter 4, however, I propose a local "tone-by-tone" scaling model to

account for *catathesis*. The local tone-by-tone scaling model correctly predicts that the "magnitude" of *catathesis* of a post-accent tone T_i diminishes as more tones intervene between T_i and the preceding pitch accent. In contrast, the global pitch range lowering model incorrectly predicts that all post-accent tones equally undergo *catathesis* regardless of the number of tones intervening between them and the preceding pitch accent.

Another important question, examined in Part II (Chapters 6, 7 and 8), is the "structural" vs. "non-structural" character of the post-FOCUS F0 compression. According to the structural view of the post-FOCUS compression, the phenomenon is a result of the absence of phonological phrase boundaries (i.e. dephrasing) after FOCUS. The non-structural view is that the phenomenon is a result of FOCUS affecting the phonetic interpretation of tones without manipulating the hierarchical organization of phonological phrase structure. I conclude that those views are both correct. Some aspects of the post-FOCUS F0 reduction are only accounted for by dephrasing while there is also a non-structural effect unexplained by dephrasing only.

LIST OF TABLES

LIST OF FIGURES

TABLE OF CONTENTS

PART II. POST-FOCUS INTONATION IN TOKYO JAPANESE:
THE STRUCTURAL VIEW VS. THE NON-STRUCTURAL VIEW

CHAPTER 1

INTRODUCTION

In this introductory chapter, I provide an overview of the dissertation (Section 1.1), the data collection and experiment methods (Section 1.2), basic assumptions and analyses of intonation related to the thesis (Section 1.3 and Section 1.4).

1.1. The Issues and Overview of the Dissertation

This dissertation provides empirical and theoretical analyses of downtrends and reduction of fundamental frequency (F0) in Tokyo Japanese. One of the well-known facts about intonation is that F0 of speech tends to decrease over the course of an utterance. The decrease in F0, however, is not a single, uniform phenomenon. There are at least three types of F0 downtrends in languages: (i) time-dependent declination (Collier and 't Hart 1971, among many others); (ii) downtrends conditioned by certain tonal sequences such as *downstep, downdrift*, or *catathesis*[1] (Clements & Ford, 1979; Pulleyblank, 1986; Clark, 1990; Liberman et al., 1993; among many others for *downstep* and *downdrift* in African tone languages; Pierrehumbert (1980), Beckman & Pierrehumbert (1986) for the English; and Poser (1984) and Pierrehumbert & Beckman

[1] *Downstep* and *downdrift* are types of downtrend patterns found in African tone languages (Clements & Ford, 1979; Pulleyblank, 1986; Clark, 1990; Liberman et al., 1993; among many others). The term *downstep* describes the phenomenon of a H (high) tone being realized at a lower pitch as a result of being preceded by another H tone. There are at least two interpretations for this phenomenon. According to Clements & Ford (1979) and Pulleyblank (1986), *downstep* is viewed as the consequence of a "floating" L (low) tone between the two high tones, which triggers lowering of the following H tone. Clark (1990), on the other hand, regards it as two distinct H tones: a higher H tone and a "downstepped" !H tone. The term *downdrift* refers to a phenomenon of progressive lowering of H and L tones when they overtly occur

(1988) for the Japanese post-bitonal accent lowering of tones, i.e. *catathesis*), and (iii) reduction/compression of F0 movements in a post-contrastive focus part of an utterance (Grønnum, 1989 for Dannish; Gårding 1993 for Swedish; Garding, 1987; Selkirk & Shen, 1990; Xu, 1999; Shih, 2000 for Chinese; Di Cristo & Jankowski, 1999 for French; Le Gac, 2002 for Somali).

One important question related to those downtrends is how local or global those downtrends are. This is the main concern of Part I (Chapters 3 and 4) of this dissertation. In that part of the thesis, I focus on the first two types of downtrends: time-dependent declination (Chapter 3) and *catathesis*, i.e. post-accent lowering of tones (Chapter 4). Those two types of downtrends have been considered to be *global* phenomena. The time-dependent declination has been considered to be *global* in the sense that it is usually formalized as a gradually declining slope of the *base line* (and the *top line*) of a tonal space which unfolds over the whole utterance or across phrases (Gussenhoven & Rietveld, 1988; among others). The post-accent downtrend, i.e. *catathesis*, has been considered to be *global* in the sense that it is formalized as the lowering of the *top line* of a tonal space but not manipulation of relative values of each of the post-accent tones (Pierrehumbert & Beckman, 1988). In Chapters 3 and 4, however, I reveal more local aspects of those downtrends than previously thought.

Another important question, examined in Part II of the thesis, is the structural vs. non-structural character of the post-contrastive focus (post-FOCUS) compression and downtrend of F0. The structural view of the post-FOCUS compression is that it is a phonetic manifestation of the absence of phonological phrase boundaries after FOCUS,

---

in sequence. *Catathesis* (from the Greek roots for "down-placement") refers to lowering of tones that follow a bitonal pitch accent, and it is *catathesis* that is found in Tokyo Japanese.

while the non-structural view of the phenomenon is that it is due to pitch range lowering or compression. Those two theories of post-FOCUS downtrend are compared in Part II, and are shown to both play a role. In the following part of this section, I provide a more detailed overview of those chapters.

1.1.1. Chapter 3 (Part I)

It is widely believed that the time-dependent declination is a global phenomenon and formalized as a declining *base line* of the pitch range unfolding over an utterance or across phonological phrases. However, very little of the previous literature on time-dependent declination has systematically investigated local aspects of time-dependent downtrend, for example effects of passage of time between two neighboring tones. In Chapter 3, I report that the F0 of the second tone (T2) gets substantially lower at the rate of –25~–60 Hz as the duration between two neighboring tones (T1 and T2) increases. The rate of this local lowering is far greater than what we usually expect from the global declination. When it comes to tones further away, they are barely affected by the durational change between T1 and T2, and only undergo a less steep global time-dependent downtrend. I interpret those findings as an independent tone-bound declination slope in addition to the global declination slope. Unlike the global declination, the slope of the tone-bound declination is reset each time the F0 target of a new tone is achieved.

1.1.2. Chapter 4 (Part I)

The other issue related to global vs. local aspects of downtrend considered in Chapter 4 is how to model *catathesis* in Tokyo Japanese, i.e. the post-accent downtrend. Pierrehumbert & Beckman (1988) propose that it should be captured as global pitch range lowering and compression after an accent. However, the phenomenon can be accounted for as a more local tone-by-tone scaling without any manipulation of pitch range height. According to the former global view, even post-accent tones that are not adjacent to the preceding pitch accent are predicted to be as equally low as those tones that are adjacent to the preceding accent. According to the latter view, however, the magnitude of post-accent lowering may be alleviated as more tones intervene between them and the preceding accent. My experimental results were consistent with the prediction made by the latter account, and conclude that the post-accent lowering in Tokyo Japanese should be accounted for by the tone-by-tone scaling model.

1.1.3. Chapters 6, 7 and 8 (Part II)

Part II of the dissertation (Chapters 6, 7 and 8) is about post-contrastive focus downtrend and F0 compression in Tokyo Japanese. Contrastive focus is expressed with the uppercase as FOCUS to distinguish it from presentational focus (new information). There are two views regarding the phenomenon. Pierrehumbert & Beckman (1988) proposed that the phenomenon should be captured as a result of the pitch range compression of the post-FOCUS part of an utterance without phonologically deleting phonological phrase boundaries in that part. Nagahara (1994), Truckenbrodt (1995) and Uechi (1997), however, take another view. According to them, the phenomena should be accounted for by post-FOCUS dephrasing, i.e. the deletion of phonological phrase boundaries after the FOCUS. I refer to the former as a non-structural view and the latter as a structural view. The main goal of Part II of the thesis, then, is to compare those two views on the post-FOCUS downtrend and compression of F0. I show in that part of the thesis that both the structural view and the non-structural view are correct. Some aspects of the post-FOCUS compression of F0 are only accounted for by deletion of phonological phrase boundaries while there is still a non-structural post-FOCUS effect that is unexplained by the phrase boundary deletion only.

All the phonological analyses presented in Part II (and in the following part of this preliminary chapter) are cast in terms of Optimality Theory proposed by Prince & Smolensky (1993). The theory regards a grammar as a set of ranked constraints on output representations. Though constraints are assumed to be universal, languages differ in terms of the ranking of constraints. The heart of the theory is that constraints are in principle violable as long as the violation results in the satisfaction of some higher ranked constraint. An input representation is mapped onto any form of output representations, and the grammatical output representation is the one that best satisfies the constraint hierarchy. That is, the grammatical output representation does not necessarily satisfy all the constraints but must be optimal (i.e. satisfy a higher ranking constraint and violate a lower ranking constraint conflicting with the higher constraint).

1.2. Data Collection Procedure

The accounts and analyses of Tokyo Japanese downtrends presented in this thesis draws on data obtained from a series of recording sessions.

There were three sets of recording experiments, and they took place between January 2001 and January 2002. Each of those three sets of recording experiments consisted of three to five recording sessions, which took place on different days. Each recording session consisted of three to five subsessions. In the subsessions of the same recording session, speakers read the same reading materials (sentences and dialogues). Usually, the number of reading materials in one subsession was between 25 ~ 36 sentences/dialogues.

Each of the reading materials were presented to the speakers typed on a card using the Japanese writing system (i.e. Japanese *kanji/kana* orthography). No commas were used in the written materials so that it was up to the speakers where to insert phrase breaks. These cards were shuffled into a random order in each subsession and given to speakers. Some of those reading materials were embedded in a dialogue. For such dialogues, the speakers and the experimenter (i.e. the author of the thesis) played a hypothetical roles in a conversational exchange. Speakers were asked to give natural renditions as much possible irrespective of whether the reading materials were in a dialogue or not.

Five female Tokyo Japanese speakers participated in those recording sessions: AS, MR, NK, RO and SK. Except for AS, who was at her late fifties, they were all at their late twenties or early thirties at the time of recording. They were born and grew up in the Tokyo area, and had no speech impairment. AS, MR and SK participated in all three sets of recording sessions. NK participated in only the first set of recording sessions, and RO participated in the second and the third set of recordings.

Recordings were made at a sound proof studio in the University of Tokyo, Komaba, using a SONY DAT recorder. The recorded utterances were re-digitized in the sample rate of 22kHz. Their F0 was analyzed by PitchWorks on a Mac PowerBook G3. Then, F0 measurements were made at points on F0 tracks obtained by the program. Details about those measurement points are given in each chapter.

Statistical analyses were carried out by SPSS on the Mac Power Book. In the appendix, description of sentences and phrases used in this thesis is provided.

1.3. On Prosodic Structure

Downtrends are the lowering of the F0 value of tones. Important tonal phenomena such as the presence/absence and distribution of tones and the F0 scaling of tones in various languages are tightly correlated with the prosodic structure, and Tokyo Japanese is not an exception. Given this, it is necessary to introduce my assumptions about prosodic structure in this preliminary chapter.

1.3.1. The Prosodic Constituents and Hierarchy

I adopt the Prosodic Structure Hypothesis proposed by Selkirk (1986, et seq) and (Nespor & Vogel (1986), and adopted by Pierrehumbert & Beckman (1988) among many others. According to the hypothesis, an utterance is parsed into a sequence of prosodic constituents at each of the different levels of the hierarchy shown in (1).

(1) Prosodic Hierarchy

Utterance (Utt)
|
Intonational Phrase (IP)
|
Major Phonological Phrase (MaP, a.k.a Intermediate Phrase)
|
Minor Phonological Phrase (MiP, a.k.a. Minor Phrase)
|
Prosodic Word (PWd)
|
Foot
|
Syllable
|
Mora

Prosodic constituents most directly relevant to this thesis are MaP (Major Phonological Phrase) and MiP (Minor Phonological Phrase). In Tokyo Japanese, an F0 rise usually referred to as "Initial Lowering" is found at the left edge of a Minor Phrase (Poser, 1984; Pierrehumbert & Beckman, 1988; Selkirk & Tateishi, 1988; Kubozono, 1993). In terms of a tonal sequence, the F0 rise consists of L and H edge tones (Pierrehumbert & Beckman, 1988).[2] I provide a more detailed discussion on those edge tones later in this chapter. The presence of the F0 rise, i.e. presence of a Minor Phrase boundary, is controlled by a variety of factors such as size of constituents, their accentedness, the morpho-syntactic context, focus status and so on. In Part II of the

[2] Just as LH edge tones appear at the left edge of Tokyo Japanese Minor Phrases, edges of prosodic constituents in various other languages also serve as the location of those phonological/phonetic phenomena. For instance, edge tone insertion at the left or right edge of prosodic constituent is frequently found in Bengali (H tone at the right edge of a phonological phrase, Hayes & Lahiri, 1991; Selkirk, 2002) and in Korean (LH tones at the left and the right edges of an Accentual Phrase, i.e a Minor Phrase, Jun, 1998; Jun & Oh, 1996).

thesis, I will show how interaction among some of those factors plays a role in determining the presence or absence of a Minor Phrase boundary.

The Major Phrase in Tokyo Japanese is associated with two intonational phenomena. One is *catathesis*; the other is upward pitch resetting (Poser, 1984; Pierrehumbert & Beckman, 1988). *Catathesis* is a lowering of tones that follow the bitonal pitch accent H*+L. Full discussions on the phenomenon are provided in Chapter 3 of the thesis. Upward pitch resetting at the left edge of a Major Phrase brings the lowered pitch to a higher level. The post-accent lowering of tones (*catathesis*) takes place only within a Major Phrase, and the onset of a new Major Phrase is where the upward pitch reset takes place. One way to account for the post-FOCUS compression of F0 movement discussed in Part II of the dissertation is that such Major Phrase boundaries are all deleted in the post-FOCUS part of an utterance (Nagahara, 1994; Truckenbrodt, 1995). As a result, no upward pitch resetting takes place there, which results in compression of F0 movement. This is "structural" account of Nagahara and Truckenbrodt for the post-FOCUS pitch compression. It is compared with Pierrehumbert & Beckman's pitch range lowering account in Part II.

1.3.2. Factors Determining the Prosodic Structure

In this subsection, I introduce some of the constraints that play a crucial role in determining prosodic structure of languages. From time to time, I introduce actual example phenomena of Tokyo Japanese to give motivation for those constraints.

The hierarchically organized structure in (1) is distinct from the input morpho-syntactic structure of the sentence, while there is partial influence from the morpho-syntactic representation on the prosodic structure. For example, there is a tendency for the left or right edge of a syntactic maximal projection (i.e. XP) to correspond to the left or edge of a phonological phrase such as a Major Phrase respectively (Selkirk, 1986; Selkirk & Tateishi, 1988; Selkirk & Shen, 1995). One instance of correspondence between those two edges is Tokyo Japanese prosodic phrase formation. The post-accent tonal lowering phenomenon, *catathesis*, tends to be cancelled (i.e. upward F0 resetting usually takes place) at the left edge of an XP (Selkirk & Tateishi 1988, 1991). Since *catathesis* cancellation (or upward F0 resetting) is a diagnostic for a Major Phrase boundary, Selkirk & Tateishi interpreted it as presence of a Major Phrase boundary at the left edge of an XP in Tokyo Japanese. They proposed that the following XP-Major Phrase alignment constraint play a crucial role in Japanese.

(2) ALIGN$_L$ (XP, MaP) (Selkirk & Tateishi, 1988, 1991)
The left edge of an XP must coincide with the left edge of a Major Phrase.

Other types of syntax-phonology interface constraints are proposed by Truckenbrodt (1994) and Selkirk (2002). One of those constraints proposed by Truckenbrodt is a WRAP constraint, which calls for items dominated by the same XP in the input to be dominated by the same phonological phrase node. Other types of interface constraints specify the relationship between syntactic constituent and the prosodic "prominence". In Part II of the thesis, I provide a detailed discussion on the relationship between a syntactic constituent with a [Focus] feature and prosodic prominence. I show that a constraint proposed by Truckenbrodt (1995), which requires items dominated by a [Focus] marked syntactic constituent to correspond to the most prominent item in an utterance, is necessary to account for aspects of post-FOCUS compression of F0 movement. Also, in a later part of that chapter, I show that a constraint on correspondence between a stem constituent in the morpho-syntactic representation and the most prominent item of a PWd explains the location of a pitch accent in Tokyo Japanese.

Another important set of constraints on prosodic structure is phonological markedness constraints on prosodic domination and constraints on the size of prosodic constituents. The constraints on prosodic domination that appear to hold universally are LAYEREDNESS and HEADEDNESS (Selkirk, 1995).[3]

(3) LAYEREDNESS
No C_I dominates a C_j, $j < I$

(4) HEADEDNESS
Any C_I must dominate a C_{I-1} (except if C_I = mora)

The constraint in (3), for example, forbids a syllable dominating a foot or a foot dominating a prosodic word. Also, the constraint in (4) calls for a prosodic word to dominate a foot, for instance.

Constraints on prosodic domination are not the only prosodic markedness constraints which regulate prosodic structure. It is also known that the weight or size of a prosodic constituent is an important factor determining phonological structure (Ghini, 1993, Selkirk & Tateishi, 1988; Selkirk, 2001a; Kubozono, 1993; Jun, 1993; among others). It has been suggested that some of those weight requirements are captured by constraints on the minimum and maximum size of possible constituents, BINARY

MINIMUM(C_I) and BINARY MAXIMUM(C_I) (Selkirk, 2000). These size constraints assess the output constituents of a particular level of prosodic structure C_I in terms of the number of constituents (C_{I-1}) that C_I immediately dominates.

(5) BINARY MINIMUM(C_I)
Prosodic constituent of level C_I must dominate at least two prosodic constituent of level $C_{I-1.}$

(6) BINARY MAXIMUM(C_I)
Prosodic constituent of level C_I may dominate at most two prosodic constituent of level $C_{I-1.}$

The BINARY MINIMUM(C_I) constraint plays a crucial role in Tokyo Japanese Minor Phrase formation. There is a tendency in the language for a sequence of two "unaccented" words to coalesce into a single Minor Phrase when there is no XP boundary between them and those two words are relatively short (Selkirk & Tateishi, 1988; Shinya, 2002; and my own observation). When those conditions are met, no initial F0 rise (the LH edge tones) are found at the left edge of the second word of the two-word sequence.

This BINARYMINIMUM(MiP), however, is not always observed. For example, when the second word of the sequence of two unaccented words is a compound word consisting of more than one root word, then the probability of having an initial F0 rise between those two words increases (Chapter 7 of this thesis). This follows a generalization made by Kubozono (1993) that the left edge of a branching X^0 usually coincides with the left edge of a Minor Phrase. I explain his fact by having an

---

[3] In optimality theoretic terms, the inviolability of these constraints implies that they are undominated in the constraint ranking of every language, i.e. they form part of Gen.

additional syntax-alignment constraint, i.e. ALIGN$_L$(X^0-branching, MiP), outranks the BINARYMINIMUM(MiP) constraint.

(7) ALIGN$_L$(X^0-branching, MiP)

The left edge of a branching X^0 coincides with the left edge of a Minor Phrase.

(8) ALIGN$_L$(X^0-branching, MiP) >> BINARYMINIMUM(MiP)

In addition, once both of the two words in the word sequence are replaced by accented words, then initial F0 rise is more likely to be present at the left edge of Word2 (observations by Selkirk, 2001ab; Shinya, 2002; Chapter 7 of this thesis). This is because there is a highly ranked markedness constraint which calls for at most one accent in one Minor Phrase, which I tentatively call MiP-Acc following Selkirk (2001a).

(9) MiP-Acc

At most one accent in a MiP.

(10) MiP-Acc >> BINARYMINIMUM(MiP)

The MiP-Acc constraint in (9) requires each Minor Phrase to dominate at most one accent. Satisfying the BINARYMINIMUM(MiP) constraint means coalescing the two accented words into a single Minor Phrase, which results in violation the outranking MiP-Acc constraint. As a result, the grammatical representation needs a Minor Phrase boundary between the two accented words and puts them into separate Minor Phrases to satisfy the MiP-Acc constraint. This MiP-Acc constraint is later replaced with an alignment constraint between the pitch accent H*+L and the most prominent mora of a Minor Phrase, which is referred to as $\mathrm{ALIGN}_{\mathrm{L}}(\mathrm{H^*{+}L}, \mathrm{DTE}_{\mathrm{MiP}})$.

Though the set of constraints and factors listed above are not the exclusive set of factors determining the phonological structure of an utterance, we move on to the next issue, i.e. the relationship between the prosodic structure and prosodic prominence.

1.3.3. Prosodic Prominence

The other important aspect of prosodic structure is prosodic "prominence". The notion of prosodic prominence or prosodic head is relevant to Part II of the thesis which is about post-FOCUS compression of F0 as well as to one of the later sections of this chapter which is about accent location.

According to the prosodic structure hypothesis, every prosodic constituent but the mora dominates a head constituent. The immediate head of prosodic constituent *C* is the most prominent daughter constituent immediately dominated by *C*. Also, the notion of head is transitive. That is, the most prominent prosodic constituent immediately dominated by the immediate head of *C* is also a head of *C*. For example, a syllable or even a mora may be a head of a phonological phrase such as a Minor Phrase as long as it is dominated by the chain of heads of that Minor Phrase. Given the notion of prosodic head and transitivity of head relationship, we can now define a notion of Designated Terminal Element (DTE, or Δ).

(11) Designated Terminal Element (DTE, or Δ) Selkirk (2000a, 2000b)
DTE of a prosodic constituent *C* is the mora (i.e. the terminal prosodic node) that is dominated by the chain of heads of *C*

Later in this thesis, the notion of DTE and prosodic prominence will play an important role. I show in 4.4 of this chapter that it is the DTE of a Minor Phrase that is associated with the pitch accent H*+L in Tokyo Japanese. Also, in Part II of the thesis which is about compression of F0 movements after a FOCUS, I introduce the FOCUS-Prominence theory of Truckenbrodt (1995) and Selkirk (2002). According to their theory, a FOCUS constituent in the syntactic representation should correspond to the highest prominence of an utterance or an Intonational Phrase. More specifically, Selkirk (2002) formalized this FOCUS-prominence relationship with a constraint that calls for a syntactic constituent bearing the [FOCUS] feature to correspond to a sequence of phonological terminal elements containing the DTE of an utterance or an Intonational Phrase.

1.4. Tones and Prosodic Structure in Tokyo Japanese

Throughout the thesis, I assume that rising and falling F0 patterns are defined by F0 maxima and minima “targets”, which are referred to as “tones”, following the autosegmental theory of Goldsmith (1976), Leben (1973, 1976) among others.

In tone-languages like most African languages and Chinese, tones are part of lexical representations: syllables are underlyingly specified for a tone or a sequence of tones. As a result, there is a contrast among syllables or words in terms of their tonal shapes. However, in “intonational” languages like most of the Indo-European languages, tones are not part of lexical representations but only appear in the surface representation for either boundary/edge marking purposes or prominence-related purposes. I refer to tones marking edges of phonological constituents as “edge tones” or “boundary tones” and those linked to prosodically prominent syllables as “pitch accents” following Pierrehumbert (1980). Pitch accents are expressed with a diacritic mark “*”.

Tokyo Japanese is in the middle of those two types of languages in the sense that a certain tones is part of lexical representations while others are inserted in the surface representation for edges/boundary marking purposes. The former, i.e. the lexically provided HL sequence surfaces as a sharp F0 fall. Though the HL sequence is part of the lexical specification, it surfaces at a prosodically prominent position in verbal paradigms as discussed more in detail in 1.4.4. Therefore, the lexically provided HL sequence in Japanese is also prominence-related tones and parallel to the pitch accents in languages like English. I refer to the lexically provided HL sequence in Tokyo Japanese as “pitch accent” and express it as “H*+L” following Pierrehumbert & Beckman (1988).

In what follows, I provide basic and essential facts and analyses of Minor Phrase edge tones and the pitch accent in Japanese.

1.4.1. The L and H Edge Tones in Tokyo Japanese

As already mentioned in Section 1 of this chapter, the left edge of a Minor Phonological Phrase boundary coincides with an F0 initial rise, and this initial rise is interpreted as a sequence of L and H edge tones. The picture in Figure 1.01 shows two examples of such an initial F0 rise.

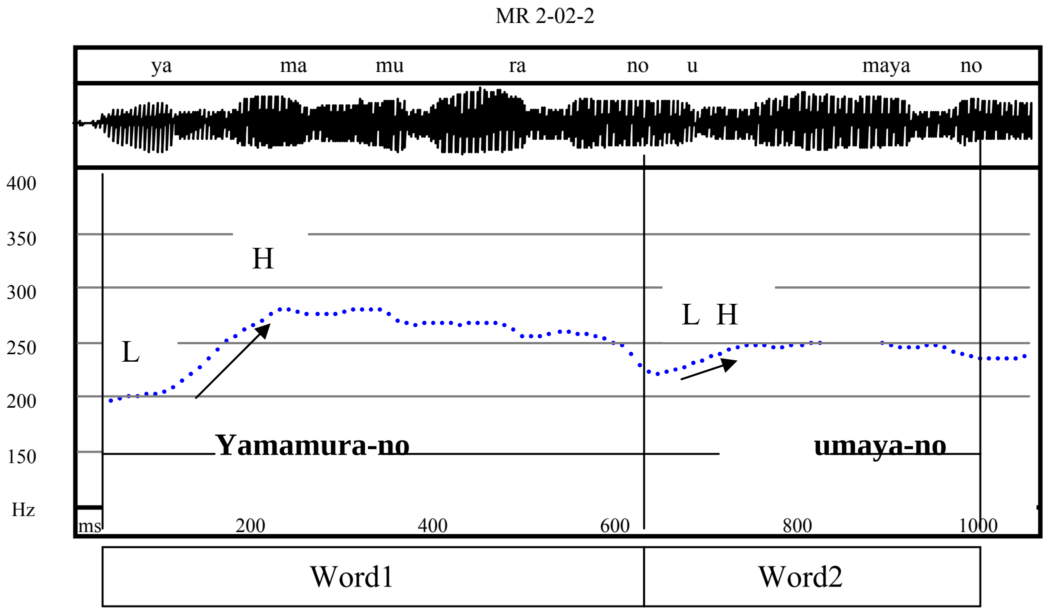


Figure 1.01. Examples of an Initial F0 Rise from L to H

The picture in Figure 1.01 is the F0 contour of Word1 and Word2 of an example from the <*Umaya*> Set introduced in the previous section of this chapter, and obtained from MR's data file. MR is different from SK and NK because she almost always had an initial F0 rise at the onset of both Word1 (*yamamura-no* "mountain-village-Gen") and Word2 (*umaya-no* 'barn-Gen") of the he <*Umaya*> Set. Those initial rises are clearly shown in Figure 1.01.

This initial F0 rise, which is usually referred to as Initial Lowering, is interpreted as a sequence of L and H edge tones (Pierrehumbert & Beckman, 1988). The L edge tone is linked to the initial mora. The H edge tone is usually aligned with the second mora of a Minor Phrase but sometimes shifts rightward especially when the

F0 excursion size from the preceding the L and the H tone is large (my own observation). Venditti (1995) makes a similar observation.[4]

The F0 excursion size between the L and the H edge tone is a function of the phonological phrase boundary strength. A good example is shown in the picture above (Figure 1.01). The F0 excursion size between the L and H edge tones of Word1 is about 50 Hz greater than that between those tones of Word2. This is a reflection of the levels of prosodic phrase boundaries of those words. Word1 in that picture is at sentence-initial position but is preceded by an adverbial expression *chikágoro* "recently". There is a pause after the adverbial expression and before Word1, and I believe that Word1 is at an Intonational Phrase-initial position as well as at a Major Phrase-initial position. On the other hand, only a Minor Phrase boundary is present at the left edge of Word2. Word1 with a stronger boundary coincides with a greater F0 excursion size between L and H than Word2 with a weaker boundary as shown in (13) schematically.

[4] A bitonal pitch accent H*+L may also fall on the initial and the second mora (syllable) of a Minor Phrase. When the accent is phonologically associated with the initial syllable of a Minor Phrase, the H* accent tone is phonetically aligned with the right edge of the initial syllable or the beginning of the following syllable, and an F0 rise from be left edge to the right edge of the initial syllable is present (my observation). I interpret this rise as evidence that the L edge tone of the Minor Phrase is undeleted and still aligned with the very beginning of the F0 rise. As for the H edge tone, I tentatively assume that it is also undeleted and aligned with somewhere between the target of the L edge tone and the following H* accent tone. However, because the distance among those three tones is exceptionally short, the target of the H edge tone is undetectable.

(12) $[_{IP}$

$[_{MaP}$

$[_{MiP}$ Word1] $[_{MiP}$ Word2]

↑ ↑

greater F0 rise smaller F0 rise

1.4.2. Spreading of the H Edge Tone

There are two types of speakers with respect to the realization of the H edge associated with an "unaccented" Minor Phrase . One type of speaker simply draws an F0 interpolation line between the H tone and the L edge tone which is associated with the initial syllable of the following Minor Phrase. One of our speakers, AS, always adopted this interpolation procedure as shown in Figure 1.02.

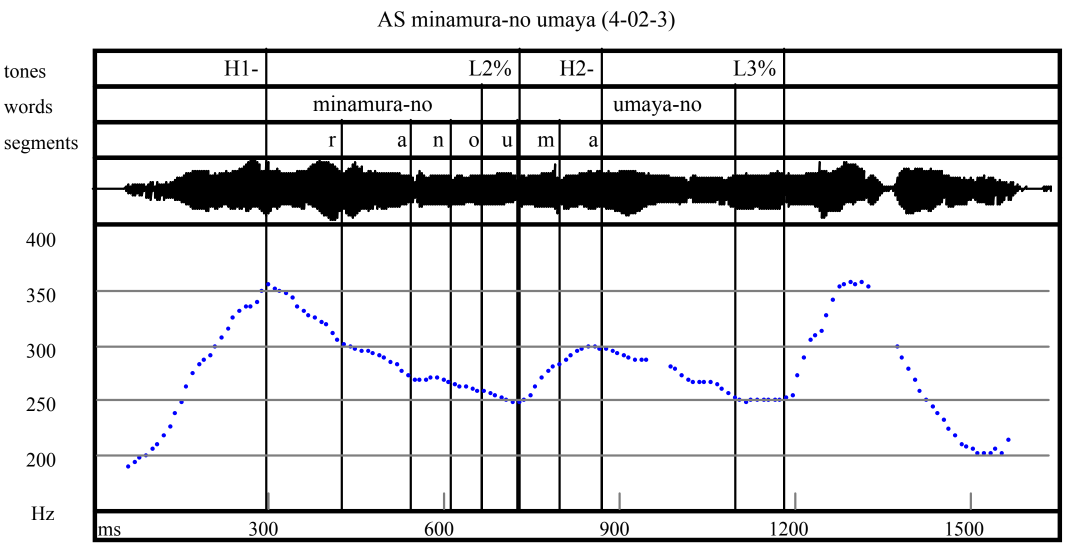


Figure 1.02. An Example of Interpolation between H to the following L

This complies with what Pirrehumbert & Beckman (1988) argued for in their book. However, this is not the only possible way to realize the H edge tone in Tokyo Japanese: there are also speakers that spread the H edge tone from the second to the final mora of the unaccented Minor Phrase. That is, they keep a high pitched plateau from the second to the final syllable of the unaccented MiP, and there is an abrupt fall from the final syllable to the initial syllable of the following Minor Phrase. This is interpreted as spreading of the H edge tone associated with the second syllable of the unaccented Minor phrase. Examples of such F0 discontinuity between the initial unaccented MiP and the following MiP obtained from two of my speakers (MR and SK) are shown in Figure 1.03 and Figure 1.04.

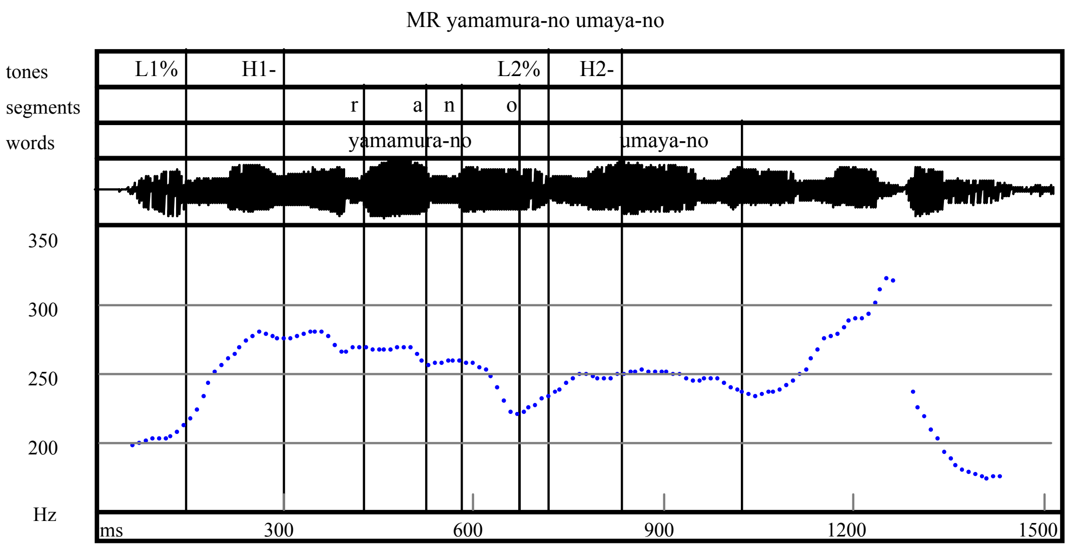


Figure 1.03. An Example of H Tone Spreading (MR)

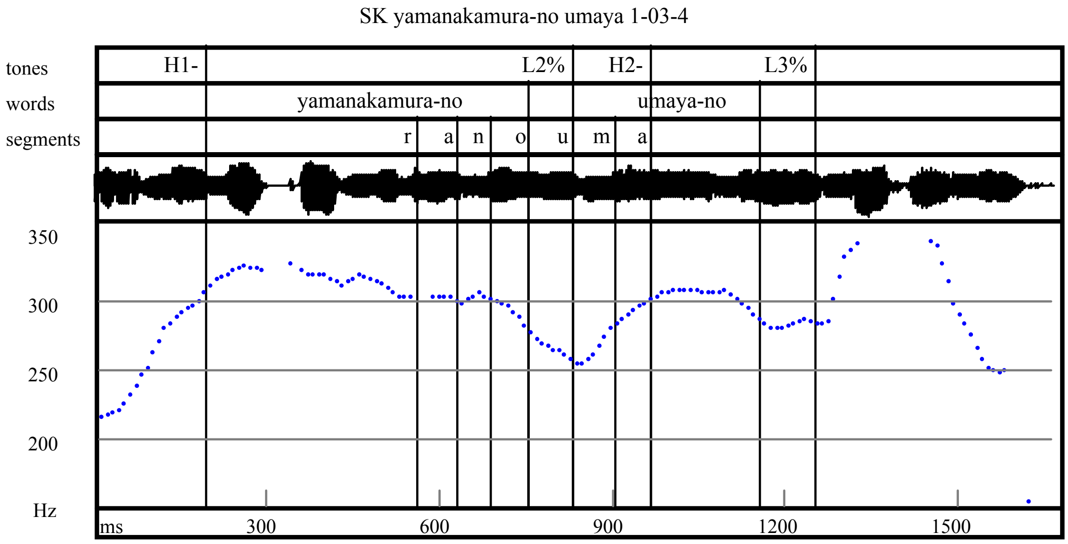


Figure 1.04. An Example of H Tone Spreading (SK)

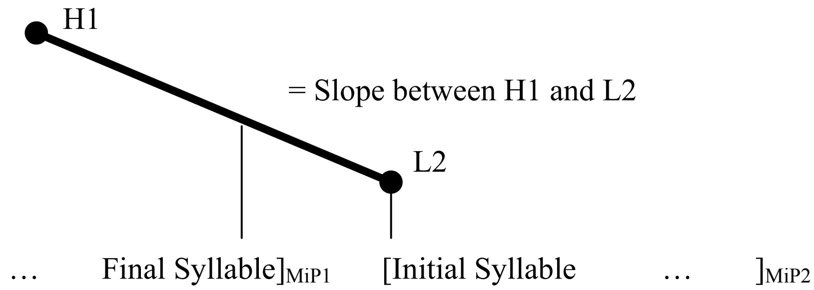
Though Pierrehumbert & Beckman (1988) argued against such H tone spreading, most traditional scholars of Japanese tonology including Haraguchi (1977) and Poser (1984) agree on presence of H spreading or high pitched plateau in an unaccented word. This discrepancy between Pierrehumbert & Beckman and the traditional scholars may be just a matter of speaker selection. Pierrehumbert & Beckman happened to have only interpolation speakers, while Haraguchi's and Poser's claim is based on their observation of spreading speakers.[5]

To support the idea that there are two distinct realizations of the plateau H, a qunantitative analysis was carried out. In that analysis, I examined the rate of F0 change from the H edge tone to the following syllables. Let us consider a sequence of two tones, H1 and L2. H1 is associated with the second syllable of the first Minor

Phrase (MiP1) and L2 is associated with the initial syllable of the second Minor Phrase (MiP2).

(13)	[σ	σ	$\sigma]_{MiP1}$	[σ	σ	$\sigma]_{MiP2}$
	L1	H1		L2	H2	

If the interpolation procedure is chosen, the rate of F0 change from H1 to the final syllable of MiP1 and that from H1 to the initial syllable of MiP2 (i.e. L2) should be the same as shown schematically in Figure 1.05. On the other hand, if the H- tone spreading takes place instead, the rate of the F0 change from H1 to the final syllable of MiP1 is smaller than that from H1 to the initial syllable of MiP2 as shown in Figure 1.06. This is shown schematically below.

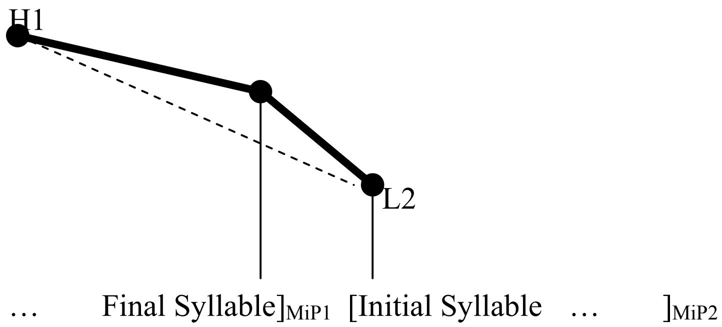


The Rate of F0 Change between H1 and Final Syllable of MiP1
= The Rate of F0 Change between H1 and L2

Figure 1.05. The "Interpolation" between H1 and L2

---

[5] Haraguchi (1977) does not provide any instrumental data. Though Poser's (1984) work is based on his instrumental data, he does not provide any empirical data to specifically show F0 plateau of the spreading H edge tone.



The Rate of F0 Change bet H1 and Final Syllable of MiP1
< The Rate of F0 Change between H1 and L2%

Figure 1.06. The "H- tone spreading" to the Final Syllable of MiP1

To test those predictions, the rate of F0 change from H1 to the final syllable of MiP1 and that from H1 to the initial syllable of MiP2 (i.e. the L2 edge tone) were compared using data obtained from the <*Umaya*> Set. The rate of F0 change from H1 to the initial syllable of MiP2 was obtained by dividing the F0 difference between those two points by the duration (in ms) between them. In the same way, the rate of F0 change from H1 to the final syllable of MiP1 was obtained by dividing the F0 difference between those two points by the duration between them.[6] Then the mean values of those two F0 change rates were compared by ANOVA. The results are summarized in the following table.

---

[6] The exact measurement points are the following:
(a) the F0 peak of MiP1 = the F0 of the H1 edge tone
(b) the F0 of the onset of the nucleus vowel of the final syllable of MiP1 = the F0 of the MiP1 final syllable.
(c) the lowest F0 associated with the initial syllable of MiP2 = the F0 of the L2 edge tone.

Table 1.01. The Results of the Mean F0 Change Rate

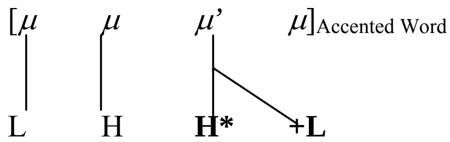
	Mean F0 Change Rate (per ms) between H1 of MiP1 and Final Syllable of MiP1	Mean F0 Change Rate (per ms) Between H1 of MiP1 and L2 of MiP2	Result of ANOVA
AS (n= 64)	.167 Hz/ms	.156 Hz/ms	$F(1,126) = 2.94$ $p = 0.08$
NK (n = 59)	.042 Hz/ms	.057 Hz/ms	$F(1,116) = 24.50$ $*p < 0.001$
MR (n=43)	.052 Hz/ms	.072 Hz/ms	$F(1,84) = 10.39$ $*p= 0.002$
SK (n=68)	.051 Hz/ms	.094 Hz/ms	$F(1,134) = 71.60$ $*p < 0.001$

As expected, speaker AS, whom I described as an interpolation speaker, has no significant difference between those two means of F0 change rate while speakers MR, SK and NK, whom I described as an H tone spreading speaker, have a significant difference between them. For those three speakers, the F0 change rate between H1 and the final syllable of MiP1 is significantly smaller than that of between H1 and L2.

In summary, there are two kinds of speakers in terms of realization of F0 associated with unaccented Minor Phrases in Tokyo Japanese. One type of speakers are those who interpolate the H edge tone and the L edge tone associated with the initial syllable of the following Minor Phrase. The other type of speaker are those who spread the H edge tone to the final syllable of the same Minor Phrase, resulting in a F0 discontinuity between the final syllable of that Minor Phrase and the initial syllable of the following Minor Phrase.

1.4.3. The Pitch Accent H*+L

In addition to the edge tones, Tokyo Japanese has a bitonal pitch accent H*+L, which is characterized as sharp F0 fall. The H* accent tone surfaces at either the mora lexically specified for an accent or the mora in some default location determined by constraint interaction and prominence assignment of prosodic structure. The +L accent tone surfaces at some mora that follows the mora phonetically aligned with the H* accent tone.[7] Those two tones form a single constituent and are both phonologically associated or aligned with the mora (syllable) that is specified for an accent or the mora in the default location.



(μ stands for a mora.)

Figure 1.07. The Association between the Pitch Accent and Accented Mora

The mora with the diacritic marker stands for either the mora lexically specified for an accent or the mora in the default location determined by constraint interactions and prominence assignment. Solid lines represent the phonological association between the accented mora and the pitch accent. Temporal alignment between the segmental tier and the tonal tier represents the phonetic alignment of tones and morae.

The tones comprising the pitch accent usually take more "extreme" F0 values than edge tones. That is, the H* accent tones tend to be realized higher than the preceding H edge tone within the same Minor Phrase (see the following picture: Figure 1.08).[8] In

[7] No previous literature on Tokyo Japanese pitch accent has systematically shown where exactly the +L trailing tone of a pitch accent is aligned. However, example F0 contours presented in Pierrehumbert & Beckman (1988): Figure 2.15, Page 50 of their book, as well as F0 contours obtained in my experiments which are presented in Chapter 3, show that the +L trail tone is aligned with the right edge of the second or the third syllable after the accented syllable.

[8] Relative F0 height of the preceding H edge tone and the following H* accent tone within the same Minor Phrase are affected by the duration between those two tones. As the duration increases, the F0 of

that picture, the H* accent tone is about 25 Hz higher than the preceding H edge tone. Also, the +L accent tone is realized lower than L edge tones preceded by no pitch accent (see Chapter 4).

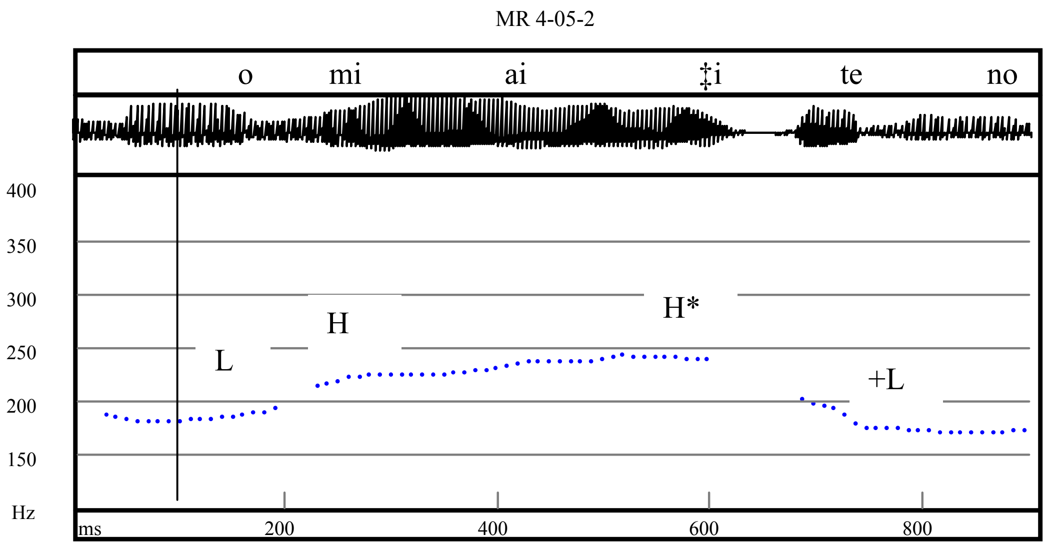


Figure 1.08. H* Realized Higher than H

Just as pitch accents fall on a prosodically prominent syllable in English (Selkirk, 2000), there is evidence that the pitch accent in Tokyo Japanese is also located at the most prominent syllable of a phonological phrase. In the following part of this section, I discuss the relation between the pitch accent H*+L and prosodic prominence.

---

the H* accent tone decreases. As a result, the H* accent tone is not necessarily higher than the H edge tone when the duration between them is long (see Chapter 2).

1.4.4. Location of the Pitch Accent: Verbal Paradigm vs. Nominal Paradigm

There is a contrast among lexical items with respect to their accentedness: some lexical items are specified for an accent but others are not. I assume that the H*+L pitch accent is present and linked to the syllable of the lexical item specified for an accent in the input representation.[9]

Table 1.02. Examples of Lexically Accented and Unaccented Verbs

	Accented	Unaccented
Verbs	*kíru* "cut" *fukaméru* "deepen"	*kiru* "wear" *yugameru* "distort"
Nouns	*áme* "rain" *namekúji* "slug"	*ame* "candy" *omikuji* "lottery"

In the noun paradigm, the bitonal pitch accent surfaces at the mora (syllable) lexically specified for an accent (McCawley, 1968; Poser 1984, Smith 1997) unless they are made into compound words.[10] Therefore, surface location of the pitch accent H*+L is unpredictable.

---

[9] I am aware of theories that tones are not necessarily present in the input representation and it is the constraint interaction that determines which mora to be associated with which form (L or H) of tones (de Lacy, 1999). However, since lexical representation and tonal association in the output are not the main theme of this thesis, I make the assumption that the H*+L pitch accent is already provided in the lexical representation to make discussions simpler.

[10] Irrespective of lexical specification of accent, the default, i.e. the most productive, nominal compound forms assign an accent to either the last mora of the first member of the compound (when the second member consists of at most two morae) or the first mora of the second member of the compound (when the second member is either three or four morae), and that is the only accent that that surfaces in those forms (Kubozono, 1995; Kubozono et al, 1997).

(14) Accented Noun Forms: Pitch Accent Location Unpredictable
(The diacritic marker indicates the location where the pitch accent H*+L appears.)

A. Final Accent

a.	*mé-ga*	"eye-Nom"
b.	*kí-ga*	"tree-Nom"
c.	*yamá-ga*	"mountain-Nom"
d.	*kotobá-ga*	"language-Nom"

A. Penultimate Accent

f.	*túma-ga*	"wife-Nom"
g.	*fúne-ga*	"ship-Nom"
h.	*kokóro-ga*	"heart-Nom"
i.	*wagamáma-ga*	"selfishness-Nom"
j.	*namekúji-ga*	"slug-Nom"

C. Antepenultimate Accent

f.	*ínochi-ga*	"life-Nom"
g.	*kárasu-ga*	"raven-Nom"
h.	*nadésiko-ga*	"a pink (a flower name)"
i.	*katatúmuri-ga*	"snail-Nom"

D. Pre-Antepenultimate Accent

j.	*kámakiri-ga*	"mantis"

When it comes to accented verbs and adjectives, the surface location of the bitonal pitch accent is predictable (McCawley, 1968; Poser 1984, Smith 1997). I show some examples of verbal forms with the past tense suffix and the progressive tense suffix in the following table. All the examples are mine.

Table 1.03. Examples of Accented Verbs: Accent Location Predictable

Accented Stem	suffix *ta* Past	suffix *tei-ru* Progressive Present
(a) *itonam* "run (a business)"	*itonánda*	*itonándeiru*
(b) *sakeb* "shout"	*sakénda*	*sakéndeiru*
(c) *shaber* "speak"	*shabétta*	*shabétteiru*
(d) *hagas* "peal"	*hagásita*	*hagásiteiru*
(e) *hamidas* "bulge out"	*hamidásita*	*hamidásiteiru*
(f) *tabe* "eat"	*tábeta*	*tábeteiru*
(g) *kakitome* "write down"	*kakitómeta*	*kakitómeteiru*

The diacritic marker indicates surface location of the pitch accent H*+L appears.

The descriptive generalization derived from the table above is that verbs of "consonant-ending" stem have the accent on the "rightmost (or final)" vowel of the stem and those of "vowel-ending" stem have the accent on the "second rightmost (or penultimate)" vowel of the stem.

(15) [[(Accented) … C V_iC V_j C]$_{\text{Verb-Stem}}$ + ta/teiru]$_{\text{Word}}$

↓

epenthetic vowel

…CV_i CV_j' C(V) ta/teiru

(16) [[(Accented) …C V_i C V_j]$_{\text{Verb-Stem}}$ + ta/teiru]$_{\text{Word}}$

↓

… CV_i' CV_j ta/teiru

In terms of prosodic structure theory, I interpret the generalization that (i) the right edge of the verb stem coincides with a foot, and (ii) the foot has a special status that it receives a pitch accent. The special status of the foot is reinterpreted as the most prominent foot of a PWd, i.e. the head of the PWd, and the H*+L pitch accent is aligned with the PWd head.

(15') [[Consonant-Ending Stem] + Suffix]$_{\text{Word}}$
(**CV_i [CV_j' C (V)]$_{\textbf{Head foot}}$ ta/teiru**)$_{\text{PWd}}$
|
H*+L

(16') [[Vowel Ending Stem] + Suffix]$_{\text{Wrd}}$
(**… [CV_i' CV_j]$_{\textbf{Head foot}}$ ta/teiru**)$_{\text{PWd}}$
|
H*+L

1.4.5. Prosodic Prominence and the Pitch Accent Location

That the verb accent location is predictable is explained by markedness outranking faithfulness. That is, a set of markedness constraints calling for the accent H*+L to fall on the foot at the stem right edge outranks a faithfulness constraint that calls for the location of the surface H*+L to be the same as that of the input H*+L.

(17) Markedness >> FAITHLOC(H*+L)

However, the ranking in (17) wrongly predicts that the accent location of noun forms should fall on the default stem edge position. To solve the problem, Smith (1997) proposed *Noun Faithfulness* constraints to account for the contrast between noun forms and verb forms. Noun faithfulness constraints are domain-specific faithfulness constraints that apply only to nouns, and what is relevant here is the constraint that forces the input specification of accent location in noun forms to be preserved in the output representation, i.e. FAITHLOC$_{\text{NOUN}}$(H*+L).

(18) FAITHLOC$_{\text{NOUN}}$(H*+L) … from Smith 1997

Output accent is faithful to its input location in *noun*s.

This noun-faithfulness constraint dominates the set of markedness constraints as shown in (19), which prevents only the noun accent from being associated with the stem edge position.

(19) FAITHLOC$_{\text{NOUN}}$(H*+L) >> Markedness >> FAITHLOC(H*+L)

… from Smith (1997)

In the following discussion, I focus on the content of the markedness constraints that play the crucial role in determining the location of the pitch accent in the verbal paradigm. I will argue that they are a morphosyntax-phonology interface constraint that requires the stem of a word to correspond to the prosodic prominence and a phonological markedness constraint which requires the prosodic head of a PWd to be the rightmost position of a PWd.

The markedness constraints responsible for pitch accent location in the verbal paradigm should capture the fact that the PWd head (i.e. the most prominent foot of a PWd) is aligned with the right edge of the stem. As already shown, the right edge of verbal stems is at the PWd medial position in the verbal paradigm. I interpret the fact that the pitch accent falls on the "PWd-medial but stem-final" position as a compromise between two conflicting constraint STEM:HEAD$_{\text{Pwd}}$, and EDGEMOST-RIGHT. EDGEMOST-RIGHT was proposed by Prince & Smolensky (1993) to capture the tendency in languages that prosodic prominence (stress) of a PWd falls on the syllable close to the rightmost edge of the PWd. Also, I propose a STEM:HEAD$_{\text{Pwd}}$ constraint, which is a morphology-phonology interface constraint and requires the input stem correspond to the sequence of terminal elements in the output representation which contains the most prominent mora (DTE) dominated by the head Foot of a PWd.

(20) STEM:HEAD$_{Pwd}$ (Stem-Prominence)

The terminal string of a stem *S* in the input representation must correspond to the terminal string of the phonological representation which contains the DTE a PWd.

(21) EDGEMOST-RIGHT$_{PWd}$ … or ALIGN$_R$ (DTE, PWd)

DTE of PWd should be the final mora of PWd.

Because STEM:HEAD$_{wd,}$ outranks EDGEMOST-RIGHT$_{PWd}$, the most prominent Foot of a PWd does not have to be at the PWd-final position but may be somewhere "within" the stem. EDGEMOST, however, is violated gradiently, and the violation count is made by the number of morae between the accented mora and the right edge of the prosodic constituent. As a result, the optimal representation is something that satisfies STEM:HEAD$_{wd}$ and minimally violates EDGEMOST, i.e. the representation in which the right edge of the head Foot coincides with the stem-final edge.

(22) STEM:HEAD$_{wd,}$ >> EDGEMOST-RIGHT$_{PWd}$

Association between the H*+L pitch accent and the prosodic prominence is the next issue to be considered. One way to capture the association is an alignment constraint between the H*+L accent tone and the DTE of a PWd. In Part II of the thesis, however, I argue that it is the alignment constraint between the H*+L accent and the DTE of a MiP, which is shown in (23).

(23) ALIGN_L (H*+L, DTE_{MiP})

Align the left edge of H*+L with the left edge of DTE of a Minor Phrase.

Since the DTE of a Minor Phrase is also the DTE of the most prominent PWd, alignment between the H*+L accent tone and the DTE of the MiP complies with the intuition that the H*+L is linked to the most prominent mora of a PWd. In Part II, I argue that the ALIGN_L (H*+L, DTE_{MiP}) constraint accounts for the observation made in that part that a post-FOCUS accented word forms its own Minor Phrase even in an environment where coalescence of the post-FOCUS accented word and the immediately preceding FOCUS word is preferred. Because the ALIGN_L (H*+L, DTE_{MiP}) is highly ranked and outranks the constraint which calls for such coalescence, the post-FOCUS accented word must have its own Minor Phrase because at most one DTE_{MiP} is allowed in a single Minor Phrase.

In summary, I argued in this subsection that the bitonal accent tone H*+L in Tokyo Japanese has correlation with prosodic prominence. That is, its default location is the DTE of a PWd (and a Minor Phrase). Also, the relationship between the pitch accent and prosodic prominence is consistent with the fact introduced in 4.3. of this chapter that the pitch accent tones receives more *extreme* F0 values than edge tones (i.e. the H* accent tone is realized higher than the preceding H edge tone, and the +L tone receives lower F0 value than non-post-accent L edge tones). This is because the phonetics distinguishes those two types of tones on the basis of whether they are associated with prosodic prominence or not, and assigns more *salient* value to the pitch accent tones, i.e. tones associated with the prosodic prominence. In Chapter 4, I give more detailed discussions of what those phonetic rules are.

PART I

THE GLOBAL VS. LOCAL VIEWS OF DOWNTRENDS

CHAPTER 2

INTRODUCTION TO PART I

The main concern of this part of the thesis is global vs. local aspects of downtrends, especially focusing on time-dependent declination (Chapter 3) and post-accent downtrend, i.e. *catathesis* (Chapter 4). Those two types of downtrends have been considered to be phenomena triggered by declining slope or lowering of an abstract level of phonetic representation, i.e. the *base line* or the *top line* of a pitch range, but not manipulation of relative values of tones within the pitch range.

More concretely, the time-dependent declination in languages is formalized as a globally declining slope of the base line (and the top line) of a pitch range, which unfolds over the whole utterance or across phrases (Gussenhoven & Rietveld, 1988; among others). Also, *catathesis*, the post-accent downtrend in Tokyo Japanese, is considered to be due to lowering and compression of the pitch range after each pitch accent (Pierrehumbert & Beckman, 1988). This pitch range lowering model of *catathesis* is also global in a sense that it affects all the tones following a pitch accent as long as they are within the *catathesis* domain.[11]

[11] The domain of *catathesis* is a Major Phonological Phrase.

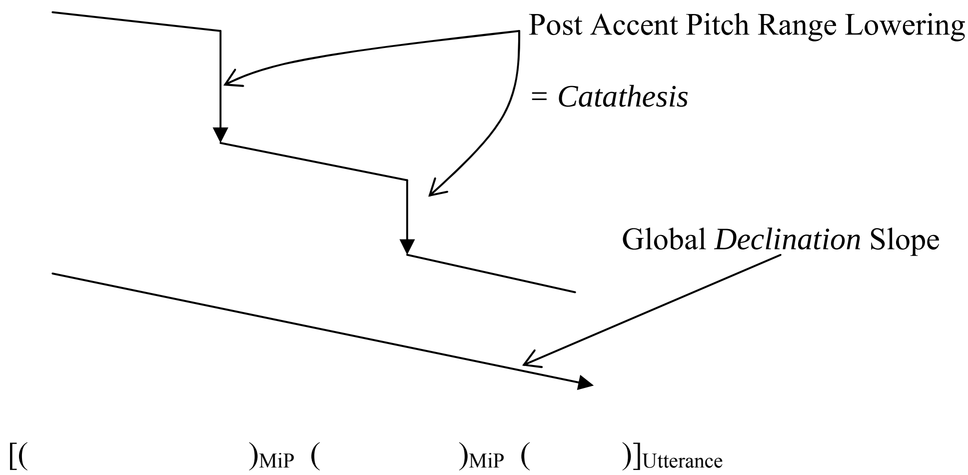


Figure 2.01. The Previous View of Declination and Catathesis

However, I show in Chapter 3 that though the global declination slope model captures one aspect of time-dependent downtrend, the model is both too simple and also incomplete because it ignores the significant effect of time on lowering of F0 within a more local domain, i.e. between two neighboring tones. Also, I show in Chapter 4 that the global model of *catathesis* resorting to pitch range lowering is inadequate and argue for a competing model of local tone-by-tone scaling.

The proposal in those two chapters, then, is a more complex and non-uniform picture of downtrend in Tokyo Japanese. That is, an adequate model is more complex that there are two types of time-dependent declination (i.e. a global declination slope unfolding over an utterance and a more local tone-bound declination slope). At the same time, this adequate model is not uniform in a sense that *catathesis* is not pitch range phenomenon while the time-dependent declination is a phenomenon determined by an aspect of the pitch range, i.e. by the slope of the pitch range.

CHAPTER 3

THE GLOBAL & LOCAL TIME-DEPENDENT DECLINATION

It is known that time has the effect of lowering F0 in many languages such as in Dutch (Collier & t' Hart, 1971; 't Hart & Cohen, 1973; Collier, 1975), in English (Maeda, 1976; Pierrehumbert, 1980), in Danish (Thorsen (Grønnum), 1980), in Swedish (Bruce, 1977) and in Japanese (Poser, 1984; Pierrehumbert & Beckman, 1988; Banno, 1999). Following the predecessors, I refer to the time-dependent downtrend of F0 as declination. The time-dependent declination is known for its global nature: the F0 value of tones gradually lowers (about -10 ~ -15 Hz per second) as the duration between them and the onset of an utterance increases. Given this nature, it is widely accepted that the time-dependent declination is due to subglottal air pressure decrease during the course of an utterance (Gelfer et al., 1983, 1985; Collier, 1975, 1985, Collier & Gelfer, 1984). With respect to speech perception of tonal values and prominence, the declination is formalized as a gradually declining slope of the *base line* of a pitch range unfolding over the whole utterance or across phonological phrase boundaries (Gussenhoven & Rietveld, 1988; among others). According to them, tones are superimposed on the gradually declining slope without changing their relative tonal values. Henceforth, I refer to the time-dependent declination of this character as "global declination".

However, there is little literature on time-dependent downtrend that systematically investigated its more local aspects, for example effects of passage of time on a local domain such as between two neighboring tones. My experimental work presented in this chapter reveals that the effect of time-passage within the local domain of two neighboring tones is far greater than the downtrend expected by the global declination. I argue for an additional “tone-bound” declination, which I refer to as “local declination” from time to time. The adequate model of downtrend should be able to provide a formal representation of the local declination as well as the global declination. I propose that tones be superimposed on tone-bound declination slope, a slope that expands only two neighboring tones, T1 and T2, which is reset every time the F0 target of T2 is achieved. This tone-bound slope is superimposed on the global declination slope.

The organization of this chapter is the following. In Section 3.1, I first introduce Pierrehumbert & Beckman's (1988) study of global declination in Tokyo Japanese, which shows that time is indeed one of the factors to contribute to downtrend in that language. They observed that the H edge tone of a following phrase underwent a time-dependent declination of about -10 Hz/sec as the time duration of the preceding phrase increased. The results of their experiment supporting global declination were also replicated by my own experiment, presented in Section 3.2. In that section, I will also show that a L edge tone which immediately precedes the H edge tone undergoes a more substantial time-dependent downtrend. I interpret this finding as evidence for an additional time-dependent factor, i.e. a more substantial downtrend than the global declination, which unfolds only between two neighboring tones. Based on that observation, a new downtrend model, which I call a "tone-bound" declination model, will be proposed in Section 3.3. In Section 3.4 and Section 3.5, I will test predictions made by the new declination model by examining the time-dependent downtrend of tones associated with accented Minor Phrases.

3.1. Global Declination: Pierrehumbert & Beckman's Result

Pierrehumbert & Beckman (1988) observed in Japanese that the F0 value of a tone decreases as the syllable number preceding the tone increases, even when there is a phonological phrase boundary between the target tone and the stretch of variable length. This observation is interpreted as evidence for a global time-dependent downtrend that unfolds across phonological phrases. This across-phrase, or utterance-level downtrend is usually referred to as declination. In this section, I briefly introduce the evidence for declination observed by Pierrehumbert and Beckman.

The form in (1) is a schematic representation of actual forms used in Pierrehumbert & Beckman's study to derive such a global time-dependent lowering effect. It consists of at least two Minor Phrases, MiP1 and MiP2. As mentioned in Chapter 1, each Minor Phrase has a sequence of L and H edge tones at its left edge. Pierrehumbert & Beckman assume that the L edge tone is associated with the initial mora of the following Minor Phrase. The H edge tone is associated with the second mora of each Minor Phrase. In Pierrehumbert & Beckman's experiment, the number of syllables in MiP1 was varied from three to nine syllables, more specifically the number of syllables between the second syllable and the final syllable of MiP1 was varied while the number of syllables in MiP2 was left unchanged. This is summarized schematically in (1).

(1) [σ σ --- syllable number varied --- σ]$_{MiP1}$ [σ σ σ]$_{MiP2}$

L1 H1 L2 H2

Pierrehumbert & Beckman found that the F0 peak of the second Minor Phrase (i.e. H2 of MiP2) decreased as the number of syllables between H1 and L2 (i.e. the number of syllables in MiP1) increased. They used this finding as evidence for a global declination which unfolds over Minor Phrase boundary. In the following part of this section, I briefly introduce their experimental procedures and results.

In their experiment, speakers were instructed to vary their pitch ranges when reading target forms. Given this, it was necessary to factor out any effect of the pitch range variation when analyzing the pure effect of the duration (or syllable number) variation on the F0 value of H2. In order to do so, they employed a residual analysis. Assuming that the F0 of H1 represents what the maximum height of the pitch range of a given utterance is, they first ran a regression analysis in which H2 is a dependent variable and H1 is a predictor. Figure 3.01 shows a plot of the regression analysis in Pierrehumbert & Beckman's study.

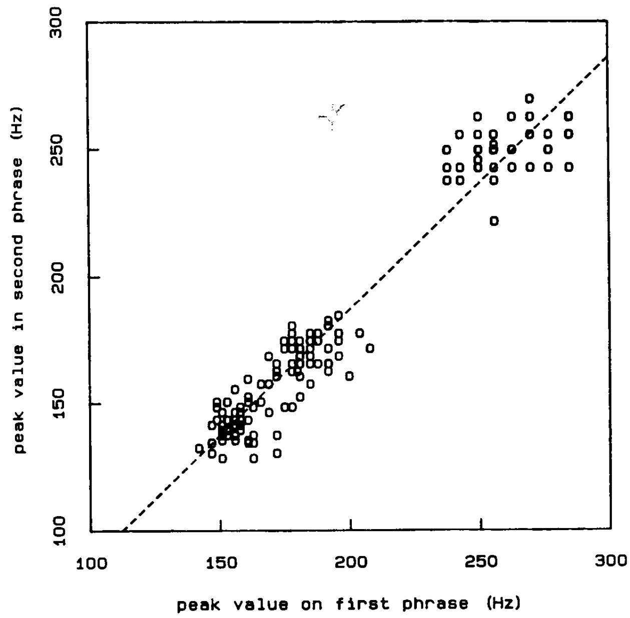


Figure 3.01. Pierrehumbert & Beckman's (1988) Figure (Figure 3.7: page 71)

Pierrehumbert & Beckman's (1988) figure (Figure 3.7: page 71) for the regression analysis between the F0 peak of the preceding phrase (the horizontal axis) and the F0 peak of the second phrase (the vertical axis).

The residual values obtained from this regression analysis are those from which the effect of H1, i.e. the effect of pitch range, are already factored out. If the variations among those residual values are well predicted by the syllable number variation (or the duration variation) preceding H2, then we can conclude that the duration has an effect on the F0 of H2. In order to see whether there is any correlation between those residual values and the duration (or syllable number) variation, they found the mean of those residuals by different syllable numbers. This is shown in Figure 3.02 (Pierrehumbert & Beckman's (1988) Figure 3.7, page 71).

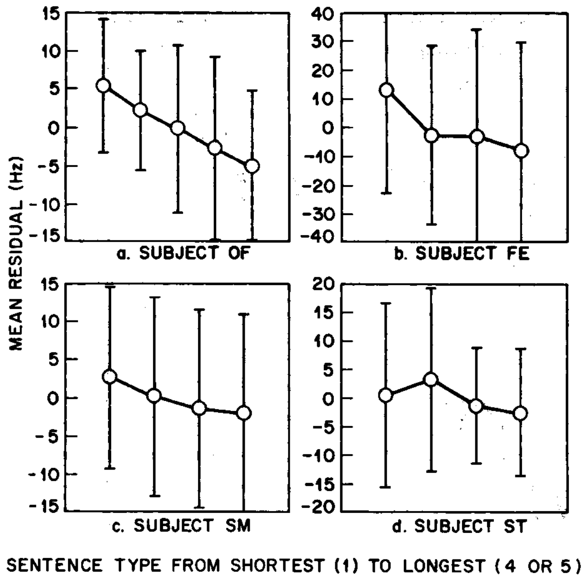


Figure 3.02. Pierrehumbert & Beckman's (1988) Figure (Figure 3.8: page 72) Pierrehumbert & Beckman's (1988) figure (Figure 3.8: page 72) for the relationship between the mean residual values (vertical axis) and phrasal length by syllable number (horizontal axis).

For the smallest syllable number in Figure 3.02 (Pierrehumbert & Beckman's (1988) Figure 3.8, page 72), the mean residual is higher than the expected F0 value of H2 which is represented by the regression on the pooled data. On the other hand, for the largest syllable number case, the mean residual is lower then the expected F0 value of H2. They estimated the declination rate by comparing the spread of the residuals with the corresponding time differences. According to them, this estimated rate is -10 Hz/sec for one of their male speakers.

In my experiment, too, I also used forms parallel to theirs and derived a similar result: H2 decreases about -8~-18 H/sec in female voice as the number of syllables between H1 and L2 (i.e. the syllable number in MiP1) increases. In the same experiment, I also found evidence for a more substantial time-dependent downtrend of L2, which was more than 25 Hz/sec for all of the speakers. I will then propose in Section 3.3 that this substantial downtrend of L2 is an instance of a "tone-bound" declination, which is reset each time the target of the tone with which it is associated is achieved

3.2. Global vs. Local Time-Dependent Downtrend in Tokyo Japanese

In this section, I provide results of my own production experiment related to time-dependent declination slope in Tokyo Japanese. The experiment was originally carried out to replicate the result of global declination obtained by Pierrehumbert & Beckman introduced in Section 1 of this chapter. As a byproduct of that experiment, however, a new finding was made: passage of time has not only correlation with the global declination but also with a steeper declining slope that unfolds within a smaller domain.

3.2.1. Experimental Materials

As for the reading material, the <*umaya*> set was used (see Appendix). This set contains sequences of two target words, which are shown in (2). Those target words are unaccented, and syntactically form an immediate constituent, i.e. there is no large syntactic boundary at the left edge of the second word. The number of syllables of the initial word was varied from three to twelve syllables. Except for the shortest one consisting of only three syllables, compound words are used for the initial word. It is because compound formation makes it easier to increase the number of syllables within a single word. Also, presence of a compound initial word is considered to prevent those two words coalesced into a single Minor Phrase (Section 1.4.3 in Chapter 1; Section 7.1 in Chapter 7).

(2) **The <*Umaya*> Set**

	Word1	**Word2**	
a,	*yama-no*	*umaya-no*	
	mountain-Gen	barn-Gen	
b.	*yamamura-no*	*umaya-no*	…
	(village name)-Gen	barn-Gen	…
c.	*yamanakamura-no*	*umaya-no*	…
	(village name)-Gen	barn-Gen	…
d.	*yamanakagawamura-no*	*umaya-no*	…
	(village name)-Gen	barn-Gen	…
e.	*minamiyamanakagawamura-no*	*umaya-no*	…
	(village name)-Gen	barn-gen	…

Speaker AS and Speaker MR almost always inserted a Minor Phrase boundary between those two words regardless of the length and compound/non-compound status of the initial word (Word1). That is, LH edge tones are present both at the left edge of the Word1 and at the left edge of the second word (Word2) in those two speakers' speech as shown schematically in (3).

(3) [σ σ --- syllable number varied --- $\sigma]_{\text{MiP1}}$ [σ σ $\sigma]_{\text{MiP2}}$
L1 H1 L2 H2

However, presence of the Minor Phrase boundary between those two words was not consistent in the other two speakers' speech, i.e. SK and NK, in spite of the fact most of the instances of the initial word (Word1) were compound forms. I excluded tokens without the Minor Phrase boundary between those two words because absence of LH edge tones at the left edge of Word2 makes it impossible to ask the core question, i.e. whether the H edge tone of Word2 undergoes time-dependent declination as the duration of the preceding word increases.[12] In the following sections, I refer to Word1 and Word2 as MiP1 and MiP2 respectively.

3.2.2. The Rate of Global Declination

In this section, I present an analysis of data and results of the global declination slope. The measurement points for this analysis are the peak F0 of MiP1 ([H1]) which corresponds to the H1 edge tone of that phrase, and the peak F0 of MiP2 ([H2]) which corresponds to the H2 edge tone of that phrase. At the same time, the duration between those two H tones in ms.

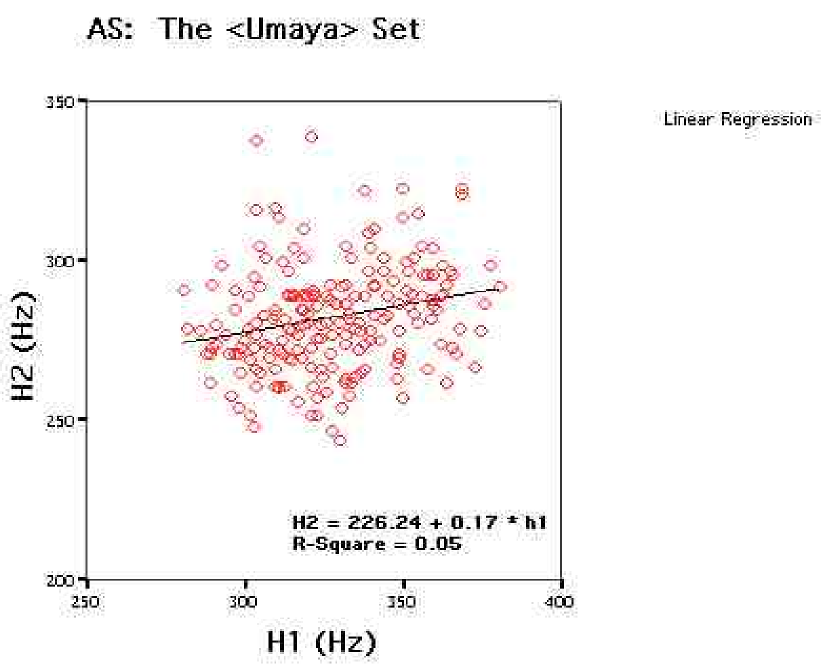
Though I did not ask speakers to vary their pitch ranges, it is still possible that there is a substantial variation in the F0 height of [H1]. Also, it is likely that the F0 value of [H1] has an effect on that of [H2]. Such effect of [H1] on [H2] needs to be factored out to obtain the effect of duration on the F0 of [H2]. Therefore, the same residual analysis procedure as Pierrehumbert & Beckman adopted was carried out.

[12] <The number and the percentage of tokens with an MiP boundary at the onset of Word 2 *umaya* >

# of syllables within the 1st Word	NK	SK
3	0/10 (0%)	1/20 (5%)
5	13/34 (38%)	18/32 (56%)
7 ~ 8	20/42 (32%)	32/50 (64%)
10	11/22 (50%)	8/12 (67%)
12	15/22 (68%)	9/12 (75%)

3.2.2.1. Obtaining Residual Values

The effect of [H1] on [H2] was first factored out by obtaining residual values from the regression analysis where [H2] is the dependent variable and [H1] is the predictor. Those residual values are, then, compared with the duration between [H1] and [H2]. If the residual values decline as the duration between those two tones increases, then it is a good indication that [H2] undergoes time-dependent declination. In the Figure 3.03, I present scatter plots showing the relation between those two measurement points.

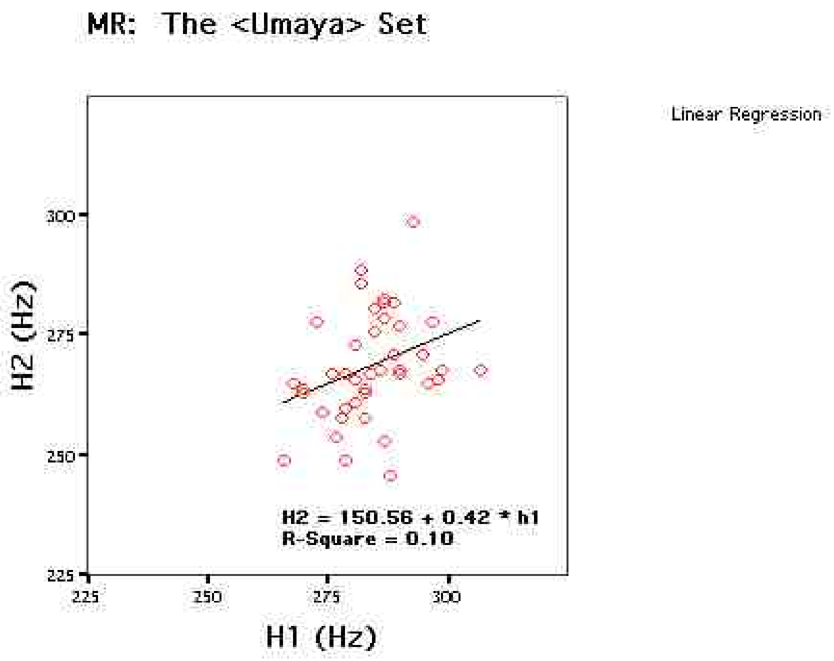


a. Speaker AS

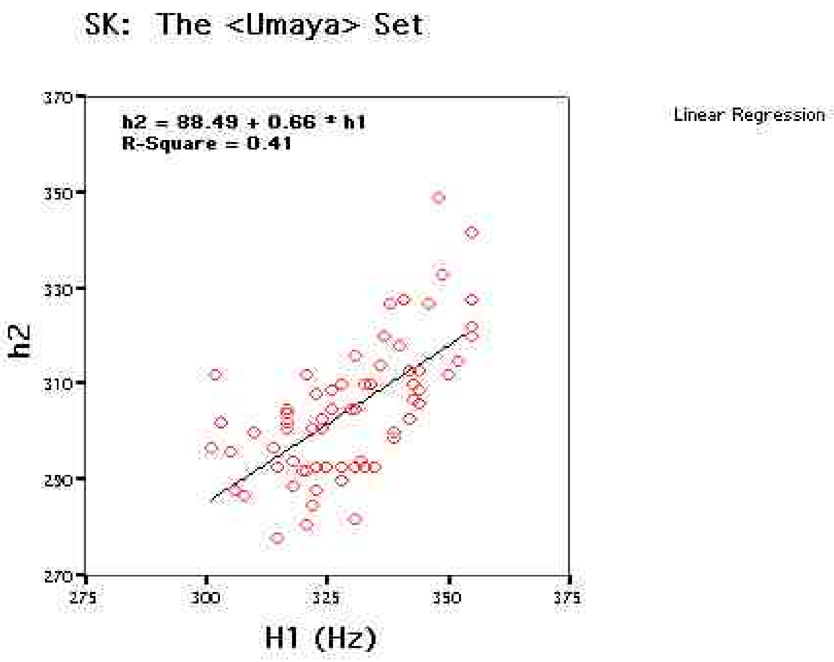
Figure 3.03. The Relation between H1 and H2

Continued next page

Figure 3.03 continued



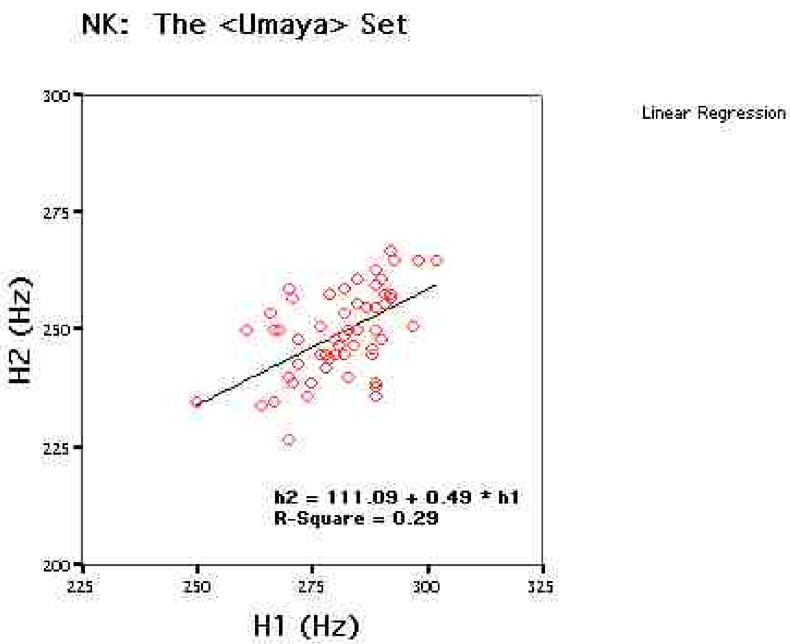
b. Speaker MR



c. Speaker SK

Continued next page

Figure 3.03 continued

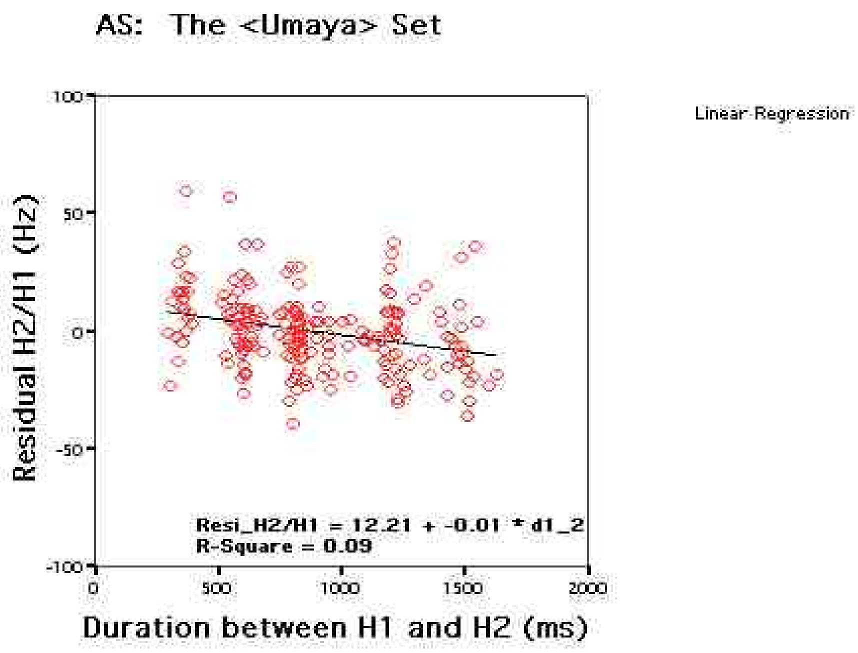


d. Speaker NK

For all the speakers, there is a significant positive correlation between [H1] and [H2] ([AS: $F(1, 199) = 11.258$, $p = 0.001$], [MR: $F(1, 41) = 4.66$, $p = 0.037$], [SK, $F(1, 66) = 46.24$, $p < 0.001$], [NK: $F(1, 57) = 23.79$, $p < 0.001$]). This outcome further motivates the necessity of the residual analysis. Unstandardized residual values of [H2] were, then, obtained. Those residual values were compared with the duration between [H1] and [H2] by another regression analysis.

3.2.2.2. Residual Values, Duration and Regression Analysis

In the subsequent regression analysis, the residual values of [H2] were made into the dependent variable and the duration between [H1] and [H2] were made into the predictor. The relationship between those two variables is shown in scatter plots of Figure 3.04.

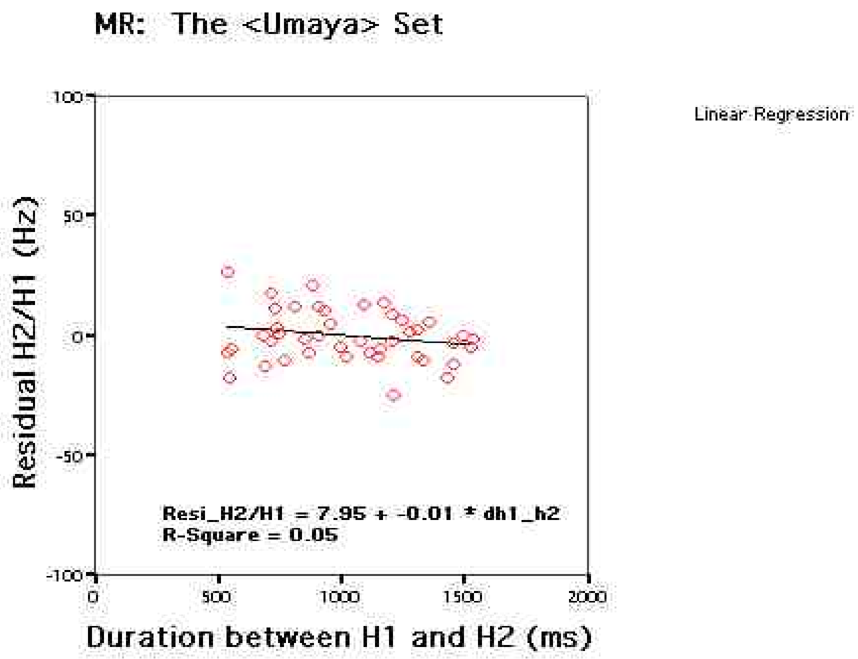


a. Speaker AS

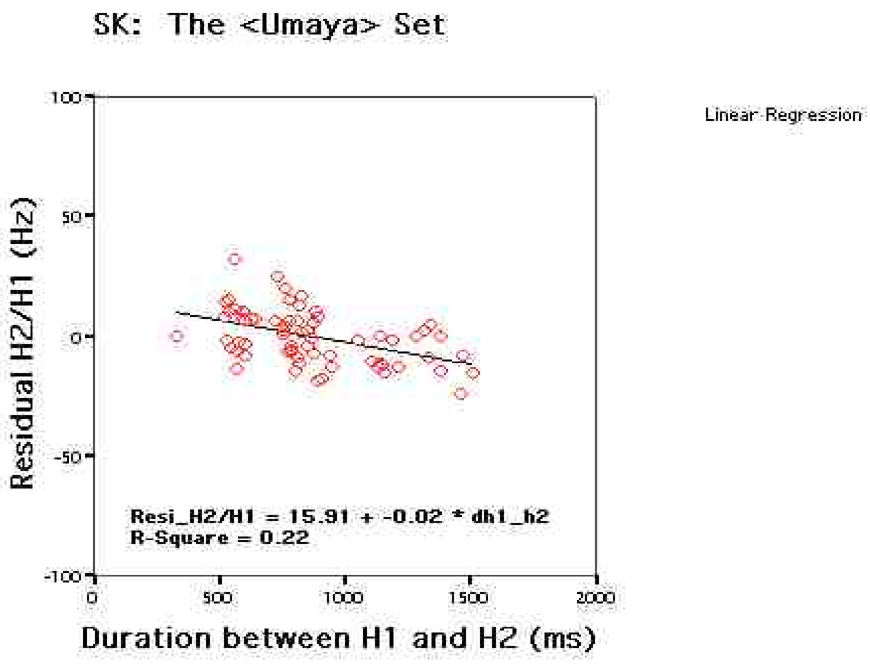
Figure 3.04. The Relationship between the Duration and Residual Values

Continued next page

Figure 3.04 continued



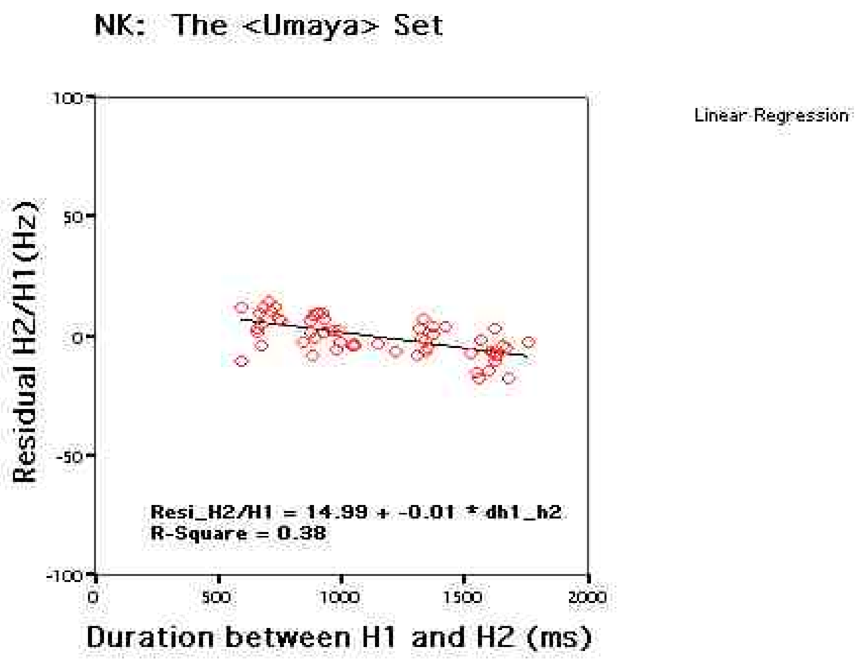
b. Speaker MR



c. Speaker SK

Continued next page

Figure 3.04 continued



d. Speaker NK

According to the scatter plots in Figure 3.04, for the shortest duration between [H1] and [H2] the residual values are far above zero, while for the longest duration those values are below zero. The declining regression slope is significant for three of the four speakers ([AS: coefficient = -0.014, Std Error = 0.0032, t = -4.33, p < 0.001], [SK: coefficient = -0.018, Std Error = 0.0043, t = -4.27, p < 0.001], [NK, coefficient = -0.013, Std Error = 0.0022, t = -5.92, p < 0.001]). This result confirms presence of the time-dependent declination slope across a Minor Phrase boundary. The only exception is MR. Though her [H2] residual data points are also fit to a negative regression slope, the slope is not statistically significant (coefficient = -0.008, Std Error = 0.0054, t = -1.41, p = 0.17).

The rate of the time-dependent declining slope of the residual values of [H2] is represented by the coefficient of the regression slopes obtained in the analysis (AS -0.014; MR -0.008; SK -0.018; NK -0.013. Those declination rates are, however, expressed in terms of *ms* (millisecond), and they are translated into the *sec* (second) terms. The rate per second is -14 Hz/sec for AS; -8 Hz/sec for MR; -18 Hz/sec for SK; and -13 Hz/sec for NK. This result is similar to the rate obtained by Pierrehumbert & Beckman's, i.e. about -10 Hz/sec.[13]

This is not, however, the only observation made in this experiment. The additional finding is provided in Section 3.2.3.

3.2.3. A Local Time-Dependent Downtrend

In addition to the global declination of the H2 edge tone between -8 and -18 Hz/sec, I found more substantial declination slope associated with “tonal alignment points” that precede the H2 edge tone. The declination rate of those points is –25 ~ -50 Hz/sec, which was far greater than the rate expected by the global declination factor only. I will interpret this finding as evidence for a more local declination slope on which each tone is superimposed and reset each time the target of that tone is achieved. The details of this "tone-bound" declination will be presented in Section 3.3.

[13] This across-phrase time-dependent declination rate of -8 Hz/sec to -18 Hz/sec is consistent with the downtrend of F0 triggered by subglottal air pressure decrease during the course of an utterance. In one of their experiments, Collier and Gelfer (1984) found that the rate of F0 declination induced by subglottal air pressure decrease was between 7 Hz/sec and 14 Hz/sec, which is parallel to the declination rate obtained in this experiment and Pierrehumbert & Beckman's experiment.

3.2.3.1. Time-Dependent Downtrend of the Offset of a Local F0 Trajectory Slope

In this subsection, I first provide results related to the time-dependent declination of "tonal alignment points" that have more local relationship with the F0 peak of MiP1 ([H1]). The previous section was concerned with the time-dependent declining slope of [H2]. However, this tonal target is not immediately adjacent to the preceding peak because there is always at least one tonal alignment point between those two points, i.e. the L2 edge tone. Depending on speakers, there is another tonal alignment point in addition to the L2 tone: the "end point" of the spreading of the preceding H1 edge tone (see Section 1.4.2 of Chapter 1 for more detailed discussions on spreading of H edge tones). The main concern of this section is the time-dependent downtrend of those tonal alignment point that immediately follows the F0 peak [H1] of MiP1.

As mentioned above, there are two types of speakers in terms of which point to be count as the tonal alignment point immediately following the F0 peak of MiP1. For those speakers who "interpolate" the H edge tone of the preceding phrase and the L edge tone of the following phrase, the L edge tone is the tonal alignment point that immediately follows the F0 peak of MiP1. AS belongs to this group of speakers. However, the rest of our speakers are all "spreading" speakers, who spread the H edge tone from the second syllable to the final syllable of the same Minor Phrase. As a result, there is a tonal alignment point at the final syllable of the Minor Phrase, i.e. the right edge of the H edge tone is aligned with the final syllable. This is graphically shown in Figure 3.05.

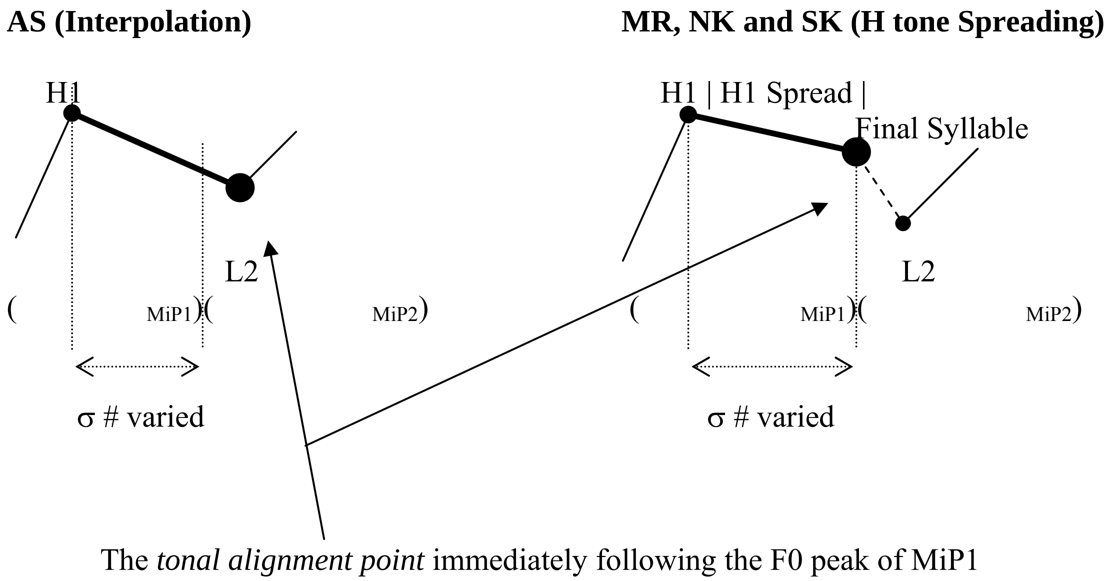


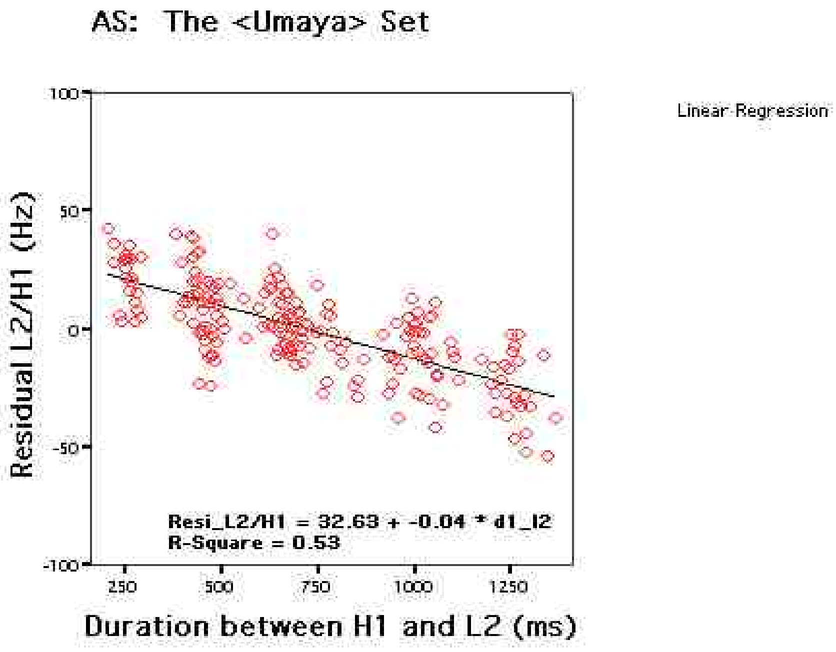
Figure 3.05. The Interpolation Speaker vs. H Tone Spreading Speaker

My interest in this subsection is how the tonal alignment points adjacent to the preceding F0 peak are affected by the syllable number variation that takes place between the F0 peaks of those points.

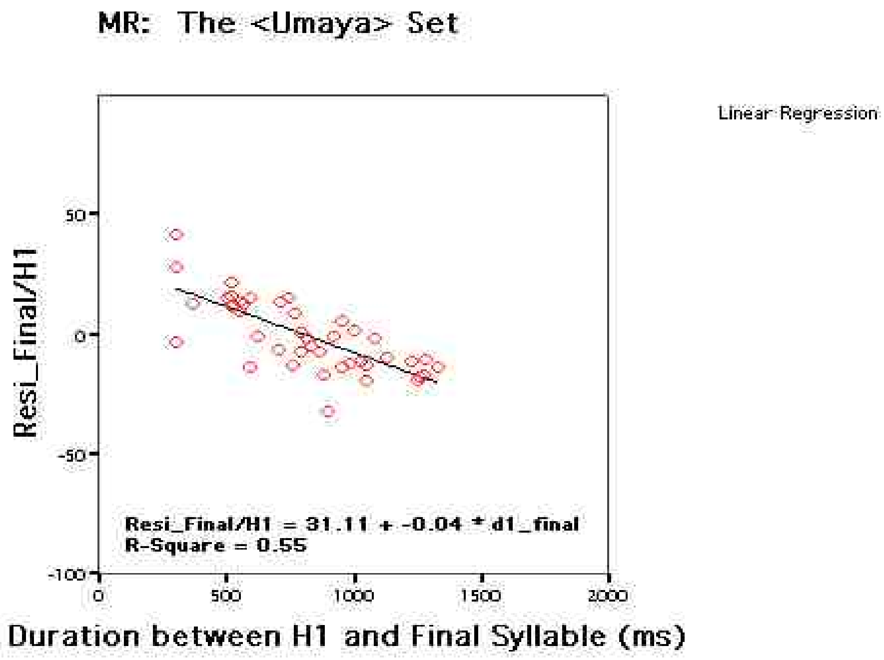
To investigate the question, a residual analysis comparable to those employed in the global declination analysis of the previous section was again adopted. Residual values of those tonal alignment points relative to the preceding F0 peak, which I call [H1], were first obtained. Second, another regression analysis was carried out in which those residuals were the dependent variable and the duration between [H1] and those tonal alignment points was the predictor. The result of this second regression analysis is discussed below and summarized in Figure 3.06. Henceforth, I refer to those tonal alignment points as [L2] for AS and as [Final Syllable] for the rest of the speakers.

In the scatter plots of Figure 3.06, the residual values of those tonal alignment points ([L2] and [Final Syllable]) are plotted on the vertical axis and the duration between the immediately preceding peak [H1] and those points are plotted on the horizontal axis. The data points of the residual values are fit to a declining regression slope. The coefficients of those regression slopes are all significant ([AS: coefficient = -0.045, Std Error = 0.003, t = -15, $p < 0.001$], [MR: coefficient = -0.039, Std Error = 0.001, t = -7.014, $p < 0.001$], [SK: coefficient = -0.046, Std Error = 0.005, t = -9.94, $p < 0.001$], [NK: coefficient = -0.025, Std Error = 0.002, t = -10.23, $p < 0.001$]). Those slope coefficient values are translated into the declination rates in terms of second: -45 Hz/sec for AS; -39 Hz/sec for MR; -46 Hz/sec for SK; -25 Hz/sec for NK.[14] Those declination rates are far greater than the declination rates obtained for [H2] in the previous section.

[14] The coefficients of the regression slopes shown in Figure 2.06 are round off, and are not optimally accurate.



a. Speaker AS

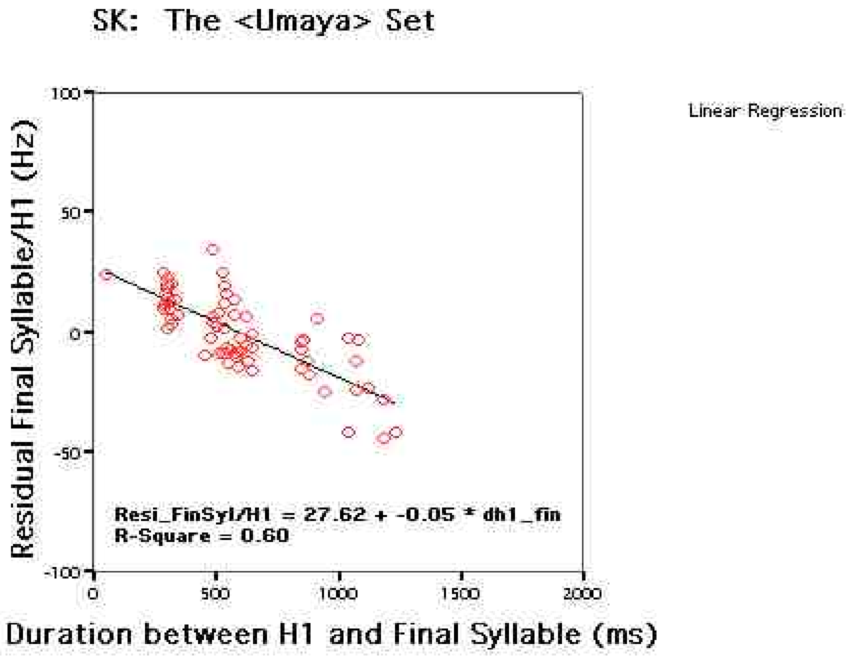


b. Speaker MR

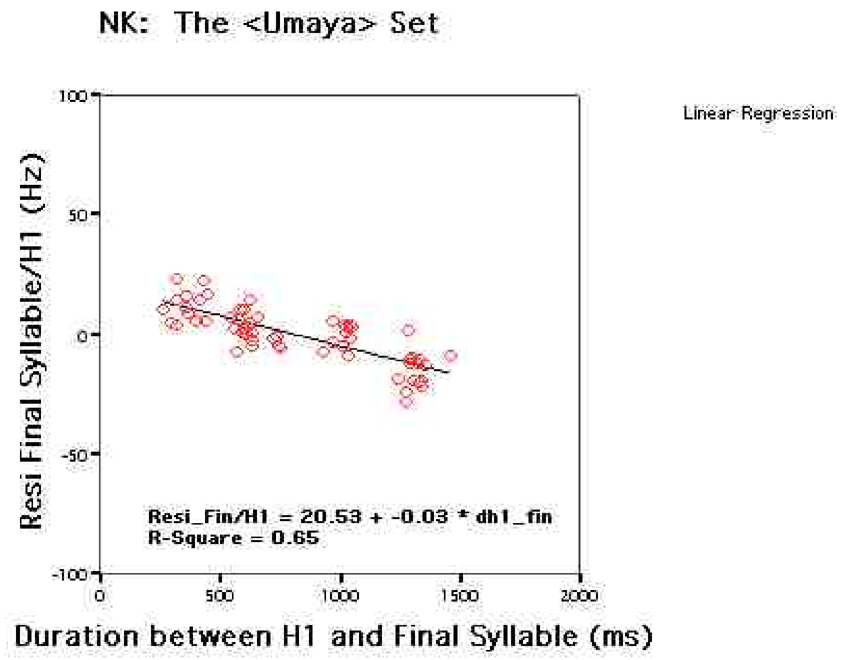
Figure 3.06. The Relationship between Duration and Residual Values

Continued next page

Figure 3.06 continued



c. Speaker SK



d. Speaker NK

Those steep declining regression slopes indicate that the time-dependent downtrend of those tonal alignment points considered in this subsection are qualitatively different from the time-dependent declination of the [H2] tone. The [H2] tone only underwent a time-dependent declination of -8~-18 Hz/sec, which is consistent with the subglottal air pressure decrease account of Collier (1984) and her colleagues. However, the time-dependent downtrend of -25~-40 Hz/sec obtained in this section is too large to be accounted for in that way. Also, it needs an account for why the [H2] edge tone that follows those tonal alignment points does not undergo such steep declination. In the later section of this chapter, I propose a new declination model with a “tone-bound” declination slope that keeps constantly resetting to the original point. Before proceeding to that section, let us consider the time-dependent downtrend of the L2 tone of the H tone spreading speakers.

3.2.3.2. The Time-Dependent Downtrend of L2 of the "Spreading Case"

In this section the downtrend of the L2 edge tone of Speaker MR, NK and SK, i.e. the H tone spreading speakers, is considered. In the last subsection, we only focused on the F0 value of the final syllable of MiP1 for those three speakers. In this subsection, however, we ask whether the L2 edge tone associated with the initial syllable of MiP2 of those speakers also undergoes such an acute time-dependent downtrend.

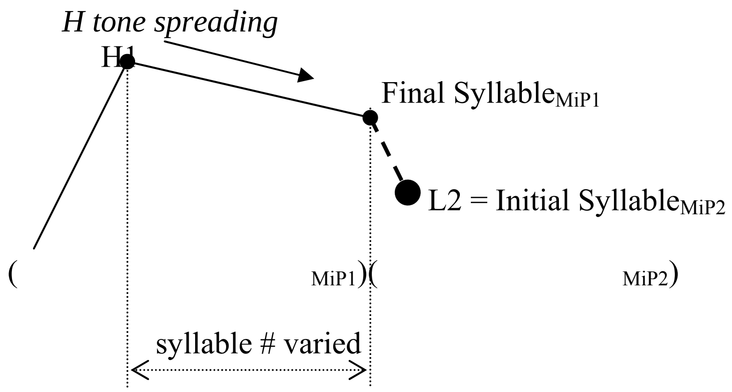
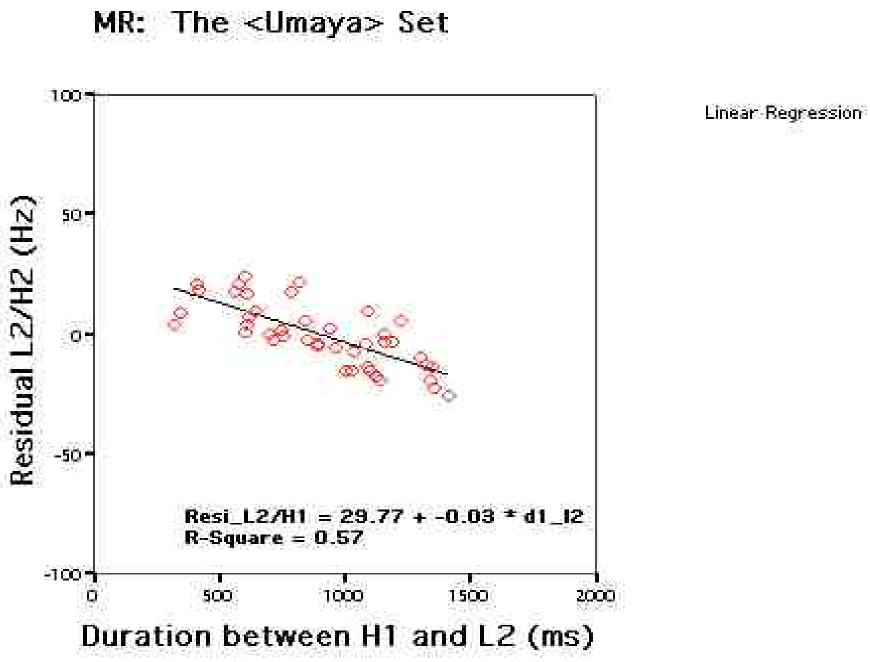
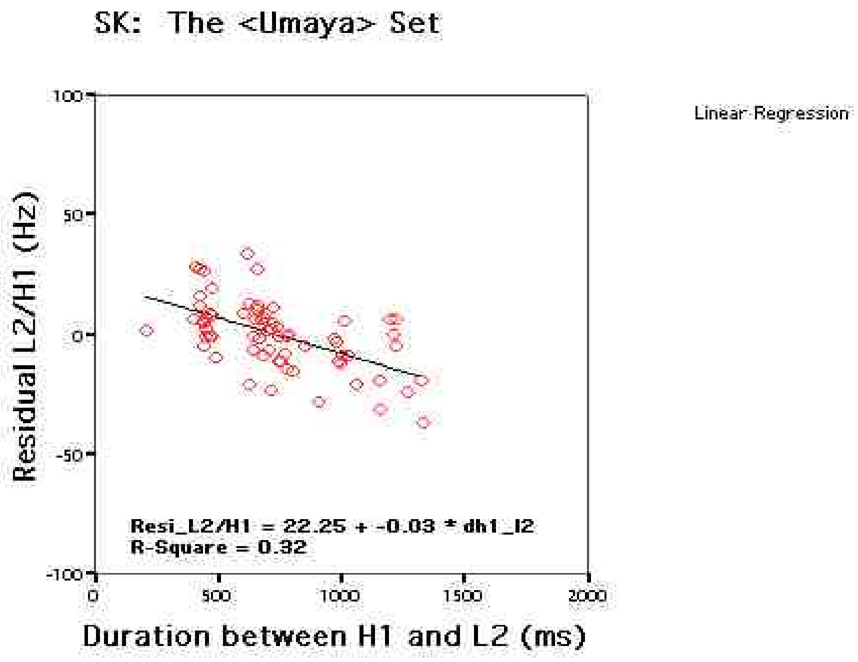


Figure 3.07. The H Tone Spreading Speakers (MR, NK and SK)

Again, a residual analysis comparable to those adopted in the previous sections was adopted. That is, the residual values of the L2 tone were first obtained in a regression analysis in which [L2] was the dependent variable and [H1] was the predictor. Then, another regression analysis was carried out to examine the relationship between those residual values of [L2] and the duration between [H1] and [L2]. In the second regression analysis, I found that the residual values of [L2] were fitted to a declining regression slope as the duration between those two points increases. This is graphically shown in Figure 3.08.



a. Speaker MR

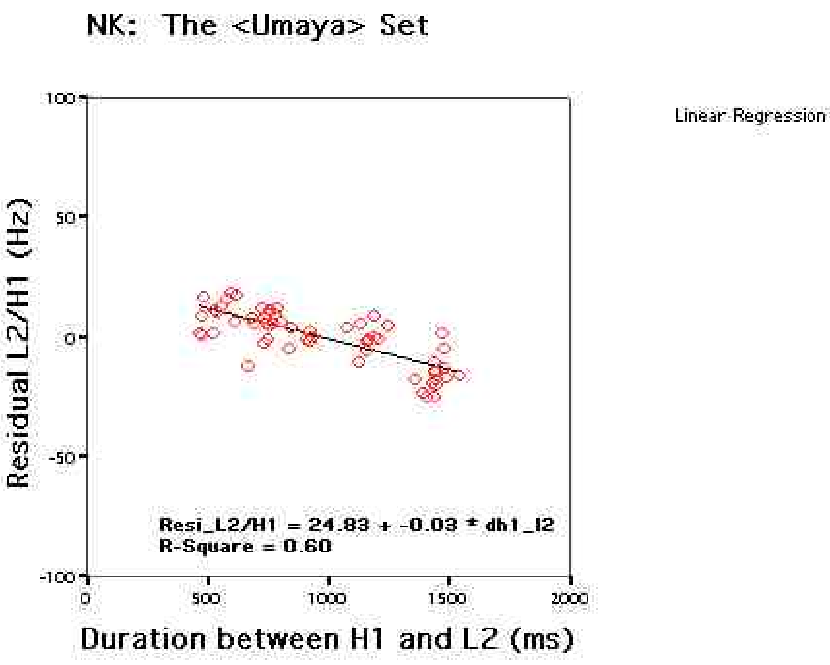


b. Speaker SK

Figure 3.08. The relationship between Duration and Residual Values

Continued next page

Figure 3.08 continued



c. Speaker NK

Those regression slopes were again steep, and the coefficients of those regression slops are all significant ([MR: coefficient = 0.033, Std Error = 0.005, t = -7.3, p < 0.001], [SK: coefficient = 0.030, Std Error = 0.005, t = -5.6, p < 0.001], [NK: coefficient = 0.026, Std Error = 0.003, t = -9.16, t < 0.001]). Those coefficient values of the declining regression slopes are translated into the declination rates in terms of a second: -33 Hz/sec for MR; -30 Hz/sec for SK; -26 Hz/sec for NK, and they are far greater than the declination rates obtained for H2. Given those results, I conclude that the L2 edge tone of those H tone spreading speakers also undergoes a substantial time-dependent downtrend qualitatively different from the global declination.

3.3. The "Tone-Bound" Declination Model

In the preceding section, I showed that for both the interpolation speaker and the H tone spreading speakers the L2 edge tone undergoes an acute time-dependent downtrend as the duration between the H1 edge tone and the L2 tone increases. Since there is an Minor Phrase boundary between those two tones, the domain of this substantial time-dependent downtrend is not confined to a single Minor Phrase. However, this substantial downtrend is not observed on the H2 edge tone, which immediately follows the L2 edge tone and it only undergoes a relatively moderate global declination. This is shown schematically in (4).

(4) [σ σ --- syllable number varied --- $\sigma_{Final}]_{MiP1}$ [σ σ $\sigma]_{MiP2}$

L1 H1 L2 H2

[substantial declination] [moderate declination]

The challenge, then, is to explain why both L2 and the offset of the spread H1 undergo this additional time-dependent lowering, while the H2 tone does not. In what follows, I propose a new model that captures both the acute local downtrend and the more moderate global declination in Tokyo Japanese.

The idea is that a pitch range undergoes both global declination and a more local downtrend. The global downtrend unfolds over an Utterance or an Intonational Phrase while the latter, the local downtrend, lasts only a small interval and is constantly reset. This local downtrend is reset immediately after the F0 target of some tone is achieved. After the reset, the pitch range undergoes a new local downtrend until the target of the next tone is achieved. I am calling this local downtrend “tone-bound” declination. More specific proposals about this tone-bound declination in this model are the following, and are depicted in Figure 3.09.

- Both the top line and the lower line of a pitch range undergo a tone-bound declination.
- This tone-bound declination of the top line and the lower line is reset to the original point whenever the F0 target of some tone is achieved.
- The original points to which the local downtrend slope of the top line and the lower line are reset undergo a global declination.

Based on those proposals, a visual representation of a pitch range which has undergone both the global declination and the local downtrend is presented in Figure 3.09.

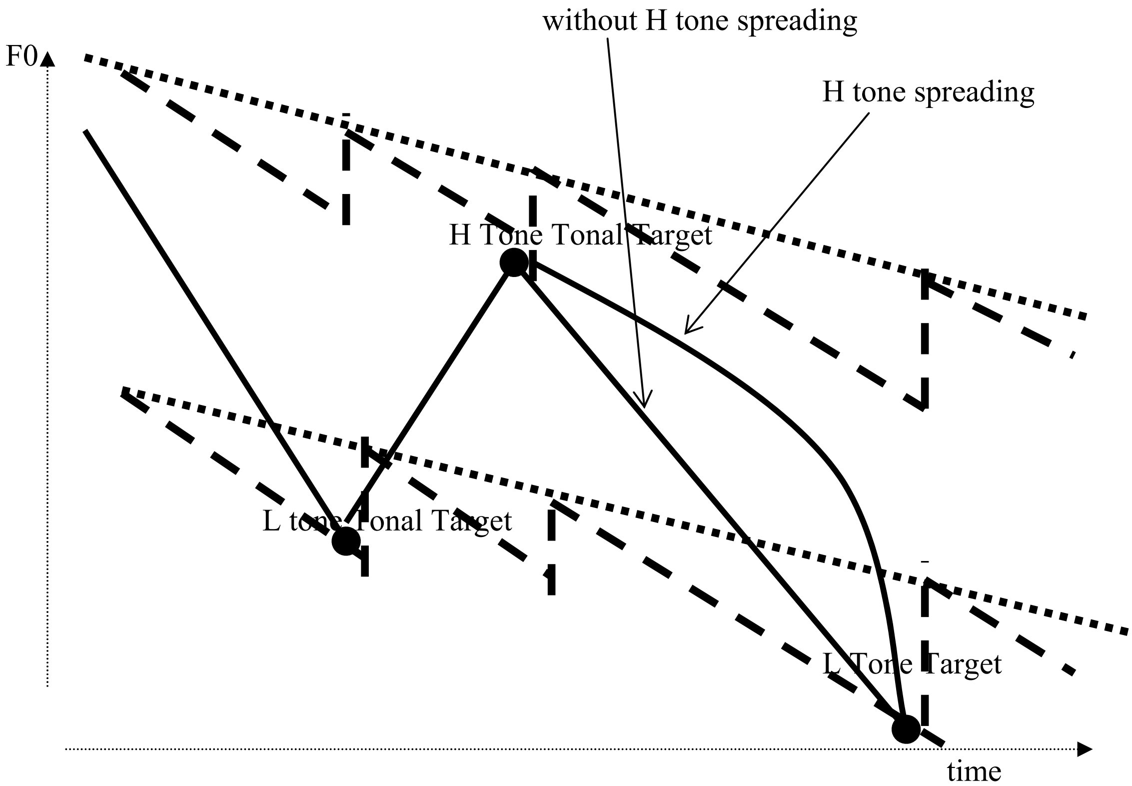


Figure 3.09. The Tone Bound Declination Model

Dotted lines represent points to which the local downtrend slope of the top line and that of the lower line are reset, and it undergoes a global declination. Dashed lines represent the local downtrend of the top line and the lower line of a pitch range. Filled dots represent tonal targets within a given pitch range, and solid lines represent F0 trajectories that connect successive two tonal target to each other.

Now imagine that there are three tones and two Minor Phrases: a H1 edge tone associated with the second syllable of MiP1, a L2 edge tone associated with the initial syllable of MiP2 and a H2 edge tone associated with the second syllable of MiP2. Also imagine that there are three different linguistic forms carrying those tones and Minor Phrases. One of those forms has a relatively shorter duration, say 300 ms, between H1 and L2; the second form has an intermediate duration, say 600 ms, between those two tones; and the other form has the longest duration, say 1000 ms, between them. I call the first the "shortest form", the second the "intermediate form" and the last the "longest form". For the shortest form, the local downtrend slope that starts immediately after the realization of the H1 edge tone lasts only 300 ms until the L2 edge tone appears. As a result, the L2 edge tone is realized relatively higher. On the other hand, the local downtrend slope of the intermediate form and the longest form lasts for 600 ms and 1000 ms respectively. As a result, the L2 edge tones of those two forms are realized relatively lower. Between those two forms, their L2 tone values are different: the L2 tone of the longest form is lower than that of the intermediate form. When it comes to the H2 edge tone that immediately follows the L2 edge tone, it is not affected by the local downtrend. This is because the local downtrend is reset to the "original" point once the L2 edge tone is realized. That original point to which the local downtrend slopes are reset undergoes a global declination. As a result, H2 edge tone only undergoes a global declination.

The remaining question is how the H tone *spreading* observed in the three of the four speakers' speech is dealt with in this model. The model should be able to capture the fact depicted in Figure 3.10 that both the [offset of the H tone *spreading*] and the immediately following L target undergo a similarly acute time-dependent downtrend, i.e. downtrend of -25 ~ -45 Hz/sec as the duration between the preceding F0 peak (i.e. the onset of the H tone) and the offset of the H tone spreading increases.

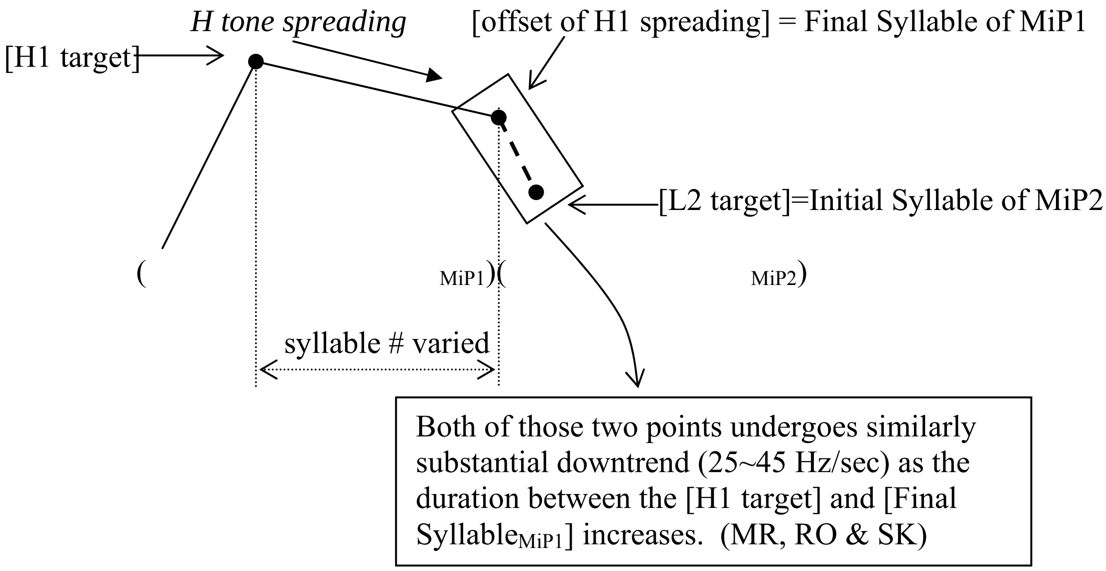
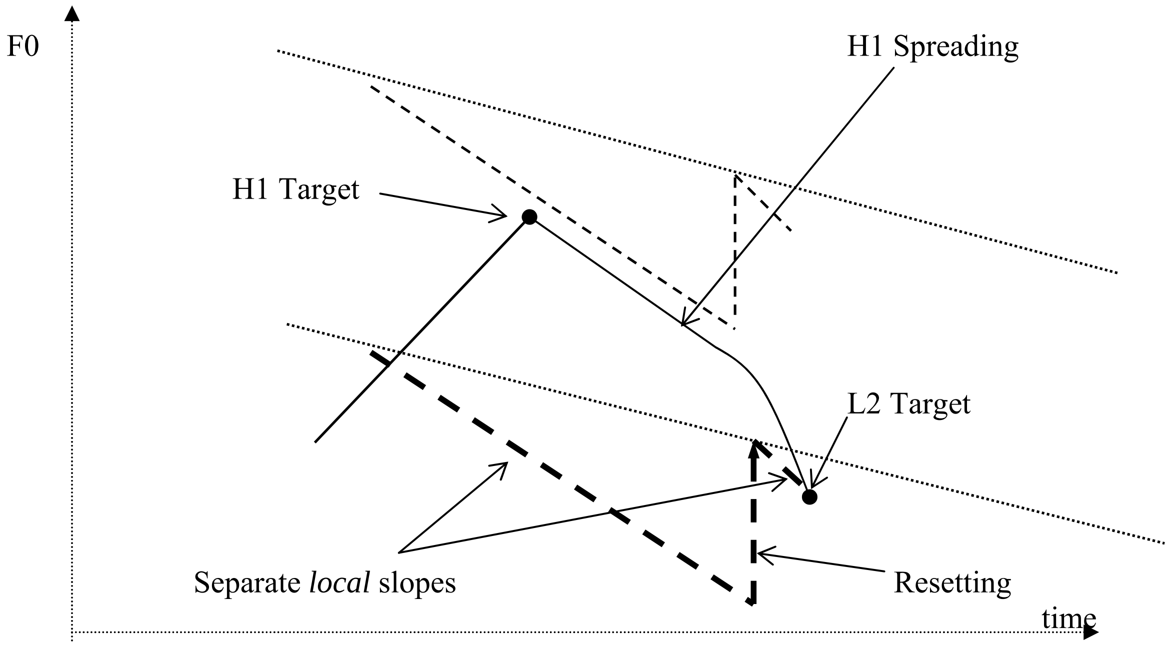


Figure 3.10. The Offset of H Tone Spreading and the Following L Target

The fact those two points undergo similarly substantial downtrend is <u>not</u> predicted correctly if the spreading part of H1 and the following L2 target are superposed on separate *local* declining slopes as visually shown in Figure 3.11-a.

If the spreading part of H1 and the following L2 target are superimposed on separate local slopes as shown in Figure 3.11-a, we wrongly predict that the offset of H1 *spreading* undergoes substantial declination (i.e. the more acute *local* downtrend) as the duration between the H1 target (i.e. the onset of H1 *spreading*) and the offset of H1 *spreading* increases, while the L2 target undergoes only a trivial declination (i.e. the *global* declination). It is because the *local* declining slope associated with the durational change between the H1 target and the offset of H1 *spreading* is reset to the *original* point immediately before the L2 target is achieved. To avoid such wrong prediction to be made, I propose that the spreading part of H1 and the L 2 target be superimposed on the same *local* declining slope as shown in Figure 3.11-b.

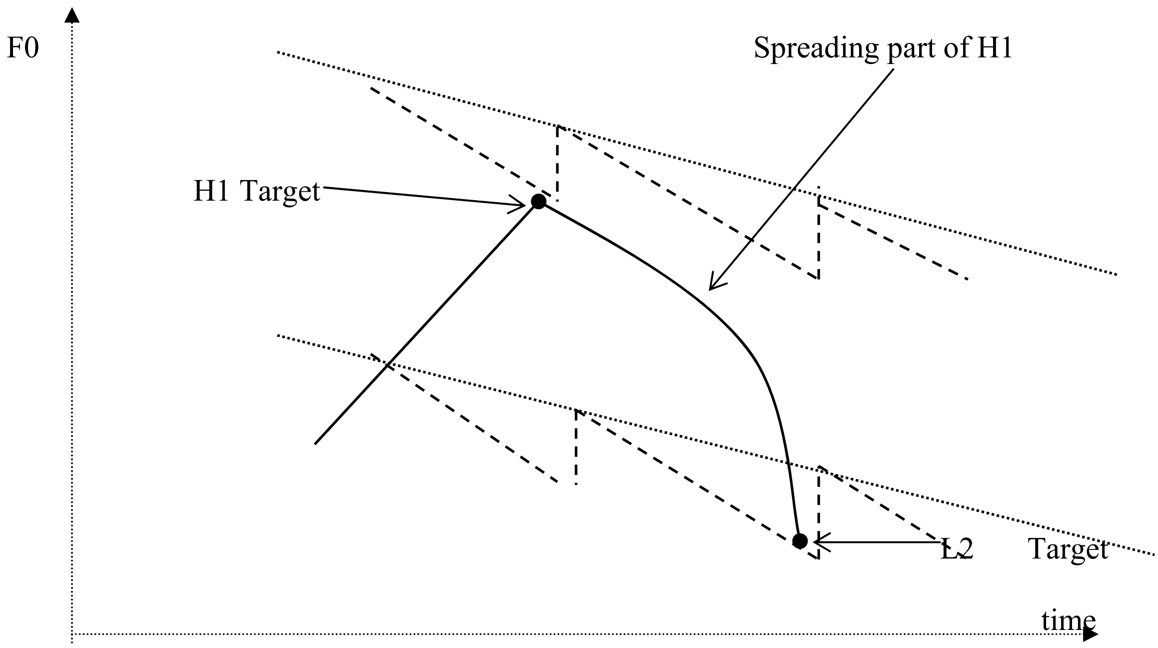


a. The spreading part of H1 and the L2 target are superposed on separate *local* declination slopes. (**Unwanted representation**)

Figure 3.11. The H Tone Spreading and the Tone Bound Declination

Continued next page

Figure 3.11 continued



b. Both the [offset of the *spreading* of H1] and the [L2 tonal target] are superposed on the same *local* declination slope.

In Figure 3.11-b, the *local* slope on which the target of H1 edge tone (i.e. the left edge of H1) is superimposed is reset to the *original* point as soon as the target is achieved. Then, a new *local* declining slope starts from there. It is this new *local* slope on which the spreading part of H1 and the target of the following L2 are both superimposed. The local slope on which both the spreading part of H1 and the following L2 are superimposed undergoes substantial lowering as the duration between the H1 target and the offset of H1 spreading increases. As a result, both the [offset of H1 spreading] and the [L2 target] undergo similarly substantial time-dependent lowering as the duration between the onset of H1 and the offset of the H1 spreading increases.

3.4. Testing Predictions 1: When Duration between H and H* is Varied

The model proposed in Section 3.3 makes interesting predictions about downtrend of tones associated with accented Minor Phrases in which the duration between the H edge tone and the H* accent tone is varied.

As already introduced in Chapter 1, there are two types of Minor Phrases in Tokyo Japanese: one is accented and the other is unaccented. An accented MiP may have a sequence of two H tones: the H edge tones and the H* accent tone as long as H* appears at the third or later syllable. After the H* accent tone, two L tones appear: the +L trailing tone of the pitch accent and the L edge tone that is associated with the left edge of the following Minor Phrase. If the duration (i.e. syllable number) between those two H tones (H and H*) is varied while that between H* and the following L tones is kept constant, then the declination model proposed in Section 3.3 predicts that the H* tone undergoes the substantial local time-dependent downtrend while the following L tones do not.

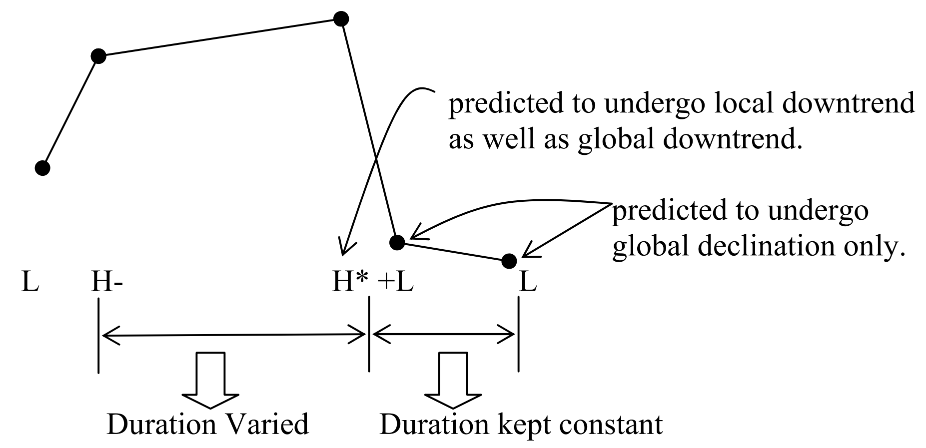


Figure 3.12. The Accent +L and the following L Edge Tone

As for the H* tone, the model predicts that it undergoes an acute time-dependent downtrend as the duration between the preceding H edge tone and the accent H* tone increases. This is because the longer the duration between those two tones is, the lower the local downtrend slope of a pitch range reaches. Once the H* accent tone is achieved, however, the local downtrend is cancelled and reset to the *original* point. As a result, following L tones (i.e. the +L tone and the L edge tone of the following Minor Phrase) are predicted undergo trivial *global* declination only.

Those predictions are tested by examining the F0 values associated with compound nouns. I use those materials because compound nouns make it possible to vary the syllable number between a H edge tone and the following H* accent tone just by adding a word consisting of different number of syllables between the initial and the final compound member. In Japanese compound nouns, the H*+L pitch accent is associated with the initial syllable of the final compound member when the final member consists of three or four morae (Kubozono, 1995; Kubozono et al., 1995). For example, if the final member of a compound noun is *mónaka* "bean cake", then its initial syllable *mo* is associated with the pitch accent H*+L because it is the initial syllable of the final member. By varying the length of the preceding member of the compound, we may increase or decrease the duration between the pitch accent H* (i.e. the initial syllable of the last member of the compound) and the preceding H edge tone. In the following, I present compound nouns from the <*Maronmónaka*> Set used in this experiment.

(5) Compound Forms from the **<*Maronmónaka*> Set**

(a) [*maronmónaka-o*]$_{MiP}$ (chestnut monaka-Acc)
H- H*+L L

(b) [*maroniromónaka-o*]$_{MiP}$ (chestnut-colored monaka-Acc)

H- H*+L L

(c) [*maronkuriimumónaka-o*]$_{MiP}$ (chestnut-cream monaka)
H- H*+L L

(d) [*maronaisukuriimumónaka-o*]$_{MiP}$ (chestnut-ice cream monaka)
H- H*+L L

In the <*Maronmónaka*> Set, those compound nouns are immediately followed by an accusative case particle [*o*]. Those compound nouns together with the accusative case marker formed a single Minor Phrase and were embedded in the following sentence.

suruto	*_____(compound form)__-o*	*kago ippai-ni mótta*	*onnánoko-ga*	*toorikakarimásita.*
then	______________________-Acc	full of basket	girl-Nom	passed by.

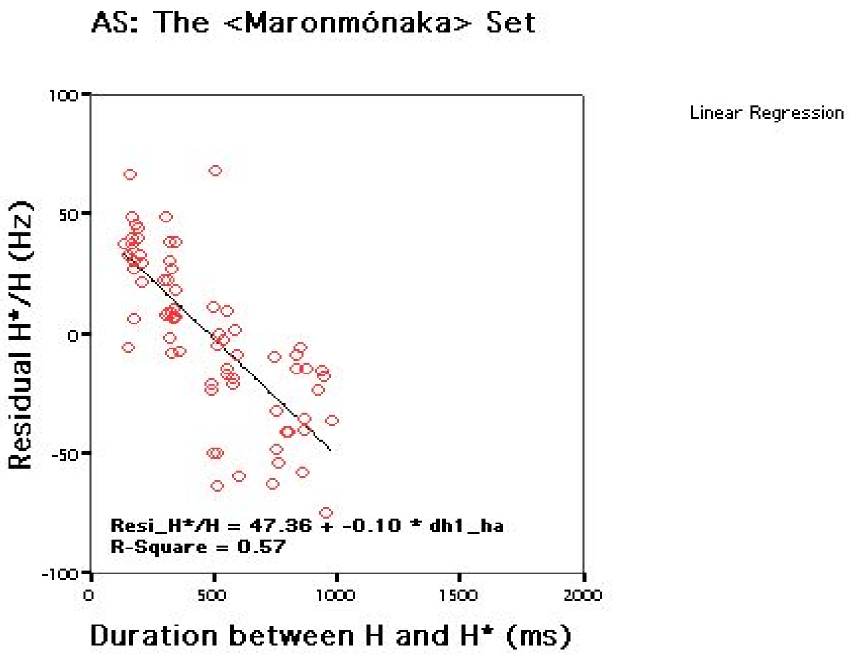
" Then, a girl with a basket full of ___(target form)___ passed by."

3.4.1. The Time-Dependent Downtrend of H*

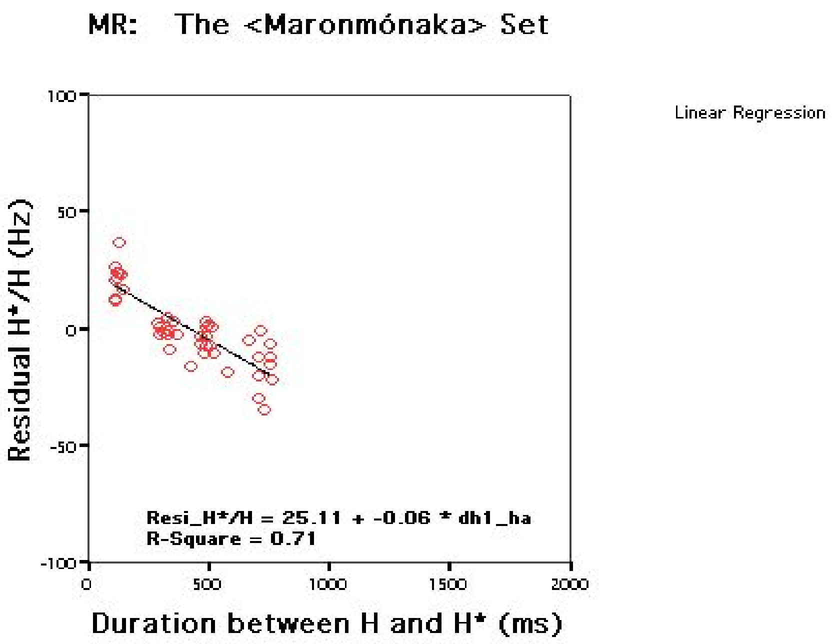
One of the predictions is that the H* accent tone undergoes a drastic downtrend as the duration between the preceding H edge tone and the H* accent tone increases. This downtrend would be comparable to the downtrend of the L2 edge tone observed in Section 3.2. There are two measurement points for testing this prediction. One is the H edge tone target which is associated with the second syllable [*ron*]. The other was the around the accented syllable [*mó*], the onset of a sharp accent fall, representing the F0 of the H* accent tone. I call the former [H] and the latter [H*]. At the same time, the duration between those two points in ms was obtained.

To examine the time-dependent downtrend of [H*], the residual analysis analogous to those adopted in the previous sections are again used. That is, the residual values of [H*] were first obtained from the regression analysis in which the preceding [H] was the predictor and [H*] was the dependent variable. As already mentioned in the preceding sections, those residual values represent relative height of the target tone, in this case [H*], i.e. whether [H*] is higher than its expected value represented by the point on the regression line given a certain value of [H]. If those residual values of [H*] decrease as the duration between the preceding [H] and [H*] increases and the rate of the decrease is far greater than that of the global declination, then it will be concluded that [H*] also undergoes a local time-dependent downtrend.

Scatter plots in Figure 3.13 graphically shows the relationship between the residual values of [H*] and the duration between [H] and [H*].



a. Speaker AS

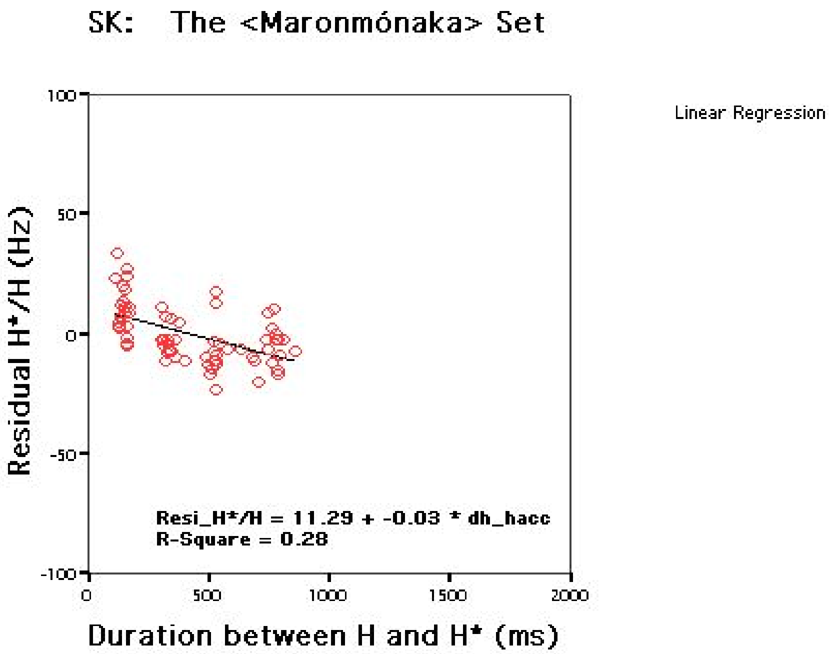


b. Speaker MR

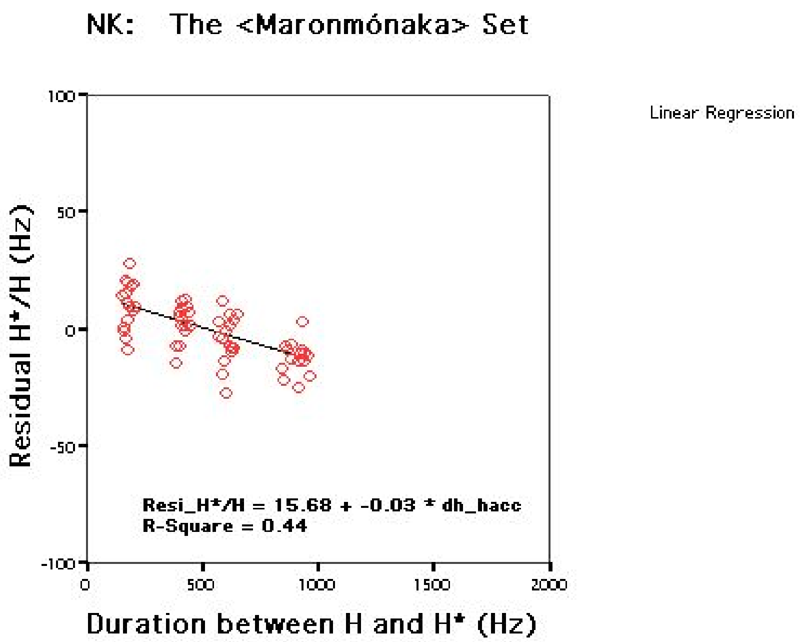
Figure 3.13. The Relation between Duration and Residual Values

Continued next page

Figure 3.13 continued



c. Speaker SK



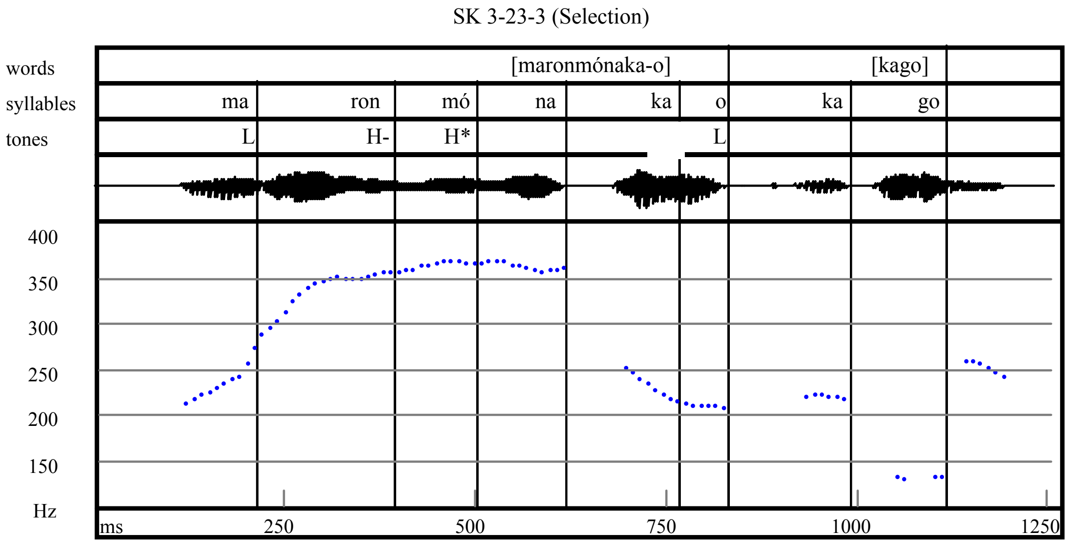
d. Speaker NK

For all speakers, the data points of the residual values are fitted to a declining regression slope. These declination slopes (i.e. the coefficients of the regression slopes) are all statistically significant ([AS: coefficient = -0.098, Std Error = 0.01, $t = -9.57$, $p < 0.001$], [MR: coefficient = -0.059, Std Error = 0.006, $t = -10.12$, $p < 0.001$], [SK: coefficient = -0.026, Std Error = 0.005, $t = -5.23$, $p < 0.001$], [NK: coefficient = -0.030, Std Error = 0.004243, $t = -6.98$, $p < 0.001$]). Those declination rates are translated into -26~-98 Hz/sec, which are far greater than the global declination. Given this, I conclude that the H* accent undergoes the local time-dependent downtrend, and this conclusion is in favor of the prediction made by the downtrend model proposed in Section 3.3.

3.4.2. The Time-Dependent Downtrend of the Post Accent L Tone

Another prediction made by the downtrend model that I proposed in Section 3.3 is that the L tones that follow the H* accent tone (i.e. +L of the bitonal pitch accent and L edge tone of the following Minor Phrase) do not undergo such a drastic downtrend because a local downtrend slope is reset to the original point immediately after the target F0 value of the H* accent tone is achieved.

Ideally, F0 of both the +L tone and the following L edge tone should be measured and their downtrend should be examined. However, as discussed more in detail in Chapter 4, there is no predetermined location for +L tone alignment and detecting the +L tone alignment point is relatively a complex task. As a result, I did not include any measurement of the +L tone in the following analysis. I measured the right edge of the accusative particle [*o*] whose F0 was approximately the F0 of the L edge tone associated with the initial syllable of the following phrase. Since initial syllable of the following phrase starts with an obstruent consonant [*k*] as already shown above, that initial syllable was not ideal for measuring F0. For this reason, the right edge of the accusative marker [*o*] was measured instead. I refer to this measurement point as [L] from now on. Our question, then, is whether [L] undergoes substantial downtrend as the syllable number between the preceding [H] edge tone and the [H*] accent tone is increased.



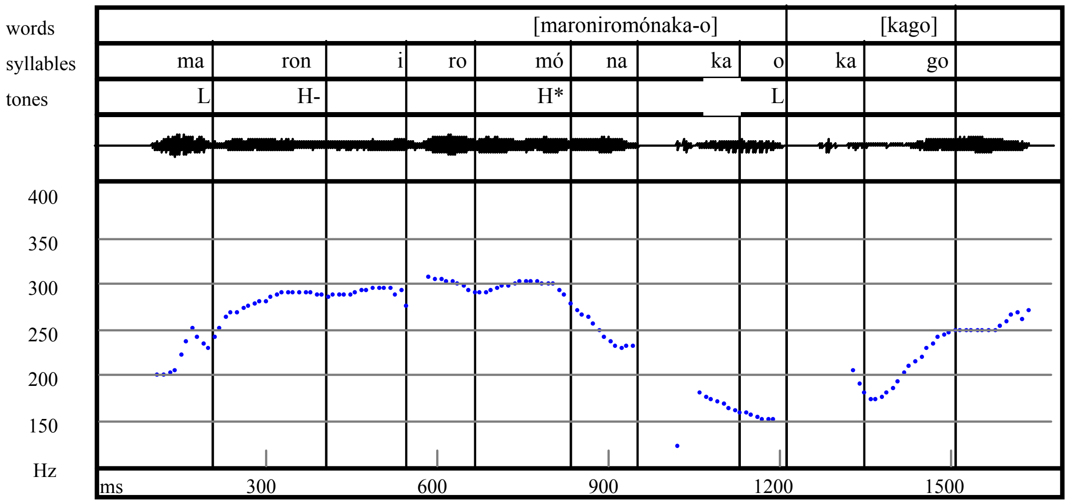
a. Speaker SK

Figure 3.14. Example F0 Contours of [maronmónaka-o]$_{MiP}$

Continued next page

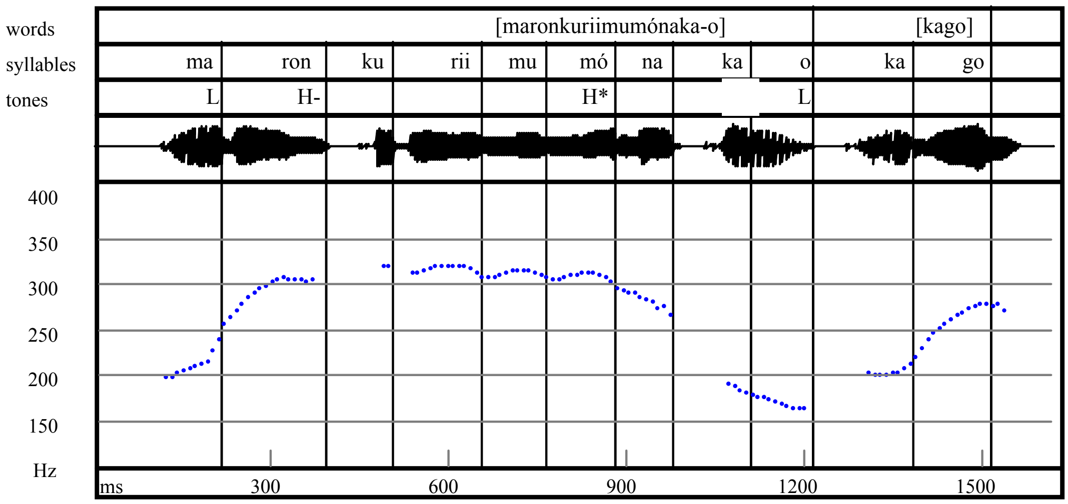
Figure 3.14 Continued

NK 3-22-3



b. Speaker NK

MR 2-24-3

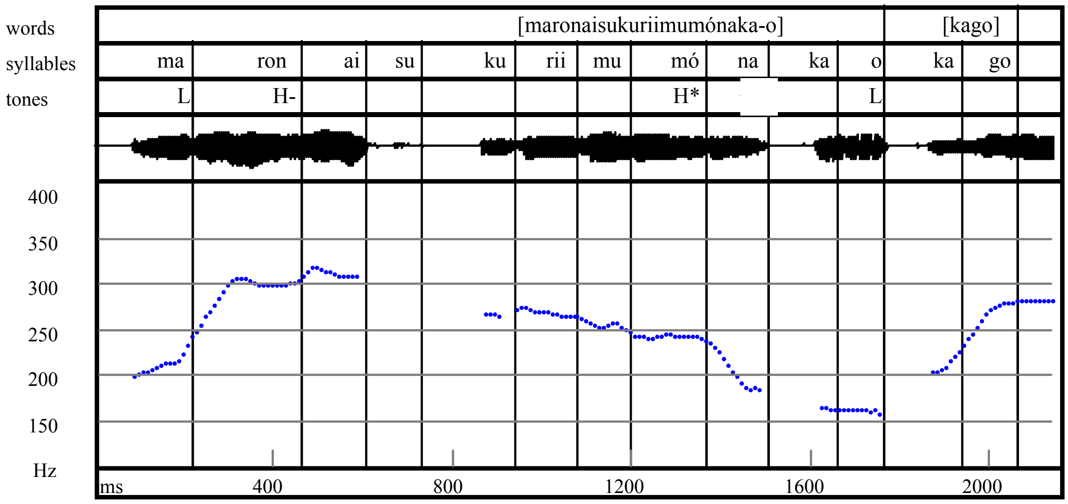


c. Speaker MR

Continued next page

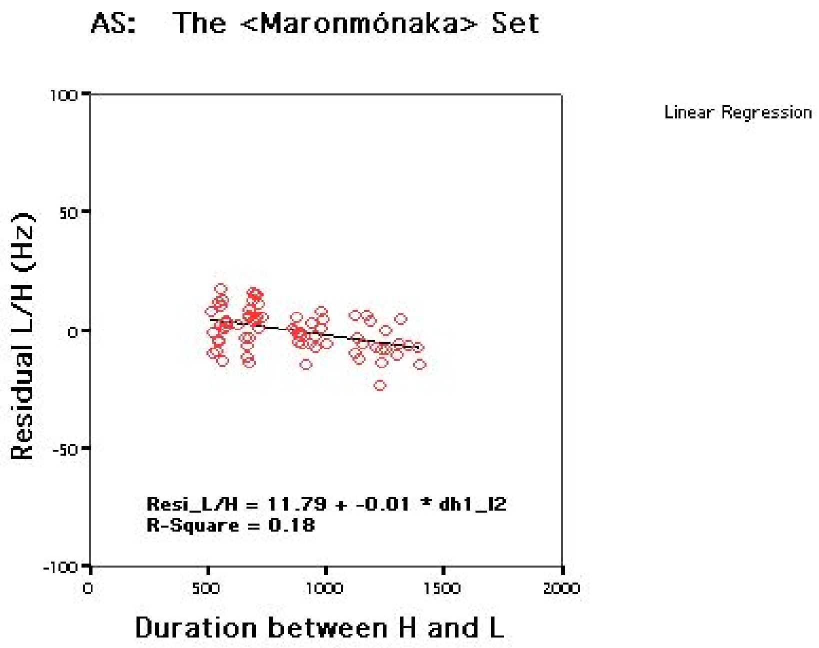
Figure 3.14 continued

AS 3-21-3



d. Speaker AS

To make the analysis in this section parallel to those already adopted in the analyses of the preceding sections of this chapter, a residual analysis was again adopted. In all of those previous analyses, the residual values of target tones were obtained by running a regression analysis in which the target tone is the dependent variable and the preceding [H] edge tone is into the predictor. In this analysis, too, residual values of [L] were first obtained in a regression analysis in which [L] was the dependent variable and the preceding [H] was the predictor. Then, those residual values of [L] were compared with the duration between [H] and [L]. If the residual values undergo a drastic downtrend as the duration increases and the magnitude of the downtrend is compatible with the downtrend of the H* accent tone already observed in Section 3.4.1, [L] also undergoes a local time-dependent downtrend as well as the global declination. In the following, I present results obtained from this residual analysis in Figure 3.15.

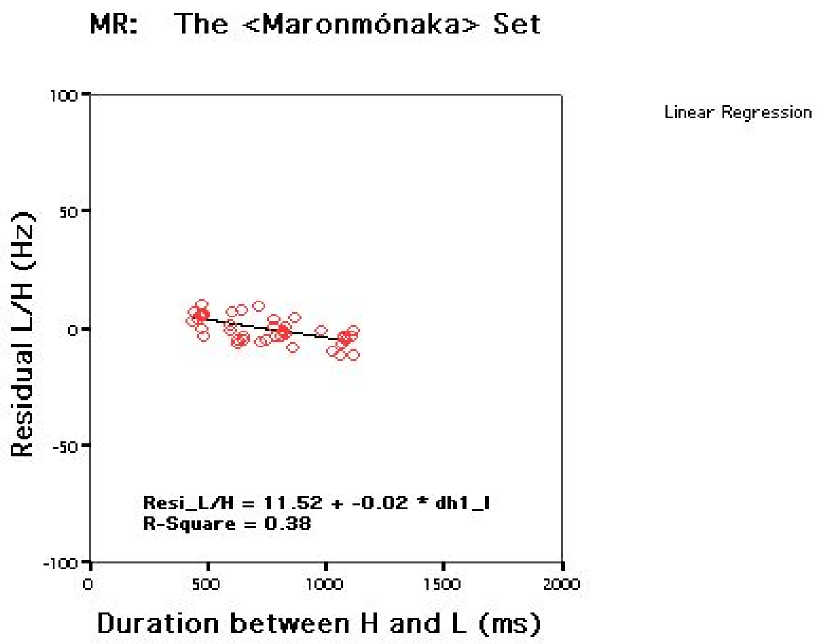


a. Speaker AS

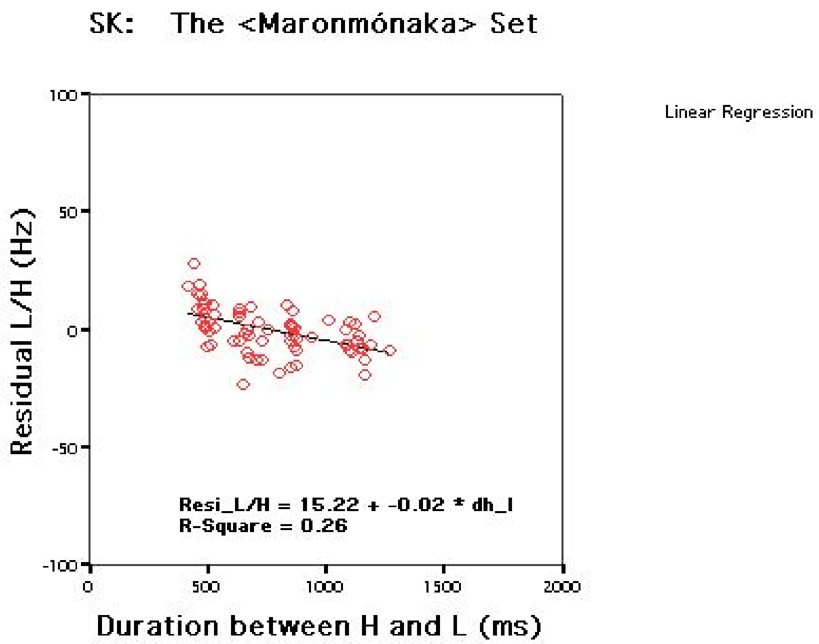
Figure 3.15. The Relation between Duration and Residual Values

Continued next page

Figure 3.15 Continued



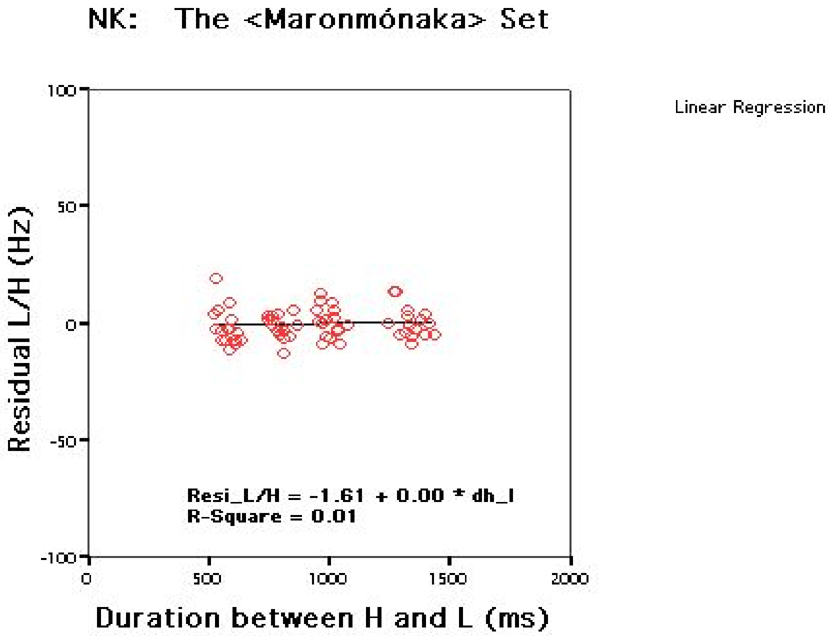
b. Speaker MR



c. Speaker SK

Continued next page

Figure 3.15 continued



d. Speaker NK

It turned out that not all the four speakers have a straightforward result though the results obtained from AS's and MR's speech perfectly comply with the predictions made by the downtrend model proposed in Section 3.3, i.e. the post-accent L boundary tone only undergoes a small amount of time-dependent declination which is below 20 Hz/sec.

Let us first consider those results obtained from AS's and MR's data which are in favor of the prediction made by my downtrend model. The coefficient value of their regression slopes vary between -0.014 and -0.015, and those coefficients are significant ([AS: Std Error = 0.003, $t = -2.96$, $p < 0.001$], [MR: Std Error = 0.003, $t = -4.19$, $p < 0.001$]). Those coefficients of the regression slops are translated into declination rate in terms of *sec* (second): -14 Hz/sec for AS and -15 Hz/sec for MR. This time-dependent downtrend of -14 ~ -15 Hz/sec is nothing more than the global declination, which is predicted to be always present by my model.

However, when it comes to the result obtained from NK's and SK's data, some additional explanations are necessary to make it fitted to the prediction made by the downtrend model. First of all, the residual values of [L] obtained from NK's data do not undergo time-dependent downtrend of any sort (Figure 3.15-d). That is to say, those residual values and the duration between the preceding [H] and [L] have no correlation. This may be explained by a possible "bottom" effect. In Tokyo Japanese, L tones that immediately follow a pitch accent H* are usually realized extremely low (Chapter 4). It is likely that all the post-accent L tones of NK hit the very bottom of her pitch range regardless of what the duration factor is. As a result, no time-dependent downtrend is any longer available. Also, this explanation is not incompatible with my downtrend model because I can simply add to the model that no downtrend slopes penetrate the very bottom limit of one's pitch range and keep going lower that line, or all the downtrend slopes stay level once they hit the very bottom limit of one's pitch range. In this way, the lack of time-dependent downtrend of the post-accent L boundary tone observed in NK's speech is not counter-evidence against my model.

The more complicated case is SK's. As shown in Figure 3.15-c, the residual values of [L] do not undergo any time dependent downtrend after the duration between the preceding [H] and [L] reaches 600 ms. This lack of any time-dependent downtrend may be again explained by the "bottom" effect, i.e. hitting the very bottom of the speaker's possible pitch range. What is problematic is that those residual values before the duration between [H] and the [L] hits 600 ms are exceptionally higher than the rest of those values. Those exceptionally high residual values of the L boundary tone are all obtained from the shortest compound form "*maronmónaka-no*", and I suspect that something special must be going on in that form. One possible factor contributing to the unexpectedly higher [L] values associated with the shortest form is an "undershoot" effect. I noticed during measuring SK's F0 targets in these compound nouns that the fall of the bitonal pitch accent H*+L starts relatively late in the shortest form, i.e. "*maronmónaka-no*". The fall tends to start at the offset of a syllable which immediately follows the accented syllable, i.e. *na*. An actual pitch track form of SK's speech that shows this delay is in Figure 3.14-a. Such a delay of an accentual fall is almost never observed in her longer forms, however. I attribute these exceptionally high residual values of the post-accent L tone to an "accentual fall delay" which is peculiar to these shortest forms. That is to say, the accentual delay results in an "undershoot" of the post-accent L boundary tone. It is explained in the following way. Because of the accentual fall delay, there are only two syllables left for both the +L tail tone of the pitch accent and the following L boundary tone to be achieved. Those two tones, however, do not provide enough time for SK's pitch movement to achieve the target values of those L tones which are close to the very bottom of her pitch range. This results in undershooting those target values and the exceptionally higher F0 of the post-accent L boundary tone. If this explanation is on the right track, then this unexpected outcome of SK's is not problematic to my downtrend model either because neither the "bottom" effect nor the "undershoot" effect is incompatible with that model.

3.5. Testing Predictions 2: When the Duration between H*+L and L% are Varied

As briefly described in the summary of the preceding section, the downtrend model proposed in Section 3.3 predicts that even a post-accent L boundary tone undergoes a local time-dependent downtrend once the duration between the accented syllable and the L boundary tone increases. That is to say, the model predicts that AS and MR will have the post-accent L boundary tone undergo a local time-dependent downtrend as well as the global declination once the duration between the L boundary tone and the immediately preceding pitch accent is varied.

To test this prediction, the <*ánnasan*> set containing constituents carrying a pitch accent followed by a L boundary tone are used. In order to maximize the variation between the pitch accent and the post-accent L boundary tone, the accent was always placed on the very initial syllable of the target forms as schematically shown in (7). Also, the actual forms used in this experiment are shown in (8-a) ~ (8-c).

(7) [σ σ --- syllable number varied --- σ]$_{\text{MiP1}}$ [σ …
H* +L L

predicted to undergo local downtrend
as well as global downtrend.

(8) **The <*Ánnasan*> Set**
MiP1 **MiP2**

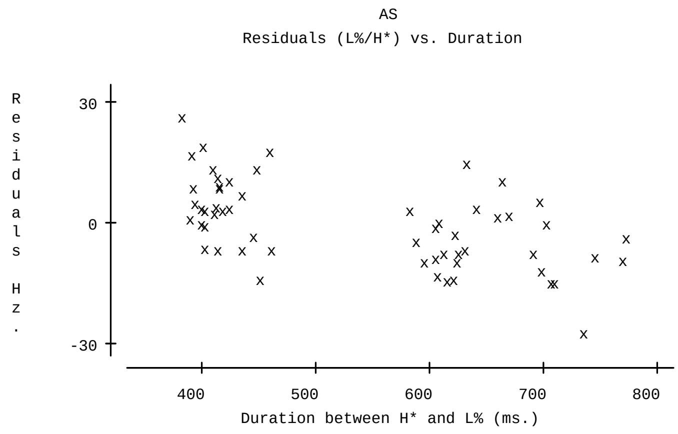
(a) [*án(na)-san-no*] [*omiaiáite…*
H*+L L
Anna-Ms.-Gen

(b) [*márinamu-san-no*] [*omiaiáite…*
H*+L L
Marinam-Ms.-Gen

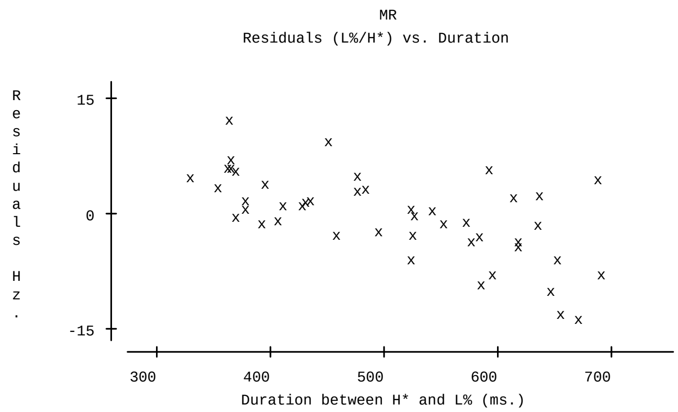
(c) [*ándoryuu-san-no*] [*omiaiáite…*
H*+L L
Andrew-Mr.-Gen

The duration between the accented syllable and the L boundary tone varies in steps from the shortest in (8-a) to the longest in (8-c).

Given those configurations, we expect that in AS's and MR's speech, the post-accent L boundary tone undergoes not only a global declination but also a local downtrend which end up in a time-dependent lowering of more than -25 Hz/sec. This prediction is again tested by residual analysis similar to those adopted in previous experiments. That is, the residual values of the L boundary tone ([L] henceforth) were obtained from a regression analysis in which [L] is the dependent variable and the preceding phrase-initial H* accent tone ([H*] henceforth) is the predictor. Those residual values represent the relative height of [L] given a certain F0 of the preceding [H*] tone. Those residual values of [L] are compared with the duration between [H*] and the [L]. If the residual values of [L] decrease as the duration between those points increases and the rate of the decrease is more than what is expected by the global declination factor only, then we conclude that the post-accent [L] undergoes the local downtrend as well. I show the results of this residual analysis obtained from AS's and MR's data in Figure 3.16.



a. Speaker AS.



b. Speaker MR.

Figure 3.16. The Relationship between Duration and Residual Values

The data points of the residual values of [L] are fitted to declining regression slopes for both AS and MR. The coefficient of those slopes varies between –0.044 (AS) and –0.032 (MR). Also, those coefficients are statistically significant ([AS: Std Error = 0.009, t = -4.84, $p < 0.001$]. [MR: Std Error = 0.006, t = -5.49, $p < 0.001$])]. Those coefficient values are translated into a declination slope rate of -44 Hz/sec for AS and -32 Hz/sec for MR. Those declination rates are far larger than those expected from the global declination only, and I conclude that the post-accent L boundary tone associated with forms in (8-a) ~ (8-c) undergoes a local downtrend as well. Remember that in the previous experiment of compound forms, those two speakers had only a global declination of a post-accent L boundary tone. In that experiment, the duration between the preceding H edge tone and the H* accent tone was varied while the duration between the pitch accent tone and the post-accent L boundary tone was kept constant across all the forms. In this experiment, however, the duration between the accent and the post-accent L boundary tone was varied, and the L boundary tone underwent a time-dependent downtrend of more than 30 Hz/sec which is enough to conclude that there is a local downtrend factor at work as well as the global declination. This is exactly what my downtrend model proposed in Section 3.3 predicts.

3.6. Chapter Conclusion

In this chapter, I have shown that there are two types of time-dependent downtrend in Tokyo Japanese. One of them is a global declination which unfolds across a wider domain, quite possibly an Intonational Phrase or an Utterance. The other is a "tone-bound" declination, which I proposed in my downtrend model and which lasts until the tonal value of each tone is achieved and is reset immediately after that achievement. This "tone-bound" declination is a new finding in the intonational studies. In addition, formalizing this "tone-bound" declination by proposing a new model and testing predictions made by that model is another contribution of this chapter. It turned out the model makes correct predictions for downtrend of not only L tones within a sequence of unaccented Minor Phrases, but also that of a H* accent tone and a post-accent L boundary tone.

CHAPTER 4

GLOBAL VS. LOCAL APPROACHES TO *CATATHESIS*

4.1. Introduction

In the last chapter, we considered the time factor contributing to downtrends in Tokyo Japanese. I found a "tone-bound" declination, whose domain was more local than the global declination that unfolds over the whole utterance or an intonational phrase. Though their domains are different, both of those two types of declination are captured as a declining slope of a base line (and the top line) of a tonal space on which tones are superimposed. However, not all downward scaling of tones is due to change in the slope, shape, height or width of a tonal space. For example, in languages such as Chinese, some downtrend may be due to a carry-over effect of the preceding low tone (Xu 1999). I will show in this chapter that Tokyo Japanese *catathesis* is also a good example for such downtrend induced by propagation of the low F0 value of the preceding tone.

In Tokyo Japanese there is a sharp F0 fall within a bitonal pitch accent from the H* accent tone to the +L trailing tone. This +L trailing tone takes an F0 value lower than L edge tones that are not preceded by a pitch accent. At the same time, subsequent tones also undergo lowering: tones that follow the H*+L bitonal pitch accent are realized lower than those tones which do not follow it. This lowering of both +L and subsequent tones is called *catathesis*, and is well-documented by Poser 1984, Pierrehumbert & Beckman 1988 and Kubozono 1993. This *catathesis* effect was also replicated in my own experimental work presented in Section 1 of this chapter.

Whether this post-accent downtrend is captured as global manipulation of a tonal space or a more local manipulation of each tonal value by tone-by-tone scaling is the question asked in this chapter. Pierrehumbert & Beckman (1988) take the former view that *catathesis* should be accounted for by lowering the top line of a tonal space soon after a H* accent tone appears (see Section 4.2). Since the tonal space as a whole is lowered, post-accent tones superimposed on the space are also lowered. The advantage of this tonal space lowering model is its simplicity. For example, they propose that both the +L accent tone and the L edge tone bear the same transformed value, a value relative to the given pitch range. Nonetheless, the +L tone is realized lower than non-post-accent L edge tone because the former is in a lowered pitch range. This simplicity, however, pays the price of wrong predictions. For example, I observed that the magnitude of *catathesis* of a post-accent tone diminished as more tones intervened between the preceding accent tone and the post-accent tone. Since tonal space lowering is a global operation, i.e. it equally lowers the F0 of the tone immediately following an accent and those tones further along, no diminishing effect of *catathesis* is expected no matter how many tones intervene between the preceding accent and the following tone.

This observation, however, is not problematic to the alternative hypothesis that I propose in Section 4.3 of this chapter, i.e. the more local tone-by-tone scaling model. According to this alternative model, *catathesis* is a consequence of (i) assignment of an extra-low F0 to the +L trailing tone of a pitch accent and (ii) propagation of the extra lowness of the +L trailing tone to subsequent tones via local tone-by-tone scaling. Since the extra-low F0 of the +L accent tone only directly influences the immediately following tone, this alternative model expects that the magnitude of *catathesis* of a post-accent tone may diminish as more tones intervene between the preceding +L accent tone and the post-accent tone.

In Section 4.4 of this chapter, I will present further evidence to support the tone-by-tone scaling hypothesis. For example, in that section I will show that there is strong correlation between the F0 of the +L accent tone and the following post-accent tones. This is exactly what the tone-by-tone scaling model predicts because F0 of a tone is computed relative to that of the preceding tone. In the tonal space lowering model, by contrast, the F0 value of each tone is computed relative to the pitch range independently of the value of the preceding tone. Therefore, a strong correlation between +L and the following tone is not predicted.

4.1.1. *Catathesis*: Basic Facts and Assumptions

McCawley (1965), Poser (1984) and Pierrehumbert & Beckman (1988) observed that tones following the accent H* tones are realized lower than those tones that do not follow it. Pierrehumbert & Beckman called this post-accent lowering of tones *catathesis*.

Let us consider a hypothetical case with two different sequences of Minor Phrases in (1) and (2). The initial Minor Phrase (MiP1, henceforth) is accented while the second Minor Phrase (MiP2, henceforth) is unaccented in (1). In (2), both MiP1 and MiP2 are unaccented.

(1) Accented

(σ σ σ' σ σ)$_{\text{MiP1}}$ (σ σ σ σ σ)$_{\text{MiP2}}$

L_1 H_1 H_1*+L_1 L_2 H_2

(2) Unaccented

(σ σ σ σ σ)$_{\text{MiP1}}$ (σ σ σ σ σ)$_{\text{MiP2}}$

L_1 H_1 L_2 H_2

Tones associated with MiP_1 bear subscript "1" and those associated with MiP_2 carry subscript "2". L and H tones without any diacritic mark such as "*" and "+" are edge tones that marks the left edge of each Minor Phrase.

According to Poser and Pierrehumbert & Beckman, L tones that follow the accent H*, i.e. +L_1 and L_2 of the sequence in (1), are lower than the L_2 tone of the sequence in (2). In the same way, the post-accent H tone, i.e. the H_2 tone of the sequence in (1), is lower than the H_2 tone of the sequence in (2). In the following section, I present some data to replicate this observation.

4.1.2. Replicating *Catathesis*

To replicate the observation that tones following the pitch accent H* are realized lower than those which are not, I compared F0 values of tones associated with the following two forms from Dataset <*omiaiáite*>. Detailed description of those forms is shown in Appendix of this chapter.

The <*Omiaiáite*> Set

(3)	[*Án(na)san -no*]$_{MiP1}$	[*o*	*miai*	*áite-no*]$_{MiP2}$	---
	H_1* +L_1	L_2	H_2	H_2*+L_2	
	Anne-Gen	date-Gen			---
	"(---) of Ms. Anna's date"				
(4)	[*Ma nae da san-no*]$_{MiP1}$	[*o*	*miai*	*áite-no*]$_{MiP2}$	---
	L_1 H_1	L_2	H_2	H_2*+L_2	
	Manae-Gen	date-Gen			---
	"(---) of Ms. Manaeda's date"				

The form in (3) contains a sequence of two accented Minor Phrases: *Ánnasan-no* "Ms. Anna's" and *omiaiáite-no* "date-Gen". The form in (4) contains a sequence of an unaccented Minor Phrase (*Manaedasan-no* "Mr. Manaeda's") and a following accented Minor Phrase (*omiaiáite-no* "date-Gen"). Four speakers (AS, MR, RO and SK) read those sentences between eight and sixteen times. The recording procedures are the same as those already reported in Chapter 1.

We expect that tones that follow the accent H_1* tone of the form in (3), i.e. tones associated with MiP_2 of the form in (3), are realized lower than those tones associated with MiP_2 of the form in (4), and this expectation was born out (see Table 4.01 and Table 4.02).

Table 4.01. The Mean F0 Value of the Post-Accent L_2 of Form (3) and that of the Non-Post-Accent L_2 of Form (4)

Speakers	**Post Accent L2**	**Non-Post Accent L2**
AS	185 Hz (n=16)	223 Hz (n=11)
MR	175 Hz (n=16)	243 Hz (n=16)
RO	167 Hz (n=13)	224 Hz (n=13)
SK	190 Hz (n= 12)	270 Hz (n=13)

The difference between those two means is significant for all of the four speakers at $p < .05/4 = .0125$ (Bonferroni adjustment for multiple comparisons).

Table 4.02. The Mean F0 Value of the Post-Accent H_2* of Form (3) and that of the Non-Post-Accent H_2* Form (4)

Speakers	**Post Accent H2***	**Non-Post Accent H2***
AS	202 Hz (n=16)	218 Hz (n=11)
MR	231 Hz (n=16)	268 Hz (n=16)
RO	254 Hz (n=13)	273 Hz (n=13)
SK	268 Hz (n= 12)	298 Hz (n=13)

The difference between those means is significant for AS, MR and SK at $p < .05/4 = .0125$ (Bonferroni adjustment for multiple comparisons). It is significant for RO at $p < .05$ without adjustment.

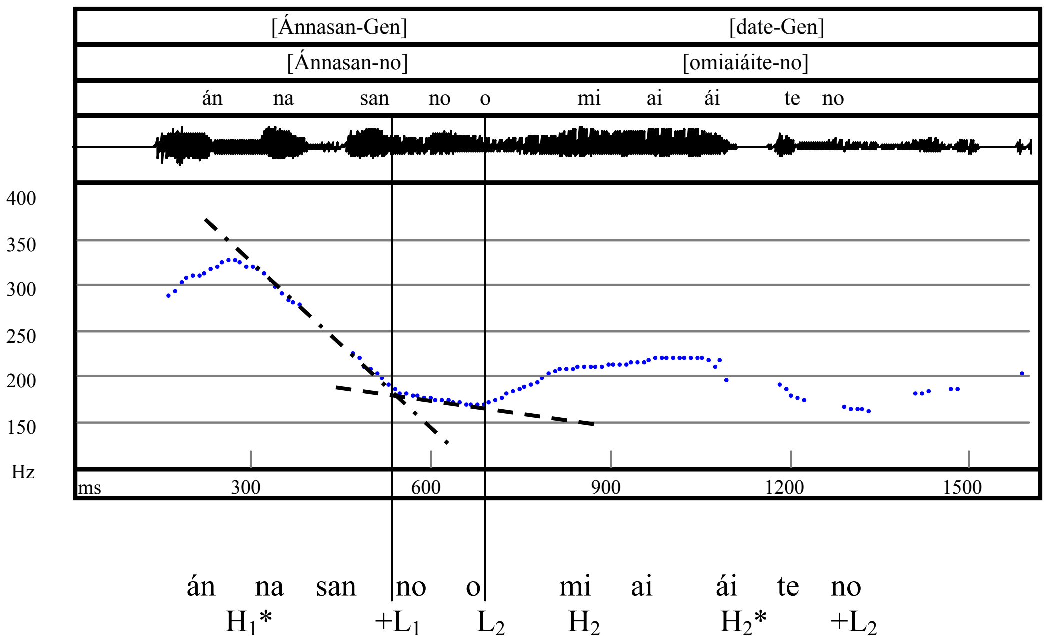
Table 4.01 and Table 4.02 show a comparison between the mean F0 values of L_2 and H_2* associated with MiP_2. For all four speakers, mean F0 of those tones of the form in (3) was lower than that of those of the form in (4).

For example, the mean F0 value of the post-accent L_2 tone of the form in (3) was about 70 Hz lower than that of the non-post-accent L_2 tone of the form in (4) in MR's speech.[15] In the same way, the mean F0 value of H_2* of the form in (3) was about 35 Hz lower than that of H_2* of the form in (4) in the same speaker's speech. Those mean differences are significant for all speakers. This result agrees with Poser and Pierrehumbert & Beckman's observation, so that presence of post-accent *catathesis* is again attested in Tokyo Japanese.

Tones realized lower are not limited to those which come after the pitch accent H*+L. The +L trailing tone of the bitonal pitch accent itself is also realized as almost low as the immediately following L edge tone. This is shown in the following example F0 contours.

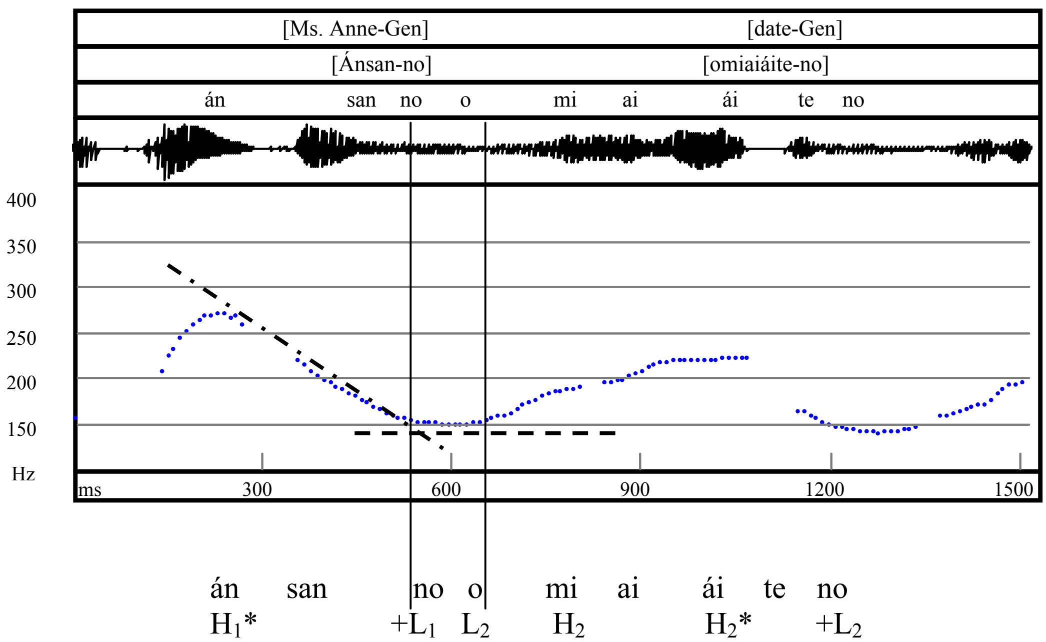
[15] I measured the right edge of the initial syllable *[o]* for the L_2 edge tone.

MR 2-05-4



a. Speaker MR

RO 4-05-3

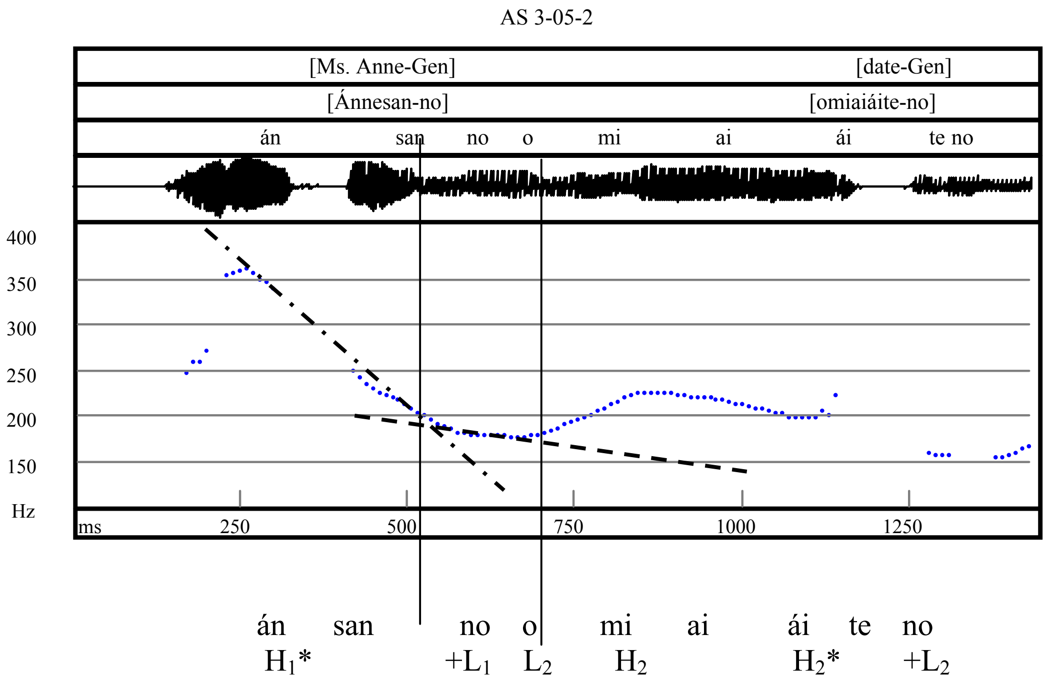


b. Speaker RO

Figure 4.01. Example F0 Contours of a Sequence of Accented Minor Phrases

Continued next page

Figure 4.01 continued

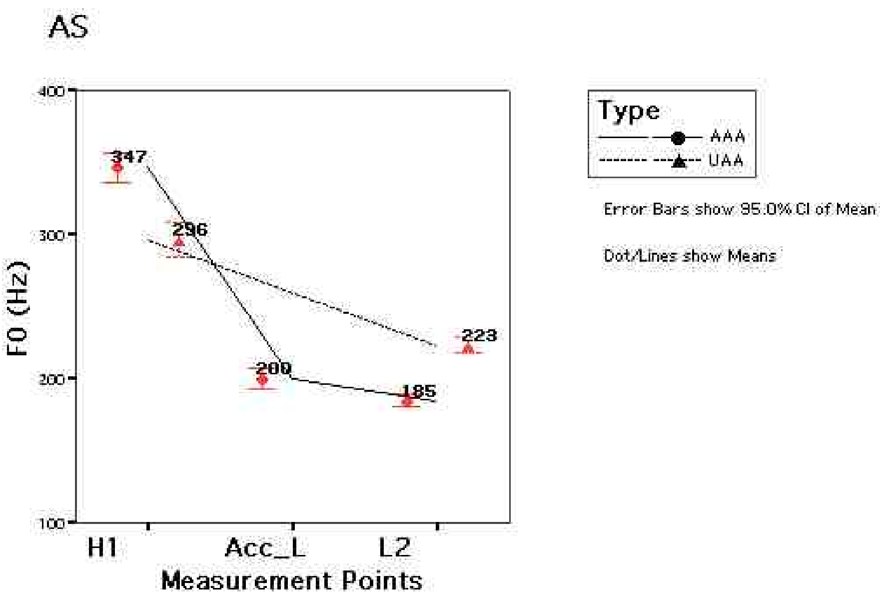


c. Speaker AS

The dash-dot line on those F0 contours represents the slope of an accent fall from H_1* to $+L_1$. The following less steep dashed slope represents the slope from the $+L_1$ trailing tone to the following L_2 edge tone. This slope intersection procedure is the only way to detect the location where a +L accent tone is aligned since there is no predetermined syllable or a mora with which the +L tone is phonetically aligned (Pierrehumbert & Beckman 1988). In Figure 4.01-b and Figure 4.01-c, for example, the location of the +L tone is one syllable away from the accented mora. When it comes to Figure 4.01-a, its location is two syllables away from the accented mora.

To obtain the F0 value of the $+L_1$ trailing tone, I adopted this slope intersection procedure. I made an estimate of the steep slope of accentual fall and the following less steep slope by hand using a ruler. Then I measured the F0 value of the intersection of those two slopes, which is considered as the F0 of the $+L_1$ accent tone. Data from SK are not considered because all her tokens was associated with a single sharp falling slope all the way from the accented mora of MiP_1 to the onset of MiP_2.

I compared the mean F0 of $+L_1$ of the form in (3) with the mean F0 of the subsequent L_2 edge tone of the same form. Also, I compared it with the mean F0 of the non-post-accent L_2 edge tone of the form in (4). Those comparisons are presented in Figure 4.02.

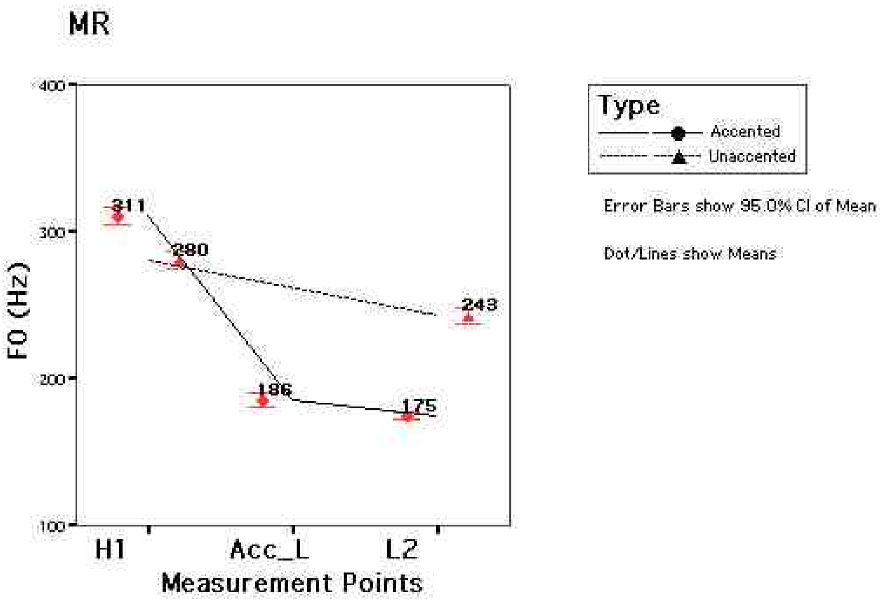


a. Speaker AS

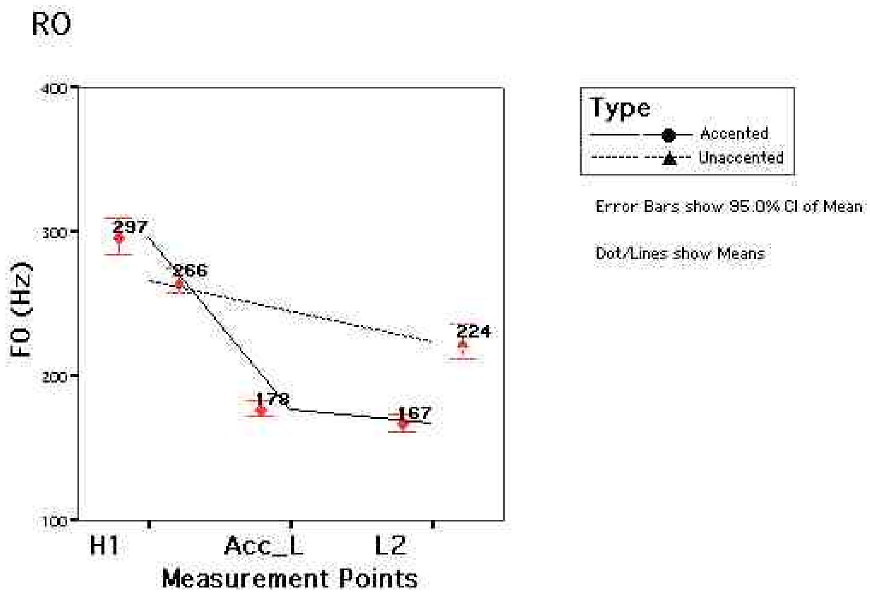
Figure 4.02. The Mean F0 Values of Tones associated with Accented MiP and Those associated Unaccented MiP

Continued next page

Figure 4.02 continued



b. Speaker MR



c. Speaker RO

In Figure 4.02, measurement points (i.e. the $+L_1$ trailing tone (Acc_L) and L_2 edge tone) are plotted as the horizontal axis, and their F0 values are plotted as the vertical axis. In all of those plots in Figure 4.02, the mean F0 of the $+L_1$ tone was realized almost as low as the subsequent post-accent L_2 edge tone.[16] At the same time, this $+L_1$ trailing tone was 23~46 Hz lower than the non-post-accent L_2 edge tone of the form in (4). These observations conform to the claim made by Poser (1984) and Pierrehumbert & Beckman (1988) that *catathesis* starts within the bitonal pitch accent, i.e. even an accent +L tone undergoes *catathesis*.

4.1.3. The Issue

One question related to this post-accent downtrend is whether that is captured as a global lowering of a tonal space or a more local manipulation of each tonal value by tone-by-tone scaling.

The only previous model of Tokyo Japanese *catathesis*, Pierrehumbert & Beckman's (1988), takes the former view (i.e. the more global tonal space lowering model). This view, however, is not the only possible account for *catathesis* as they admit themselves:

> *This is not the only way that catathesis can be modeled. Pierrehumbert (1980) presented a model for the homologous effect in English in which each tone's value was computed relative to the value of the previous tone in accordance*

[16] Though the following L_2 edge tone was realized slightly lower than the $+L_1$ trailing tone, the difference is small (i.e. only 11 ~ 15 Hz).

> *with its preassigned prominence value. The chief advantage of the alternative model is that it is less powerful than a (tonal space lowering) model. The (*tone-by-tone*) scaling rules refer only to F0 values in a small locality. They cannot pass information nonlocally by establishing or referring to register values for tones that are only exemplified some distance away. It is perfectly possible to model catathesis in this way in Japanese as well.* (Pierrehumbert & Beckman 1988: pp 139-140)

They argue against the alternative tone-by-tone scaling model and adopted the tonal space lowering model based on their observations related to the weak and strong alternation of L edge tones. They found that a long initial syllable of a Minor Phrase was associated with an allophonically weaker L edge tone, i.e. a higher L edge tone, than a short initial syllable. Nonetheless, according to them, the F0 value of the following H tone is unaffected by this weak-strong L tone alternation. They used this as evidence against the tone-by-tone scaling. Unfortunately, however, they do not present any supporting statistical data for this claim. In addition, the crucial difference between the weak and strong L edge tones is relatively small, approximately 10 Hz (Pierrehumbert & Beckman 1988: pp 31-33). It is likely that such a small F0 difference was not large enough to derive a significant difference in the F0 of the following H tone because of a possible large variance of the following H tone. Given this, Pierrehumbert & Beckman's evidence against the tone-by-tone scaling hypothesis is still weak, and it is too early to exclude the alternative tone-by-tone scaling hypothesis to account for the catathesis in Tokyo Japanese.

In the following part of this chapter (Section 4.2, Section 4.3 and Section 4.4), I compare those two possible models of *catathesis*, i.e. the tonal space lowering model and the tone-by-tone scaling model. I will conclude in those sections that *catathesis* should be accounted for by the alternative tone-by-tone hypothesis based on the following observations: (i) the magnitude of *catathesis* of a post-accent tone diminishes as more tones intervene between the preceding pitch accent and the post-accent tone and (ii) the F0 value of a post-accent tone and that of the preceding +L accent tone have a strong correlation.

4.2. The Pitch Range Lowering Model

In this section, I introduce Pierrehumbert & Beckman's (1988) tonal space lowering model of *catathesis*. According to them, the top line of a pitch range is lowered by some fixed ratio each time a H* accent tone appears in the pitch range, while the relative values of tones are unchanged. Tones in the lowered range bear lower F0 simply because their pitch range as a whole is lower. In the following part of this section, I give an overview of this global hypothesis for *catathesis*. However, I do not provide all the mathematical details of their model because those details are not necessarily relevant to the discussions that follow this section.

4.2.1. Basic Notions and Assumptions of the Pitch Range Lowering Model

As a first step to understand Pierrehumbert & Beckman's *catathesis* model, it is necessary to be acquainted with the assumptions and notions adopted in their model, such as pitch range, top line of a pitch range and transformed value of a tone.

4.2.1.1. Pitch Range

Pierrehumbert & Beckman assume a pitch range or a transformed space, a band of F0 values relative to which the tonal events are scaled. It is delimited by a top line and a bottom line. The top line and the bottom line correspond to the highest F0 value and the lowest value a H tone can take respectively.

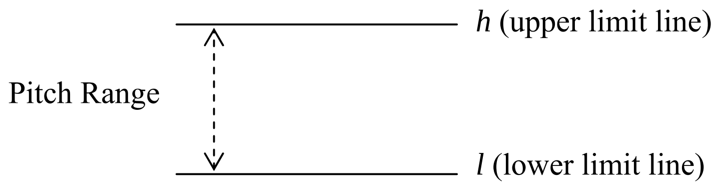


Figure 4.03. The Top Line and the Bottom Line of a Pitch Range

4.2.1.2. Transformed Values

Another notion that needs to be introduced to understand Pierrehumbert & Beckman's pitch range model is the transformed value of tones, which is a normalized value that denotes the relative height of a tone in a given pitch range. A transformed value varies between 0 and 1. For example, a H tone bears a transformed value of 0.5 if the tone 's height is half of the given pitch range. This is graphically shown in the following picture.

<for H tones>

h (upper limit line)

● 3/4 from the bottom line = 0.75

● 1/2 from the bottom line = 0.5

l (lower limit line)

Figure 4.04. Transformed Values of H Tones (Pierrehumbert & Beckman's Model)

For H tones, those transformed values increase as their F0 values increase. In contrast, it is the other way round for L tones: their F0 decreases as their transformed value increases. For example, if the height of a L tone is three fourths of a given pitch range, then its transformed value is computed by subtracting 0.75 from 1, which yields a transformed value of 0.25. This inverse relation between transformed values and actual F0 values of L tones comply with our intuition that the lower L tones are more salient than higher L tones are.

<for L tones>

h (upper limit line)

● 3/4 from the bottom line = 1-0.75= 0.25

● 1/2 from the bottom line = 1-0.5 = 0.5

l (lower limit line)

Figure 4.05. Transformed Values of L Tones (Pierrehumbert & Beckman's Model)

4.2.2. PB's Catathesis Model

Using the assumptions and notions introduced above, Pierrehumbert & Beckman's *catathesis* model is considered in this section. The core of their proposal is that the *h* limit line of a pitch range is lowered by some fixed ratio each time a H* accent tone appears.

According to them, this is the sole effect of *catathesis*. The transformed values (i.e. relative values) of tones are unchanged even in the lowered pitch range. For example, non-post-accent H edge tones and post-accent H edge tones bear the same transformed value. In the same way, Major Phrase medial L edge tones bear the same transformed value regardless of whether they appear in a post-accent region or non-post-accent region. Post-accent tones bear lower F0 simply because their pitch range as a whole is lower.

In addition, there is nothing special about a +L trailing tone of a pitch accent. It also bears the same transformed value as a Major Phrase medial L edge tone. However, it is realized lower than non-post-accent Major Phrase medial L edge tones because it is realized in a lowered pitch range.

In the pictures below, both a post-accent and a non-post accent H edge tone bear the same transformed value 0.8. In the same way, post-accent +L and L edge tones bear the same transformed value (i.e. 0.5) as a non-post-accent L edge tone does. Nonetheless, those post-accent tones are realized lower than the non-post-accent tones because they are in a lowered pitch range.

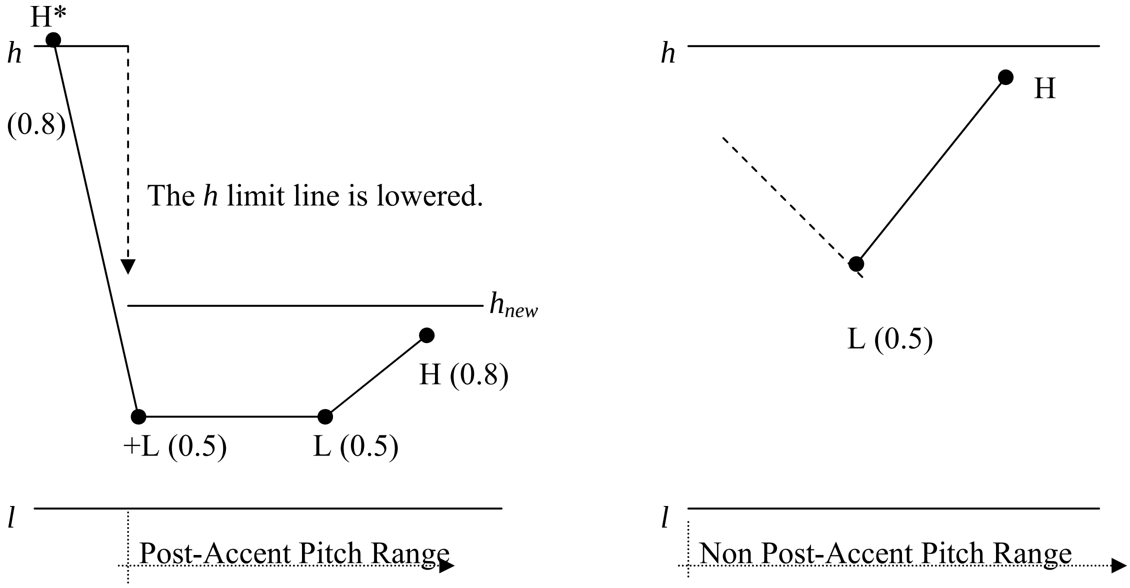
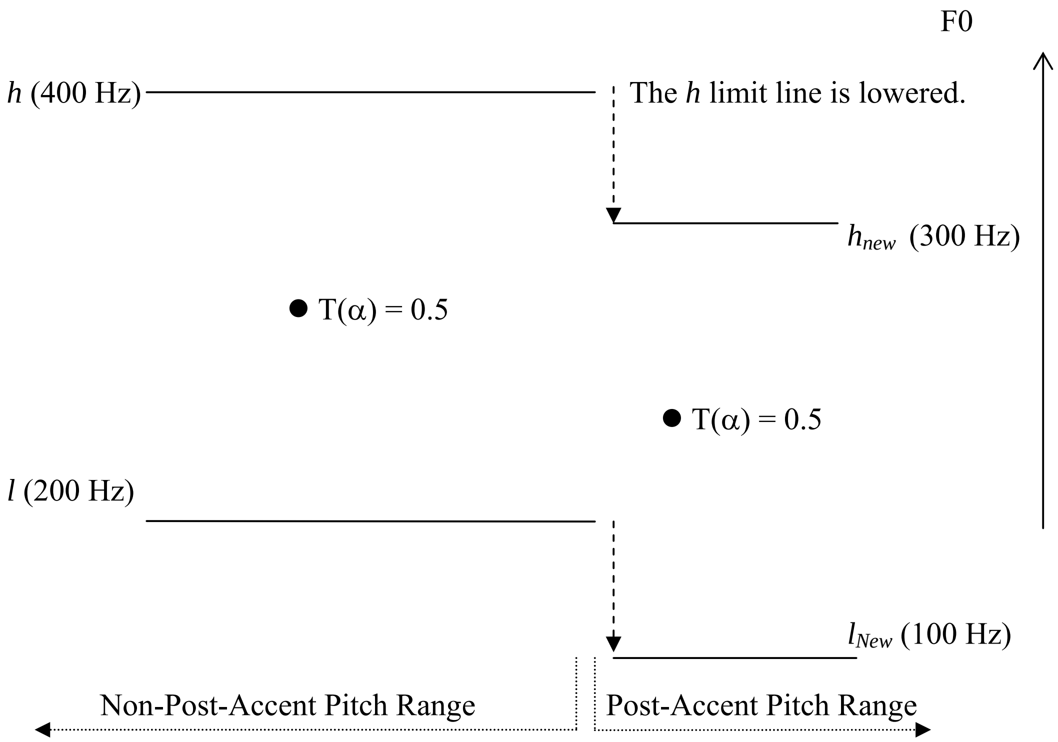


Figure 4.06. Transformed Values and Pitch Ranges of Post-Accent Tones and Non-Post-Accent Tones (Pierrehumbert & Beckman's Pitch Range Lowering Model)

A variant of Pierrehumbert & Beckman's pitch range lowering model is also possible: not only the *h* top limit line but also the *l* bottom line are slid down by the same ratio. This is proposed by Ladd (1992) and van den Berg et al. (1992) for downstep of other languages such as English and Dutch. This alternative model can be also applied to Tokyo Japanese *catathesis*, which is graphically shown below.



T(α) = a transformed value of a tone α.

Figure 4.07. A Variant of Pierrehumbert & Beckman's Pitch Range Lowering Model

For the purpose of the discussion in the later part of this chapter, it does not matter whether it is only the *h* top line that is lowered or it is both the *h* and *l* line that is lowered. More importantly, both of those pitch range lowering hypotheses for *catathesis* share the common view that *catathesis* is induced by manipulation of a tonal space (i.e. tonal space lowering) but not by manipulation of the relative value of each tone.

4.2.3. The Problem: Adjacency Effect of *Catathesis*

Since tonal space lowering is globally applied to the whole post-accent pitch range according to their model, Pierrehumbert & Beckman predict no diminishing of *catathesis* of post-accent tones irrespective of how many tones intervene between the preceding accent and the post-accent tones. However, I found that the magnitude of lowering of post-accent tones was diminished as those tones get farther way from the preceding accent.

Again consider the forms in (3) and (4) from Dataset <*omiaiáite*>. MiP_2 in form (3) is preceded by an accented MiP_1 while that in form (4) is preceded by an unaccented MiP_2.

Dataset <*omiaiáite*>

(3) **AA** (Sequence of Accented Words)

[*Án(na)san*	*-no*]$_{MiP1}$	[*o*	*miai*	*áite-no*]$_{MiP2}$	---
H_1*+L_1		L_2	H_2	H_2*+L_2	
Anne-Gen		date-Gen			---

"(---) of Ms. Anna's date"

(4) **UA** (Sequence of Unaccented and Accented Words)

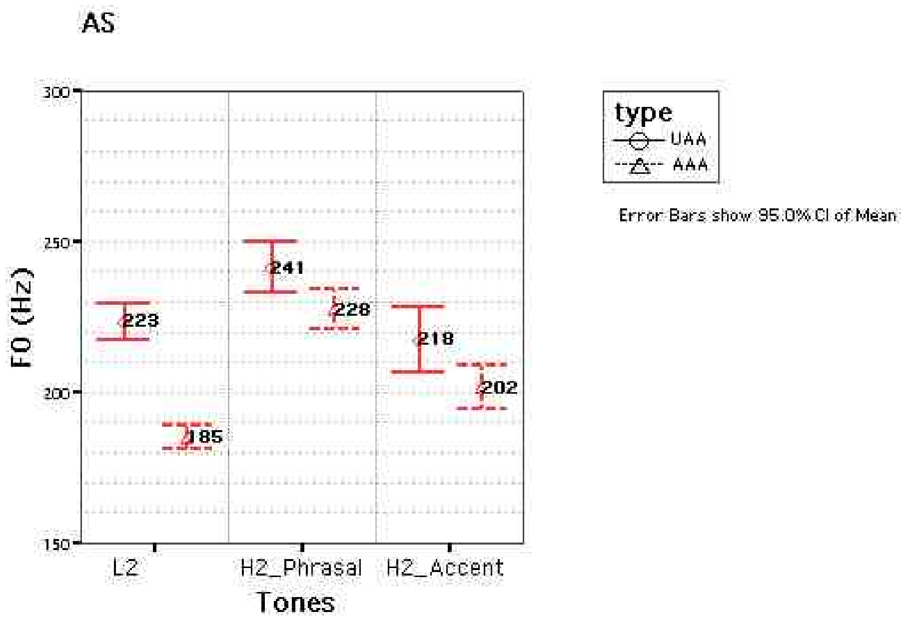
[*Ma*	*nae da san-no*]$_{MiP1}$	[*o*	*miai*	*áite-no*]$_{MiP2}$	---
L_1	H_1	L_2	H_2	H_2*+L_2	
Manae-Gen		date-Gen			---

"(---) of Ms. Manaeda's date"

I have already shown in Section 1 that not only the post-accent L_2 in (3) but also the post-accent H_2 and H_2* in the same form underwent *catathesis*, i.e. their F0 values were consistently lower than that of their non-post-accent counterpart in (4).

However, when it comes to the magnitude of *catathesis* of those post-accent tones, not all of those post-accent tones were equal. The magnitude of *catathesis* of the post-accent L_2 edge tone was the greatest, that of the post-accent H_2* accent tone was the smallest. That is, the magnitude of *catathesis* diminished as the post-accent tone got farther away from the preceding accent.

To see such *catathesis* diminution, let us again consider the summary of data obtained from the form in (3) and (4), which is shown in Figure 4.08. In that figure, the mean F0 values and of tones associated with MiP_2 of the form (3) and form (4) and their 95% confidence intervals are shown.

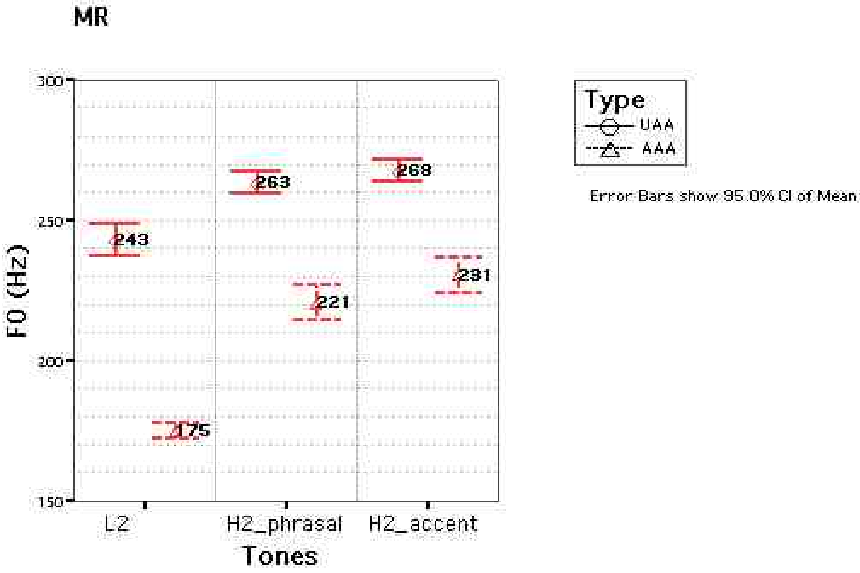


a. Speaker AS

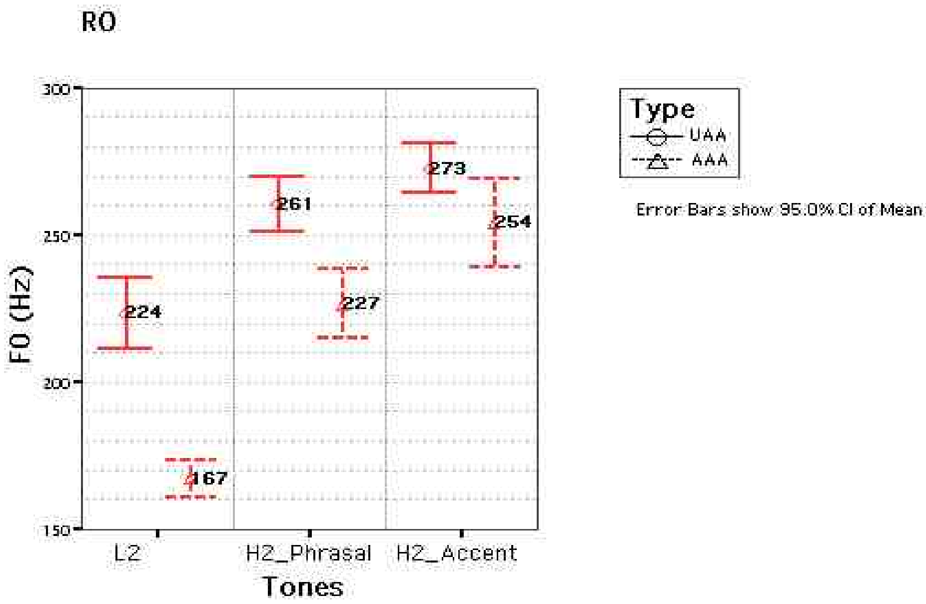
Figure 4.08. The Mean F0 Values and of Tones Associated with MiP_2 of "AA" in (3) and "UA" in (4), and Their 95% Confidence Intervals

Continued next page

Figure 4.08 continued



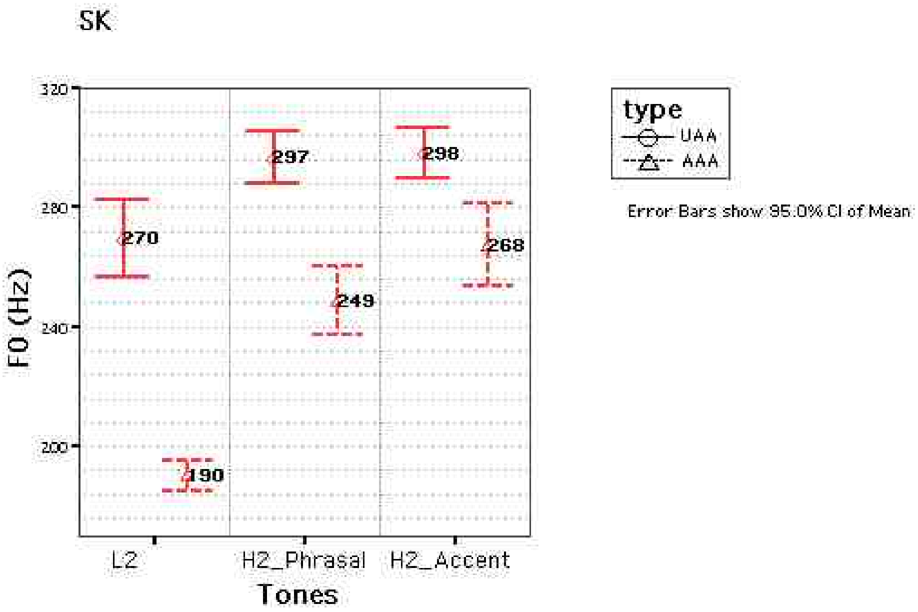
b. Speaker MR



c. Speaker RO

Continued next page

Figure 4.08 continued



d. Speaker SK

Plots in this figure show that the magnitude of post-accent lowering was the greatest at L_2 and the smallest at H_2*. For example, consider speaker SK (Figure 3.08-d). The mean F0 of the post-accent L_2 edge tone was 80 Hz lower than that of the non-post-accent L_2 tone. As for the mean F0 of the post-accent H_2 edge tone, it was 48 Hz lower than the mean F0 of its non-post-accent counterpart. This post-accent lowering was further diminished into 30 Hz at H_2* accent tone. The same "diminishing" effect of *catathesis* was observed in not only SK's speech but also in MR and RO's speech (Figure 4.08-b and Figure 4.08-c). This diminishing effect of *catathesis* is visually summarized in the following picture.

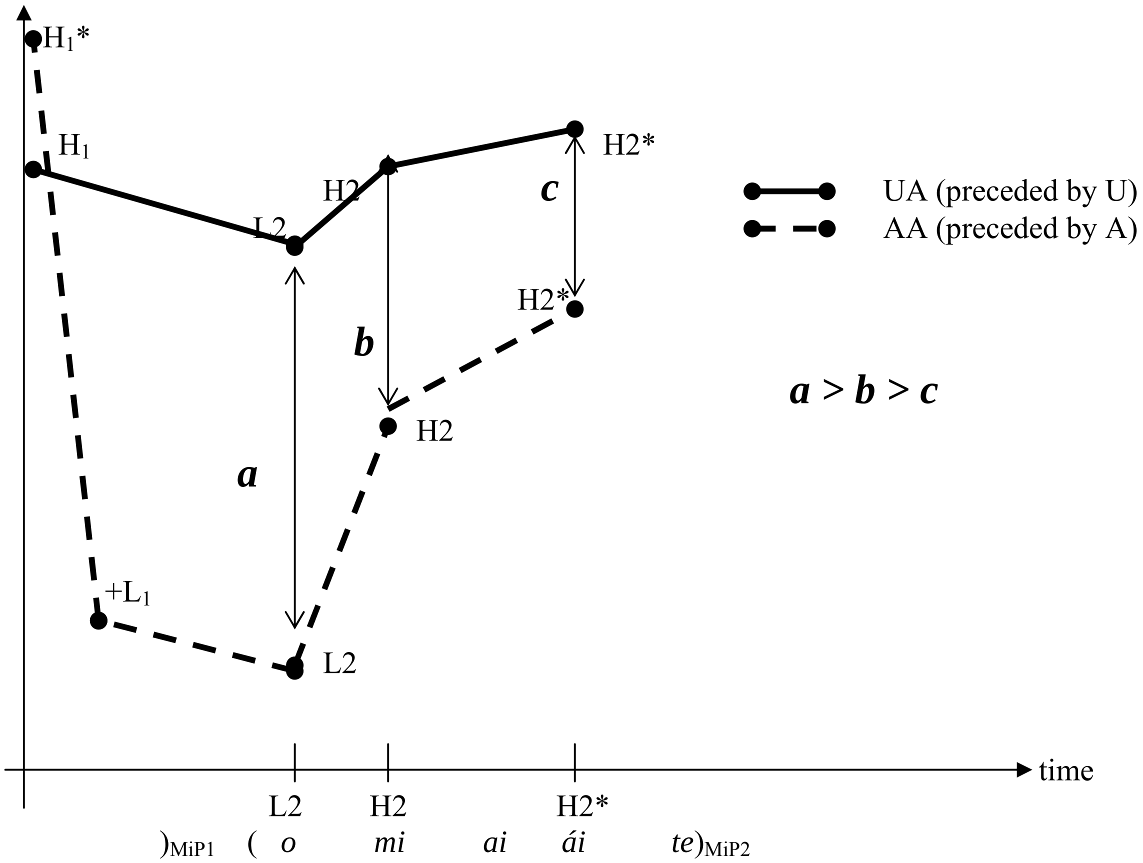


Figure 4.09. The Diminishing Effect of *Catathesis*

The fact that the magnitude of post-accent lowering diminishes as more tones intervene between the preceding pitch accent and the post-accent tones weakens the validity of Pierrehumbert & Beckman's global top line lowering model. Instead, I propose in Section 4.3 that *catathesis* in Japanese should be captured as a result of a more local tone-by-tone scaling which easily allows the *catathesis* diminution.

4.3. The Alternative Hypothesis: Tone-by-Tone Scaling Model

In this section, I propose the alternative tone-by-tone scaling model of *catathesis* which allows the magnitude of *catathesis* to be diminished as more tones intervene between the preceding pitch accent and the following tone.

One of the assumptions different from Pierrehumbert & Beckman's global pitch range lowering hypothesis is that presence of a pitch accent does not affect the height of the top line and the base line of a pitch range. Rather, *catathesis*, i.e. post-accent lowering of F0, is nothing but propagation of the extreme lowness of the +L accent tone to the following tones via tone-by-tone scaling. According to this local tone-by-tone scaling model, the F0 value of a tone (T_i) preceded by another tone (T_{i-1}) is a function of the F0 of T_{i-1}. However, only a fraction of the F0 value of T_{i-1} is actually reflected in the F0 of T_i due to certain phonetic interpretation rules. As a result, the effect of the extremely low F0 of the +L accent tone on the following tone is alleviated as more tones intervene between the +L tone and the following tone. This is the origin of the diminution of *catathesis* observed in the previous section.

In 4.3.1, I first introduce our L tone scaling model. In 4.3.2, our H tone scaling model is presented as well as the discussion of how the *catathesis* diminution is achieved by the local tone-by-tone scaling.

4.3.1. About L Tone Scaling

In this subsection, I introduce the tone-by-tone scaling model of L tone values. The general proposals are first presented in 4.3.1.1, and those proposals are tested in 4.3.1.2 and 4.3.1.3.

4.3.1.1. The General Ideas of L Tone Scaling

Crucial to this tone-by-tone scaling model of *catathesis* is the extra lowness of the accent +L tone. The claim is that *catathesis* is nothing but propagation of the extra low value of the accent +L tone to the following tones via local tone-by-tone scaling.

In Section 4.2, I showed that the accent +L tone is realized lower than a non-post-accent L edge tone. An accent +L tone bearing a lower F0 value than a non-post-accent L edge tone indicates that the phonetics treats the accent +L tone differently from the non-accent L edge tones. This is not surprising given the fact that the accent +L tone is part of the bitonal pitch accent H*+L which is lexically provided (i.e. already present in the lexical representation of a word) and phonologically associated with the most prominent mora of a PWd. In contrast, L edge tones are not lexical and not necessarily associated with the most prominent mora of a PWd, and this contrast is visible to the phonetics.[17]

[17] The phonetics treating a pitch accent tone differently from other types of tones is not limited to the +L tone in Tokyo Japanese. The H* tone of the bitonal pitch accent H*+L is also realized higher than a H edge tone that precedes i (Pierrehumbert & Beckman 1988, Kubozono 1993, Warner 1997, and this thesis). However, the H* tone may be realized as high as or even lower than the preceding H edge tone when the distance between those two tones is long because of tone-bound declination. (See Chapter 2 of this thesis).

That the phonetics treats the accent +L tone differently from other L edge tones means that the +L tone and the L edge tones follow separate phonetic interpretation rules. Though they follow separate rules, I propose that both of those two different rules are expressed by the following linear function formula:

(5) $L_i = a * (T_{i-1} - \text{base line}) + \text{base line}$

$= (1-a)*\text{base line} + a*T_{i-1} \qquad 0 < a < 1$

The *base line* is close to the bottom F0 value of a speech-related pitch range. Speakers refer to the F0 of the *base line* when computing the F0 values of L tones. The F0 of the *base line* constantly undergoes the global and the local declination as already introduced in Chapter 3. For the meantime, however, those declination factors are ignored to make our discussions simpler. The coefficient of the equation in (i), i.e. *a*, is less than 1 but greater than zero, and the coefficient value may vary according to various factors, such as the type of the target L_i tone (e.g. whether that is part of a pitch accent or is an edge tone), phonological, syntactic and information structure factors, etc. The F0 value obtained by the function $a * (T_{i-1} - \textit{base line})$ in the above formula denotes the F0 excursion size from the *base line* to L_i, and the F0 excursion size rises as the coefficient *a* increases. By adding the *base line* F0 value to the F0 excursion size, the absolute F0 value of L_i is obtained.

Let us imagine a hypothetical case shown in Figure 4.10, where the *base line* is 150 Hz, the coefficient *a* is 0.3 and the F0 value of T_{i-1} is 350 Hz. Given this, we predict that the F0 of L_i is 210 Hz.

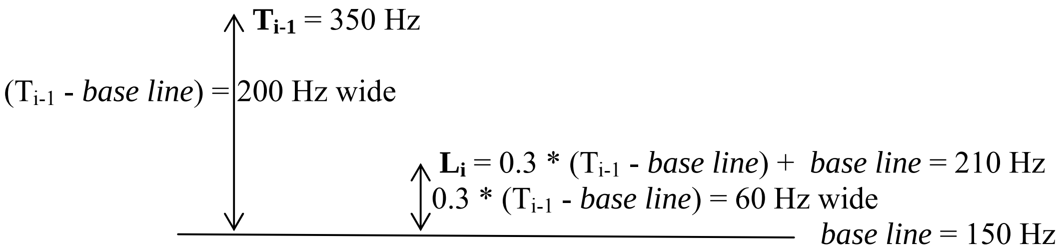


Figure 4.10. A Hypothetical L Tone Scaling (base line is 150 Hz, the coefficient a is 0.3 and the F0 value of T_{i-1} is 350 Hz.)

Let us next consider another hypothetical case where the coefficient *a* is greater, i.e. 0.8, while the F0 value of the *base line* and that of T_{i-1} are the same as the previous case shown in Figure 4.10. Such a greater coefficient value gives a larger F0 excursion size from the *base line* to the target tonal point (L_i), which leads to a higher F0 value of the L_i tone: we predict that the F0 value of L_i should be 310 Hz. This is shown in Figure 4.11.

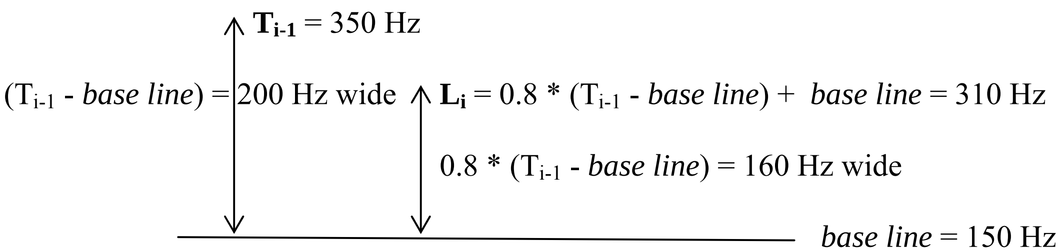


Figure 4.11. A Hypothetical L Tone Scaling (base line is 150 Hz, the coefficient a is 0.8 and the F0 value of T_{i-1} is 350 Hz.)

Now let us consider actual L tones in Tokyo Japanese, i.e. the L edge tone and the +L accent tone. The +L tone immediately following the H* accent tone is always lower than the non-post-accent L edge tone immediately following a H edge tone. I propose that this is because the coefficient of the equation formula associated with +L accent tones, a_{+L}, is smaller than that associated with L edge tones, a_{Ledge}. This smaller coefficient of the +L accent tone leads to a smaller F0 excursion size from the baseline to the +L tone, which ultimately leads to a lower absolute F0 value of the +L tone. Consider the formulae shown in (6) and (7): the formula in (6) is for the +L accent tone and that in (7) is for the L edge tone. The part of the equation in (6) $a_{+L} * (T_{preceding} - \textit{base line})$ denotes the F0 excursion size from the base line to the +L accent tone and the same part in (7) $a_{Ledge} * (T_{preceding} - \textit{base line})$ denotes the excursion size from the *baseline* to the L edge tone. Since the coefficient a_{+L} is smaller than a_{Ledge}, the excursion size obtained by $a_{+L} * (T_{preceding} - \textit{base line})$ is smaller than that obtained by $a_{Ledge} * (T_{preceding} - \textit{base line})$.

(6) $+L = a_{+L} * (T_{preceding} - \text{base line}) + \text{base line}$

(7) $L_{edge} = a_{Ledge} * (T_{preceding} - \text{base line}) + \text{base line}$

where $0 < a_{+L} < a_{Ledge} < 1$

These equations in (6) and (7) are also converted into the following formulae in (6') and (7') respectively.

(6') $+L = (1-a_{+L}) * \text{base line} + a_{+L} * T_{preceding}$

(7') $L_{edge} = (1-a_{Ledge}) * \text{base line} + a_{Ledge} * T_{preceding}$

where $0 < a_{+L} < a_{Ledge} < 1$

According to the equations in (6') and (7'), the relationship between each of these L tones and preceding tone is expressed by a "linear" slope as graphically presented below.

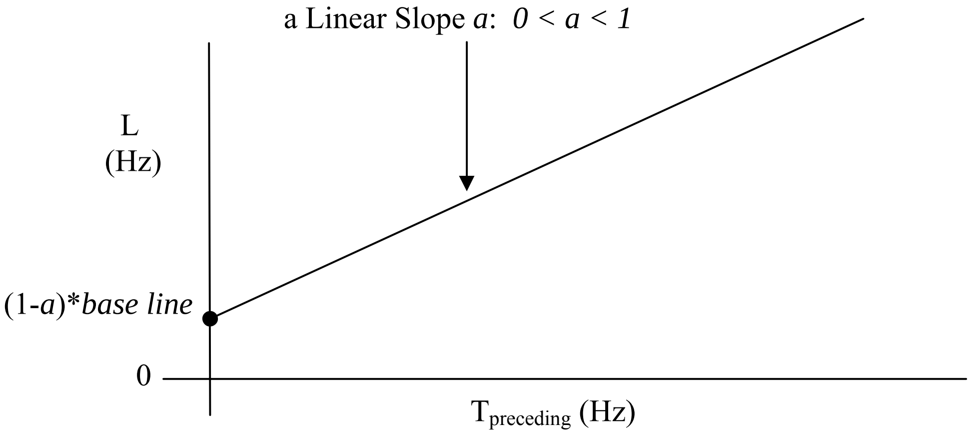


Figure 4.12. A Hypothetical Linear Relationship between L and the Preceding T

Those equations in (6') and (7') also predict that the intercept value of the linear slope should be different between the L edge tone and the +L accent tone. Since the coefficient of the L edge tone, a_{Ledge}, is greater than that of the +L accent tone, $a_{\text{+L}}$, the intercept of the L edge tone, which corresponds to $(1-a_{Ledge})*base\ line$, should be smaller than the intercept of the +L accent tone, which corresponds to $(1-a_{+L})*base\ line$.

In the following, we examine whether the linear relationship expressed by the equations in (6') and (7') is really present and test relevant predictions made by them.

4.3.1.2. A Regression Analysis between +L and the Preceding H*

To test the equation in (6'), a regression analysis between the +L tone and the immediately preceding H* accent tone was carried out. Data from sentences shown in (8) and (9) were used for this analysis.

from the <*Maronkéeki Set*>

(8) **UAA**

[*nijukko-iri-no*]	[*yuuháimu-no*]$_{\mathrm{MiP2}}$	[*maronkéeki*]
H_1	$\mathbf{H_2*}$ $\mathbf{+L_2}$	
twenty-pieces- -Copula	Juheim-Gen	chestnut cake

"twenty pieces of chestnut cake of Juheim in a single box"

(9) **AAA**

[*nihyakúen-no*]	[*yuuháimu-no*]$_{\mathrm{MiP2}}$	[*maronkéeki*]
$H_1* +L_1$	$\mathbf{H_2*}$ $\mathbf{+L_2}$	
two hundred yen-Copula	Juheim-Gen	chestnut cake

"chestnut cake of Juheim that are two hundred yen"

We examine the correlation between the H_2* accent tone of the second Minor Phrase (i.e. MiP2) and the immediately following +L accent tone of the same phrase. MiP2 is preceded by an unaccented MiP1 in (8) while the same phrase is preceded by an accented MiP1 in (9). Since the H_2* accent tone in (9) is post-accent, it is predicted to be realized lower than the non-post-accent H_2* accent tone in (8). Such variation in the F0 of H_2* is necessary in order to examine the nature of the correlation between the F0 of the H_2* accent tone and that of the immediately following $+L_2$ accent tone. The expected linear relationship between $+L_2$ and H_2* is graphically presented in Figure 4.13.

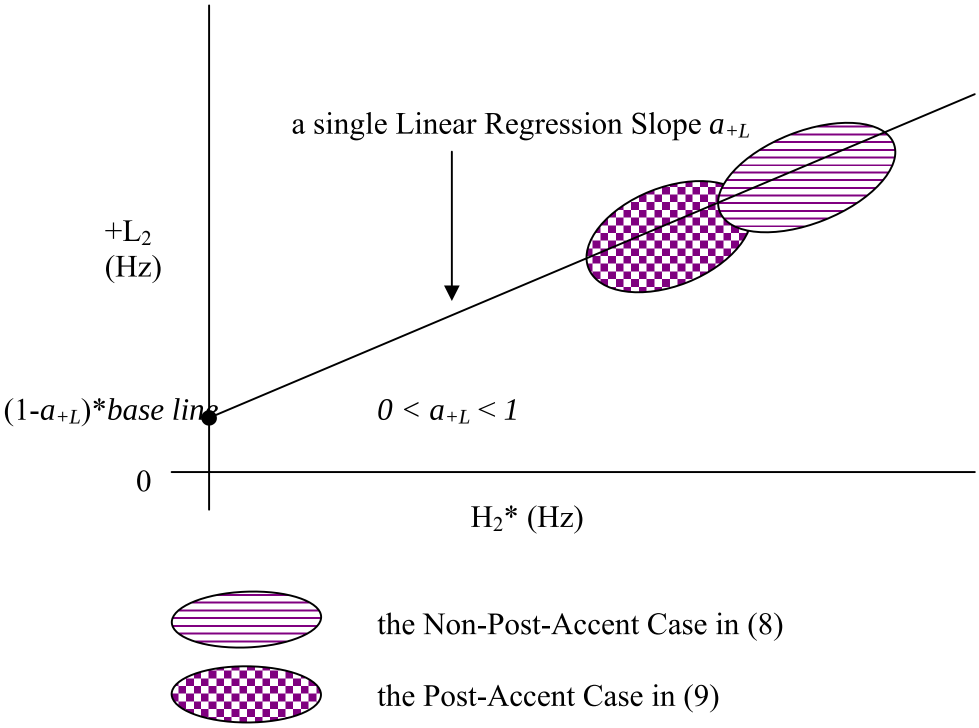


Figure 4.13. The Expected Relationship between +L_2 and the Preceding H_2*

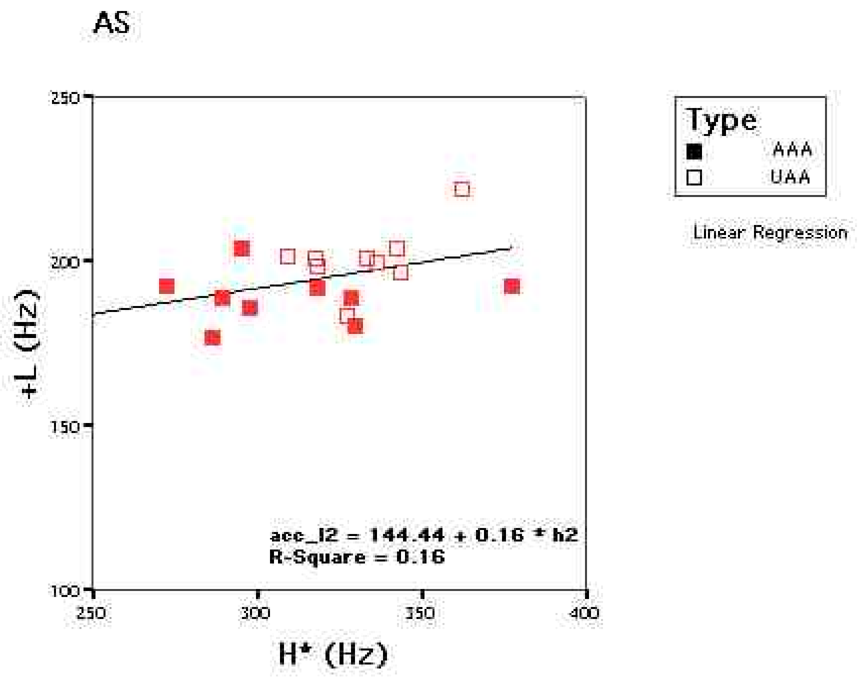
Though the post-accent case in (9) and the non-post-accent case in (8) share different mean F0 values of H_2* and +L_2, their data points are expected to be fitted to a single positive regression slope as in Figure 4.13. To test this expectation, a regression analysis was carried out. The F0 value of the H_2* was made into a predictor and that of the +L_2 was made into a dependent variable.[18] The results of the regression analysis, i.e. regression coefficients and R^2, are shown in Table 4.03 and in the scatter plots of Figure 4.14.

[18] The F0 value of +L_2 was measured adopting the "intersection" procedure introduced in Section 1.2 of this chapter.

Table 4.03. Regression Coefficients

	Intercept	Standard Error	Slope	Standard Error	R^2	n
AS	144!	28.72	0.16	0.89	0.16	18
MR	143!	20.59	0.17*	0.06	0.29	20
RO	60#	29.48	0.39!	0.09	0.51	19
SK	77*	31.49	0.36!	0.09	0.54	18

! $p < 0.01$
* $p < 0.05$
# only marginally significant ($p < 0.06$)

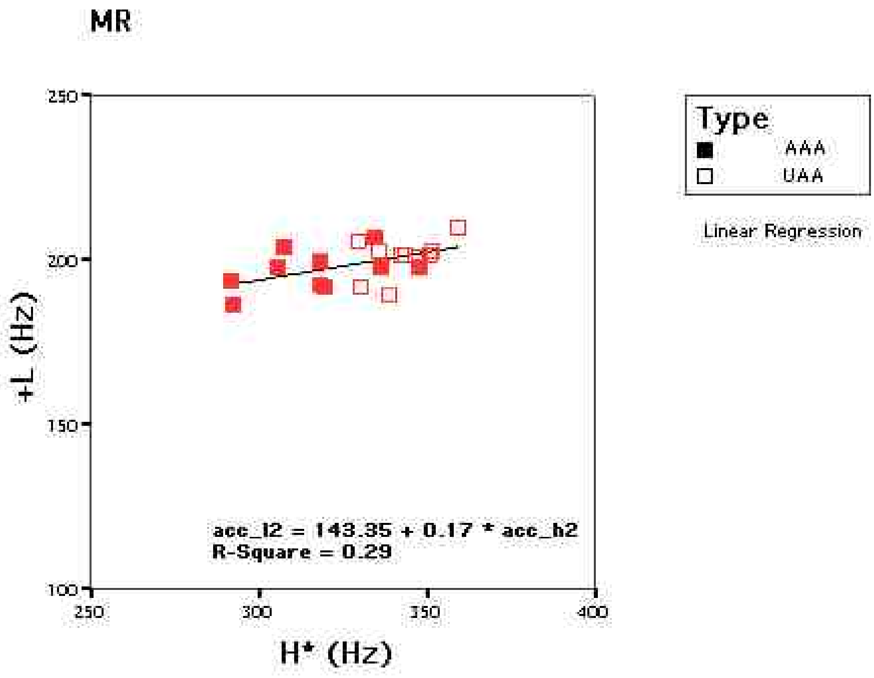


a. Speaker AS

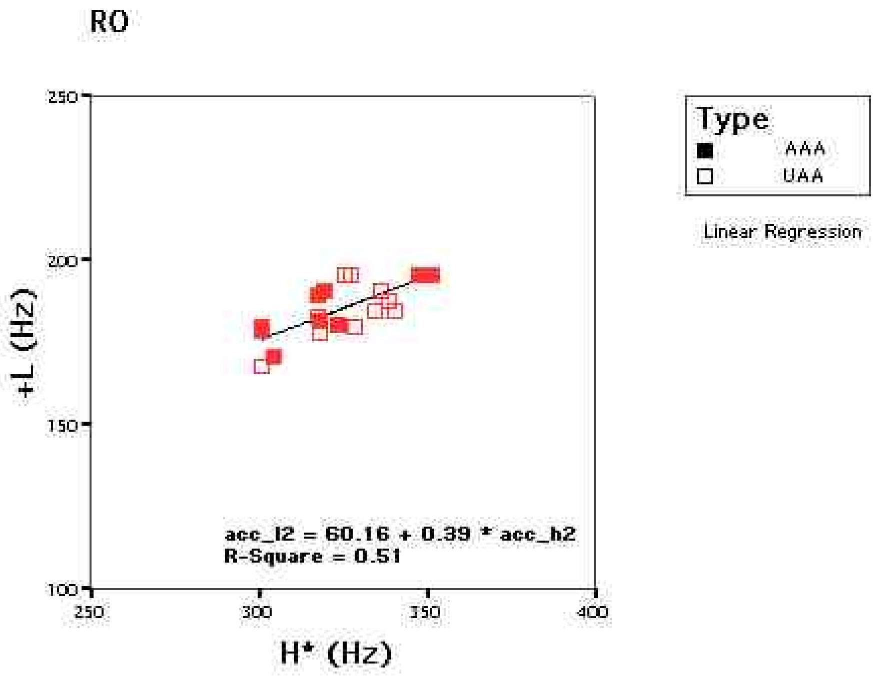
Figure 4.14. The Relationship between the preceding H* and the following +L

Continued next page

Figure 4.14 continued



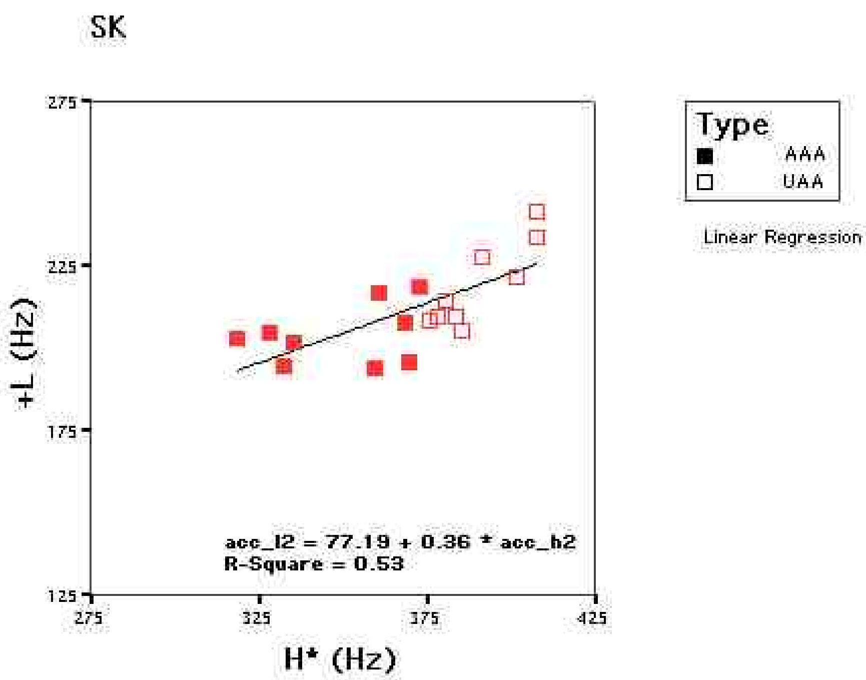
b. Speaker MR



c. Speaker RO

Continued next page

Figure 4.14 continued



d. Speaker SK

Table 4.03 shows that three of the four speakers (MR, RO and SK) had a significant positive regression slope fitted to the data points of both the non-post-accent H_2*~+L_2 relationship and the post-accent H_2*~+L_2 relationship. Their regression slope varies between 0.17 and 0.39 (MR: 0.17, RO: 0.39, SK: 0.36). Also they had a statistically significant or marginally significant intercept (MR: 143, RO: 60, SK: 77). Those intercept values represent the *base line* multiplied by (*1-slope*).

Those results comply with the prediction of our model of +L tone scaling, namely that the relationship between the +L and the preceding H* should fall on a positive slope less than 1 with a non-zero intercept. In the next subsection, we test another L tone scaling of our model, i.e. the scaling of the L edge tone.

4.3.1.3. A Regression Analysis between the L Edge tone and the Preceding Tone

In this subsection, the relationship between the L edge tone and the preceding tone is examined. According to our L tone scaling model, the relationship between a L edge tone and the preceding tone should be also accounted for by a linear positive slope less than 1. As mentioned in Section 4.3.1, the slope associated with the L edge tone scaling is predicted to be greater than the slope associated with the +L tone scaling that we observed in Section 4.3.2. At the same time, our model predicts that the intercept of the linear slope associated with the L edge tone scaling should be smaller than that associated with the +L tone scaling (see Section 4.3.1 for reasoning).

4.3.1.3.1. Reading Materials

Data obtained from the full paradigm of the <*Maronkéeki Set*> are used to examine the relationship between the L edge tone and the preceding tone.

The <*Maronkéeki* Set>

(8) **UAA**

[*nijukko-iri-no*]$_{\text{MiP1}}$ [*yuuháimu-no*]$_{\text{MiP2}}$ [*maronkéeki*]$_{\text{MiP3}}$

H_2* **+L_2** **L_3**

twenty-pieces- -Copula Juheim-Gen chestnut cake

"twenty pieces of chestnut cake of Juheim in a single box"

(9) **AAA**

[*nihyakúen-no*]$_{\text{MiP1}}$ [*yuuháimu-no*]$_{\text{MiP2}}$ [*maronkéeki*]$_{\text{MiP3}}$

H_2* **+L_2** **L_3**

two hundred yen-Copula Juheim-Gen chestnut cake

"chestnut cake of Juheim that are two hundred yen"

(10) **UUA**

[*nijukko-iri-no*]$_{\text{MiP1}}$ [*imuraraya-no*]$_{\text{MiP2}}$ [*maronkéeki*]$_{\text{MiP3}}$

H$_2$ **L$_3$**

twenty-pieces- -Copula Imuraya-Gen chestnut cake

"twenty pieces of chestnut cake of Imuraya in a single box"

(11) **AUA**

[*nihyakúen-no*]$_{\text{MiP1}}$ [*imuraya-no*]$_{\text{MiP2}}$ [*maronkéeki*]$_{\text{MiP3}}$

H$_2$ **L$_3$**

two hundred yen-Copula Imuraya-Gen chestnut cake

"chestnut cake of Imuraya that are two hundred yen"

All four forms in the <*Maronkéeki Set*> share the same third word, i.e. *maronkéeki*, and the F0 scaling of the L_3 edge tone associated with that word is our concern. Depending on whether the preceding second word (i.e. MiP_2) is accented or unaccented, the L_3 edge tone is immediately preceded by either a $+L_2$ accent tone or a H_2 edge tone. Henceforth, I refer to those immediately preceding tones as $T_{\text{preceding}}$. Therefore, the $T_{\text{preceding}}$ tone in (8) and (9) is $+L_2$ accent tone, while it is the H_2 edge tone in (10) and (11).

(8) and (9) [Acc/Unacc]$_{\text{MiP1}}$ [Accented]$_{\text{MiP2}}$ [ ]$_{\text{MiP3}}$

H_2* **+L$_2$** **L$_3$**

↓

T$_{\text{preceding}}$

(10) and (11) [Acc/Unacc]$_{\text{MiP1}}$ [Unaccented]$_{\text{MiP2}}$ [ ]$_{\text{MiP3}}$

H$_2$ **L$_3$**

↓

T$_{\text{preceding}}$

Also, the full paradigm of this <*Maronkéeki Set*> allows us to substantially vary the F0 value of each of these $T_{preceding}$ tones which are associated with the second word/phrase (i.e. MiP_2) by varying the accent condition of the first word (i.e. MiP_1). The $T_{preceding}$ tone may be realized lower when it is preceded by an accented MiP_1 as in (9) and (11), while it may be realized higher when it is preceded by an unaccented MiP_1 as in (8) and (10). Such variation in the F0 value of the $T_{preceding}$ tone is necessary to obtain the relationship between the F0 of the $T_{preceding}$ tone and that of the following L_3 edge tone.

4.3.1.3.3. Measurement Points of $T_{preceding}$

The measurement points of $T_{preceding}$ need to be clarified. When $T_{preceding}$ is the $+L_2$ tone, i.e. when the L_3 edge tone was preceded by an accented MiP_2, I simply measured the F0 value of the point where the $+L_2$ accent tone is aligned (see 3.1.2. for how to find the point). When $T_{preceding}$, is the H_2 edge tone of the preceding MiP_2, i.e. when the L_3 edge tone was preceded by an unaccented MiP_2, the measurement point of the H_2 edge tone was the "right" edge of the H_2 edge tone. The "right edge of the H_2 tone", however, was not unitary across all speakers. Remember that there are two types of speakers: the "H edge tone spreading" and the "non-spreading" speakers. Three of the four speakers (MR, RO and SK) belong to the former group and AS belongs to the latter. That is, those three "spreading" speakers have the H edge tone spread from the second syllable to the final syllable of a Minor Phrase when the Minor Phrase is unaccented, as graphically shown in Figure 4.15.

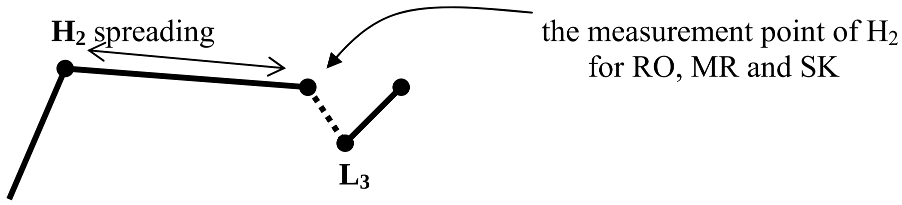


Figure 4.15. The H_2 edge tone ($T_{preceding}$) measurement point for the "H-tone spreading" speakers RO, MR and SK

For those three speakers, it was the right edge of the "spread" H_2 tone, i.e. the right edge of the high plateau, that I chose as the measurement point of $T_{preceding}$. When it comes to AS, the only speaker without the H spreading, I simply measured the highest F0 of the second word (i.e. MiP_2), assuming that the peak F0 of that Minor Phrase corresponds to the right edge of the H_2 edge tone.

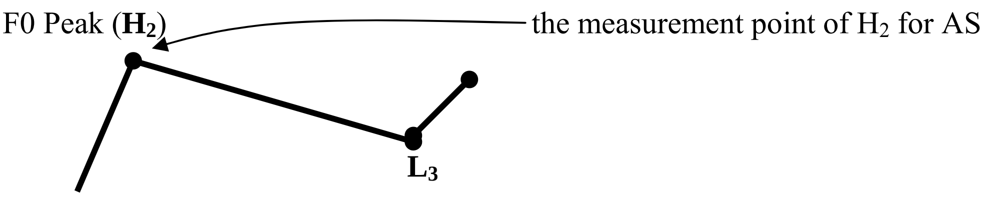


Figure 4.16. The H_2 edge tone ($T_{preceding}$) measurement point for Speaker AS

4.3.1.3.3. Predictions

According to our L tone scaling model, the relationship between the L_3 edge tone and those immediately preceding tones should be accounted for by a single positive linear slope a_{Ledge}. Such an expected relationship between the L_3 edge tone and $T_{preceding}$ is graphically represented in Figure 4.17.

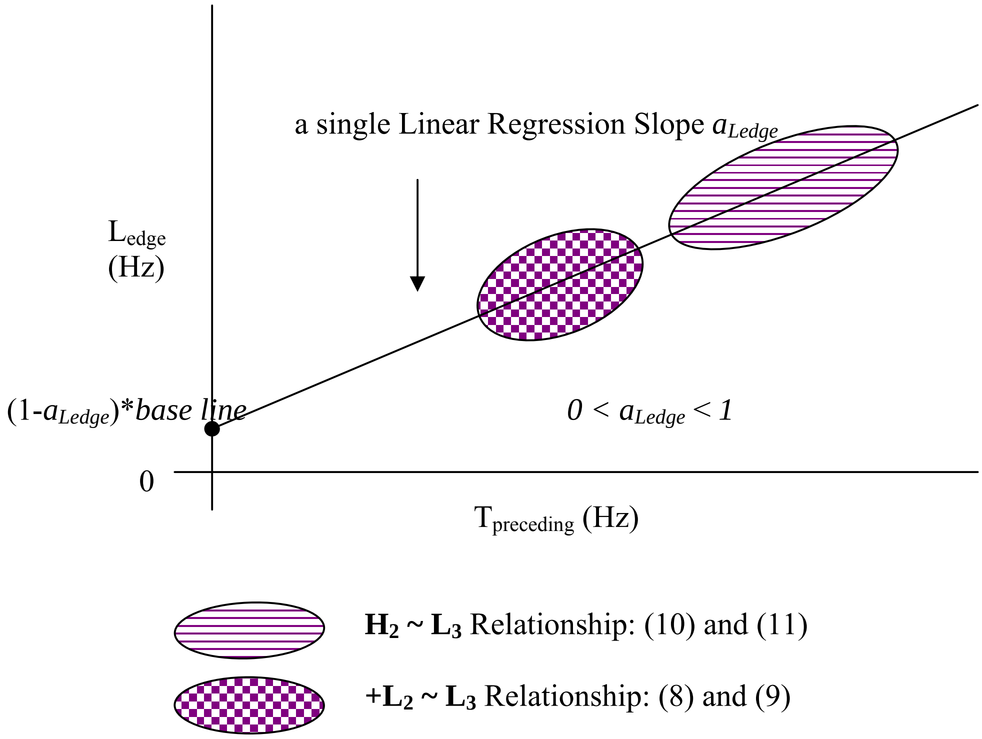


Figure 4.17. The expected relationship between L_3 edge tone and the preceding tone ($T_{preceding}$)

At the same time, our model predicts that the slope a_{Ledge} should be greater than the slope a_{+L} associated with the +L tone scaling that we observed in the previous section (see 4.3.1.2). Our model also expects that the intercept of the slope associated with the L edge tone, i.e. $(1-a_{Ledge})$ * *base line*, should be smaller than the intercept of the slope associated with the +L tone scaling, i.e. $(1-a_{+L})$ * *base line*. It is because a_{Ledge} is greater than a_{+L}, and $(1-a_{Ledge})$ is smaller than $(1-a_{+L})$ as a result. To test those predictions, a regression analysis between the F0 value of the L_3 edge tone (the dependent variable) and that of $T_{preceding}$ (the predictor) was carried out.

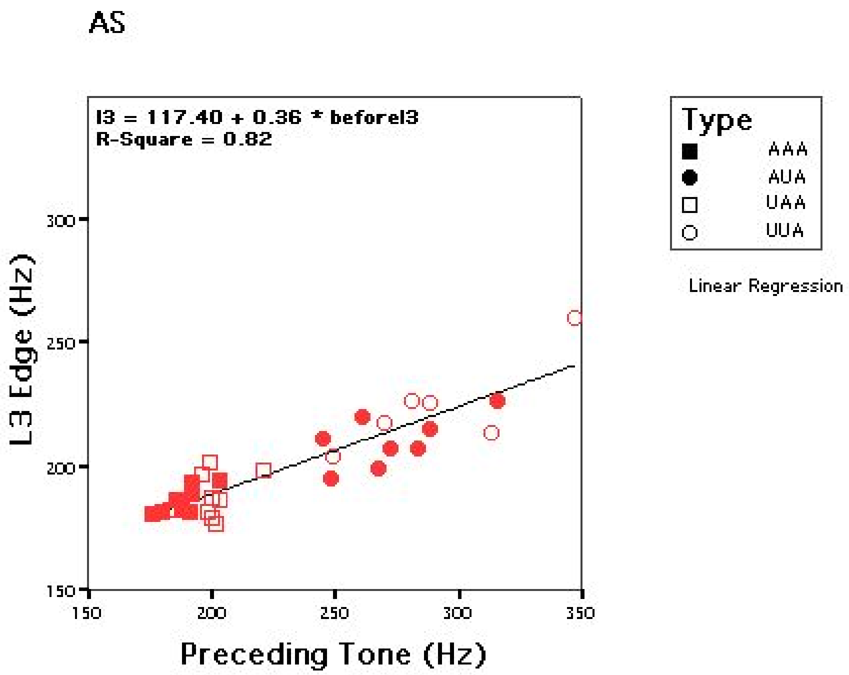
4.3.1.3.4. Results

The results of the regression analysis (i.e. regression coefficients and R^2) are presented in Table 4.04 and Figure 4.18.

Table 4.04. Regression Coefficients

	Intercept	Standard Error	Slope	Standard Error	R^2	n
AS	117!	7.34	0.36!	0.03	0.82	32
MR	1	5.24	0.93!	0.02	0.98	40
RO	69!	7.75	0.57!	0.03	0.89	40
SK	41!	5.86	0.80!	0.02	0.98	36

! $p < 0.01$

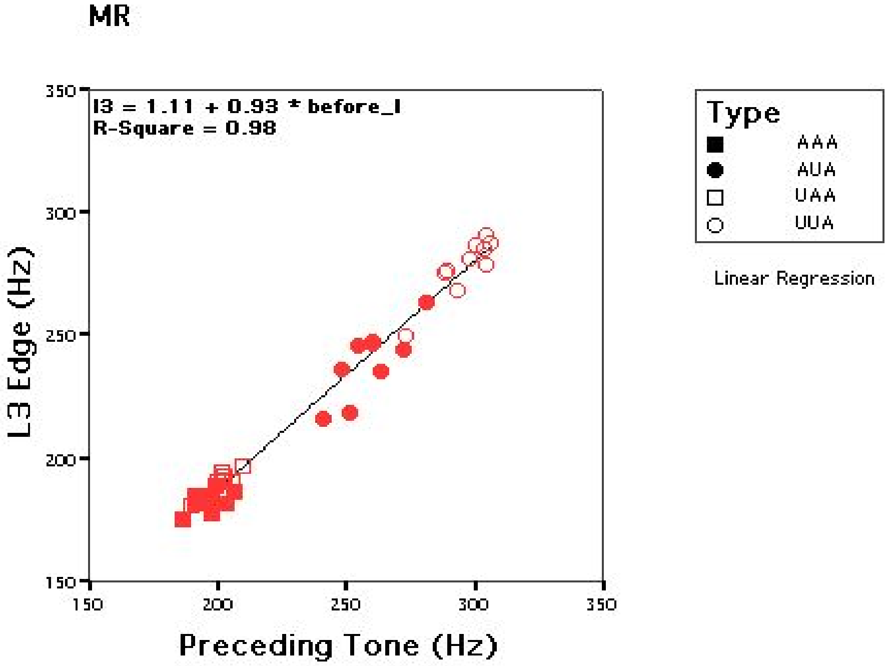


a. Speaker AS

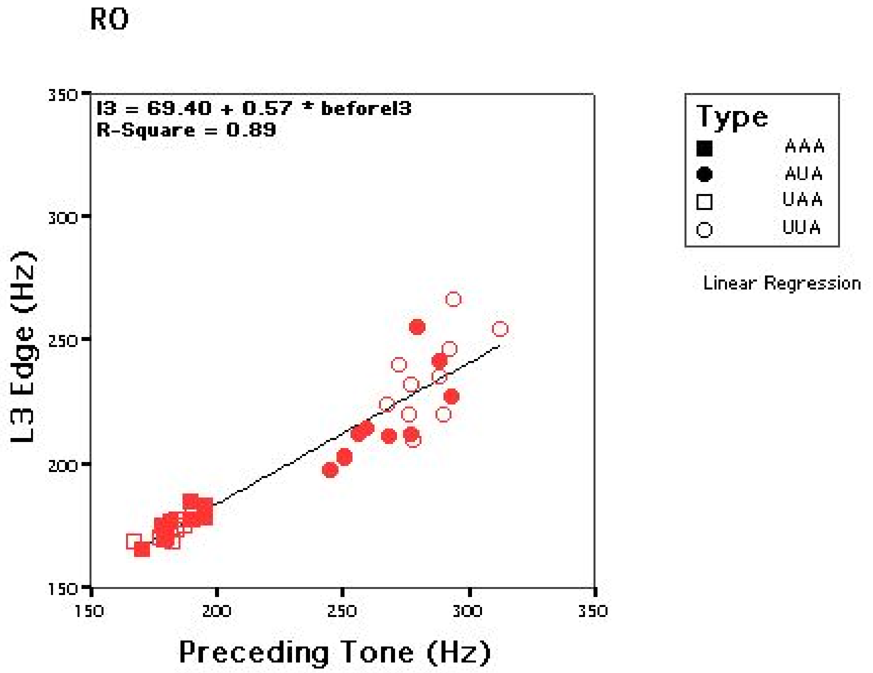
Figure 4.18. The Relationship between $T_{preceding}$ and L3

Continued next page

Figure 4.18 continued



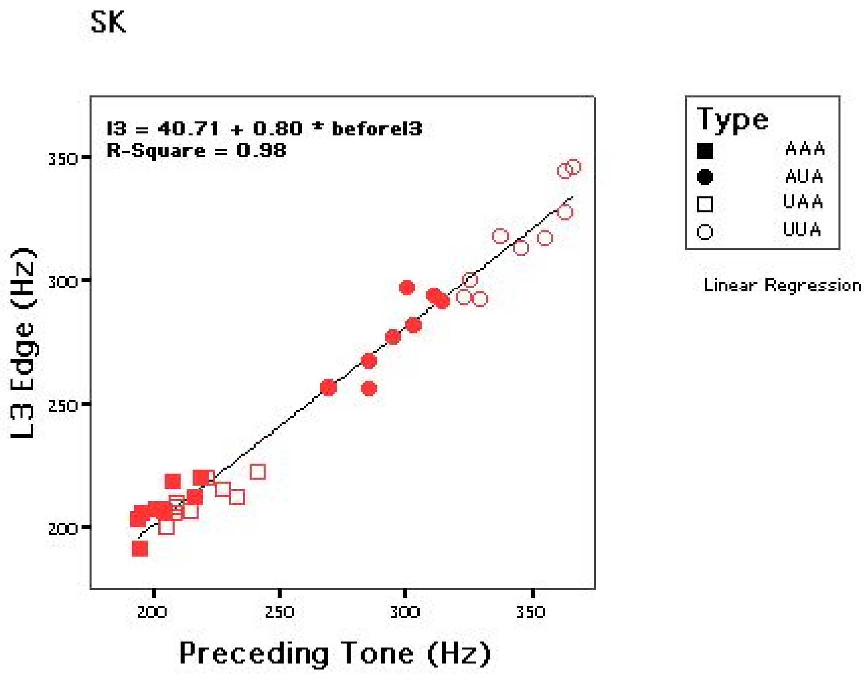
b. Speaker MR



c. Speaker RO

Continued next page

Figure 4.18 continued



d. Speaker SK

The first of our predictions, namely the L_3~$T_{preceding}$ relationship should fall in a linear positive slope less than 1 was confirmed: the slope varied between 0.36 and 0.93 and all those slopes are significantly greater than zero. At the same time, there was a very tight correlation between $T_{preceding}$ and L_3: R^2 (the coefficient of determination) was between 0.92 and 0.98.

Our second prediction that those slopes should be greater than the slopes associated with the $+L_2$~H_2* relationship that we observed in 4.3.1.2 was also supported: for each of the four speakers, the slope of L_3~$T_{preceding}$ relationship was greater than that of the $+L_2$~H_2* relationship as shown in Table 4.05.

Table 4.05. The L_3~$T_{preceding}$ Relationship Slope and the $+L_2$~H_2* Relationship Slope

Speaker	$\boldsymbol{a_{+L}}$ $+L_2$~H_2* relationship (from Table 3.03)		$\boldsymbol{a_{Ledge}}$ L_3~$T_{preceding}$ relationship (from Table 3.04)
AS	0.16	<	0.36!
MR	0.17*	<	0.93!
RO	0.39!	<	0.57!
SK	0.36!	<	0.80!

! $p < 0.01$
* $p < 0.05$

The third prediction of our L tone scaling model is that there should be a positive "non-zero" intercept for the slope of the L_3~$T_{preceding}$ relationship. Three of the four speakers satisfied this prediction. However, MR's intercept was only *1*, which was not statistically significant from zero. This result indicates that MR actually does not refer to the *base line* when computing the F0 of the L_3 tone. Rather, she computes it by simply multiplying the F0 value of $T_{preceding}$ by the slope coefficient 0.93 as shown below.

(12) MR's L edge tone scaling:

$$L_{edge} = 0.93 * T_{preceding}$$

Given this abnormal behavior of MR, I exclude her results of the L_3~$T_{preceding}$ relationship from the discussions that follow.

The fourth prediction of our L edge tone scaling model is that the intercept of those slopes should be smaller than that of the $+L_2$~H_2* relationship obtained in 4.1.2. This prediction was supported by two of the three speakers, AS and SK (see Table 4.06).

Table 4.06. The $L_3{\sim}T_{preceding}$ Relationship Slope Intercept and the $+L_2{\sim}H_2*$ Relationship Slope Intercept

Speaker	Intercept $+L_2{\sim}H_2*$ relationship (from Table 3.03)		Intercept $L_3{\sim}T_{preceding}$ relationship (from Table 3.04)
AS	144	>	117
RO	60	<	69
SK	77	>	41

RO did not comply with this prediction. It may be because of the large Standard Error of RO's estimated intercept of the $+L_2{\sim}H_2*$ relationship. The error of the estimated intercept is 29.48 (see Table 4.03) and it is almost the half of the estimated intercept itself: the contradictory result of RO may be due to the large Standard Error.

4.3.1.4. The L Tone Scaling Model and the Base Line

According to our L tone scaling model, the results shown in Table 4.03 and Table 4.04 allow us to derive an estimate of the *base line* value. Remember that our L tone scaling phonetic rules are expressed by the formulae in (6) and (7)

(6) $+L = a_{+L}*(T_{preceding} - \text{base line}) + \text{base line}$

$= (1-a_{+L})*\text{base line} + a_{+L}*T_{preceding}$

(7) $L_{edge} = a_{Ledge}*(T_{preceding} - \text{base line}) + \text{base line}$

$= (1-a_{Ledge})*\text{base line} + a_{Ledge}*T_{preceding}$

Following those equation formulae, the intercepts of the slopes that we obtained in the previous subsections are all expressed by the form ***(1-slope)*base line***. Since slope values are also known, we can derive an estimate of the *base line* values from the intercept values as shown in (13).

(13) base line = intercept / (1-slope)

If those phonetic rules shown in (6) and (7) are adequate, then we expect that the *base line* value estimated from the intercept and the slope obtained in our previous regression analyses should comply with the reality that our female speakers' speech-related voice never reaches below 140 Hz.

Table 4.07. The Estimated Base Line Values

Speaker	Estimation from the $+L_2$~H_2* relationship	Estimation from the L_3~$T_{preceding}$ relationship
AS	171 Hz	182 Hz
MR	172 Hz	N.A.
RO	98 Hz (unrealistic)	160 Hz
SK	120 Hz (unrealistic)	205 Hz

AS and MR's *base line* value estimated from the slope and intercept of their $+L_2$~H_2* relationship is approximately 170 Hz, and AS, RO and SK's *base line* value estimated from the slope and intercept of their $L_3 \sim T_{preceding}$ relationship varies between 160 Hz and 200 Hz. Those estimated values of the *base line* are plausible.

However, there are problematic cases as well: the estimated *base line* value of RO and SK's $+L_2 \sim H_2^*$ relationship is 98 Hz and 120 Hz respectively. It is unlikely that those speakers refer to such low F0 values that they even cannot produce when they compute the F0 values of those L tones. One account for the unrealistic *base line* value of RO and SK's $+L_2 \sim H_2^*$ relationship is that some additional factors are at work in their +L tone scaling. More specifically, I propose that they have an additional constant "- *c*" in their equation of the +L tone as shown below.

(14) Speaker RO and SK

$+L = a_{+L} * (T_{preceding} - \text{base line}) + \text{base line} \mathbf{- c}$

$= \mathbf{-c} + (1 - a_{+L}) * \text{base line} + a_{+L} * T_{preceding}$

$\therefore$ intercept $= \mathbf{-c} + (1 - a_{+L}) * \text{base line}$

base line = (intercept - (-c)) / $(1 - a_{+L})$

At this moment, I have no explanation for where this constant -*c* comes from. However, we could derive an appropriate *base line* value by having the additional constant. For example, let us assume that the F0 value of the additional coefficient -*c* is -50. According to Table 3.03, SK's intercept of the $+L_2 \sim H_2^*$ relationship is 77 and its slope is 0.36. Given this additional constant -50, the *base line* value is estimated to 167 (= [77 - (-50)] / [1 - 0.36]). This estimated value, i.e. 198 Hz, is within the speakers' speech-related voice pitch range, and a plausible *base line* value.

4.3.1.5. Summary of Section 4.3.1.

In summary, there was good correlation between those L tones and the immediately preceding tones. All of those relations were fitted to a positive slope less than 1, which means that (i) the F0 value of L tones is a function of the F0 value of the immediately preceding tone, and (ii) only a fraction of the F0 of the preceding tone is reflected on the F0 value of those L tones. In the following section (Section 4.3.2), I show that it is not only the F0 of L tones but also the F0 of H tones that fall in a similar relationship with that of the immediately preceding tone.

4.3.2. The Tone-by-Tone Scaling Model of H Tones

In this subsection, I introduce the tone-by-tone scaling model of H tones. The general proposals are first presented in 4.3.2.1, and those proposals are tested in 4.3.2.3 and 4.3.2.4. Also, a discussion of how the *catathesis* diminution is achieved via the tone-by-tone scaling proposed in our model will be provided in 4.3.2.2.

4.3.2.1. The General Ideas of the H Tone Scaling

I propose that there is a set of phonetic rules that determines the F0 value of H tones based on the F0 value of the preceding tone and that of the abstract *top line*. Those phonetic rules take the form of the following equation formula.

(15)
$$\begin{aligned} H_i &= d^*(\text{top line} - T_{i-1}) + T_{i-1} \\ &= d^*\text{top line} + (1-d)^*T_{i-1} \qquad 0 < d < 1 \end{aligned}$$

I assume that the *top line* is approximately the highest speech-related F0 of speakers' voice range to which speakers refer when they compute H tone values. Though Pierrehumbert & Beckman (1988) also assumed the presence of those abstract lines, the major difference between this alternative model and theirs is that the F0 value of the top line is kept constant even in the post-accent region in this tone-by-tone scaling model.

The coefficient of the equation in (15), i.e. *d*, is always less than 1 but greater than zero. The F0 value of H_i increases proportional to the value of the coefficient (or the slope) *d*. The value of *d* may vary according to the type of H tones (i.e. whether H* accent tone or a H edge tone), phonological, syntactic and information structure factors, etc. Also, I later show that this positive coefficient (or slope) is responsible for the *catathesis* diminution observed in Section 4.2.

Let us imagine a hypothetical case shown in Figure 4.19, where the *top line* is 350 Hz, the coefficient *d* is 0.8 and the F0 value of T_{i-1} is 200 Hz. Given this, we predict that the F0 of H_i is 320 Hz.

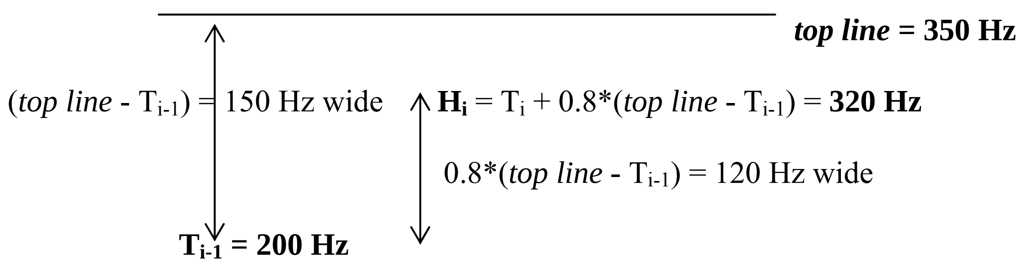


Figure 4.19. A Hypothetical H Tone Scaling

4.3.2.2. The *Catathesis* Diminution and Tone-by-Tone Scaling of H Tones

Given the equations in (15), the relationship between H tones and preceding tones is expressed by a positive linear slope (*1-d*) with an intercept of (*d*top line*) as graphically presented below.

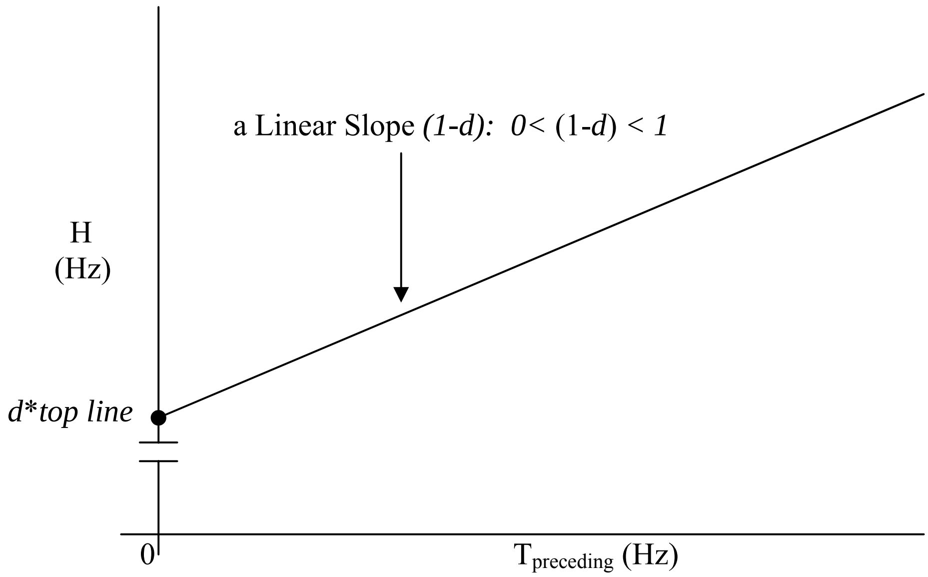


Figure 4.20. A Hypothetical Linear Relationship between H and the Preceding T

According to this linear relationship between the following H tone and the preceding tone, only a fraction of the F0 value of the preceding tone is reflected on the F0 value of the following H tone because the slope (*1-d*) is always less than 1. This is why the *catathesis* diminution takes place, which is already introduced in Section 4.2.

Let us imagine a more concrete hypothetical case of a H edge tone scaling and a H* accent tone scaling to see how the *catathesis* diminution is achieved by our tone-by-tone scaling model. In the hypothetical case, the coefficient d_{Hedge} of the equation of H edge tone scaling is 0.6 and the coefficient d_{H*} of the equation of the H* accent tone scaling is 0.2. At the same time, the top line is 350 Hz. Given those conditions, the equation formulae of those H tones are expressed as those in (16) and (17).

(16) A Hypothetical H_{edge} Tone Scaling (H_{edge} preceded by L_{edge})

$$\begin{aligned} H_{edge} &= d_{Hedge} * \text{top line} + (1- d_{Hedge}) * L_{edge} \\ &= 0.6*350 + (1-0.6)*L_{edge} \\ &= 210 + 0.4 * L_{edge} \end{aligned}$$

(17)A Hypothetical H_{accent} Tone Scaling (H_{accent} preceded by H_{edge})

$$\begin{aligned} H_{accent} &= d_{H*} * \text{top line} + (1- d_{H*}) * H_{edge} \\ &=0.2*350 + (1-0.2)* H_{edge} \\ &= 70 + 0.8 * H_{edge} \end{aligned}$$

Since the slope $(1\text{-}d_{Hedge})$ in (16) is less than 1, i.e. 0.4, *x* Hz difference in the F0 of the preceding L_{edge} tone results in only four tenths of the *x* Hz difference in the F0 of the following H_{edge} tone.

For example, let us imagine that a non-post-accent L edge tone is 250 Hz while a post-accent L edge tone is 160 Hz. That is, there is a 90 Hz difference between those L tones. According to the hypothetical equation function in (16), the H tone that immediately follows the non-post-accent L edge tone is realized as 310 Hz (= 210 + 0.4*250), while the H edge tone that immediately follows the post-accent L edge is realized as 274 Hz (= 210 + 0.4*160). That is, the 90 Hz difference between the post-accent L edge tone and the non-post-accent L edge tone is diminished into only 36 Hz difference between the post-accent H edge tone and the non-post-accent H edge tone. This is graphically shown in Figure 4.21.

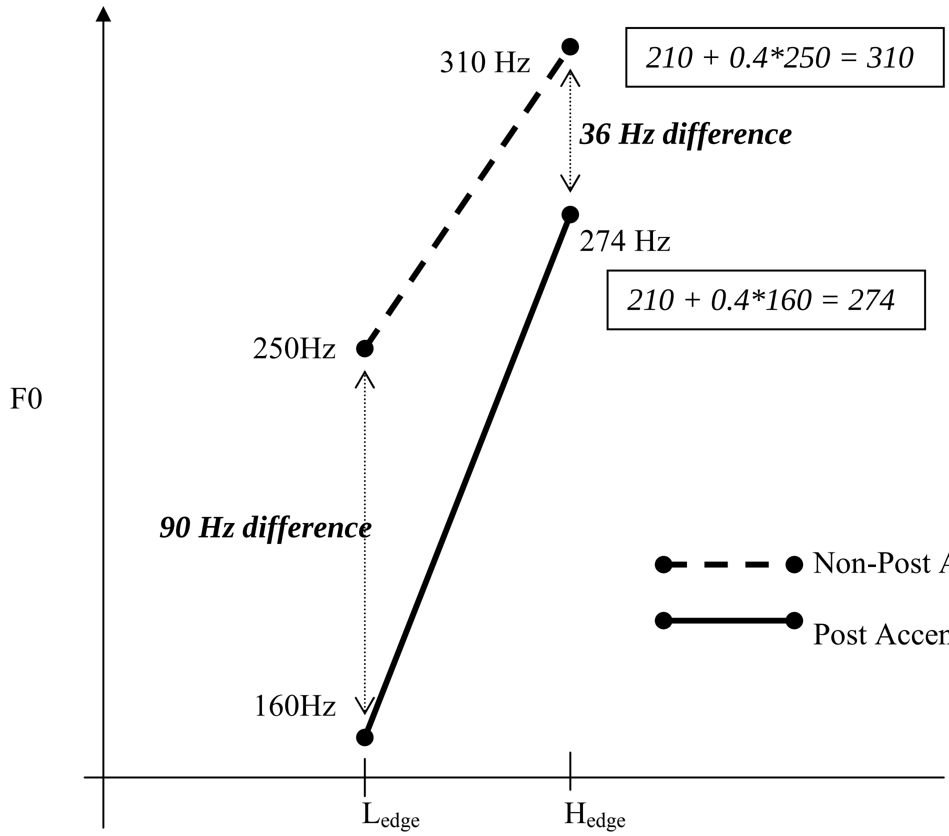


Figure 4.21. A Hypothetical Case of Catathesis Diminution

In the same way, the 36 Hz difference between the post-accent H edge tone and the non-post-accent H edge tone is further diminished into a smaller difference between the post-accent H* accent tone and the non-post-accent H* accent tone. Given that the post-accent H edge tone is 274 Hz and the non-post-accent H edge tone is 310 Hz, the following post-accent H* accent tone bears an F0 value of 289 Hz (= 70 + 0.8* H_{edge}) and the non-post-accent H* accent tone bears an F0 of 318 Hz according to the hypothetical equation formula in (17). That is, the 36 Hz difference between the post-accent and the non-post-accent H edge tones is diminished into 29 Hz difference between the post-accent and the non-post-accent H* accent tones. This is how the *catathesis* diminution observed in Section 4.2 is achieved by the local tone-by-tone scaling.

4.3.2.3. The H Edge Tone Scaling and Regression Analyses

To test the H tone scaling model proposed in 4.3.2.1, regression analyses between H tones and the immediately preceding tones were carried out. Results of the regression analysis between the H edge tone and the preceding L edge tone are provided in this subsection.

4.3.2.3.1. Reading Materials and Predictions

Again, data obtained from the full paradigm of the <*Maronkéeki* Set> are used, which are shown below. This time, our main concern is the relationship between the H_3 edge tone of the third word/MiP_3 and the immediately preceding L_3 edge tone of the same word.

The <*Maronkéeki* Set>

(8) **UAA**

[*nijukko-iri-no*]$_{\text{MiP1}}$	[*yuuháimu-no*]$_{\text{MiP2}}$	**[*maronkéeki*]$_{\text{MiP3}}$**
	$H_2* + L_2$	$\mathbf{L_3}$ $\mathbf{H_3}$
twenty-pieces- -Copula	Juheim-Gen	chestnut cake

"twenty pieces of chestnut cake of Juheim in a single box"

(9) **AAA**

[*nihyakúen-no*]$_{\text{MiP1}}$	[*yuuháimu-no*]$_{\text{MiP2}}$	**[*maronkéeki*]$_{\text{MiP3}}$**
	$H_2* + L_2$	$\mathbf{L_3}$ $\mathbf{H_3}$
two hundred yen-Copula	Juheim-Gen	chestnut cake

"chestnut cake of Imuraya that are two hundred yen"

(10) **UUA**

[*nijukko-iri-no*]$_{\text{MiP1}}$	[*imuraraya-no*]$_{\text{MiP2}}$	**[*maronkéeki*]$_{\text{MiP3}}$**
	H_2	$\mathbf{L_3}$ $\mathbf{H_3}$
twenty-pieces- -Copula	Imuraya-Gen	chestnut cake

"twenty pieces of chestnut cake of Imuraya in a single box"

(11) **AUA**

[*nihyakúen-no*]$_{\text{MiP1}}$	[*imuraya-no*]$_{\text{MiP2}}$	**[*maronkéeki*]$_{\text{MiP3}}$**
	H_2	$\mathbf{L_3}$ $\mathbf{H_3}$
two hundred yen-Copula	Imuraya-Gen	chestnut cake

"chestnut cake of Imuraya that are two hundred yen"

If our H tone scaling model is on the right track, then we expect that the data points representing the relationship between the H_3 tone and the preceding L_3 tone of those forms in (8) ~ (11) should be fitted to a single positive slope less than *1* as graphically shown in Figure 4.22.

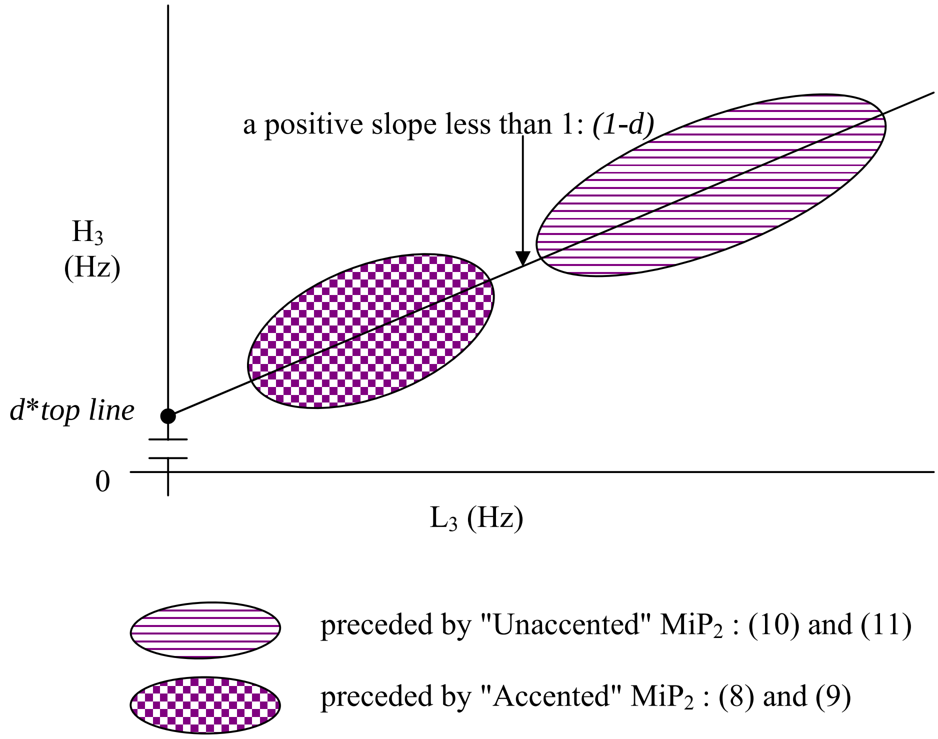


Figure 4.22. The Predicted Relationship between L_3 and H_3

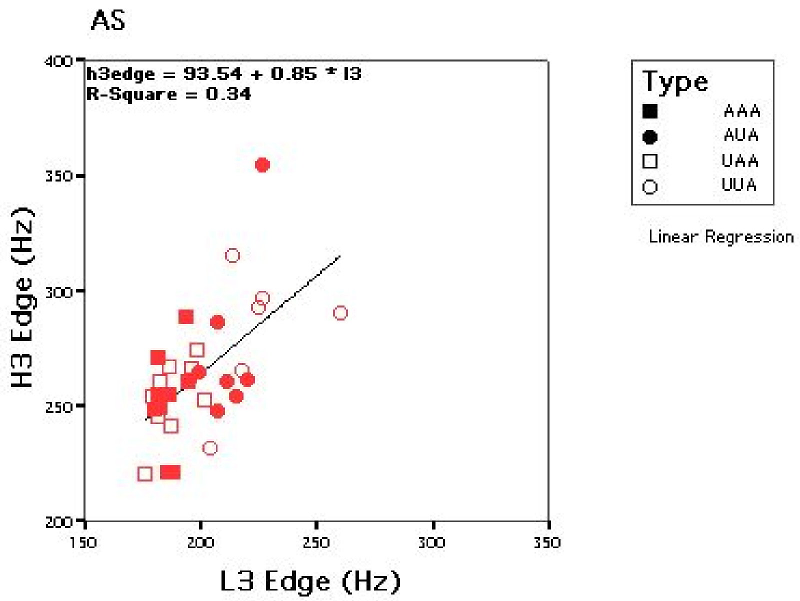
4.3.2.3.2. Results

Results of the regression analysis between the preceding L_3 edge tone (i.e. the predictor) and the following H_3 edge tone (i.e. the dependent variable) are presented in Table 4.08 and Figure 4.23.

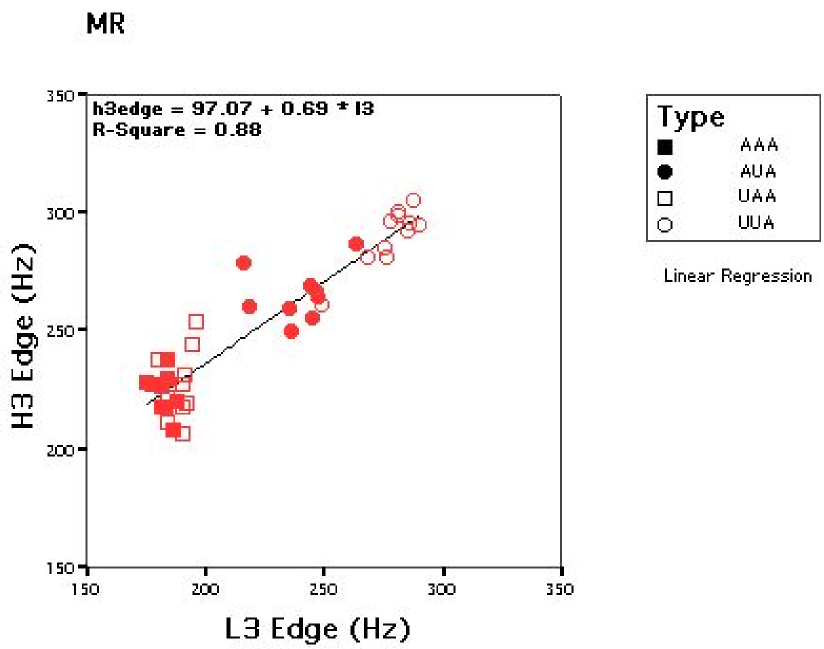
Table 4.08. Regression Coefficients

	Intercept (*d*top line*)	Standard Error	Slope (*1-d*)	Standard Error	R^2	n
AS	94*	43.15	0.85!	0.21	0.35	32
MR	97!	9.50	0.69!	0.04	0.88	41
RO	125!	9.40	0.65!	0.05	0.84	41
SK	111!	9.80	0.73!	0.04	0.91	38

! $p < 0.01$
* $p < 0.05$



a. Speaker AS

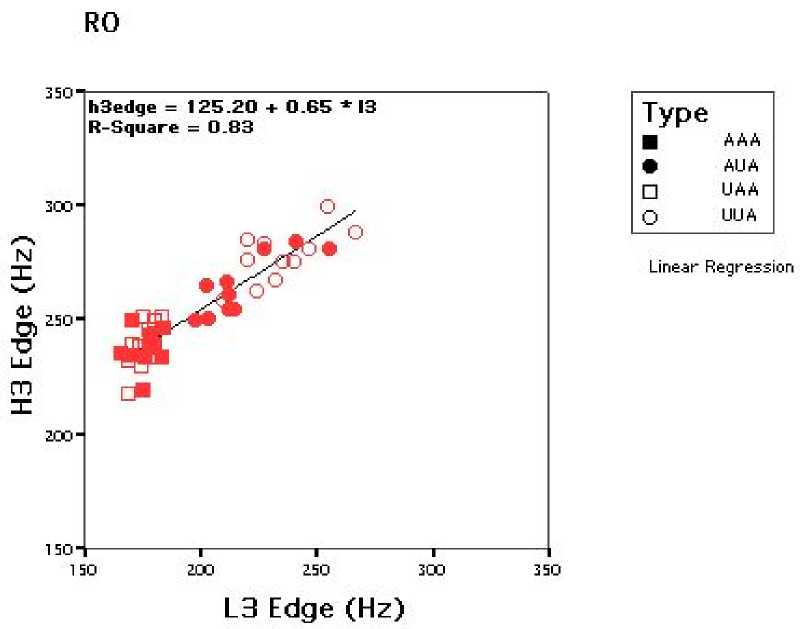


c. Speaker MR

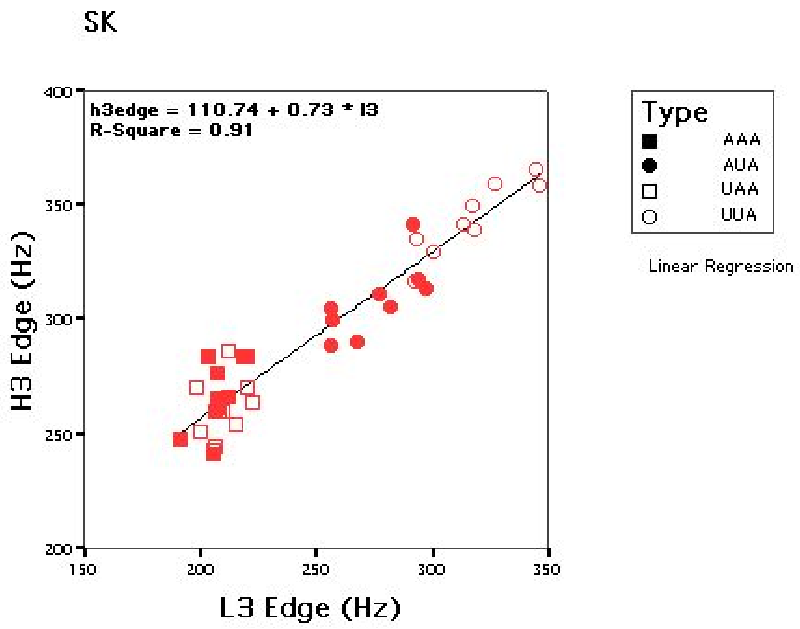
Figure 4.23. The Relationship between L3 and H3

Continued next page

Figure 4.23 continued



c. Speaker RO



d. Speaker SK

Three of our four speakers (except for AS) have a tight correlation between the H_3 edge tone and the preceding L_3 edge tone. The coefficient of determination *(R^2)* of the correlation is 0.88 for MR, 0.84 for RO, and 0.91 for SK. Also, the intercept values and the slopes of those three speakers are similar to each other: their intercept values vary from 97 to 125 (Standard Error: 9.40~9.80), and their slopes vary from 0.65 to 0.73 (Standard Error: 0.04 ~ 0.05). Those results confirm that the F0 value of the H edge tone is a function of that of the preceding L edge tone.

At the same, we could estimate the F0 value of the *top line* from those results according to our H edge tone scaling model. In our model, the H_3~L_3 relationship is a function of the *top line* values, and rewritten as (*d*top line*). The coefficient *d* is obtained by subtracting the slope value from *1*, because the slope associated with the H_3~L_3 relationship is rewritten as (*1-d*). Therefore the F0 value of the *top line* can be estimated by dividing the intercept value by (*1-slope*).

(18) $H_3 = \underbrace{\text{d*top line}}_{\text{intercept}} + \underbrace{\text{(1- d)}}_{\text{slope}}\text{*}L_3$ $\quad 0 < d = (1\text{-slope}) < 1$

(19) top line = intercept / (1 - slope)

If our H edge tone scaling model is adequate, then the F0 values of the *top line* estimated from the intercept values and slope values obtained here should match the reality that the highest speech-related pitch of those female speakers' voice is between 300 Hz and 400 Hz. The estimated *top line* values are presented in Table 4.09.

Table 4.09. The Estimated Top Line Values

	Top Line (estimate)
AS	627 Hz (implausible)
MR	313 Hz
RO	357 Hz
SK	411 Hz

The estimated *top line* value of each of our three speakers (except for AS) is plausible because it ranges from 313 Hz and 411 Hz, which is approximately the actually observed highest F0 value of their speech-related voice range. This confirms the adequacy of our H edge tone scaling model that the F0 value of the H edge tone is a function of both the *top line* and the preceding L edge tone.

The only problematic case is AS's. Her estimated *top line* value is 627 Hz, which is an implausibly high F0 value for a human speech-related voice. It is unlikely that speakers refer to such a high F0 value when computing the H edge tone values. Such an implausible estimated value may have been derived because of high Standard Errors of her intercept and slope. Her intercept value is 94 Hz while its SE is 43 Hz, almost half of the intercept value. Also, her slope is 0.85 while its SE is 0.21, one fourth of the slope and almost five times as large as the SE of other speakers' slopes. That is, her intercept values and her slope are unreliable due to those high Standard Errors, and it is not surprising that the *top line* value estimated from such unreliable values is implausible.

In summary, the results obtained from the regression analysis of the H_3~L_3 relationship revealed the adequacy of our tone-by-tone scaling model. The F0 value of the H edge tone is determined by fractions of the F0 of top line and that of the preceding

L edge tone. In the following section, another H tone scaling, i.e. the scaling of the H* accent tone is examined.

4.3.2.4. The H* Accent Tone Scaling and Regression Analyses

In this section, the tone-by-tone scaling of the H* accent tone, i.e. the relationship between the F0 of the preceding H edge tone and the H* accent tone, is examined. Again, data obtained from the <*Maronkéeki* Set> were used for our regression analysis. This time, the target F0 values were those of the preceding H_3 edge tone and those of the following H_3* accent tone associated with the third word (MiP_3). The F0 values of the following H_3* accent tone was made into a dependent variable and those of the preceding H_3 edge tone was made into a predictor. Results of the regression analysis are presented in Table 4.10 and Figure 4.24.

The <*Maronkéeki* Set>

(8) **UAA**

[*nijukko-iri-no*]$_{\text{MiP1}}$ [*yuuháimu-no*]$_{\text{MiP2}}$ **[*maronkéeki*]$_{\text{MiP3}}$**

H_2* +L_2 **H_3 H_3***

twenty-pieces- -Copula Juheim-Gen chestnut cake

"twenty pieces of chestnut cake of Juheim in a single box"

(9) **AAA**

[*nihyakúen-no*]$_{\text{MiP1}}$ [*yuuháimu-no*]$_{\text{MiP2}}$ **[*maronkéeki*]$_{\text{MiP3}}$**

H_2* +L_2 **H_3 H_3***

two hundred yen-Copula Juheim-Gen chestnut cake

"chestnut cake of Imuraya that are two hundred yen"

(10) **UUA**

[*nijukko-iri-no*]$_{\text{MiP1}}$ [*imuraraya-no*]$_{\text{MiP2}}$ **[*maronkéeki*]$_{\text{MiP3}}$**

H_2 **H_3 H_3***

twenty-pieces- -Copula Imuraya-Gen chestnut cake

"twenty pieces of chestnut cake of Imuraya in a single box"

(11) **AUA**

[*nihyakúen-no*]$_{\text{MiP1}}$ [*imuraya-no*]$_{\text{MiP2}}$ **[*maronkéeki*]$_{\text{MiP3}}$**

H_2 **H_3 H_3***

two hundred yen-Copula Imuraya-Gen chestnut cake

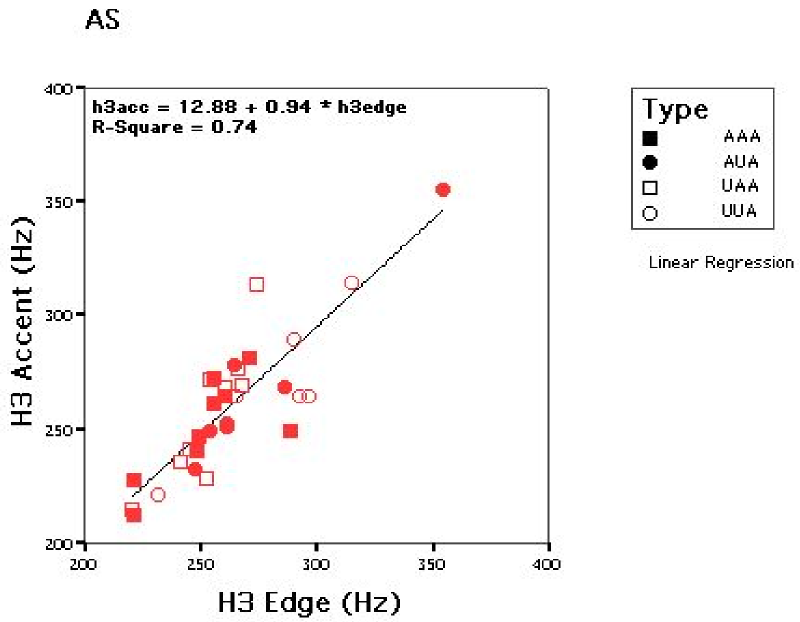
"chestnut cake of Imuraya that are two hundred yen"

Table 4.10. Regression Coefficients

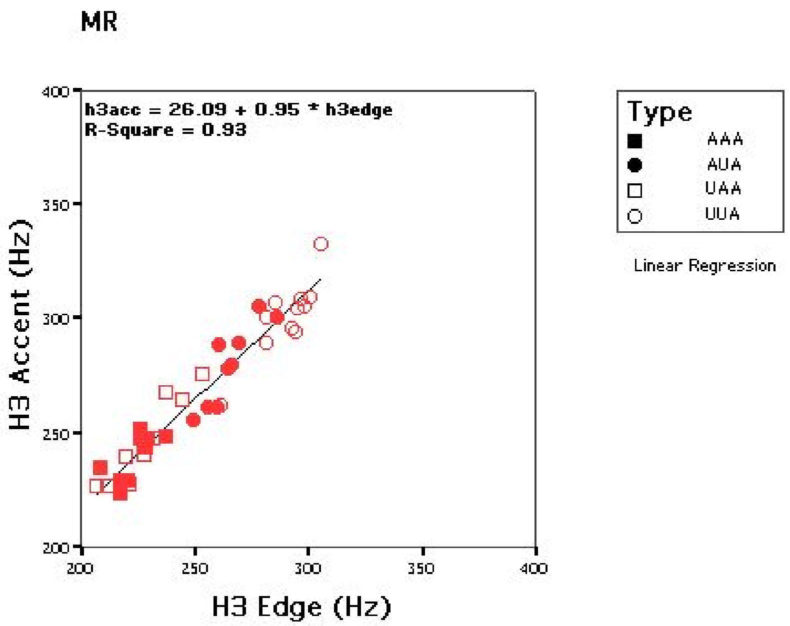
	Intercept (*d**top line*)	Standard Error	Slope (*1-d*)	Standard Error	R^2	n
AS	13	27.20	0.94!	0.10	0.74	32
MR	26*	10.56	0.95!	0.04	0.93	41
RO	161!	18.98	0.59!	0.07	0.62	41
SK	120!	14.01	0.67!	0.05	0.85	38

! $p < 0.01$

* $p < 0.05$



a. Speaker AS

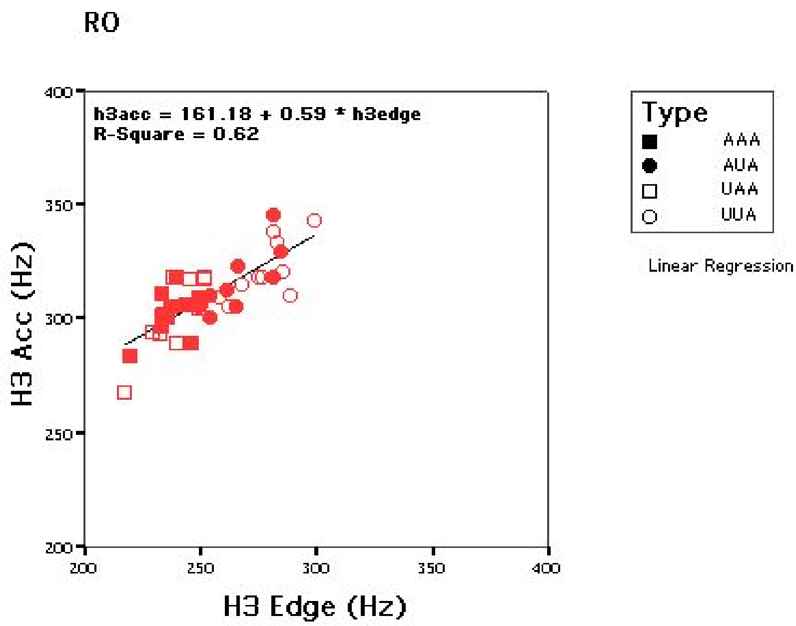


b. Speaker MR

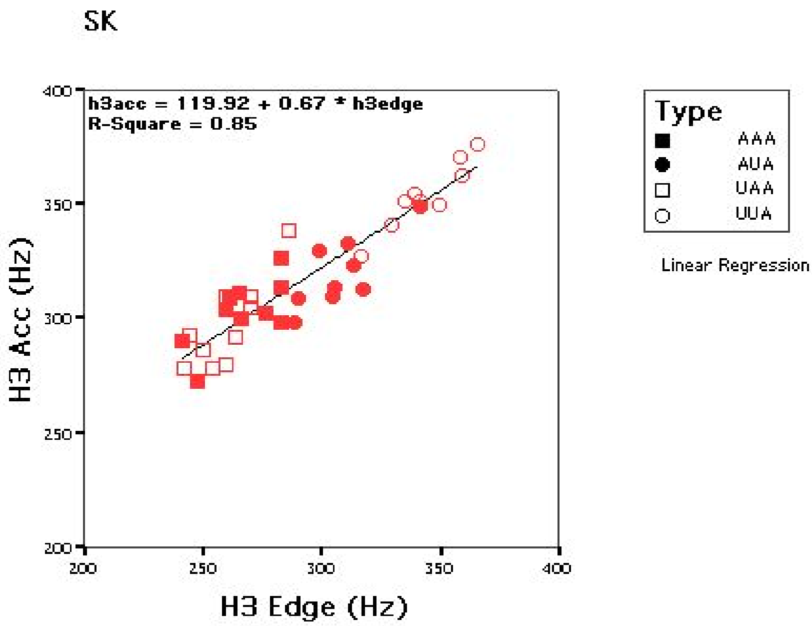
Figure 4.24. The Relationship between H3 and H3*

Continued next page

Figure 4.24 continued



c. Speaker RO



d. Speaker SK

There was good correlation between the H_3 accent tone and the H_3 edge tone across all the four speakers (i.e. R^2 varied from 0.62 to 0.93). However, those speakers were divided into two groups: a group with a small intercept and a large slope (AS and MR); a group with a large intercept and a smaller slope (RO and SK). The intercept values of AS and MR were 13 and 26 respectively, and the AS intercept was even not significantly different from zero. Their slopes were 0.94 and 0.95, which are approximately *1*. On the other hand, the intercept values of RO and SK were 161 and 120 respectively, and their slopes were 0.59 and 0.67.

Such a split between two groups may indicate that those two groups of speakers adopted different phonetic rules for scaling of the H* accent tone. The intercept and the slope values of RO and SK seem to comply with our H tone scaling model because those values yield plausible estimates of the *top line* values as shown in Table 4.11.

Table 4.11. The Estimated Top Line Values

	Top Line (estimate)
RO	161/(1-0.59) = 393 Hz
SK	120/(1-0.67) = 363 Hz

Though those estimated *top line* values in Table 4.11 were either 40~50 Hz greater or smaller than the estimate of the *top line* values obtained from the H_3 edge tone scaling in Table 3.09, they are still plausible because the highest F0 of speech-related voice of those speakers ranges between 300 and 400 Hz. The 40~50 Hz deviation from the estimated *top line* values obtained from the previous H_3 edge tone scaling indicates that some additional factors are at work either in the H_3 edge tone scaling or the H_3* accent tone scaling. That is, some additional constant should be added to either the equation of the H edge tone scaling or to that of the H* accent tone scaling to make both equations to share the same *top line* values.

The problematic cases are the outcomes of AS and MR. As mentioned above, the intercept value of AS was not statistically significant from zero. From this, I propose that AS's H* accent tone scaling should be expressed by the following formula.

(20) AS's H* accent tone scaling:

$$H_{Acc} = d * H_{edge} \qquad 0.9 < d < 1$$

According to the formula in (20), the F0 value of the H* accent is computed only based on the absolute height of the preceding H edge tone without referring to the height of the *top line*. Also, we expect from the formula in (20) that the F0 value of the H* accent tone should not be greater than that of the preceding H edge tone. This expectation is on the right track. The mean F0 value of her H_3 edge tone of the <*Maronkéeki* Set> is 264 Hz, while that of the H_3* accent tone is 261 Hz. In addition, the formula in (20) complies with our finding that AS's F0 value of the H* accent tone may be even 13~25 Hz lower than that of the preceding H edge tone (see Figure 3.10 of Section 4.2, data obtained from the <*Omiaiáite* Set>).

I propose that MR's H* accent tone value should be derived from something similar to AS's equation in (20) without referring to the *top line* F0 value, and it should be expressed as in (21).

(21) MR's H* accent tone scaling:

$H_{Acc} = c + d * H_{edge}$ $\quad\quad$ $0.9 < d < 1, \quad 15 < c < 35$

4.3.2.5. The Summary of H Tone Scaling

In summary, all the results obtained in 4.3.2.3 and 4.3.2.4 supported our proposal that the F0 value of a H tone is a function of the F0 value of the preceding tone. In addition, results obtained from two of our speakers (RO and SK) consistently supported our additional claim that the phonetic rules of H tone scaling should refer to the F0 value of the *top line* of a pitch range as well as that of the preceding tone.

Our tone-by-tone scaling model of H tones not only adequately captures the local relation between the preceding tone and the following H tone but also makes a correct prediction for the *catathesis* diminution because only a fraction of the F0 of the preceding tone is reflected on the F0 of the following H tone according to our model.

4.3.3. Further Issues to be Investigated

In 4.3.1. and 4.3.2. of Section 4.3, we considered the relationship between neighboring tones within the same Major Phrase, i.e. in the domain of *catathesis*. *Catathesis*, however, is cancelled at the left edge of a new Major Phrase according to Pierrehumbert & Beckman 1988). They observed that the F0 value of a H* accent tone associated with a Major Phrase-initial Minor Phrase was unaffected by the presence or absence of a preceding pitch accent. Our tone-by-tone scaling model should be able to handle this cancellation of *catathesis*.

However, this issue is out of the scope of this chapter because there are still quite a few things to be investigated before asking how our tone-by-tone scaling model should treat scaling of Major Phrase-initial tones. For example, we still do not know whether the Major Phrase-initial L edge tones are consistently realized lower or higher than that of the Major Phrase-medial L edge tones. Also, it is not yet known whether the scaling of those Major Phrase-initial L and H edge tones are also affected by the presence or absence of the preceding pitch accent as Major Phrase-medial L and H edge tones are. I leave for future investigation those questions and the issue of how our tone-by-tone scaling model handles Major Phrase initial tones.

4.4. Further Comparisons

In Section 4.2 and Section 4.3, I revealed empirical problems associated with the tonal space lowering model of *catathesis* proposed by Pierrehumbert & Beckman (1988), and argued that the Tokyo Japanese *catathesis* should be accounted for by an alternative and more local tone-by-tone scaling model. In this section, I present more evidence to support this alternative model.

4.4.1. Prediction 1. More Adjacency Effects on *Catathesis*

The global pitch range lowering hypothesis and the local tone-by-tone scaling hypothesis make different predictions for F0 scaling of tones associated with two of the forms in the <*Maronkéeki* Set>, again shown below.

(8) **UAA**

[*nijukko-iri-no*]$_{\text{MiP1}}$ [*yuuháimu-no*]$_{\text{MiP2}}$ **[*maronkéeki*]$_{\text{MiP3}}$**

(L_1 H_1)$_{\text{MiP1}}$ (L_2 H_2 $\mathbf{H_2{*}{+}L_2}$)$_{\text{MiP2}}$ (L_3 H_3 $H_3{*}{+}L_3$)$_{\text{MiP3}}$

twenty-pieces- -Copula Juheim-Gen chestnut cake

"twenty pieces of chestnut cake of Juheim in a single box"

(11) **AUA**

[*nihyakúen-no*]$_{\text{MiP1}}$ [*imuraya-no*]$_{\text{MiP2}}$ **[*maronkéeki*]$_{\text{MiP3}}$**

(L_1 H_1 $\mathbf{H_1{*}{+}L_1}$)$_{\text{MiP1}}$ (L_2 H_2)$_{\text{MiP2}}$ (L_3 H_3 $H_3{*}{+}L_3$)$_{\text{MiP3}}$

two hundred yen-Copula Imuraya-Gen chestnut cake

"chestnut cake of Imuraya that are two hundred yen"

The form UAA in (8) consists of an unaccented MiP_1 and accented MiP_2 and MiP_3. The form AUA in (11) consists of an initial accented MiP_1, and unaccented MiP_2 and accented MiP_3. In both of those forms, the third Minor Phrase (MiP_3) is preceded by exactly one accented Minor Phrase. However, the position of the preceding accented Minor Phrase varies in those forms. The preceding accent is associated with the second Minor Phrase (MiP_2) in the UAA form while that is associated with the initial Minor Phrase (MiP_1) in the AUA form.

UAA: (L_1 H_1)$_{\text{MiP1}}$ (L_2 H_2 $\mathbf{H_2{*}{+}L_2}$)$_{\text{MiP2}}$ **(L_3 H_3 $H_3{*}{+}L_3$)$_{\text{MiP3}}$**

preceding Acc ↑

AUA: (L_1 H_1 $\mathbf{H_1{*}{+}L_1}$)$_{\text{MiP1}}$ (L_2 H_2)$_{\text{MiP2}}$ **(L_3 H_3 $H_3{*}{+}L_3$)$_{\text{MiP3}}$**

preceding Acc ↑

Given such configurations, those two different *catathesis* models (i.e. the global pitch range lowering model and the local tone-by-tone scaling model) make different predictions for the F0 values of the post-accent tones associated with the third Minor Phrase (MiP_3).

4.4.1.1. Prediction by the Global Pitch Range Lowering Hypotheses

According to the more global pitch range lowering hypothesis, pitch range lowering equally takes place once before MiP_3 in both of those two forms though where such a lowering occurs varies. It takes place soon after H_2* of MiP_2 appears in UAA and soon after H_1* of MiP_1 appears in AUA. What is crucial to them, however, is not the position of pitch range lowering but presence of a pitch range lowering in both of those two forms.

Given the presence of pitch range lowering in both of those two forms, the global model predicts that the tones associated with MiP_3 of the UAA form and those of the AUA form to be realized in an equally lowered pitch range. In addition, since the relative values of tones are the same in both UAA and AUA, this model predicts that the F0 of tones associated with MiP_3 of the UAA form and that of the AUA form should be equal.

Pitch Range Lowering

UAA: $(L_1\ H_1\quad)_{MiP1}\ (L_2\ H_2\ \ H_2{*}{+}L_2)_{MiP2}\quad (\boxed{L_3\ H_3\ \ H_3{*}{+}L_3})_{MiP3}$

Same Relative Values
Equally Low Pitch Range

Pitch Range Lowering

AUA: $(L_1\ H_1\ \ H_1{*}{+}L_1)_{MiP1}\ (L_2\ H_2\quad)_{MiP2}\quad (\boxed{L_3\ H_3\ \ H_3{*}{+}L_3})_{MiP3}$

<Prediction by the Pitch Range Lowering Hypothesis>

T_3 of UAA = T_3 of AUA

4.4.1.2. Prediction by the Local Tone-by-Tone Scaling Hypothesis

On the other hand, the local tone-by-tone scaling hypothesis predicts that tones associated with MiP_3 of the UAA case and those associated with MiP_3 of the AUA case are not necessarily the same. More concretely, this alternative hypothesis predicts that the former are realized lower than the latter. In the following part of this section, let us consider why such prediction is made.

In the UAA case, MiP_3 is directly preceded by the $+L_2$ trailing tone which bears an extra low F0 because of the phonetic rule of +L tone scaling (see Section 4.3.1 for more detailed discussion on the +L tone scaling). Because of tone-by-tone scaling, this low F0 propagates to the L_3 edge tone of MiP_3. In the same way, the following H_3 edge tone and H_3 pitch accent are also influenced by the lowness of the immediately preceding L_2 edge tone. When it comes to the AUA case, MiP_3 is not directly preceded by a +L trailing tone. Rather it is immediately preceded by a H_2 edge tone of MiP_2. Since the H_2 edge tone is realized higher than a +L trailing tone, those tones that follow the H_2 edge tone are realized higher than the tones that follow the +L trailing tone. Therefore, this approach predicts that tones associated with MiP_3 of the UAA case are realized lower than those of the AUA case.

UAA: (L_1 H_1)$_{MiP1}$ (L_2 H_2 H_2*+**L_2**)$_{MiP2}$ (L_3 H_3 H_3*+L_3)$_{MiP3}$

"extra" low

directly preceded by an "extra" low +L.

AUA: (L_1 H_1 H_1*+L_1)$_{MiP1}$ (L_2 **H_2**)$_{MiP2}$ (L_3 H_3 H_3*+L_3)$_{MiP3}$

"extra" low

directly preceded by a H tone but not +L.

<Prediction by the Tone-by-Tone Scaling Hypothesis>

T_3 of UAA < T_3 of AUA

4.4.1.3. Testing Prediction 1

In this additional test, F0 of tones associated with MiP_3 of the following four forms from Dataset <*maronkéeki*> are compared.[19]

(8) **UAA**

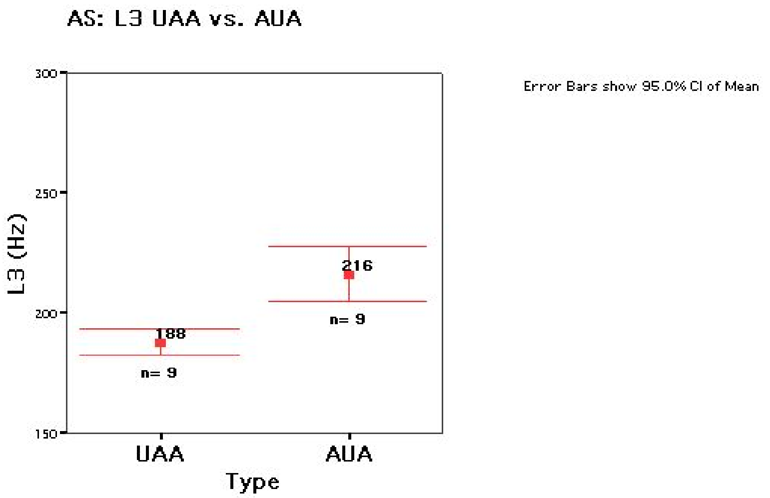
[*nijukko iri* *-no*]$_{MiP1}$	[*yuu hái mu-no*]$_{MiP2}$	[*ma ron kée ki-o*]$_{MiP3}$
L_1H_1	$L_2H_2H_2$*+L_2	L_3H_3 H_3*+L_3
twenty pieces-Gen	Juheim-Gen	hazel nut cake-Acc

(11) **AUA**

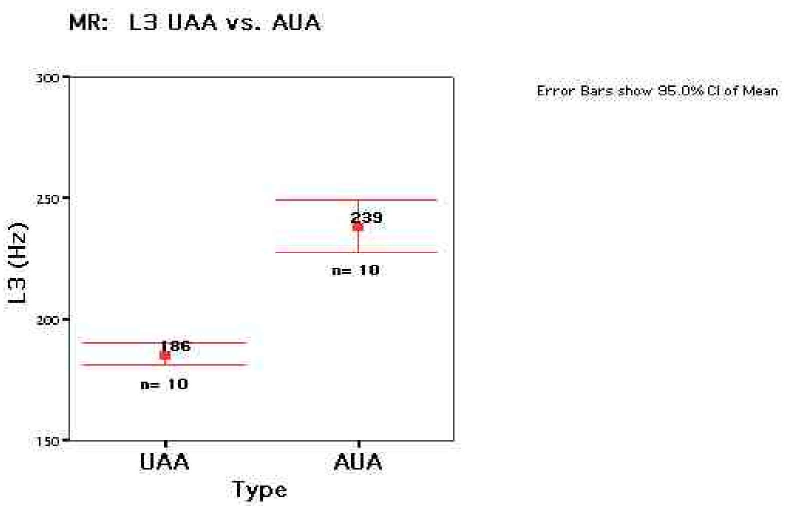
[*nihyakúen* *-no*]$_{MiP1}$	[*i mu ra ya-no*]$_{MiP2}$	[*ma ron kée ki-o*]$_{MiP3}$
L_1H_1 H_1*+L_1	L_2H_2	L_3H_3 H_3*+L_3
two hundred yen-Gen	Imuraya-Gen	hazel nut cake-Acc

The F0 of L_3 and the F0 peak of MiP_3 of those four forms (i.e. the F0 of H_3*) were measured. Mean F0 of L_3 and that of H_3* of those four forms obtained from all four speakers are shown in Figure 4.25 and Figure 4.26.

[19] See the appendix of this chapter for more information about Dataset <*maronkéeki*>.



a. Speaker AS

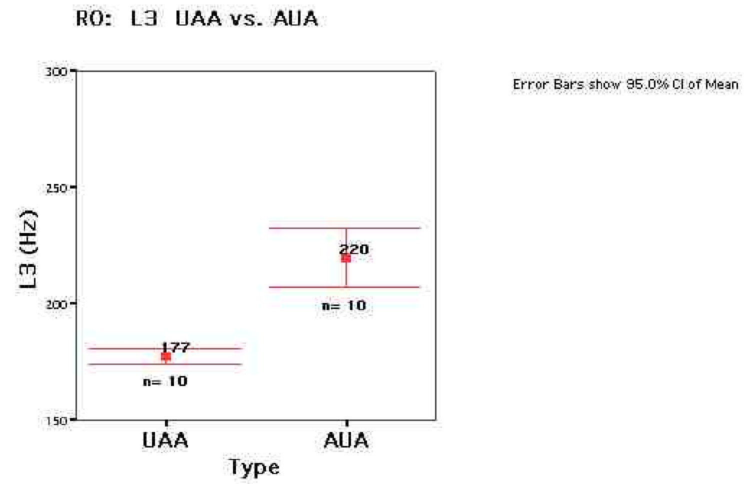


b. Speaker MR

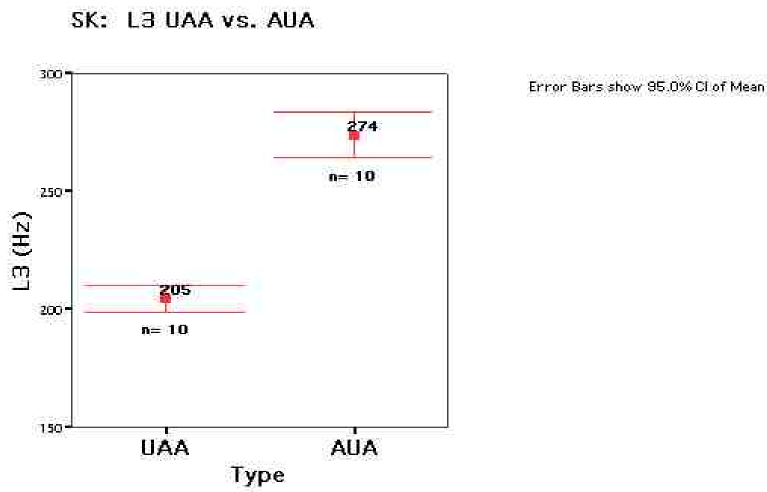
Figure 4.25. The Mean L3 Values of the UAA and the AUA Cases

Continued next page

Figure 4.25 continued



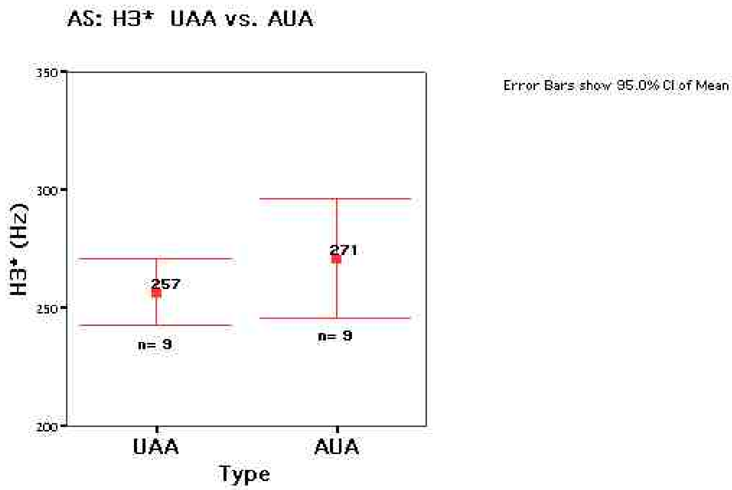
c. Speaker RO



d. Speaker SK

The mean F0 value of L_3 of the UAA case turned out to be consistently much lower than that of the AUA case in all four speakers' speech, and because their 95% confidence intervals did not overlap, the difference between the UAA and the AUA cases was significant.

There was also a parallel difference between the mean F0 value of H_3* of the UAA case and that of the AUA case: the former was lower than the latter in all four speakers' speech.

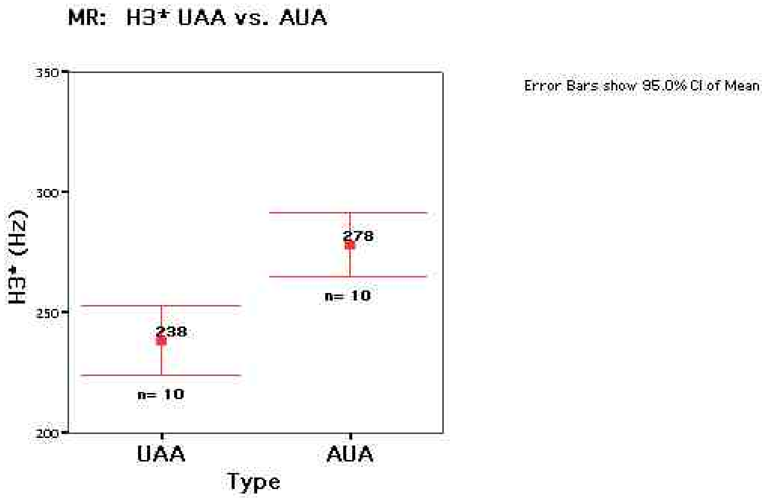


a. Speaker AS

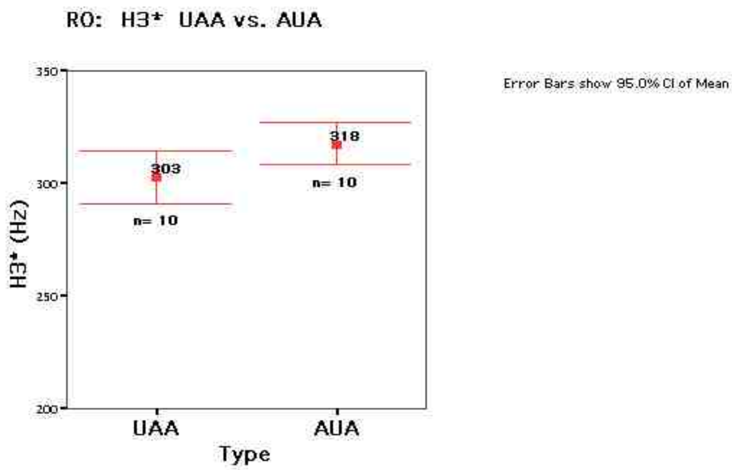
Figure 4.26. The Mean H3* Values of the UAA and the AUA Cases

Continued next page

Figure 4.26 continued



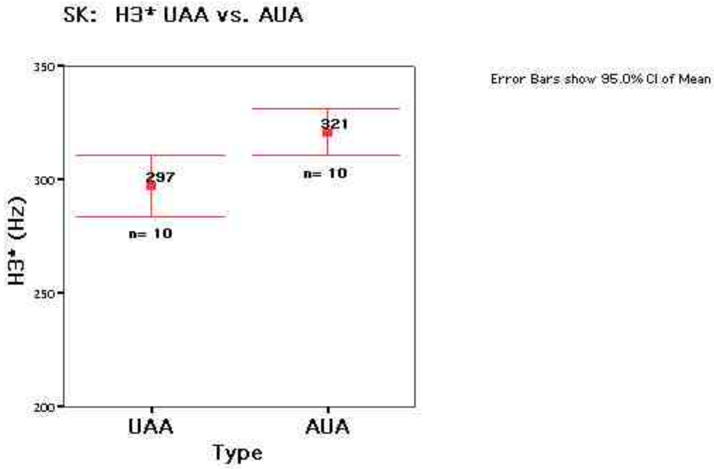
b. Speaker MR



c. Speaker RO

Continued next page

Figure 4.26 continued



d. Speaker SK

As for MR and SK, the 95% confidence intervals of the former and that of the latter did not overlap, and the difference between them was interpreted as significant. When it comes to RO, though there was a slight overlap between their confidence intervals, the mean difference between those two cases was significant according to an ANOVA test ($F(1,18)=4.80$, $^*p < 0.05$). It was only AS who did not have any significant difference between the mean F0 of H_3* of the UAA case and that of the AUA case ($F(1,16)=1.32$, $p = 0.268$). Though the difference between the F0 of H_3* of those two forms was not significant in AS's speech, she still follows the tendency that H_3* of the UAA form was lower than that of the AUA form.

In summary, we found that tones associated with MiP_3 of the UAA case had lower F0 value than those associated with MiP_3 of the AUA case. Those results comply with the prediction made by the local tone-by-tone scaling hypothesis that I proposed in Section 4.3.

<Results> T_3 of UAA < T_3 of AUA

→ Prediction made by the tone-by-tone model was supported.

4.4.2. Prediction 2. Tonal Value Effects on *Catathesis*

Given the test results presented in 4.4.1 and 4.4.2, the more local tone-by-tone scaling hypothesis is already more promising than the more global pitch range lowering hypothesis. To further support this hypothesis, another test was carried out.

Those two models, make different predictions for the *catathesis* of tones associated with the second Minor Phrase (MiP_2) of the following sequences of two accented Minor Phrases shown in (22) and (23) from the dataset <*Yonjúuen*>.

(22) [*san* *mán*]$_{MiP1}$ [*yon* *júu* *en* *-no*]$_{MiP2}$ …
L1 H1 H_1*+L_1 L_2 H_2 H_2*+L_2
thirty thousand forty yen-Gen
" … of thirty thousand and forty yen"

(23) [*sán* *bya* *ku*]$_{MiP1}$ [*yon* *júu* *en* *-no*]$_{MiP2}$ …
H_1* +L_1 L_2 H_2 H_2*+L_2
three hundred forty yen-Gen
" … of three hundred and forty yen"

In (22), the pitch accent of the first Minor Phrase (MiP_1) is associated with the final syllable (*mán*) of that phrase. Though both the H_1* accent tone and the +L_1 trailing tone are phonologically associated with the accented nucleus vowel of that syllable, i.e. [*á*], only the H_1* accent tone is phonetically aligned with it and the +L_1 trailing tone is phonetically aligned with the following sonorant coda segment [*n*]. However, only one segment is not enough for the +L tone to achieve its extra low F0 target because it usually takes more than two syllables (or morae) for a +L tone to achieve its extra low target. Given this, we expect that the target of the +L_1 trailing tone in (22) should undergo "target undershoot" and the tone will be realized higher than in (23).[20] The pitch accent of the MiP_3 in (23) is associated with the first syllable of that phrase, and there is enough space for the +L_1 tone to achieve its F0 target because two syllables intervene between the accented syllable and the end of that phrase. The target of +L_1 in (23) will not undergo undershoot.

(22) **+L_1 Undershoot**

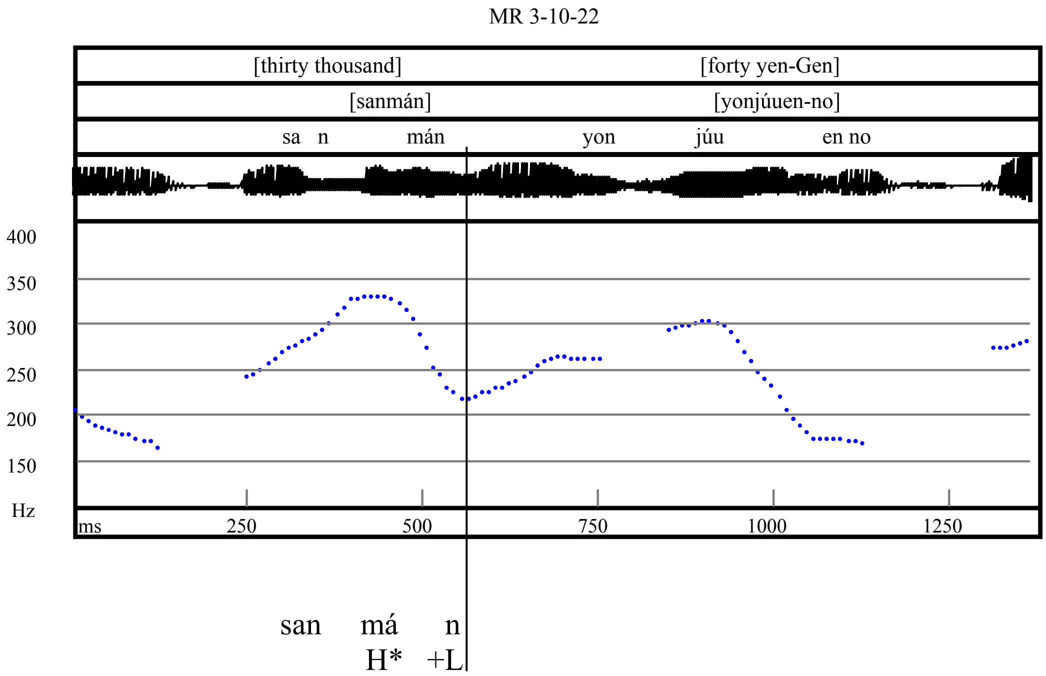
[*san* *mán*]$_{MiP1}$ [*yon* *júu* *en* *-no*]$_{MiP2}$ …
L1 H1 H_1*+L_1 L_2 H_2 H_2*+L_2
↑
Higher (undershoot)

(23) **No +L_1 Undershoot**

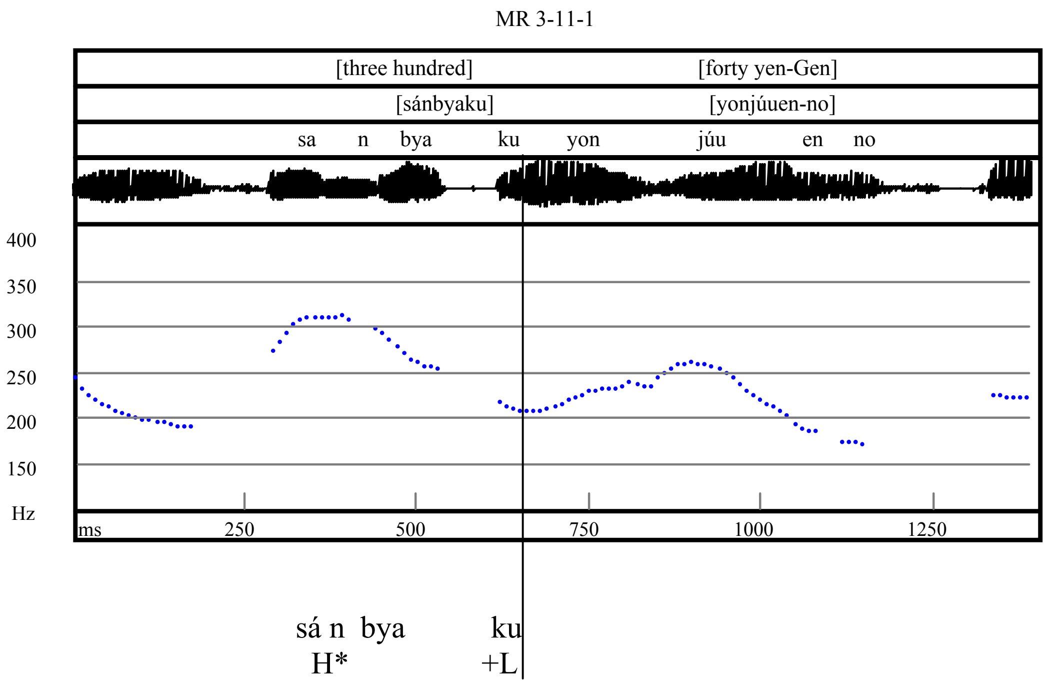
[*sán* *bya* *ku*]$_{MiP1}$ [*yon* *júu* *en* *-no*]$_{MiP2}$ …
H_1* +L_1 L_2 H_2 H_2*+L_2
↑
Lower

To confirm the presence of such target undershoot of the +L_1 trailing tone of the form in (22), I measured the F0 of the right edge of the final syllable of MiP_1 of both the form in (22) and the form in (23) because that was where the +L trailing tone was aligned. Some exemplar cases of +L tone alignment are presented below.

[20] For target undershoot, see Lindblom (1963, 1964), Fougeron (1998), Truckenbrodt (1998).



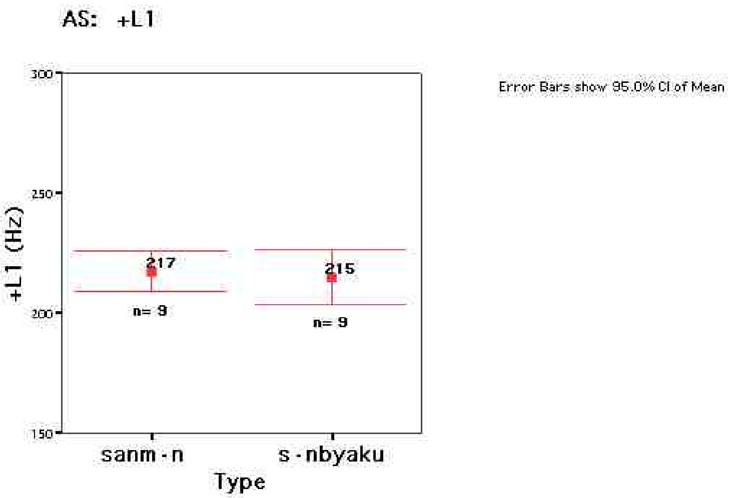
a. Speaker MR (+L Undershoot)



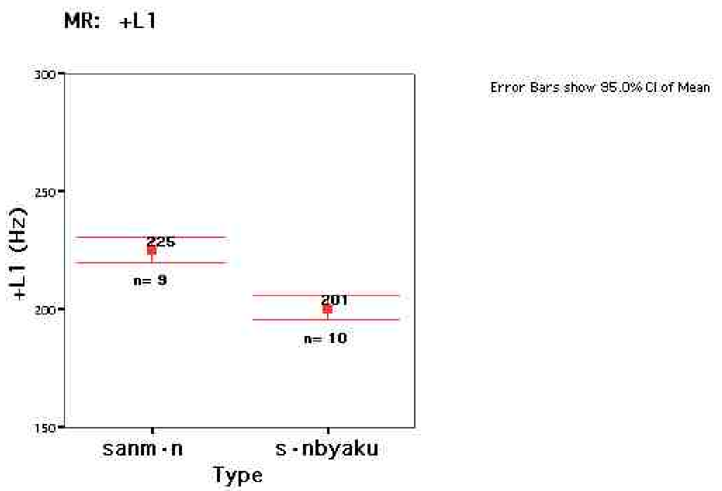
b. Speaker MR (No +L Undershoot)

Figure 4.27. Examples of +L Tone Alignment

If the +L_1 tone of the form in (22) is really undershot, then the MiP_1 final syllable in (22) should be scaled higher than that in (23). Figure 4.28 shows the comparison between the +L_1 tone values of those two forms.



a. Speaker AS

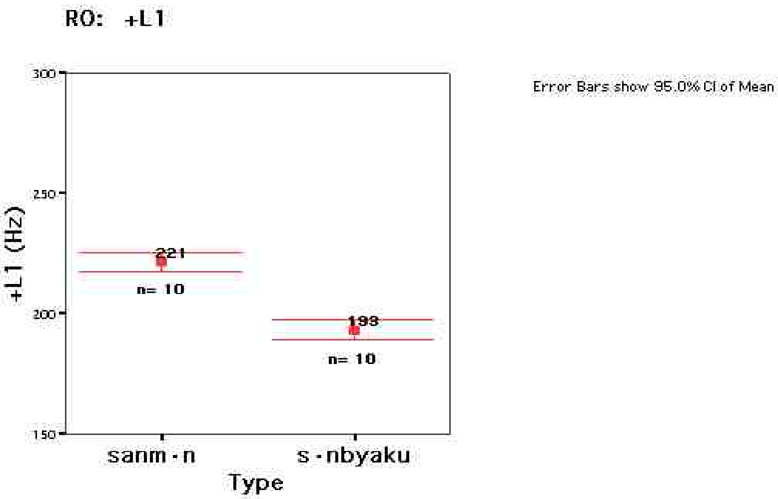


b. Speaker MR

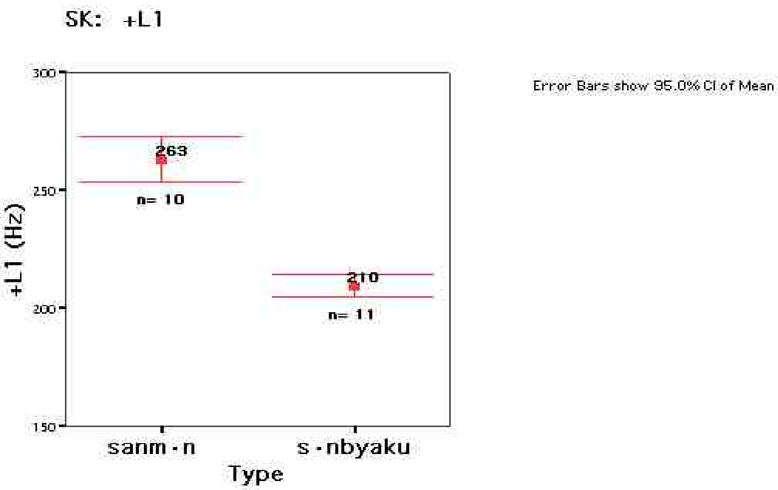
Figure 4.28. The Mean +L1 Values of Form (22) and Form (23)

Continued next page

Figure 4.28 continued



c. Speaker RO



d. Speaker SK

The comparison revealed that three of the fours speakers (MR, RO and SK) had target undershoot of the $+L_1$ trailing tone of the form in (22). That is, the MiP_1 final syllable of the form in (22) was realized higher than that of the form in (23) in those three speakers' speech, and the difference between those two forms was significant because the 95% confidence intervals of the mean F0 value of $+L_1$ of those two forms did not overlap (see Figure 4.28-b~d).

AS was the only exception (see Figure 4.28-a). There was almost no difference between the mean F0 values of $+L_1$ of those two forms: no target undershoot of $+L_1$ of the form in (22) was present. In the rest of the discussions and comparisons, AS's data are not considered because her data do not satisfy the premise that the $+L_1$ trailing tone of the form in (22) undergoes target undershoot.

Given the fact that the $+L_1$ of the form in (22) undergoes target undershoot while that of the form in (23) does not, the global pitch range lowering hypothesis and the local tone-by-tone scaling hypothesis make different predictions for the F0 value of tones associated with the second Minor Phrase (MiP_2) that follow the $+L_1$ trailing tone. In the following subsection, those different predictions are considered.

4.4.2.1. Prediction by the Pitch Range Lowering Hypothesis

According to the more global pitch range lowering hypotheses, pitch range lowering equally takes place in both (22) and (23) soon after H_1* of the first Minor Phrase (MiP_1) appears. That is, tones associated with the second Minor Phrase (MiP_2) of the form in (22) and those of the form in (23) should both be realized in an equally lowered pitch range regardless of whether the F0 target of the preceding $+L_1$ is undershoot or not.

(22) **Pitch Range Lowering**

[*san* *mán*]$_{MiP1}$ [*yon* *júu* *en* *-no*]$_{MiP2}$ …
L1 H1 H_1*+L_1 L_2 H_2 H_2*+L_2

Realized in Equally Lowered Pitch Range
& Sharing the Same Transformed Values

(23) **Pitch Range Lowering**

[*sán* *bya* *ku*]$_{MiP1}$ [*yon* *júu* *en* *-no*]$_{MiP2}$ …
H_1* +L_1 L_2 H_2 H_2*+L_2

<Prediction by the Pitch Range Lowering Hypothesis>

T_2 of the form in (22) = T_2 of the form in (23)

4.4.2.2. Prediction by the Tone-by-Tone Scaling Hypothesis

In contrast, F0 values of tones following a pitch accent are dependent on the F0 value of the immediately preceding +L trailing tone according to the "local" tone-by-tone scaling hypothesis. Therefore, this local approach predicts that tones associated with MiP_2 of the form in (22) should be realized higher than those of the form in (23).

According to this hypothesis, the F0 value of the L_2 edge tone of MiP_2 is computed relative to the F0 of the immediately preceding +L_1 trailing tone, and the undershooting +L_1 tone of the form in (22) raises the F0 value of the following L_2 tone. In turn, the higher F0 of the L_2 tone further propagates to the following H_2 and H_2* tone via tone-by-tone scaling.

(22) **+L_1 Undershoot**

[*san* *mán*]$_{MiP1}$ [*yon* *júu* *en* *-no*]$_{MiP2}$ …
L1 H1 H_1*+L_1 L_2 H_2 H_2*+L_2

Higher

preceded by a higher +L = higher F0 of T_2

(23) **No +L_1 Undershoot**

[*sán* *bya* *ku*]$_{MiP1}$ [*yon* *júu* *en* *-no*]$_{MiP2}$ …
H_1* +L_1 L_2 H_2 H_2*+L_2

Lower

preceded by a lower +L = lower F0 of T_2

<Prediction by the Pitch Range Lowering Hypothesis>

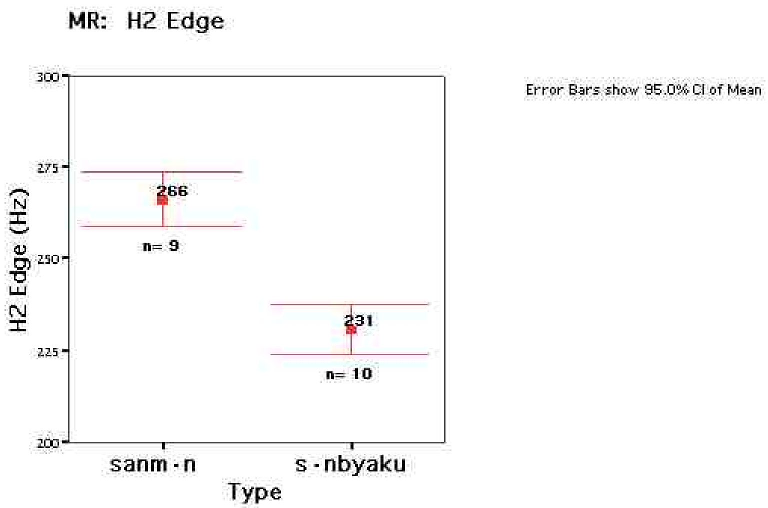
T_2 of the form in (22) $>$ T_2 of the form in (23)

Comparison between those two different predictions made by the pitch range lowering hypothesis and the tone-by-tone scaling hypothesis is presented in the following subsection.

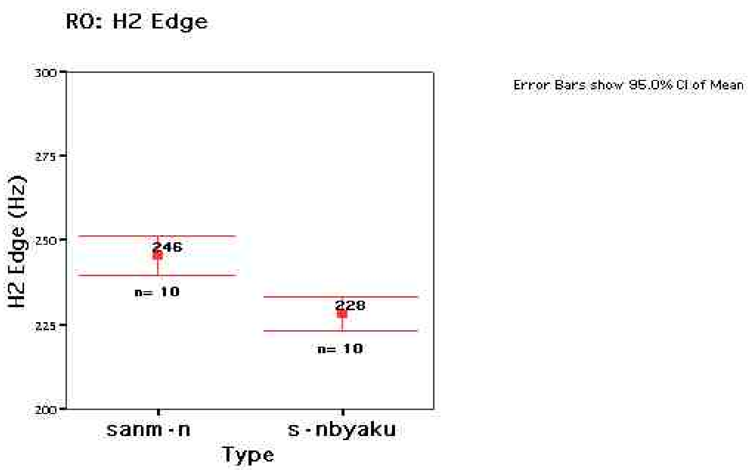
4.4.2.3. Testing Prediction 2

F0 of the H_2 edge tone and the H_2* accent tone of the form in (22) and that of those H tones of the form in (23) were compared. For this purpose, F0 values of the first syllable (more specifically the right edge of the nucleus vowel of the first syllable) and the second syllable of MiP_2 were measured.

Since the first syllable of MiP_2, i.e. *yon*, is heavy, and the right edge of the nucleus vowel of that syllable, i.e. *o*, was where a H_2 edge tone was aligned. The second syllable of MiP_2 (i.e. the accented syllable of MiP_2) is where a H_2* accent tone is aligned. Therefore, F0 of the first syllable nucleus and that of the second syllable stand for F0 of the H_2 edge tone and that of the H_2* accent tone respectively. The comparison between the mean F0 of H_2 edge tone of (22) and that of (23) is presented in Figure 4.29 and that between the H_2* accent tone of (22) and the accent tone of (23) are presented in Figure 4.30.



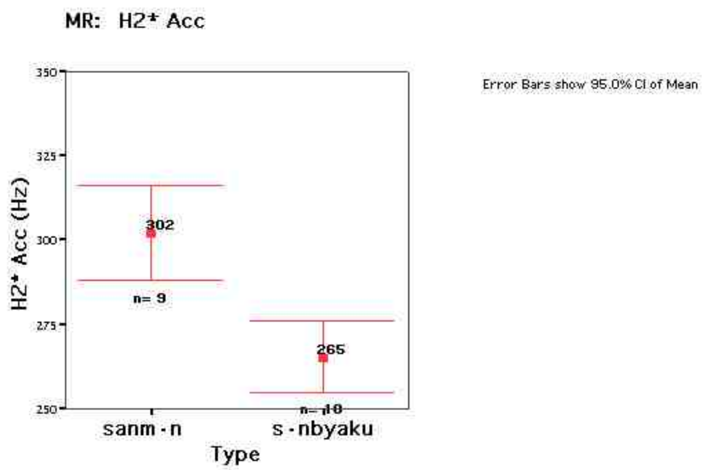
a. Speaker MR



b. Speaker RO

Figure 4.29. The Mean H2 Values of Form (22) and Form (23)

For all three speakers, the mean F0 value of the H_2 edge tone of the form in (23), i.e. the H_2 edge tone that is preceded by the non-undershot +L, was lower than that of the form in (22), i.e. the H_2 edge tone preceded by the undershot +L. At the same time, their 95% confidence intervals did not overlap. That is, the former was significantly lower than the latter. This result complies with the prediction made by the tone-by-tone scaling model. The same result was obtained from the comparison of the H_2* accent tones as shown in Figure 4.30.

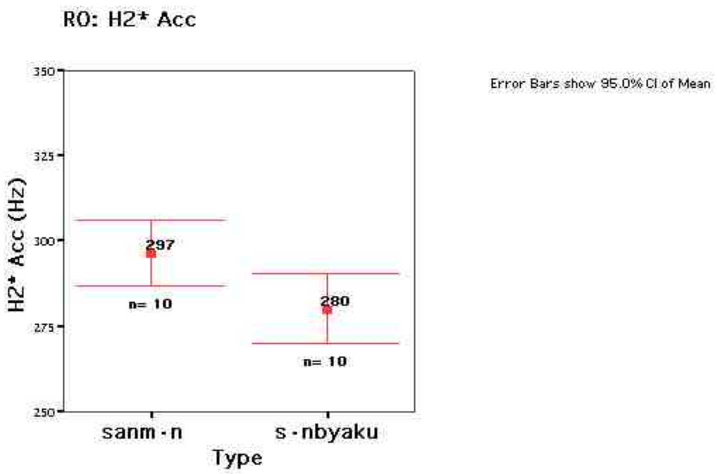


a. Speaker MR

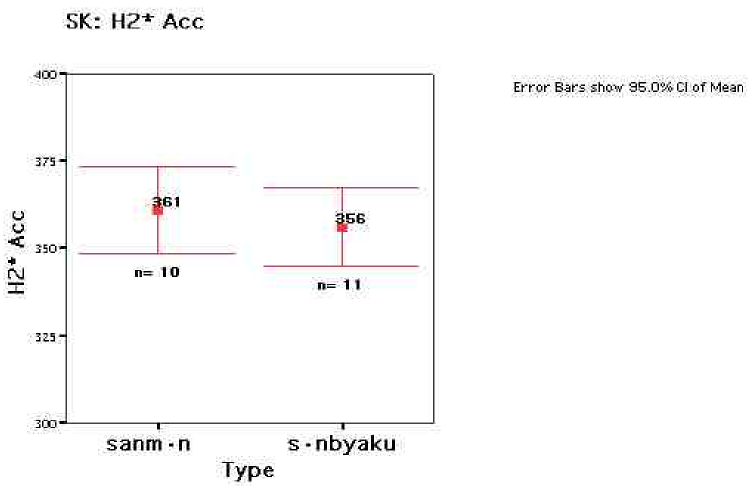
Figure 4.30. The Mean H2* Values of Form (22) and Form (23)

Continued next page

Figure 4.30 continued



b. Speaker RO



c. Speaker SK

For all three speakers, the mean F0 value of the H_2* accent tone of the form in (22), i.e. the accent tone preceded by the non-undershoot L, was lower than that of the accent tone of the form in (23), i.e. the accent H tone preceded by the undershoot L.

Two of the speakers (RO and SK) have the confidence intervals of those two means overlapping, while MR did not have such overlap. Nonetheless, the mean difference was statistically significant for RO ($F(1,18) = 6.72$, $*p < 0.05$). It was only SK whose mean difference was not significant ($F(1,19) = 0.43$, $p > 0.05$).

In summary, we observed that H tones following a higher undershot-L tone tended to be realized higher than those following a lower non-undershot-L tone, in spite of the fact that those H tones are equally realized in a post-accent region. Those results support our tone-by-tone scaling model, and they are not expected by the global pitch range lowering model of *catathesis.*

<Result> T_2 of the form in (22) > T_2 of the form in (23)

→ Supports the Tone-by-Tone Scaling Hypothesis

4.5. Chapter Conclusion

I argued that the Tokyo Japanese *catathesis* is a consequence of local tone-by-tone scaling. This tone-by-tone scaling hypothesis itself is not a new idea. Pierrehumbert (1980) proposed that English downstep should be captured by tone-by-tone scaling, which is usually cited as a local approach to intonation and contrasted to global approaches, i.e. approaches that reduces tonal scaling to global change in pitch range (Beckman 1995, Grønnum 1995, Ladd 1995, Möbius 1995). This local tone-by-tone scaling hypothesis, however, was never seriously considered in the context of the Tokyo Japanese *catathesis*. The major contribution of this chapter is to provide new support for this local approach to account for the Tokyo Japanese *catathesis*.

In contrast, Pierrehumbert & Beckman (1988) proposed that Tokyo Japanese *catathesis* should be captured as a global tonal space lowering. According to their approach, the magnitude of *catathesis* of a post-accent tone never diminishes no matter how many tones intervene between that post-accent tone and the preceding H*+L pitch accent. This is because tonal space lowering is a global operation equally affecting all the post-accent tones (at least those post-accent tones are all in the same *catathesis* domain, i.e. the same Major Phonological Phrase). Secondly, according to their pitch range lowering approach, the F0 of each tone is not directly affected by the F0 of the immediately preceding tones. This is because each tone is independently computed relative to the pitch range but not relative to the F0 of the neighboring tones. Those predictions, however, turned out to be wrong: the magnitude of *catathesis* diminished as more tones followed the preceding accent and there was a strong correlation between the F0 of the preceding L tone and the following H tone. Those results disapprove the pitch range-lowering model while supporting the tone-by-tone scaling model.

One issue related to *catathesis* which was not discussed in this chapter is the resetting phenomenon. According to Pierrehumbert & Beckman, the domain of *catathesis* is a Major Phrase (i.e. an Intermediate Phrase). That is, the propagation of *catathesis* is blocked at the left boundary of a new Major Phrase. The tonal space lowering treats this resetting phenomenon simply as shoving up of the ceiling of a tonal space at the left edge of a new Major Phrase boundary. On the other hand, the tone-by-tone scaling approach needs additional phonetic rules to assign higher F0 values to the tones that appear at the left edge of each Major Phrase. A further investigation is necessary to reveal what exactly this resetting phenomenon is like and what those additional phonetic rules should be.

PART II

POST-FOCUS INTONATION IN TOKYO JAPANESE:

THE STRUCTURAL VIEW VS. THE NON-STRUCTURAL VIEW

CHAPTER 5

INTRODUCTION TO PART II

Even when its truth conditions are constant, a sentence may carry more than one interpretation with respect to its information structure. For example, the sentence may consist of all new information, or one of the words in that sentence may be interpreted as a contrastive focus (henceforth FOCUS) and the rest of the items may be interpreted as background, presupposed or already given. Speakers disambiguate those different interpretations of a sentence by assigning different prosodic patterns corresponding to those different interpretations. For instance, when a contrastive FOCUS is present in a sentence, prosodic reduction of post-FOCUS items (i.e. items that come after the FOCUS) takes place in various languages. Examples of such post-FOCUS prosodic reduction are (a) absence of pitch accent in English (Ladd, 1980), Bengali (Hayes & Lahiri, 1991), Greek (Botinis, 1998), French (Jun & Fougeron, 2000), (b) deletion of phonological phrase boundaries in Korean (Cho, 1990; Jun 1993; Jun & Oh, 1996), in Hungarian (Vogel & Kenesei, 1990), and (c) downstepping and compression of F0 in Danish (Grønnum, 1989); Swedish (Garding, 1993), Chinese: (Garding, 1987; Selkirk & Shen, 1990; Xu, 1999; Shih, 2000), French (Di Cristo & Jankowski, 1999), Somali (Le Gac, 2002).

Tokyo Japanese is not an exception. In this language, too, compression and reduction of F0 movement take place in the post-FOCUS part of an utterance. Some examples are shown in Figure 5.01 and Figure 5.02. Both of those figures present F0 contours of the same sentence consisting of almost the same lexical items organized into the same syntactic structure. The sentence is shown in (1). The difference between those two contours, then, comes from the information structure of that sentence. Figure 5.01 shows the F0 contour of the sentence with new information only (i.e. *neutral* interpretation). Figure 5.02 also shows the F0 contour of the same sentence. However, the F0 contour of Figure 5.02 is associated with a different interpretation: the third word (Word3) is interpreted as a contrastive FOCUS and the rest of the words are interpreted as given.

There are two major differences between those two F0 contours. One is the F0 peak level of Word3 and the other is the F0 peak level of words following Word3. The F0 peak level of Word3 is downstepped (i.e. undergoes *catathesis*) in the neutral contour (Figure 5.01) while it is raised to a higher level in the FOCUS contour (Figure 5.02). Also, the F0 peak level of the following words (Word4 and later) in the neutral contour (Figure 5.01) is realized in a higher level without any downstepping while that of the post-FOCUS contour in Figure 5.02 is realized lower. Our main concern in this part of the dissertation is the latter contrast between those two contours, i.e. reduction/compression of F0 movement in the post-FOCUS part of an utterance.

(1)

Word1 Word2 Word3 Word4 Word5 Word6

[*Dókoka/Háyama*]-*de* [*gurabiamóderu*]-*no* [*orandájin/amerikajin*]-*ga* [*uwakiáite*]-*kara* [*mayak*]-*o* [*koonyuushita*]

[Somewhere/Hayama]-at [fashion model]-Gen [Dutch/American]-Nom [lover]-from [marijuan]-Acc [bought]

"A Dutch/ fashion model bought marijuana from her lover somewhere"

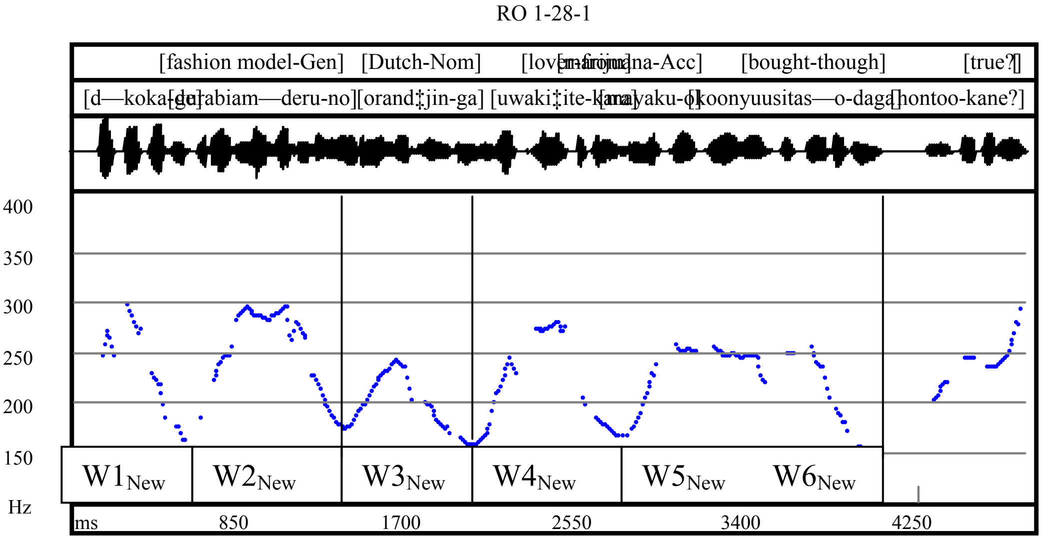


Figure 5.01. An F0 Contour Example of the Sentence with New Information Only (XP Boundaries at Even Numbered Words)

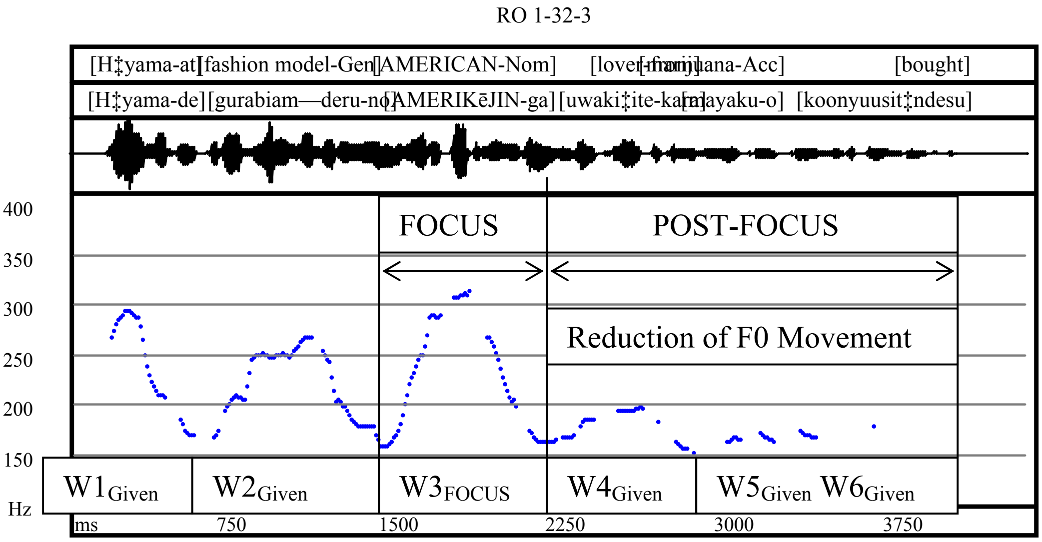


Figure 5.02. An F0 Contour Example of the Sentence with New Information Only (XP Boundaries at Odd Numbered Words)

There are two views with respect to the post-FOCUS lowering/compression of F0 movement in Tokyo Japanese: the "non-structural view" adopted by Pierrehumbert & Beckman (1988) and the "structural view" adopted by Nagahara (1994), Truckenbrodt (1995) and Uechi (1997). According to the non-structural view, the post-FOCUS reduction/compression of F0 movement is derived from pitch range lowering (or pitch range compression) without adding any change to the hierarchical organization of prosodic constituents of the post-FOCUS part of an utterance. In contrast, the structural view regards the post-FOCUS phenomenon as "deletion of phonological phrase boundaries" or "dephrasing" in the post-FOCUS part of an utterance.

Results obtained in my experiments are consistent with the structural view. That is, there is ample evidence for dephrasing in the post-FOCUS part of an utterance. However, non-structural effects are also present: some aspects of the post-FOCUS phenomena are unexplained by post-FOCUS dephrasing only and need to be accounted for by either pitch range compression or some other mechanisms.

The organization of this part of dissertation is the following. In Chapter 6, I first introduce theories behind those two views of the post-FOCUS phenomenon and different predictions made by those two views. In Chapter 7 and Chapter 8, I present results of experiments and analyses of those results.

CHAPTER 6

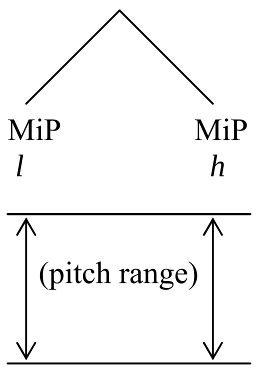
THEORIES BEHIND STRUCTURAL AND NON-STRUCTURAL VIEWS

In this chapter, I introduce theories behind the two views of post-FOCUS compression and reduction of F0 movement, the structural and non-structural views. Both of those views agree that the compression is a result of the greater prominence assigned to FOCUS in prosodic structure. However, the consequences of assigning greater prominence to the FOCUS item are different. For Pierrehumbert & Beckman (1988), the proponents of the non-structural view, the prominence assignment simply changes the "labeling" of the prosodic structure representation. This special labeling of the representation is ultimately interpreted as pitch range lowering/compression after FOCUS. For Truckenbrodt, a proponent of this structural view the FOCUS item corresponds to the sequence of terminal elements in the prosodic structure representation that contains the most prominent mora of an utterance, i.e. the DTE (Δ) of an utterance. Then, no phonological phrase boundaries appear after the prominent mora, in order to satisfy a constraint which requires the most prominent mora of an utterance to be as close as possible to the right edge of an utterance. Lack of phonological phrase boundaries after FOCUS, then, results in reduction and compression of F0 movement due to the lack of boundary tones or the lack of F0 resetting.

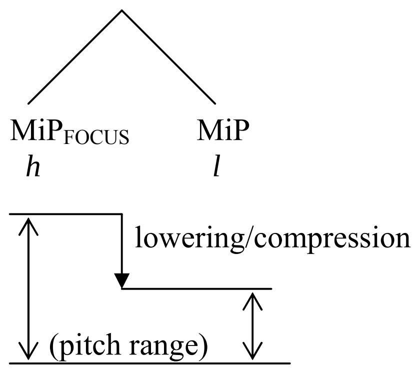
6.1. The Non-Structural View

According to Pierrehumbert & Beckman (1988), prosodic constituents are labeled as either *h* or *l*. The default labeling always follows the *l-h* order as shown in (1). This *l-h* labeling is phonetically interpreted as no change in the pitch range height and both the preceding *l* constituent and the following *h* constituent are realized in the same pitch range. However, when FOCUS is present, the FOCUS constituent is labeled as *h* and the following constituent is labeled as *l*. This *h-l* order is interpreted as pitch range lowering after the *h* node (i.e. FOCUS) as shown in (2).[21] As a result, post-FOCUS items are associated with more compressed and reduced F0 movement.

(1) The Default Labeling *l-h*



(2) The FOCUS Labeling *h-l*



The strongest version of this pitch range lowering approach is that phonological contrasts such as the presence/absence of edge tones and presence/absence of *catathesis* should be preserved even in a post-FOCUS part of an utterance. That is, the

[21] Ladd (1990) also proposes a similar metrical tree-like representation with *h-l* labels to give a formal analysis to English downstep. He also proposes that the *l-h* order be interpreted as no pitch range lowering (no register shift in his term) and the *h-l* order is interpreted as pitch range lowering (lowering of a register) at the node marked with *l*.

hierarchical organization of post-FOCUS items and all the post-FOCUS phonological phrase boundaries are kept intact.

For example, let us imagine a sequence of two words. In a *neutral* context (i.e. a FOCUS-free context) both of those words are mapped onto two independent Minor Phrases with an initial F0 rise at the onset of the second word as shown in (3). The strongest version of this approach predicts that the Minor Phrase boundary between those two words should be always kept intact between Word1 and Word2 even when Word1 is made into FOCUS as shown in (4). The only difference is that the F0 movement associated with Word2 of the FOCUS context is more compressed and lower because of pitch range lowering/compression.

(3) Neutral Context (without FOCUS)

$[\text{Word1}]_{\text{MiP}}$	$[\text{Word2}]_{\text{MiP}}$
LH	LH

(4) FOCUS Context (with FOCUS)

$[\text{Word1}_{\text{FOCUS}}]_{\text{MiP}}$	$[\text{Word2}]_{\text{MiP}}$
LH	LH

Realized lower and more compressed.

In the same way, this approach predicts that not only a Minor Phrase boundary but also a Major Phrase boundary should be kept intact after FOCUS. Imagine that two words Word1 and Word2 are mapped onto different Major Phrases as well as different

Minor Phrases in a neutral context as shown in (5). They predict that the same Major Phrase formation takes place even when Word1 is made into a FOCUS.

(5) Neutral Context (without FOCUS)

[ ]$_{\text{MaP}}$ [ ]$_{\text{MaP}}$
[Word1]$_{\text{MiP}}$ [Word2]$_{\text{MiP}}$

(6) FOCUS Context (with FOCUS)

[ ]$_{\text{MaP}}$ [ ]$_{\text{MaP}}$
[Word1$_{\text{FOCUS}}$]$_{\text{MiP}}$ [Word2]$_{\text{MiP}}$

Realized lower and more compressed.

6.2. The Structural View

In this section, I introduce the other view of post-FOCUS compression of F0 movement in Tokyo Japanese, the structural view.

6.2.1. Post-FOCUS Dephrasing

According to the structural view, post-FOCUS compression and reduction of F0 movement is induced by the absence (or deletion) of phonological constituent boundaries in a post-FOCUS position (Nagahara, 1994; Truckenbrodt, 1995; Uechi, 1997).

For example, let us imagine that there is a sequence of two words, Word1 and Word2, and those words are mapped onto separate Major Phrases as well as separate Minor Phrases in a FOCUS-free context as shown in (7). As a result, there are LH edge tones at the left edge of Word2. At the same time, the F0 excursion size between those LH tones at the left edge of Word2 is relatively large and no *catathesis* takes place there because a Major Phrase boundary is present. Once Word1 is made into a FOCUS, however, the Major Phrase boundary at the left edge of Word2 is deleted as shown in (8a). Also, according to this view, even the Minor Phrase boundary between the FOCUS Word1 and the post-FOCUS Word2 may be deleted in principle as in (8b).

(7) Neutral Context (without FOCUS)

[ ]$_{MaP}$ [ ]$_{MaP}$
[Word1]$_{MiP}$ [Word2]$_{MiP}$
LH LH

Initial F0 rise,
large F0 excursion size between L and H, no *catathesis*

(8) FOCUS Context (with FOCUS)

(a) [ ]$_{MaP}$
[Word1$_{FOCUS}$]$_{MiP}$ [Word2]$_{MiP}$
LH LH

Initial F0 rise,
smaller F0 excursion size between L and H, *catathesis*

or

(b) [ ]$_{MaP}$
[Word1$_{FOCUS}$ Word2]$_{MiP}$
LH

No initial rise,

Deletion (or absence) of phonological phrase boundaries in the post-FOCUS position results in compression and reduction of F0 movement. Lack of a Major Phrase boundary leads to smaller F0 excursion size[22] between L and H and *catathesis*, and lack of a Minor Phrase boundary leads to absence of initial F0 rise. In what follows, the theory behind this structural view is presented.

6.2.2. The Focus-Prominence Hypothesis

"Deletion" of post-FOCUS phonological phrase boundaries (i.e dephrasing) is derived from the "Focus-Prominence Hypothesis" of Truckenbrodt (1995), which is further developed by Selkirk (2000a, 2002ab). The theory of Focus-Prominence Hypothesis is part of the Prosodic Structure Theory introduced in Chapter 1 and assumes that an utterance is analyzed in terms of two types of structural representations: morpho-syntactic and phonological. The surface morpho-syntactic representation is mapped onto a phonological structure representation. A morpho-syntactic constituent interpreted as FOCUS (contrastive focus) is marked with a feature [FOCUS] (Jackendoff, 1972; Selkirk, 1984; Rooth, 1992; among others).

Truckenbrodt (1995) originally proposed a syntax-phonology interface constraint that maps the FOCUS-marked constituent to the highest prosodic prominence within a Focus domain. Selkirk (2002ab) further formalized Truckenbrodt's FOCUS-Prominence correspondence into a constraint which I call FOCUS-PROMIENCE or FOCUS-ΔIP. Selkirk's FOCUS-PROMIENCE constraint calls for the FOCUS-marked constituent in the syntactic representation to correspond to a terminal string in the

[22] See Chapter 1, Section 4.

phonological representation that contains the most prominent mora (DTE, Δ) of an Intonational Phrase.

(9) FOCUS-ΔIP (FOCUS-PROMIENCE) (Selkirk, 2002ab)

> The FOCUS-marked constituent in the morpho-syntactic representation should correspond to a string of the phonological representation which contains the highest prominence (DTE,Δ) of an Intonational Phrase.

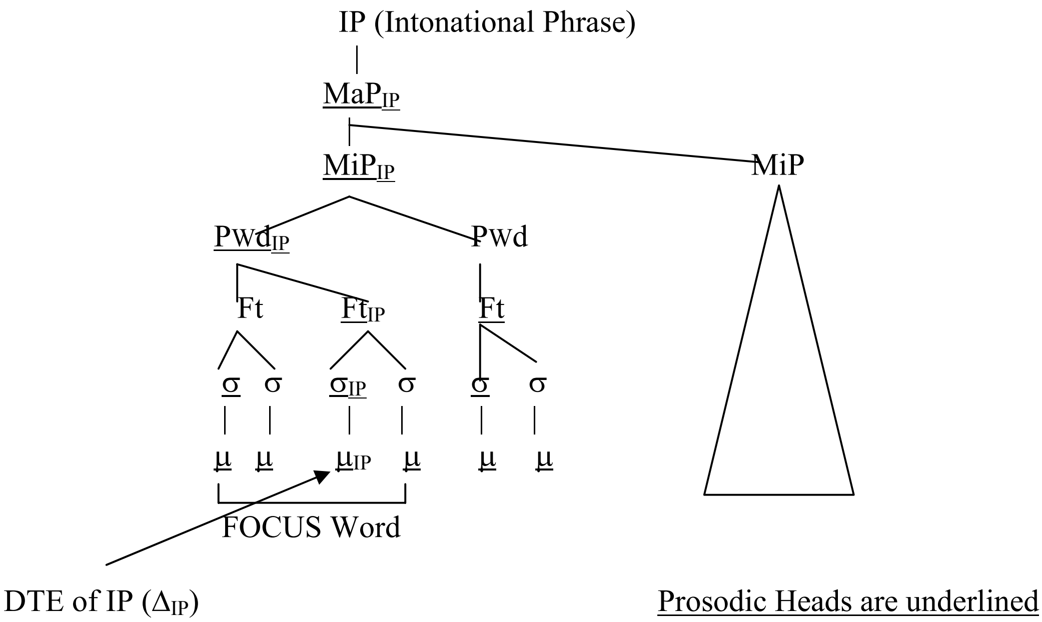


Figure 6.01. The DTE of IP and FOCUS

In addition to the syntax-phonology interface FOCUS-PROMINENCE constraint in (9), there are also prominence-related phonological markedness constraints which call for the DTE (Δ) of a phonological constituent to be close to the edges of that constituent. Such prominence-edge alignment is already well attested in many languages in word-level phonology (Prince & Smolensky, 1993; Hayes, 1995). Following Truckenbrodt (1995), I propose that the prominence-edge alignment constraint relevant here is ALIGN$_R$(Δ_{IP}, IP), which requires the DTE (Δ) of an Intonational Phrase to be as close as possible to the right edge of the Intonational Phrase.

(10) ALIGN$_R$(Δ_{IP}, IP)

The DTE (Δ) of an Intonational Phrase must coincide with the right edge of an Intonational Phrase.

According to Truckenbrodt, this constraint is gradiently violated, and the violation count is made on all the structures that follow the DTE (Δ) of an Intonational Phrase. The output representation that contains no phonological phrase constituents after FOCUS is the optimal. This is shown in the following tableau.

Tableau 6.01.

	[Wd1$_{FOCUS}$] [Wd2] [Wd3]	FOCUS-ΔIP	ALIGN$_R$ (Δ_{IP}, IP)
a.	$_{IP}$( ) $_{MaP}$(_____) $_{MaP}$( ) $_{MaP}$( ) $_{MiP}$(_____) $_{MiP}$( ) $_{MiP}$( ) Wd1 Wd2 Wd3 Δ_{IP}		*!* (2MaPs) ** (2MiPs)
b.	$_{IP}$( ) $_{MaP}$(____________________) $_{MiP}$(_____) $_{MiP}$( ) $_{MiP}$( ) Wd1 Wd2 Wd3 Δ_{IP}		*!* (2 MiPs)
c. ❍	$_{IP}$( ) $_{MaP}$(____________________) $_{MiP}$(____________________) Wd1 Wd2 Wd3 Δ_{IP}		

Heads of prosodic constituents are underlined.

Under the Focus-Prominence Hypothesis phonological phrase boundaries may still appear after FOCUS as long as a markedness constraint that calls for such boundaries outranks the prominence-edge alignment constraint ALIGN$_R$(Δ_{IP}, IP). An example of such a markedness constraint is something that calls for at least a Minor Phrase boundary at the left edge of each word under a certain condition, and I tentatively call the constraint MiP-INSERTION. Let us imagine that the MiP- INSERTION constraint outranks ALIGN$_R$(Δ_{IP}, IP) as shown in (11).

(11) MiP-INSERTION >> ALIGN$_R$(Δ_{IP}, IP)

According to this ranking, the optimal candidate must have Minor Phrase boundaries but does not have Major Phrase boundaries after FOCUS, as shown in the following tableau.

Tableau 6.02.

[Wd1$_{FOCUS}$] [Wd2] [Wd3]	MiP-INSERTION	ALIGN$_R$ (Δ_{IP}, IP)
a. $_{IP}$( ) $_{MaP}$( ) $_{MaP}$( ) $_{MaP}$( ) $_{MiP}$( ) $_{MiP}$( ) $_{MiP}$( ) Wd1 Wd2 Wd3 Δ_{IP}		**! (2MaPs) ** (2MiPs)
b. ❍ $_{IP}$( ) $_{MaP}$( ) $_{MiP}$( ) $_{MiP}$( ) $_{MiP}$( ) Wd1 Wd2 Wd3 Δ_{IP}		** (2 MiPs)
c. $_{IP}$( ) $_{MaP}$( ) $_{MiP}$( ) Wd1 Wd2 Wd3 Δ_{IP}	*!*	

I assume that FOCUS-PROMINENCE$_{IP}$ is undominated. Heads of prosodic constituents are underlined.

In summary, the Focus-Prominence Hypothesis derives dephrasing of a Major Phrase boundary and a Minor Phrase boundary from two types of constraints: the FOCUS-Prominence interface constraint (FOCUS-PROMIENCE) and the prominence-edge alignment constraint ALIGN$_R$(Δ_{IP}, IP). *Deletion* of those phrase boundaries, however, may not be mandatory as long as some markedness constraint which calls for such boundaries outranks the ALIGN$_R$(Δ_{IP}, IP) constraint.

6.2.3. Further Development of the Focus-Prominence Theory by Selkirk

Selkirk (2000b, 2002ab) further developed the theory of Focus-Prominence. Her theory recognizes existence of different types of Focus, and posits different constraints on the phrasing. She suggests that not only contrastive FOCUS but also presentational focus (i.e. new information, henceforth focus) is related to prosodic prominence. Though contrastive FOCUS is related to an Intonational Phrase-level prominence, focus is related to a lower level of prominence, presumably Major Phrase-level prominence or Minor Phrase level prominence (Selkirk 2000b, 2002ab). For example, the following Focus-Prominence constraints in (12) and (12’) are possible under her theory.

(12) focus(XP)- ΔMiP (2000b)

The terminal string of a focus-marked XP (i.e. new XP) in the input syntactic representation must correspond to a terminal string in the output phonological representation which contains the DTE (Δ) of a prosodic constituent Minor Phrase.

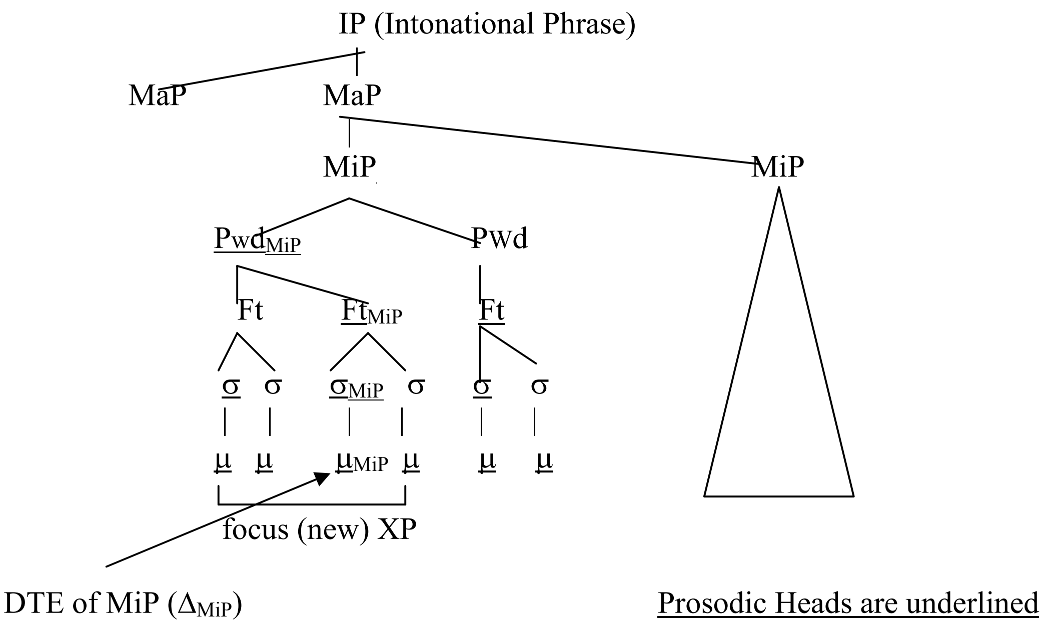


Figure 6.02. The DTE of MiP and focus-Marked XP

(12’) focus(XP)-ΔMaP Selkirk (2002b)

The terminal string of a focus-marked XP (i.e. new XP) in the input syntactic representation must correspond to a terminal string in the output phonological representation which contains the DTE (Δ) of a prosodic constituent Major Phrase.

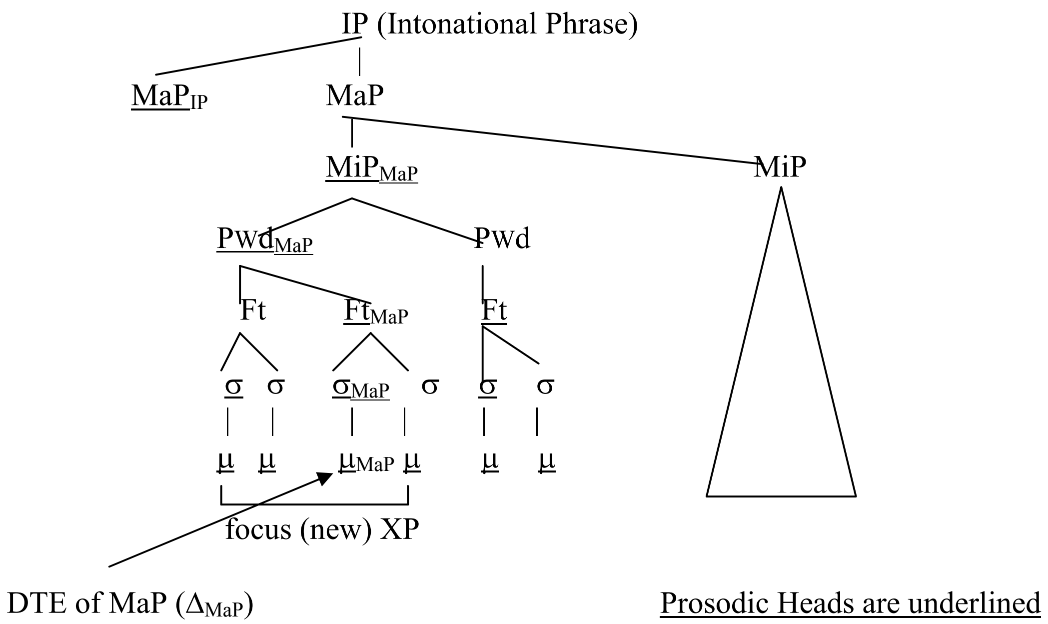


Figure 6.03. The DTE of MaP and focus-Marked XP

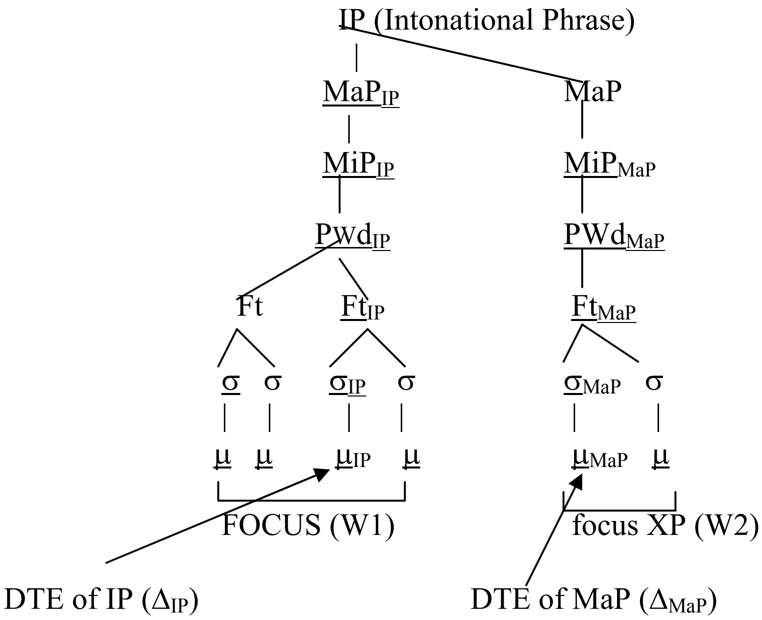
A motivation for the focus XP-Prosodic Prominence constraint comes from the distribution of H* pitch accent tone in English. In English, a H* pitch accent tone appears on either a FOCUS constituent or a new (presentational focus) constituent, while no H* tone appears on already given or presupposed constituents (Selkirk 1995b, Selkirk 2000b, 2002a, among many others). According to Selkirk (2000b), the H* accent tone is associated with the DTE (Δ) of a Minor Phrase, and that is why no H* accent tone appears on non-FOCUS given items. However, not all new (focus) constituents obligatorily bear the pitch accent H*. For example, non-XP *stage level* predicates interpreted as focus do not necessarily bear the pitch accent, while focused XP *individual level* predicates always occur with the H* pitch accent. Given this, Selkirk suggested a constraint that calls for correspondence between a focus XP and the DTE of a Minor Phrase, which is already shown in (12). Selkirk (2002b) suggests that the constraint in (12') which calls for correspondence between a focus XP and the DTE of a <u>Major Phrase</u> is also a possible formulation of a focus-prominence constraint because she assumes a family of Focus-Prominence constraints in which an each F-marked constituents (both focus and FOCUS) is associated with different levels of prosodic constituents.

One of the main concerns of the following chapters is whether the focus XP-Prominence constraints in (12) and (12') play any significant role in Tokyo Japanese. If those constraints outrank the post-FOCUS dephrasing constraint introduced above, i.e. ALIGN$_R$ (Δ_{IP}, IP), the theory predicts absence of dephrasing at post-FOCUS new XPs. To see how such a prediction is made, let us consider a hypothetical case in which the focus XP-<u>MaP</u> Prominence constraint in (12') outranks the dephrasing constraint, ALIGN$_R$ (Δ_{IP}, IP). Also imagine that the input sequence consists of a FOCUS word (Word1) and an immediately following focus XP (Word2) as shown in (13).

(13) [Word1$_{FOCUS}$] $_{XP}$[Word2$_{focus}$]

Word1$_{FOCUS}$ must correspond to the terminal string in the phonological representation that contains the DTE of an Intonational Phrase to satisfy the FOCUS-

PROMIENCE (FOCUS-ΔIP) constraint in (9). By definition, the DTE (Δ) of an Intonational Phrase is also the DTE (Δ) of the head Major Phrase of the Intonational Phrase. At the same time, the focus XP (Word2) should correspond to the DTE (Δ) of a Major Phrase to satisfy the focus(XP)-ΔMaP constraint in (12'). Given that both Word1 (FOCUS) and Word2 (focus XP) are required to correspond to the DTE (Δ) of a Major Phrase, those two words must form separate Major Phrases. It is because only one DTE is allowed within a single Major Phrase. That is, a Major Phrase boundary is called for at the left edge of Word2 in spite of the fact it is preceded by FOCUS. This is shown schematically in Figure 6.04.



Prosodic Heads are underlined

Figure 6.04. The DTE of IP (FOCUS) and the DTE of MaP (focus XP)

As long as the focus(XP)-ΔMaP constraint in (12’) outranks the prominence-edge alignment constraint (ALIGN$_R$(Δ_{IP}, IP)) in (10) as shown in (14), the output representation in (a) of the following tableau (Tableau 6.03) will be the optimal one.

(14) focus(XP)-ΔMaP >> ALIGN$_R$(Δ_{IP}, IP)

Tableau 6.03.

$_{XP}$[Wd1(*FOC*)]$_{XP}$ $_{XP}$[Wd2(*foc*)]$_{XP}$	FOCUS-PROM$_{IP}$	focus(XP)-PROM$_{MaP}$	ALIGN$_R$ (Δ_{IP}, IP)
a. ❍ $_{IP}$( ) $_{MaP}$( ) $_{MaP}$( ) $_{MiP}$( ) $_{MiP}$( ) Wd1 Wd2 Δ_{IP} Δ_{MaP}			* (1 MaP) * (1 MiP)
b. $_{IP}$( ) $_{MaP}$( ) $_{MiP}$( ) $_{MiP}$( ) Wd1 Wd2 Δ_{IP}		*!	* (1 MiP)
c. $_{IP}$( ) $_{MaP}$( ) $_{MiP}$( ) $_{MiP}$( ) Wd1 Wd2 $\Delta_{IP(MaP)}$	*!		

Heads of constituents are underlined.

In summary, according to Selkirk's version of Focus-Prominence theory, post-FOCUS dephrasing may not be a necessary property of an utterance with a FOCUS. It may be conditioned by the information status of post-FOCUS items (i.e. whether focus or given) and their syntactic category (i.e. whether XP or not). When those post-FOCUS items are focus XPs, then her theory predicts that a Phonological Phrase boundary may appear at edges of those items even in the post-FOCUS position.

6.3. Overview of the Following Chapters

Chapter 7 is about XP-internal post-FOCUS effects. In that chapter, I focus on sequences of a FOCUS word (Word1) and a post-FOCUS word (Word2) that form an immediate syntactic constituent without having any XP boundary at the left edge of

Word2. I show that there is dephrasing of a Minor Phrase boundary when both the FOCUS (Word1) the post-FOCUS Word2 are unaccented, irrespective of the information status of the following word. In that sense, the structural view is right. However, at the same time, I show that there are some non-structural aspects of XP-internal post-FOCUS effects, too. When both FOCUS (Word1) and the following word (Word2) are accented, the Minor Phrase boundary at the left edge of Word2 is kept intact. Nonetheless, the post-FOCUS Word2 is associated with a lower F0 peak and a smaller F0 excursion size.

Chapter 8 is about post-FOCUS effects across XP boundaries. In that chapter, there is an XP boundary at the left edge of the post-FOCUS Word2. In a neutral context without any FOCUS, the left edge of an XP is where a Major Phrase boundary is expected to appear (Selkirk & Tateishi 1991), and my question is whether a Major Phrase boundary appears there even when the XP boundary is preceded by a FOCUS. More specifically, the question is whether the information status difference of the post-FOCUS XP (i.e. given vs. new) results in any difference in terms of preservation of a Major Phrase boundary is one of my main concerns.

CHAPTER 7

XP-INTERNAL POST-FOCUS EFFECTS

In this chapter, we investigate XP-internal post-FOCUS effects. We focus on sequences of a FOCUS word (Word1) and a post-FOCUS word (Word2) that form an immediate syntactic constituent without having any XP boundary at the left edge of Word2. In a sequence of two words with no XP boundary at the left edge of the second word, we usually expect a Minor Phrase boundary to appear at the left edge of the second word in a *neutral* context (i.e. FOCUS-free context). Our main concern is whether that Minor Phrase boundary is kept intact even when the preceding word is made into a FOCUS. Results of my experiments show that there is dephrasing (deletion) of a Minor Phrase boundary when both the FOCUS (Word1) the post-FOCUS Word2 are unaccented, irrespective of the information status of the following word. This is used as evidence for the structural view. At the same time, however, I also found non-structural aspects of XP-internal post-FOCUS effects. When both FOCUS (Word1) and the following word (Word2) are accented, the Minor Phrase boundary at the left edge of Word2 is kept intact. Nonetheless, the post-FOCUS Word2 was associated with a lower F0 peak and smaller F0 excursion size.

In Section 7.1, I first introduce reading materials and their contexts. In Section 7.2 and Section 7.3, experimental results related to the structural effect of FOCUS on the following word are provided. In Section 7.5, I ask whether there is any non-structural effect of FOCUS.

7.1. Sequences of Two Words and Information Structure

In this section, I introduce sequences of two words whose F0 patterns are under concern. The first word (Word1) is interpreted as FOCUS and the second word is a non-FOCUS word (Word2). More specifically, Word2 is interpreted as either new or given. In addition, the accent status of those two words was varied. In one accent condition, both Word1 and Word2 were unaccented. In the other accent condition, they were both accented. Word1 and Word2 form a single XP and there is no XP boundary at the left edge of Word2. Since Japanese is a head-final language, the rightmost word of a branching constituent is always the head of that constituent and no XP boundary appears at its left edge.

X_iP

X_jP

X_j	X_i
Word1	Word2
FOC	*New/Given*
U [-Acc]	U [-Acc]
A [+Acc]	A [+Acc]

Figure 7.01. The Sequence of Word1 and Word2 Constituting an XP

I refer to the sequence of an accented FOCUS and the following accented new/given Word2 as "AAFN" and "AAFG" respectively. In the same way, the sequence of an unaccented FOCUS and the following unaccented new/given Word2 is referred to as "UUFN" and "UUFG" respectively.

Table 7.01. The Sequences of FOCUS Word1 and Post-FOCUS Word2

FOCUS Sequences	FOC$_{Word1}$ New$_{Word2}$	FOC$_{Word1}$ Given$_{Word2}$
Accented	AAFN	AAFG
Unaccented	UUFN	UUFG

We are interested in the presence or absence of a phonological phrase boundary at the left edge of the post-FOCUS Word2. At the same time, we are also interested in whether there is any non-structural post-FOCUS effect on Word2.

The phonological phrase patterns of those constituents consisting of a FOCUS and the following new or given word are compared with those of *control* constituents. The control constituents consist of either new words only or given words only without FOCUS on the first word.

Table 7.02. The Sequences of Control Word1 and Word2

Control Sequences	New$_{Word1}$ New$_{Word2}$	Given$_{Word1}$ Given$_{Word2}$
Accented	AANN (… compared with AAFN)	AAGG (… compared with AAFG)
Unaccented	UUNN (… compared with UUFN)	UUGG (… compared with UUFG)

If any phonological phrase boundaries at the left edge of Word2 of those control cases are deleted at the left edge of the post-FOCUS Word2 of the FOCUS cases, then we will conclude that there is post-FOCUS dephrasing.

A sequence of accented words and that of unaccented words used in this section are shown in (1) and (2). They all come from a dataset referred to as the "<*Yunyuu*> Set". Detailed descriptions of the dataset are provided in the Appendix.

(1) UU (from the <*Yunyuu*> Set)

Yokohama-no *yunyuu-daikooya-de*
Yokohama-Gen importing-agency-At
"At an importing agency of Yokohama"

(2) AA (from the <*Yunyuu*> Set)

Aóyama-no *yunyuu-dairíten-de*
Aoyama-Gen importing-agency-At
"At an importing agency of Aoyama"

The UU form in (1) consists of two unaccented words, while the AA form in (2) consists of two accented words. In both (1) and (2), the second word (Word2) is a compound noun and relatively long (i.e. consisting of six syllables).

For the sequence of unaccented words, it was necessary to make the second word as long as possible in order to guarantee at least a Minor Phrase boundary between Word1 and Word2 in the control cases. Previous to recording the sequences of unaccented words in (1), I recorded a sequence of two unaccented words whose second word was a simple noun consisting of only three syllables, *moderu* ("model"), as shown in (3).

(3) UU

Yokohama-no ***moderu-ga*** **Short and Simple**
Yokohama-Gen moderu-Nom

It turned out that two of the three speakers (MR and SK) had no Minor Phonological Phrase boundary at the left edge of this simple word even in the NN control context. When it comes to the other speaker RO, she allowed both absence and presence of a MiP boundary in that context. Even when both of those two words are interpreted as new, the presence of a MiP boundary at the left edge of the second word is not guaranteed as long as the second word is simple and consists of a small number of syllables. This makes it impossible to ask the core question, i.e. whether there is any post-FOCUS deletion of a Minor Phonological Phrase break. Once the second word is made into a long compound noun like *yunyuu-daikooya*, however, a MiP boundary is more likely to appear at the left edge of the second word.

As already mentioned in Section 1.3.3. of Chapter 1, a Minor Phrase boundary is present at the left edge of the unaccented compound Word2 because of an alignment constraint, ALIGN$_L$(X^0-Branching, MiP), dominating a prosodic markedness constraint which requires coalescence of two words into a single Minor Phrase, i.e. BINARY MINIMUM(MiP). This is again shown below.

(4) ALIGN$_L$(X^0-Branching, MiP)
The left edge of a branching X^0 coincides with the left edge of a Minor Phrase.

(5) BINARY MINIMUM(MiP)
A prosodic constituent of level C_I must dominate at least two prosodic constituents of level MiP.

(6) ALIGN$_L$(X^0-branching, MiP) >> BINARYMINIMUM(MiP)
… from Section 3.3., Chapter 1

In order to make things parallel to the UU cases, a compound word was also used for Word2 of the AA cases.

I made the first two syllables of Word2 share the same vowel quality, i.e. a high back vowel to avoid any F0 variation induced by vowel intrinsic F0 height. There is a universal tendency for high vowels such as [i] and [u] to have higher F0 than low vowels such as [a] (Whalen & Levitt, 1995). In Japanese female speech, high vowels are intrinsically 15~20 Hz higher than low vowels (Homma, 1973). Since the presence or absence of an F0 rise from the initial syllable to the second syllable of Word2 is to be used as evidence for presence of a Minor Phrase boundary, it was necessary to make both of those syllables share the same vowel quality. I have not been concerned with the issue of intrinsic F0 values of vowels before this chapter because the presence or absence of a Minor Phrase boundary, i.e. the presence or absence of F0 rise from the initial syllable to the second syllable of a word, has not been the main concern of those preceding chapters.

The sequences of words in (1) and (2) were embedded in a sentence, and that sentence was further embedded in a dialogue (see Appendix). The information structure of those dialogues was manipulated so that target words obtain desired interpretations. These dialogues are shown below.

Dialogue NN: Word 1 = new, Word 2 = new

Speaker:	*chotto kiiteyo.* "Hey, just listen to me."		
Experimenter:	*nani?* "What?"		
Speaker:	**[*Yokohama-no*]$_{\mathbf{Wd1}}$** Yokohama-Gen	**[*yunyuu-daikooya-de*]$_{\mathbf{Wd2}}$** importing-agency-At	*maneejaa-no* manager-Gen

yuujin -ga *moderu-ni* *mayaku-o* *uttarashíiyo.*
friend-Nom model-To marijuana-Acc sold-I heard.

"I've heard that at an importing agency office in Yokohama, a friend of the manager (of that company) sold marijuana to a model."

Dialogue GG: Word1 = given, Word2 = given

Experimenter: *Yokohama-no yunyuu-daikooya-de juugyóoin-no yuujin-ga moderu-ni mayaku-o uttarashíi-ga, hontookane?*
"Is it true that at an importing agency office in Yokohama, a friend of an employee sold drugs to a model?"

Speaker: *iie, chigaimásu.*
"No, that's not correct."

[*Yokohama-no*]$_{\mathbf{Wd1}}$	**[*yunyuu-daikooya-de*]$_{\mathbf{Wd2}}$**	*MANEEJAA-NO*
Yokohama-Gen	importing-agency-At	MANAGER-GEN

yuujin -ga	*moderu-ni*	*mayaku-o*	*uttándesu.*
friend-Nom	model-To	marijuana-Acc	sold.

"At the importing agency office in Yokohama, a friend of a MANAGER sold drugs to a model."

Dialogue FN: Word 1 = FOC, Word 2 = new

Experimenter: *nani?*
"What?"

Speaker: *Tookyoo-ya Oosaka-no yoona hanzai-no tahatsusuru tóshi-dewanaku,*
Tokyo-and Osaka like high crime rate city-Not,

tian-no yóikotode sirareru, *ano* **[*YOKOHAMA-NO*]$_{\mathbf{Wd1}}$**
safety-Gen good known, that YOKOHAMA-GEN,

[*yunyuu-daikooya-de*]$_{\mathbf{Wd2}}$ *maneejaa-no*
importing-agency-at manager-Gen

yuujin-ga *moderu-ni* *mayaku-o* *uttarasíiyo.*
friend-Nom model-to marijuana-Acc sold-I heard.

"I've heard that in an (office of) importing agency of YOKOHAMA, the city known to be safe unlike those cities like Osaka or Tokyo, which are notorious for their high crime rate, a friend of the manager (of that importing agency) sold marijuana to a model."

Dialogue FG: Word 1 = FOC, Word 2 = given

Experimenter: *Aóyama-no yunyuu-daikooya-de maneejaa-no yuujin-ga moderu-ni*
marifana-o uttá-to kiitaga, hontookánee?
"I've heard that in an importing agency of Aoyama, a friend of the manager (of that importing-agency) sold marijuana to a model. But is it true?"

Speaker: *chigaimásu.* *Aóyama-dewa-naku* **[*YOKOHAMA-NO*]$_{\mathbf{Wd1}}$**
"No" Aoyama-Copula-Not YOKOHAMA-GEN

[*yunyuu-daikooya-de*]$_{\mathbf{Wd2}}$ *maneejaa-no*
importing-agency-at manager-Gen

yuujin-wa *moderu-ni* *mayaku-o* *uttá-ndesu.*
friend-Top model-to marijuana-Acc sold-Copula.

"In the importing agency of YOKOHAMA but not in Aoyama, a friend of the manager (of the importing agency) sold marijuana to a model."

7.2. Post-FOCUS MiP Dephrasing in an Unaccented Context

The phonological phrase formation of sequences of two unaccented words is considered in this section. Again, the sequence of two unaccented words used in this section is introduced below.

(1) Unaccented W1 and W2 without an XP Boundary at the Left Edge of W2

[*Yokohama-no*]$_{\text{Word1}}$ [*yunyuu-daikooya-de*]$_{\text{Word2}}$
Yokohama-Gen importing-agency-At
"At an importing agency of Yokohama"

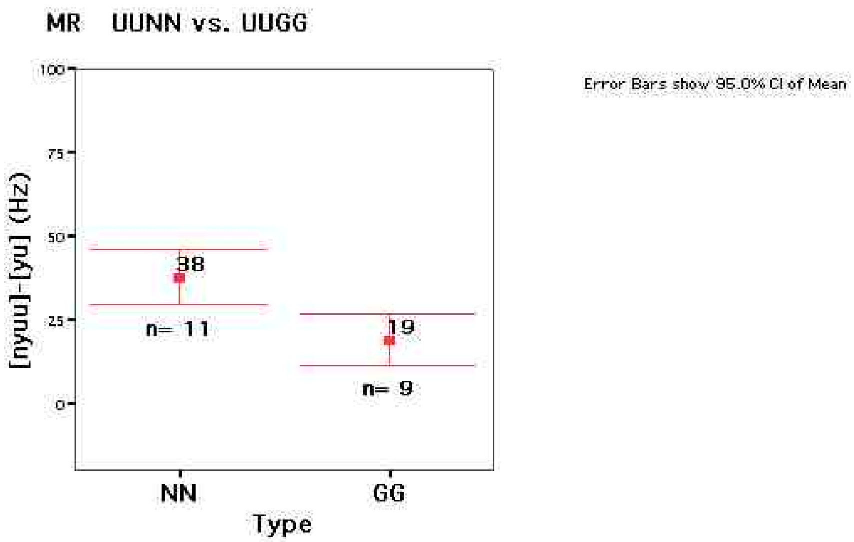
The onset of a Minor Phrase in Tokyo Japanese is marked with L and H edge tones. Those tones are usually aligned with the first and the second syllable of the Minor Phrase respectively (see Chapter 1). Given this, I use an F0 rise from the first to the second syllable of Word2 as evidence for a Minor Phrase boundary at its left edge of Word2. For this purpose, the F0 value of the first syllable [*yu*] and that of the second syllable [*nyuu*] were measured. Then, the difference between those two syllables which represents the magnitude of F0 rise from the first to the second syllable was obtained. If the F0 rise is significantly greater than zero, then I will conclude that there is a Minor Phrase boundary at the left edge of Word2.

7.2.1. The Control Cases: UUNN & UUGG

First, I show results from the control cases, i.e. UUNN and UUGG. The UUNN case consists of two unaccented new words and the UUGG case consists of two unaccented given words.

[*Yokohama-no*]Word1 [*yunyuu-daikooya-de*]Word2
Yokohama-Gen importing-agency-At
New *New* → UUNN
Given *Given* → UUGG

The cases are used as controls with which FOCUS cases were compared. We found that both of those control cases had a significant initial F0 rise from the first to the second syllable of Word2. The following figures (Figure 7.02) summarize the mean F0 difference between those two syllables ([nyuu]-[yu]) and its 95% confidence interval.

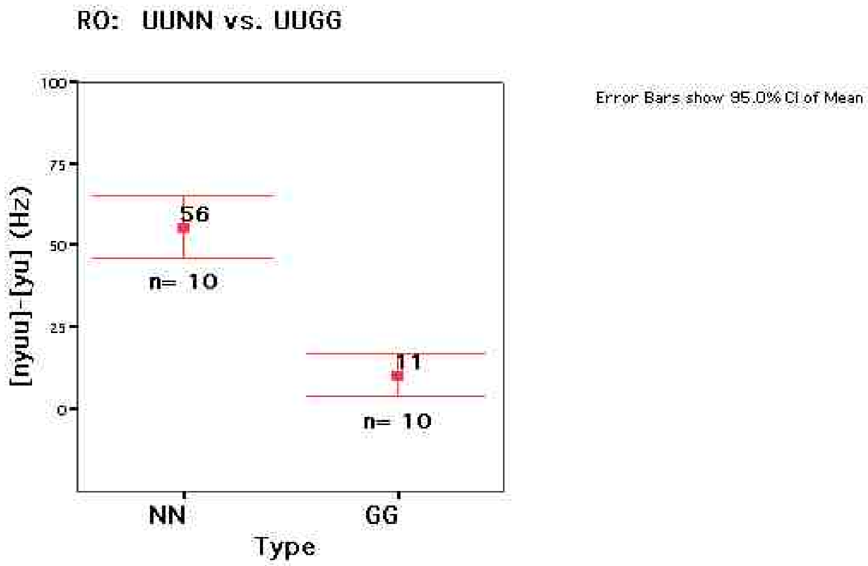


a. Speaker MR.

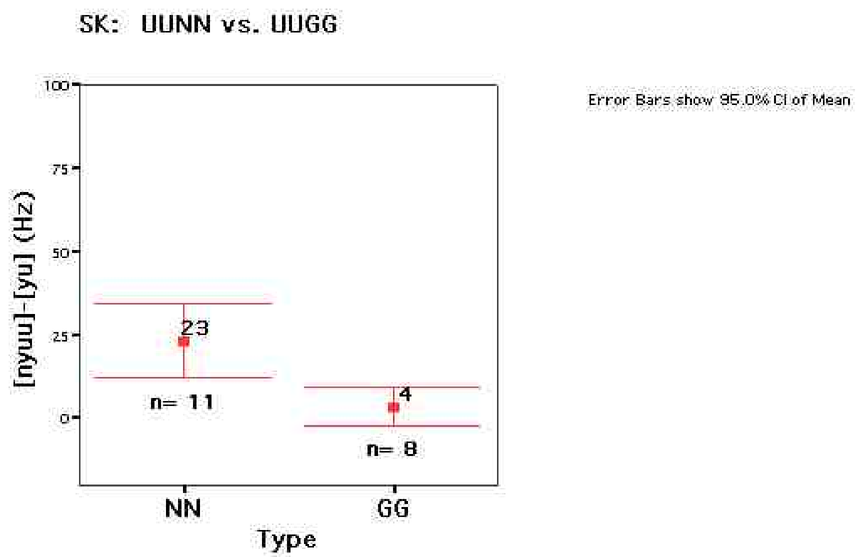
Figure 7.02. The Mean F0 Difference between the First and the Second Syllable of Word2 of the Control Cases (UU)

Continued next page

Figure 7.02 continued



b. Speaker RO



c. Speaker SK

For all three speakers and for both the NN and the GG cases, the mean value of the F0 rise from the first and the second syllable of Word2 was greater than zero. The mean values of the NN case obtained from all three speakers' data were significantly greater than zero: their 95% confidence intervals do not include zero, and this is interpreted as evidence for the presence of a Minor Phrase boundary at the onset of Word2 of the NN case.

For the GG case, too, in MR and RO's data the mean values of the F0 rise from the first to the second syllable of Word2 were statistically significant: their 95% confidence intervals did not include zero. Given this, it is reasonable to conclude that there is a L and H tonal target, i.e. a Minor Phrase boundary, at the left edge of Word2 of the GG case in those two speakers' speech. The mean of the F0 rise of SK's GG case, however, was not significantly greater than zero: the confidence interval of the mean included zero. Nonetheless, I argue for presence of a Minor Phrase boundary at the onset of Word2 based on my observation of SK's F0 contours. An exemplar case of her F0 contour is shown in Figure 7.03.

SK 4-75-22

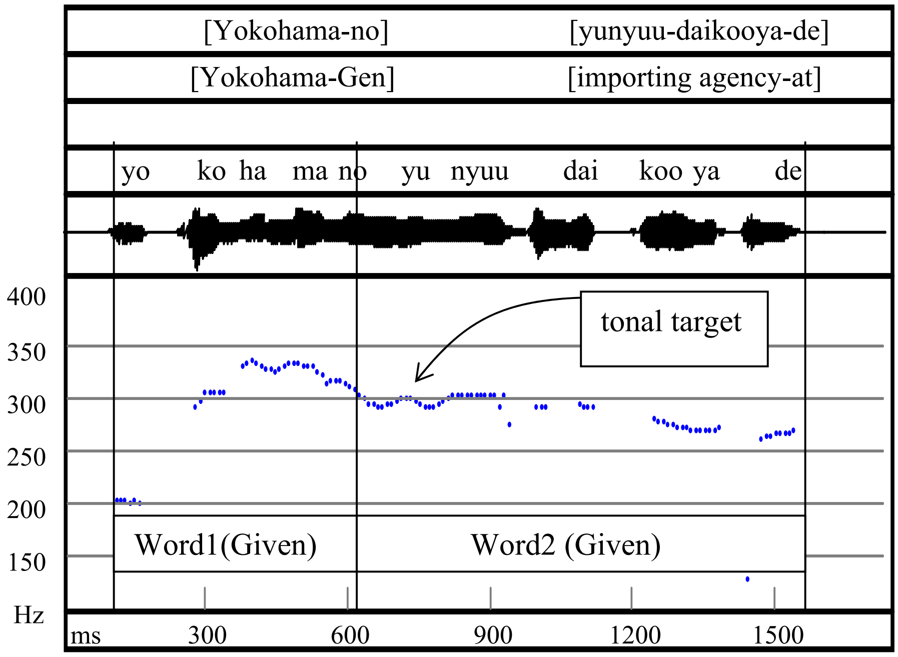


Figure 7.03. An F0 Contour Example of UUGG (Speaker SK)

The evidence on which I depend to argue for the presence of a Minor Phrase boundary at the left edge of Word2 in SK's speech is the tonal target at the onset of that word. Figure 7.03 shows an F0 contour of the sequence of a given Word1 (*yokohama-no*) and a given Word2 (*yunyuudaikooya-de*). Though there is no F0 rise from the first to the second syllable of Word2, there is still a tonal target at the onset of Word2. That is, the F0 slope associated with Word1 is cut off at the right edge of Word1, and a new F0 plateau begins at the onset of Word2. This F0 plateau lasts at least between the first and the second syllable of Word2. One way to account for this F0 shape is that there is some tonal target at the beginning of Word2. I interpret the tonal target as the L edge tone at its first syllable and a H edge tone at its second syllable, as well. Both the L edge tone and the H edge tone bear the same F0 value, which results in a F0 plateau at the beginning of Word2. I will come back to this point later in this chapter to discuss factors manipulating those edge tone values.

In summary, I conclude that both of the control cases (the NN and the GG case) without any FOCUS on Word1 have at least a Minor Phrase boundary between Word1 and Word2.

7.2.2. The FOCUS Cases: UUFN & UUFG

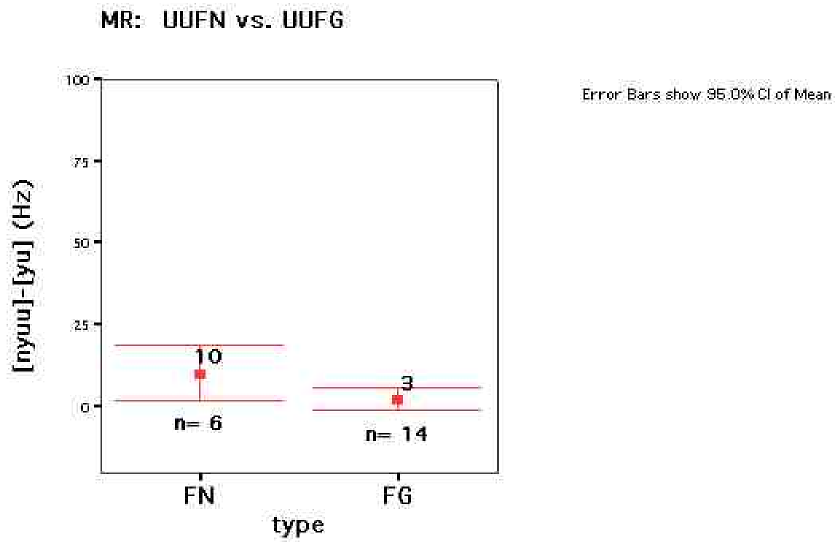
In this section, the presence or absence of a Minor Phrase boundary at the left edge of Word2 of the FOCUS cases (i.e. the UUFN and the UUFG cases) is considered. In those FOCUS cases, Word1 is a FOCUS, and Word2 is either new or given respectively.

(1) The sequence of unaccented W1 and W2 without an XP Boundary

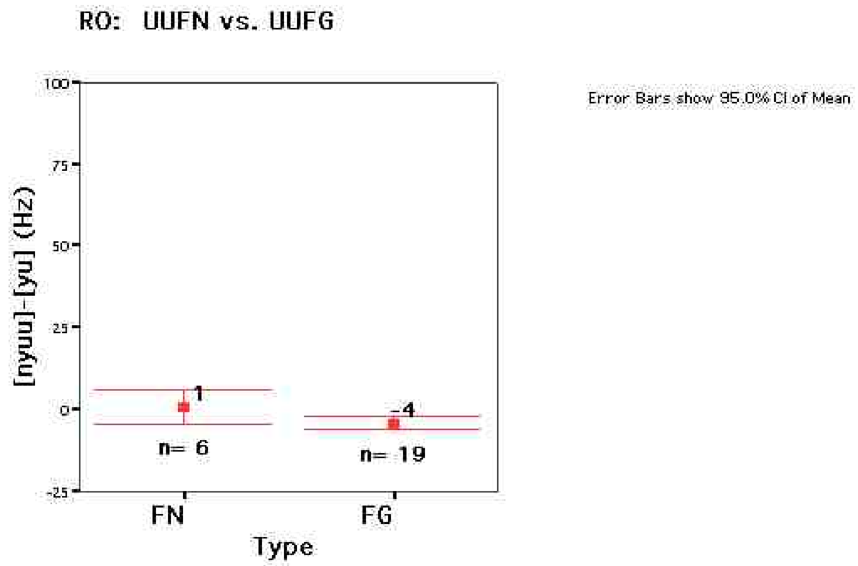
[*Yokohama-no*]$_{\text{Word1}}$	[*yunyuu-daikooya-de*]$_{\text{Word2}}$		
Yokohama-Gen	importing-agency-At		
FOCUS	*New*	→	UUFN
FOCUS	*Given*	→	UUFG

"At an importing agency of Yokohama"

Again, the F0 difference between the first syllable (*yu*) and the second syllable (*nyuu*), which represents the magnitude of an F0 rise from the former to the latter, was obtained. Its F0 mean and 95% confidence interval are shown below.



a. Speaker MR

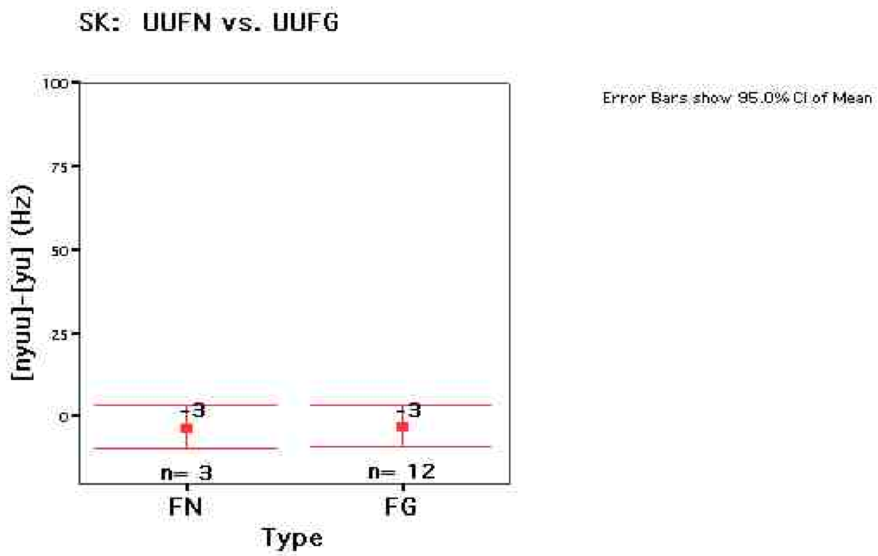


b. Speaker RO

Figure 7.04. The Mean F0 Difference between the First and the Second Syllable of Word2 of the Post-FOCUS Cases (UU)

Continued next page

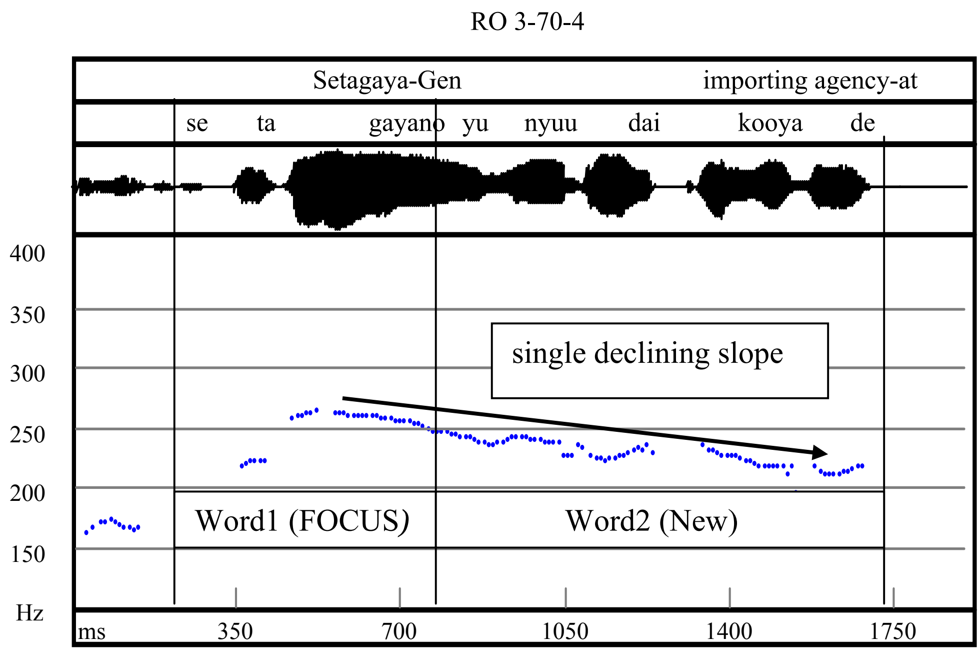
Figure 7.04 continued



c. Speaker SK

<Speaker RO & SK>

In RO and SK's speech (Figure 7.04-b and 7.04-c), both the mean of the F0 rise of the FN and that of the FG case were either below zero or close to zero. Also, those mean values were not significantly greater than zero: their 95% confidence intervals included zero. This indicates that there is no Minor Phrase boundary at the left edge of the post-FOCUS Word2 regardless of their information status. This is also verified by their F0 contours which lack any F0 discontinuity between the offset of Word1 (FOCUS) and the onset of Word2. As shown in the F0 contours (Figure 7.05), a single declining slope unfolds over both Word1 and Word2. This is good evidence that both Word2 (FOCUS) and Word2 (given or new) are coalesced into the same Minor Phrase. Though there are a few bumps in the F0 slope shown in Figure 7.05-a and Figure 7.05-b, they do not change the overall slope of the F0. Those bumps are caused by micro segmental factors such as presence of obstruents and less sonorant segments. For example, there is a dip at the left edge of the second and the third syllable of Word2 [*dai*]. The origin of this F0 dip is the voiced obstruent [*d*] and no L tone target is present at the dip.



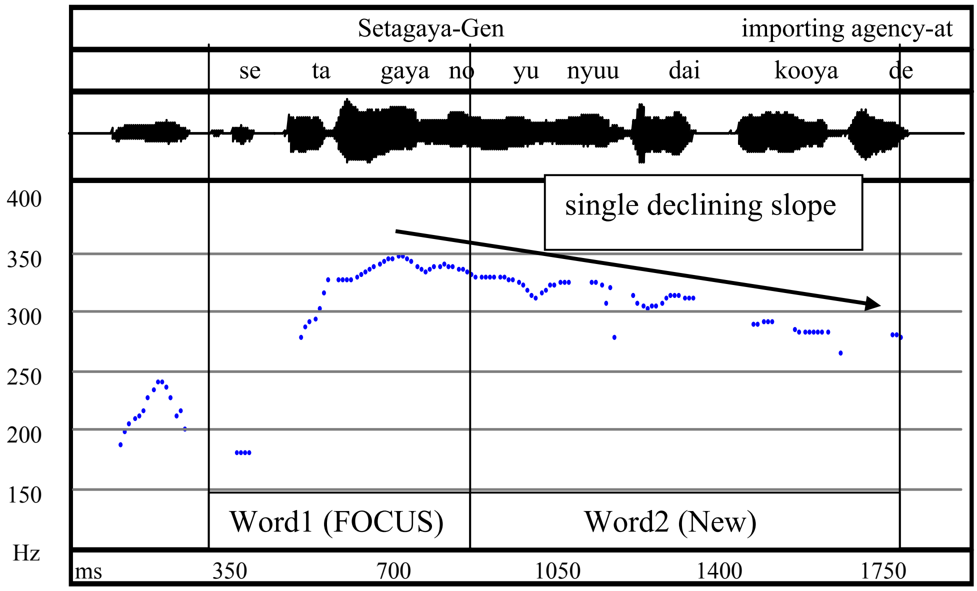
a. Speaker RO

Figure 7.05. Example F0 Contours of UUFN

Continued next page

Figure 7.05 continued

SK 4-70-4



b. Speaker SK

From this, I conclude that in RO and SK's speech FOCUS had an effect of deleting an immediately following Minor Phrase boundary regardless of the information status (new or given) of the immediately following word.

RO & SK

Control Cases: [[*New*]$_{\text{Word1}}$ [*New/Given*]$_{\text{Word2}}$]$_{\text{XP}}$
( )$_{\text{MiP}}$ ( )$_{\text{MiP}}$
LH LH

FOCUS Cases: [[*FOC*]$_{\text{Word1}}$ [*New/Given*]$_{\text{Word2}}$]$_{\text{XP}}$
( )$_{\text{MiP}}$
LH
MiP *Dephrasing*

This result is consistent with the structural view derived from the FOCUS Prominence theory of Truckenbrodt (See Chapter 6 for more detailed discussions on the FOCUS Prominence Theory). According to the FOCUS prominence theory, coalescence of the FOCUS Word1 and the post-FOCUS unaccented Word2 into a single Minor Phrase is accounted for by FOCUS-ΔIP and ALIGN$_R$(Δ$_{IP}$, IP) outranking the morphosyntax-phonology interface constraint ALIGN$_L$(X^0-Branching, MiP). The definition of those FOCUS-related constraints is given in Section 6.2.2. of Chapter 6, and the definition of ALIGN$_L$(X^0-Branching, MiP) is provided in (4), Section 7.1 of this chapter. ALIGN$_L$(X^0-Branching, MiP) requires the left edge of a compound word *yunyuu-daikooya* "importing agency" to coincide with a Minor Phrase boundary. However, the combination of the outranking constraints, FOCUS-ΔIP and ALIGN$_R$(Δ$_{IP}$, IP), requires no Minor Phrase constituents after FOCUS. As a result, the output candidate with no Minor Phrase boundary between the FOCUS Word1 and the post-FOCUS Word2 is the optimal one. This is shown in the following tableau.

(7) FOCUS-ΔIP, ALIGN$_R$(Δ$_{IP}$, IP) >> ALIGN$_L$(X^0-Branching, MiP)

Tableau 7.01.

	$_{XP}$[[Word1$_{FOC}$]$_{XP}$ Word2$_{Branching}$]$_{XP}$	FOCUS-ΔIP	ALIGN$_R$ (Δ_{IP}, IP)	ALIGN$_L$ ($X^0_{Branching}$,MiP)
a.	$_{IP}$( ) $_{MaP}$( ) $_{MaP}$( ) $_{MiP}$( ) $_{MiP}$( ) Wd1 Wd2 Δ_{IP}		*! (1MaP) * (1MiP)	
b.	$_{IP}$($_{MaP}$( ) $_{MiP}$( ) $_{MiP}$( ) Wd1 Wd2 Δ_{IP}		*! (1 MiP)	
c. ❍	$_{IP}$($_{MaP}$( ) $_{MiP}$( ) Wd1 Wd2 Δ_{IP}			*

Heads of prosodic constituents are underlined.

<Speaker MR>

Speaker MR is considered next. MR is different from the other two speakers especially in the F0 rise of the FN case. It is 10 Hz and its 95% confidence interval is (2.33, 17.94), which does not include zero. From this, it is concluded that MR has a significant F0 rise from the first to the second syllable of Word2, i.e. a L and a H edge tone that marks a Minor Phrase boundary. An example the F0 contour representing those five tokens with a small initial rise is shown in Figure 7.06.

MR 6-70-3

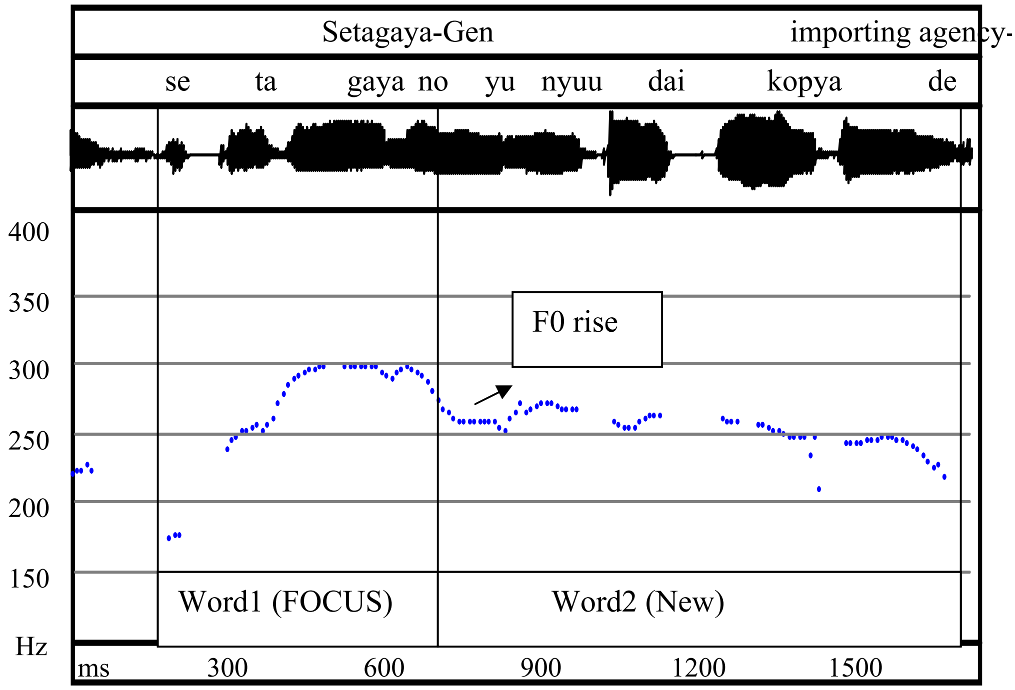


Figure 7.06. An Example F0 Contour of UUFN (Speaker MR)

When it comes to the F0 rise at the onset of a given Word2 of the FG case, its mean is 3 Hz and its 95% confidence interval includes zero: (-2.48, 7.84). Given this, the mean of the F0 rise is not significant for the FG case. Nonetheless, eleven of her thirteen tokens had an F0 discontinuity between Word1 and Word2 and a tonal target was present at the onset of Word2 like that in Figure 7.07.

MR 6-67-3

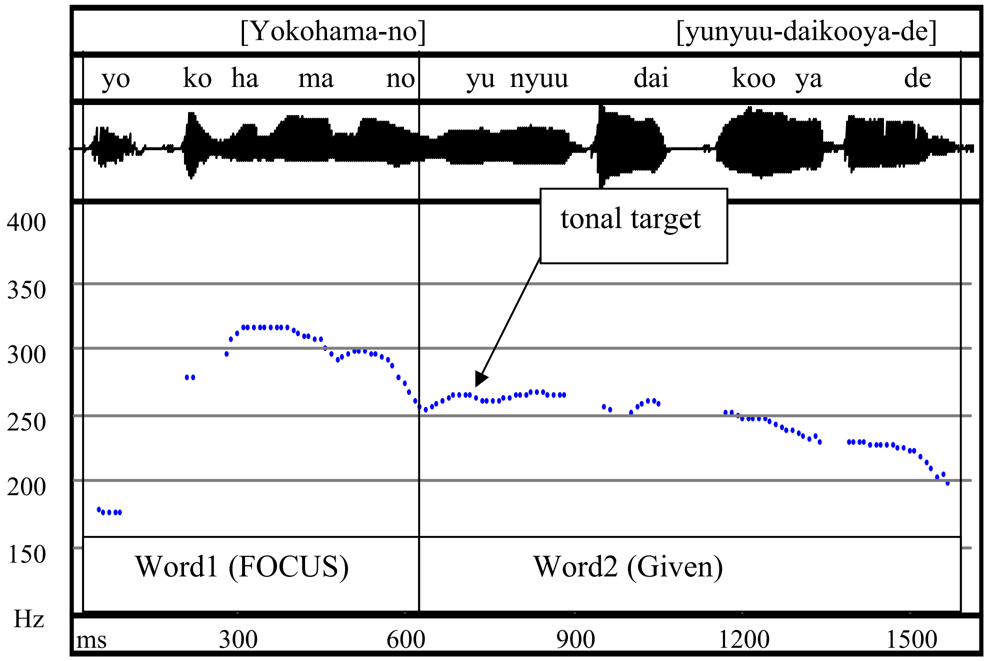


Figure 7.07. An Example F0 Contour of UUFG (Speaker MR)

One way to account for this is with a Minor Phrase boundary at the onset of Word2 even in the FG case, i.e. there are LH edge tones associated with the first and the second syllable of Word2 respectively, and those two edge tones bear the same F0 value. I conclude that MR has a Minor Phrase boundary even at the onset of the post-FOCUS given Word2 of the FG case.

In summary, both the FN and the FG case in MR's speech retained a Minor Phrase boundary at the onset of the post-FOCUS Word2 regardless of their information status. In other words, there was no post-FOCUS MiP dephrasing.

Speaker MR

Control Cases: $[[\text{New}]_{\text{Word1}}\ [\text{New/Given}]_{\text{Word2}}\ ]_{\text{XP}}$
$(\quad)_{\text{MiP}}\ (\quad\quad)_{\text{MiP}}$

$\Updownarrow$

FOCUS Cases: $[[\text{FOC}]_{\text{Word1}}\ [\text{New/Given}]_{\text{Word2}}\ ]_{\text{XP}}$
$(\quad)_{\text{MiP}}\ (\quad\quad)_{\text{MiP}}$
$\uparrow$
No *Dephrasing*

The lack of MiP dephrasing in MR's speech is still consistent with Truckenbrodt's FOCUS Prominence theory from which the structural view is derived. According to the theory, lack of MiP dephrasing is because the constraint that induces dephrasing (i.e. ALIGN$_R$(Δ_{IP}, IP)) is outranked by the conflicting constraint, ALIGN$_L$(X^0-Branching, MiP). The ranking is shown in (8) and Tableau 7.02.

(8) ALIGN$_L$(X^0-Branching, MiP) >> ALIGN$_R$(Δ_{IP}, IP)

This ranking is different from RO and SK's constraint ranking in which ALIGN$_R$(Δ_{IP}, IP) outranks ALIGN$_L$(X^0-Branching, MiP). That is, the ranking between those two constraints varies from speaker to speaker in Tokyo Japanese.

Tableau 7.02.

$_{XP}$[[Word1$_{FOC}$]$_{XP}$ Word2$_{Branching}$]$_{XP}$	FOCUS- ΔIP	ALIGN$_L$ ($X^0_{Branching}$,MiP)	ALIGN$_R$ (Δ_{IP}, IP)
a. $_{IP}$($_{MaP}$(________) $_{MaP}$( ) $_{MiP}$(________) $_{MiP}$( ) Wd1 Wd2 Δ_{IP}			* (1MaPs) *! (1MiPs)
b. ❍ $_{IP}$($_{MaP}$(____________________) $_{MiP}$(________) $_{MiP}$( ) Wd1 Wd2 Δ_{IP}			* (1 MiPs)
c. $_{IP}$($_{MaP}$(____________________) $_{MiP}$(____________________) Wd1 Wd2 Δ_{IP}		*!	

Heads of prosodic constituents are underlined.

7.2.3. Summary of Section 7.2

The results obtained from sequences of unaccented words in Section 7.2 are consistent with the structural view. RO and SK had Minor Phrase dephrasing after FOCUS. Though MR did not delete the post-FOCUS Minor Phrase boundary, it is still consistent with the FOCUS Prominence theory on which the structural view is based. In the following section (Section 7.3), we investigate whether there is any interaction between the constraints that induces the post-FOCUS Minor Phrase dephrasing and the presence of an accent on the FOCUS word and the post-FOCUS word.

7.3. The Accent and Lack of Post-FOCUS MiP Dephrasing

In this section, I ask whether the presence of an accent on the FOCUS word and the post-FOCUS word interacts with the constraint that induces dephrasing, ALIGN$_R$(Δ_{IP}, IP). The sequence of accented words used in this investigation was already introduced in (2), Section 7.1, but is again shown below.

(2) AA (from the <*Yunyuu*> Set)

[[*Aóyama-no*]$_{Word1}$ [*yunyuu-dairíten-de*]$_{Word2}$]$_{XP}$
Aoyama-Gen importing-agency-At
"At an importing agency of Aoyama"

Again, the control cases (the NN and the GG cases) and the FOCUS cases (the FN and the FG cases) were compared.

[[*Aóyama-no*]$_{Word1}$ [*yunyuu-dairíten-de*]$_{Word2}$]$_{XP}$

New	New	→	AANN (Control)
Given	Given	→	AAGG (Control)
FOCUS	New	→	AAFN
FOCUS	Given	→	AAFG

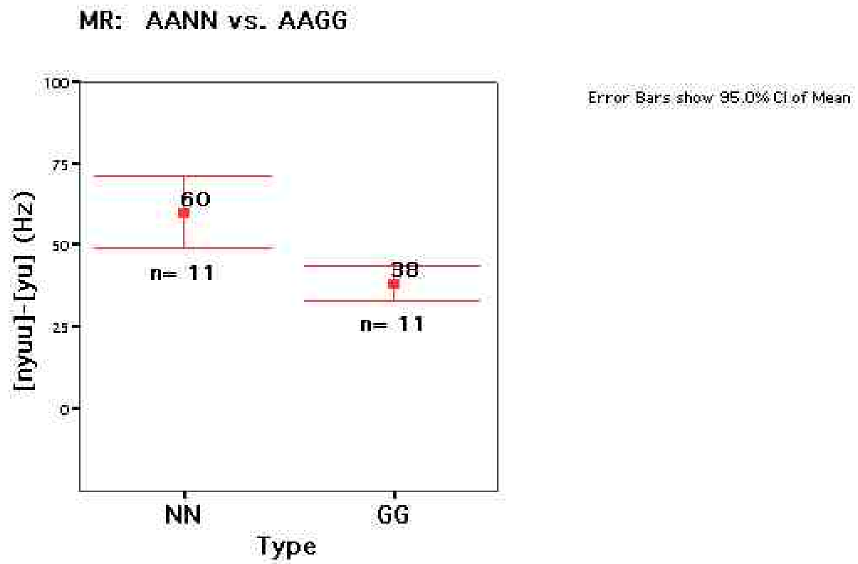
We expect a contrast between the control cases (AANN/AAGG) and the FOCUS cases (AAFN/AAFG) in RO and SK's speech unless some further constraint is at work. As already discussed in Section 7.2, RO and SK are the post-FOCUS MiP dephrasing speakers. In the grammar of those two speakers, the constraint that induces post-FOCUS MiP dephrasing, ALIGN$_R$(Δ_{IP}, IP), outranks the syntax-phonology interface constraint, ALIGN$_L$(X^0-Branching, MiP), which requires the left edge of the post-

FOCUS Word2 to coincide with a Minor Phrase boundary. Consequently, unless some further constraint is at work, we expect in those speakers' speech a Minor Phrase boundary to be present at the left edge of Word2 of the two control cases (AANN/AAGG) which is absent at the left edge of Word2 of the FOCUS cases (AAFN/AAFG).

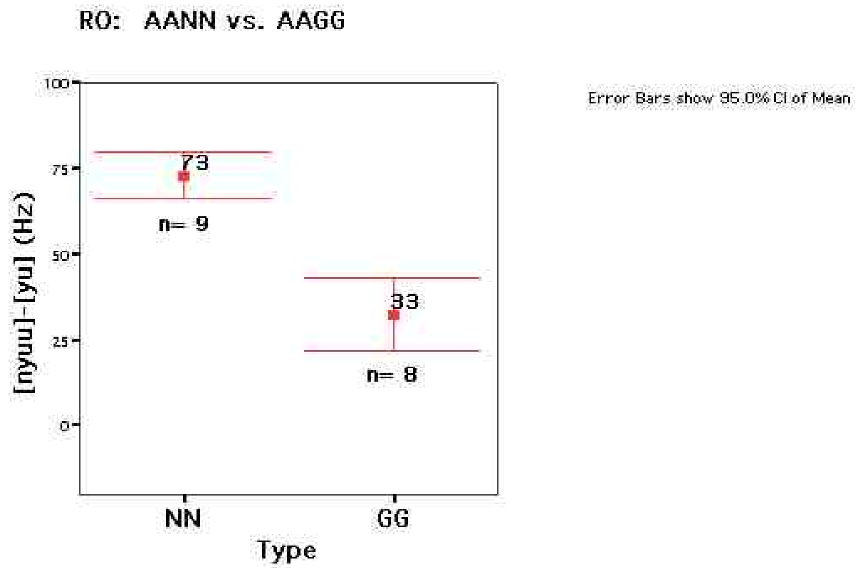
Again, the F0 difference between the first syllable (*yu*) and the second syllable (*nyuu*) of Word2 was measured. If the difference (i.e. the F0 rise from the first to the second syllable of Word2) is significant, a Minor Phrase boundary is present at the onset of Word2. Results of the control cases are presented in Section 7.3.1, and those of the FOCUS cases are presented in Section 7.3.2.

7.3.1. The Control Cases: AANN & AAGG

Just as unaccented control cases, accented control cases also have a Minor Phrase boundary at the left edge of Word2. This is evident from Figure 7.08. Plots in that figure show the mean F0 rise from the first to the second syllable. Those F0 mean were all significantly greater than zero: their 95% confidence intervals did not include zero.



a. Speaker MR

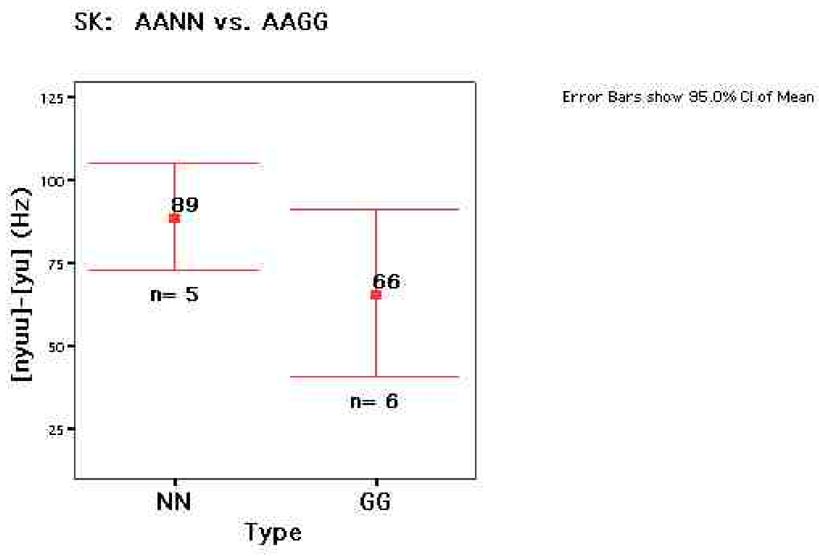


b. Speaker RO

Figure 7.08. The Mean F0 Difference between The First and the Second Syllable of Word2 of the Control Cases (AA)

Continued next page

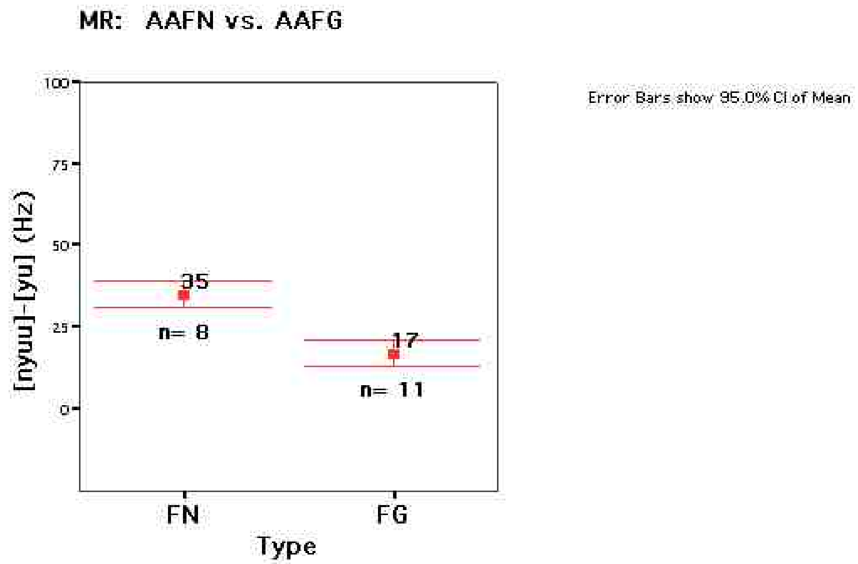
Figure 7.08 continued



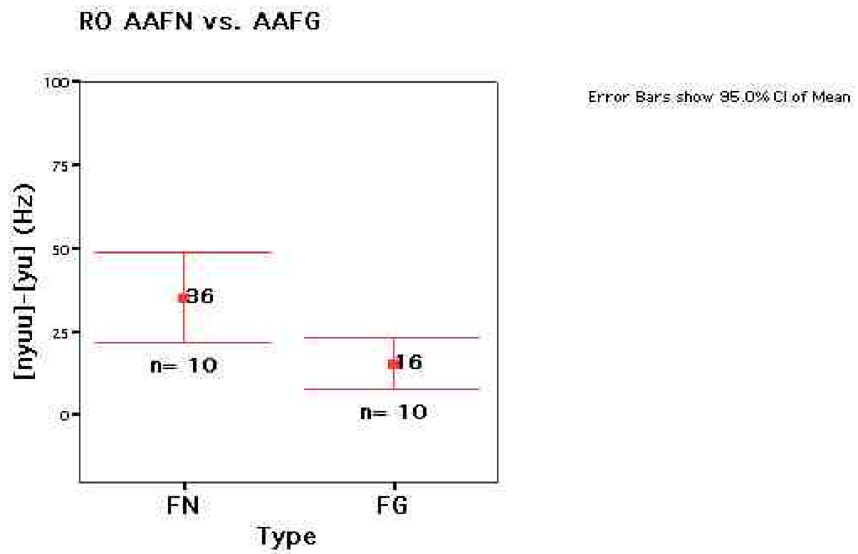
c. Speaker SK

7.3.2. The FOCUS Cases: AAFN & AAFG

The question, then, is whether this Minor Phrase boundary at the left edge of Word2 is retained even when Word1 is made into a FOCUS. According to the phonological analyses provided in Section 7.2, we expect the boundary to be absent in RO and SK's speech but to be present in MR's speech, unless some additional constraint calling for a Minor Phrase boundary at the left edge of an accented Word2 outranks the post-FOCUS dephrasing constraint in RO and SK’s grammar. These predictions are tested by examining the presence/absence of an F0 rise at the onset of post-FOCUS Word2 of the FN and the FG case. The plots in Figure 7.09 show the mean of the F0 rise from the first to the second syllable of Word2 of the AAFN and AAFG cases.



a. Speaker MR

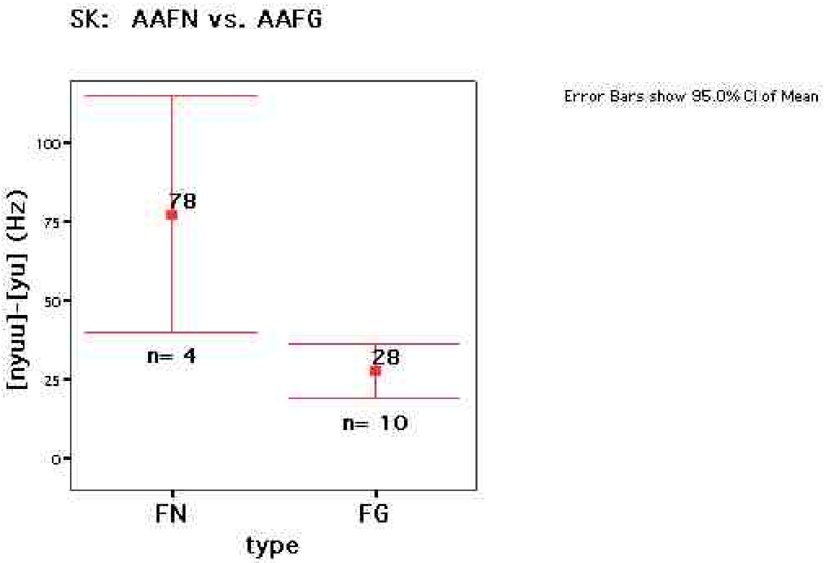


b. Speaker RO

Figure 7.09. The Mean F0 Difference between the First and the Second Syllable of Word2 of the Post-FOCUS Cases (AA)

Continued next page

Figure 7.09 Continued



c. Speaker SK

Interestingly, the mean F0 rise from the first to the second syllable of Word2 was significant not only in MR's speech but also in RO and SK's. According to the plots in Figure 5.09, both the mean F0 rise of the FN case and that of the FG case were greater than zero in all three speakers' speech. In addition, their 95% confidence intervals did not include zero. Because there was a statistically significant rise in F0 from the first to the second syllable of Word2, a Minor Phrase boundary separates it from Word1. There is no post-FOCUS dephrasing when both Word1 (FOCUS) and Word2 are accented.

<Accented Cases: All Three Speakers>

Control Cases: [[New]$_{\text{Word1}}$ [New/Given]$_{\text{Word2}}$]$_{\text{XP}}$
( )$_{\text{MiP}}$ ( )$_{\text{MiP}}$

$\Updownarrow$

FOCUS Cases: [[FOC]$_{\text{Word1}}$ [New/Given]$_{\text{Word2}}$]$_{\text{XP}}$
( )$_{\text{MiP}}$ ( )$_{\text{MiP}}$

$\uparrow$

No *Dephrasing*

7.3.3. The Phonological Analysis

We found no post-FOCUS dephrasing even in RO and SK's speech once both Word1 (FOCUS) and Word2 were accented. This shows that the phonological analysis presented in Section 7.2 is incomplete as an account for those accented cases. The presence of an accent requires the presence of a Minor Phrase boundary at the onset of Word2, while the post-FOCUS dephrasing constraint, i.e. ALIGN$_R$(ΔIP, IP), demands the absence of a Minor Phrase at the left edge of Word2. The demand of an accent for a Minor Phrase boundary, then, outranks the FOCUS's prohibition against it.

I suggest that the demand of an accent for a Minor Phrase boundary comes from a constraint which calls for a surface alignment between the pitch accent H*+L and the DTE (Δ) of a Minor Phrase, and refer to it as ALIGN$_L$(H*+L, Δ_{MiP}).

(9) ALIGN$_L$(H*+L, Δ_{MiP})

A pitch accent H*+L should coincide with the left edge of the DTE (Δ) of a Minor Phrase.

Since each prosodic constituent is allowed to dominate only one DTE, observance of ALIGN$_L$(H*+L, Δ_{MiP}) leads to at most one pitch accent within a Minor

Phrase. In turn, it leads to the demand for a Minor Phrase boundary at the left edge of each accented word. This accent-MiP prominence constraint, then, outranks the constraint which calls for the post-FOCUS dephrasing, i.e. ALIGN$_R$(Δ_{IP}, IP). As a result the most optimal output must separate two accented words into two independent Minor Phrases even in a context where post-FOCUS dephrasing is preferred. This is shown in the following tableau (Tableau 7.03).

(10) ALIGN$_L$ (H*+L, Δ_{MiP}) >> ALIGN$_R$(Δ_{IP}, IP)

Tableau 7.03.

[$_{NP}$[Word1*FOC*] $_N$[Word2]] H*+L H*+L	ALIGN$_L$ (H*+L, Δ_{MiP})	ALIGN$_R$(Δ_{IP}, IP) = Post-FOCUS Dephrasing
a. ❍ $_{IP}$($_{MaP}$( ) $_{MiP}$( ) $_{MiP}$( ) PWd1*FOC* PWd2 $\Delta_{MiP/IP}$ Δ_{MiP} H*+L H*+L		* (MiP)
b. $_{IP}$($_{MaP}$( ) $_{MiP}$( ) PWd1*FOC* PWd2 $\Delta_{MiP/IP}$ Δ_{PWd} H*+L H*+L	*!	

FOC-Δ_{IP} undominated.
Prosodic Heads are underlined.

In Section 1.4 of Chapter 1, I have already laid out the assumptions about the pitch accent H*+L. It is lexically provided and its location is also lexically specified. In a nominal paradigm, the pitch accent surfaces at the lexically specified position to satisfy a Noun-Faithfulness constraint (Smith, 1997). The lexically specified accent position in nouns becomes the DTE (Δ) of a Minor Phrase in the grammatical output representation. This happens because the accent-MiP prominence alignment constraint in (9), ALIGN$_L$(H*+L, Δ_{MiP}) outranks constraints that call for the DTE of prosodic constituents to fall on the default location for prominence (i.e. the penultimate position of the stem). This is shown in the following tableau (Tableau 7.04).

Tableau 7.04.

[*ká ra su*]$_{N\text{-}Stem}$ | H*+L	FAITH LOC-Noun (H*+L)	ALIGN$_L$ (H*+L, Δ_{MiP})	Default Prominence Location
a. ❍ $_{MaP}$( ) $_{MiP}$( ) $_{Foot}$( ) Δ_{MiP} *ká. ra. su.* | H*+L			*
b. $_{MaP}$( ) $_{MiP}$( ) $_{Foot}$( ) Δ_{MiP} *ká. ra. su.* | H*+L		*!	
c. $_{MaP}$( ) $_{MiP}$( ) $_{Foot}$( ) Δ_{MiP} *ka. rá. su.* | H*+L	*!		

7.4. Summary of Section 7.2 and Section 7.3

The predictions made by the structural view of post-FOCUS compression/reduction of F0 movement derived from Truckenbrodt's FOCUS-Prominence theory were confirmed by the post-FOCUS Minor Phrase dephrasing within the sequence of unaccented words (Section 7.2). Though the post-FOCUS dephrasing did not take place when the FOCUS and the post-FOCUS word were made into accented, the lack of dephrasing is still consistent with the FOCUS-Prominence theory. I proposed in Section 7.3 that the absence of dephrasing in an accented context should be explained by a highly ranked constraint which requires the pitch accent to be linked to the DTE of a Minor Phrase, i.e. $\text{ALIGN}_L(\text{H*+L}, \Delta_{\text{MiP}})$. Since this constraint outranks the post-FOCUS dephrasing constraint, i.e. $\text{ALIGN}_R(\Delta_{\text{IP}}, \text{IP})$, a Minor Phrase boundary separates the accented FOCUS and the accented post-FOCUS word.

In the next section, I provide evidence for additional non-structural effects of FOCUS on the following word. It turns out that the structural effect found in Section 7.2 is not the only effect of FOCUS. That is, there are still effects unaccounted for by the deletion of a Minor Phrase boundary.

7.5. The Non-Structural Effect of FOCUS

Though I concluded in the preceding sections that FOCUS has a structural effect on the following word, this conclusion does not entail absence of non-structural post-FOCUS effects. I show in this section that FOCUS has an additional non-structural effect of reducing and compressing the post-FOCUS F0 movement.

7.5.1. Comparisons and Variables

In this section, I introduce comparisons and variables used when investigating the main question of this section, i.e. whether a FOCUS has the non-structural effect of reducing or lowering the F0 movement associated with a post-FOCUS word.

7.5.1.1. Comparisons

The F0 movement of the post-FOCUS Word2 and that of the control Word2 were compared to determine whether there was any *non-structural* effect of the preceding FOCUS on the F0 movement associated with the post-FOCUS Word2.

FOCUS Cases	vs.	Control Cases
FN	-	NN
FG	-	GG

Any difference in F0 movement of Word2?

To carry out the comparison, phonological phrasing factors was required to be excluded because the comparison was about the non-structural effect of FOCUS. Unless both the FOCUS cases and the control cases share the same phonological phrase formation, no genuine non-structural effect may be observed.

Given this restriction, the comparison between the unaccented FOCUS cases and the control cases was not available for two of the three speakers (RO and SK): their FOCUS case underwent MiP dephrasing of the post-FOCUS Word2 while their control case had a Minor Phrase boundary between Word1 and Word2. For MR, however, the comparison between the unaccented FOCUS and the control cases was possible. In MR's speech, both of cases shared the same Minor Phrase formation: Word1 and Word2 formed separate Minor Phrases. As for the comparison between the accented FOCUS and the control cases, the data from all three speakers were available because the FOCUS cases and the control cases of those three speakers shared the same Minor Phrase formation: Word1 and Word2 form separate Minor Phrases.

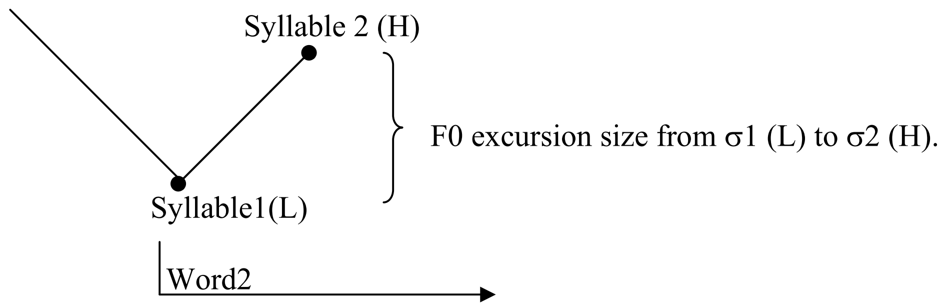
Table 7.03. The Availability of the *Comparison between FOCUS Cases vs. Control Cases*

	Comparison between FOCUS Cases vs. Control Cases	
	AA (Accented Cases)	UU (Unaccented Cases)
MR	Available	Available
RO	Available	
SK	Available	

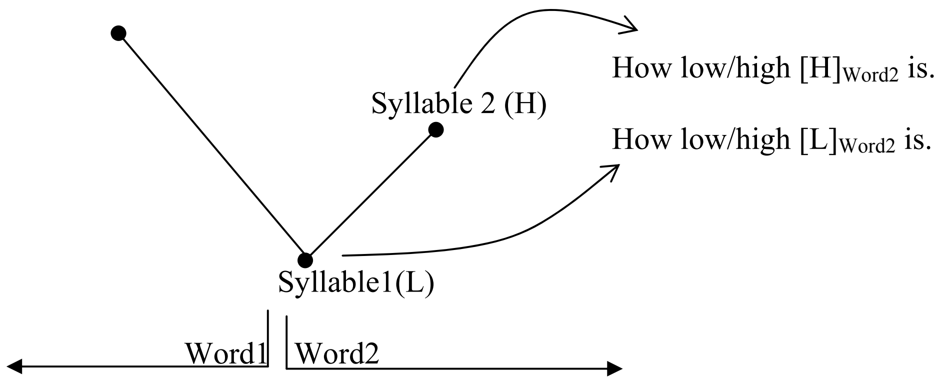
7.5.1.2. Variables

There are two types of variables associated with the F0 movement of the post-FOCUS Word2 that are under investigation. One of them is the F0 excursion size between the first to the second syllable of that post-FOCUS Word2. This measurement represents the magnitude of the initial rise from the L to the H edge tone which marks the onset of a Minor Phrase, and will help us understand whether the presence of a FOCUS reduces post-FOCUS F0 movement even when no post-FOCUS dephrasing takes place. The other type of variable is the F0 values of those L and H edge tones of Word2. These values will help us understand whether any post-FOCUS F0 movement reduction, if it exists, should be captured as a result of pitch range top line lowering as Pierrehumbert & Beckman (1988) argue.

<Variable 1: The Magnitude of the F0 Rise from $\sigma 1$ to $\sigma 2$ of Word2>



<Variable 2: The F0 of L and H Edge Tones of Word2>



If the post-FOCUS F0 movement reduction is a result of pitch range top line lowering, then the post-FOCUS tones are expected to be lower than those tones that are not preceded by FOCUS.

7.5.2. Result 1: The F0 Excursion Size from L to H

First, the F0 excursion size from the first syllable (L edge tone) to the second syllable (H edge tone) of Word2 was compared between the control cases and the FOCUS cases. Figure 7.10 (MR's UU cases) and Figure 7.11 (three speakers' AA cases) present the mean values of such F0 excursion size and their 95% confidence intervals.

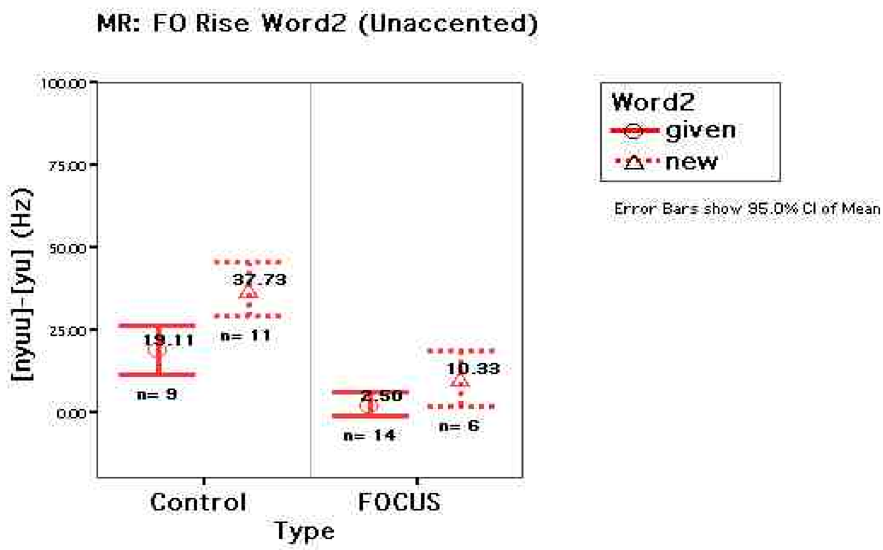
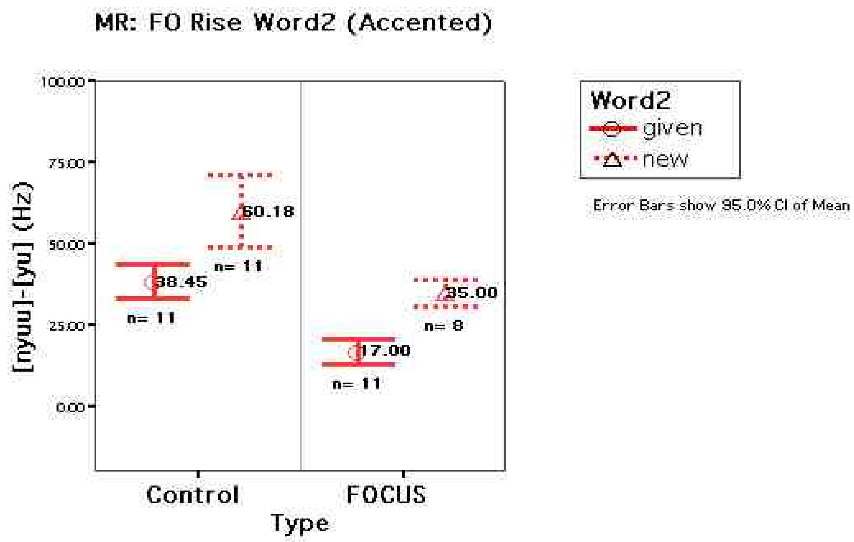
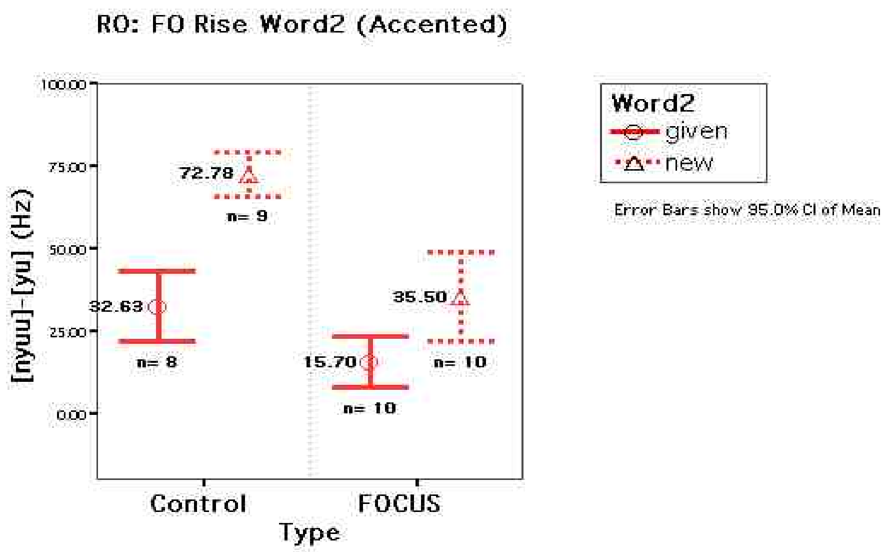


Figure 7.10. The Mean F0 Difference between the First and the Second Syllable of Word2 (Speaker MR, UU)



a. Speaker MR

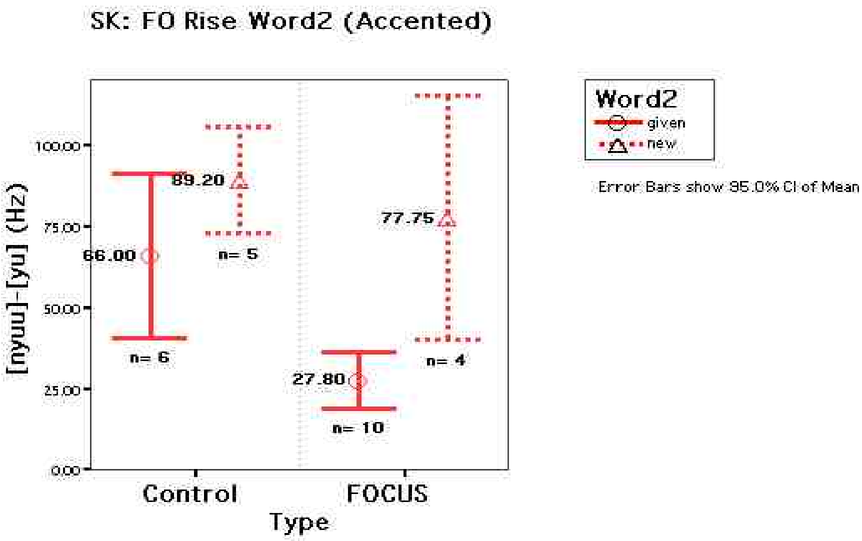


b. Speaker RO

Figure 7.11. The Mean F0 Difference between the First and the Second Syllable of Word2 (AA)

Continued next page

Figure 7.11 continued



c. Speaker SK

What is consistent across all of those figures (Figure 7.10 and 7.11) was that the mean of F0 excursion size from the first to the second syllable of Word2 of the FOCUS cases was smaller than that of the control cases (i.e. the mean of the FN cases was smaller than that of the NN cases and the mean of the FG cases was smaller than that of the GG cases).

For example, the F0 excursion size mean of MR's UUFN case was 10 Hz while that of her UUNN case was 38 Hz (Figure 7.10). In the same way, the F0 excursion size mean of her UUFG case was 3 Hz (2.50 Hz in the figure) while that of her UUGG case was 19 Hz. In addition, the F0 excursion mean differences were significant because their 95% confidence intervals barely overlapped each other. The same pattern was also observed in the accented cases (Figure 7.11-a., 7.11-b., and 7.11-c).

The only exception was SK's AANN, shown in Figure 7.11-c. Though the mean F0 excursion size of her AAFN case (78 Hz) was smaller than that of her AANN case (89 Hz), their 95% confidence intervals substantially overlapped each other. I consider this because of the small number of data points available for this comparison: only four tokens were available for her AAFN case. My conjecture is that this mean difference will be statistically significant once more tokens become available for her AAFN case. Therefore, this exceptional result of SK's does not contradict the generalization that F0 excursion size between the first and the second syllable of Word2 is smaller in the FOCUS cases than in the control cases.

In conclusion, the presence of FOCUS on Word1 diminishes the F0 excursion size between the first syllable (L edge tone) to the second syllable (H edge tone) of

Word2. I want to emphasize that this post-FOCUS reduction of an F0 excursion size is non-structural. The reduction can be only non-structural because both the FOCUS cases and the control cases share the same phonological phrase formation: a Minor Phrase boundary is present between Word1 and Word2.

One may, however, argue that the Word2 edge tone F0 excursion size difference comes from a difference in a further higher order phrase boundary, a Major Phrase boundary. It might still be possible that the control cases have a Major Phrase boundary in addition to a Minor Phrase boundary at the onset of Word2 while the FOCUS cases only have a Minor Phrase boundary there. If such a difference in phrasal organization is really present, then it is likely that this phrasal organization difference would result in a difference in the F0 excursion size of the Word2 edge tones: the presence of a Major Phrase boundary gives greater F0 excursion size to the control cases. This concern, however, is inappropriate because there is enough evidence to believe that the control cases and the FOCUS cases share exactly the same phrasal organizations. Such evidence is presented in the following part of this section.

<MR's Unaccented Control Cases: No Major Phrase between Word1 and Word2>

One argument against the claim that there is a Major Phrase boundary as well as a Minor Phrase boundary at the left edge of Word2 of MR's unaccented control cases comes from the depth of the Word2 L edge tone. According to Pierrehumbert & Beckman (1988), it is the depth of L edge tones that distinguishes the presence and absence of a Major Phrase boundary at the left edge of a word immediately following an unaccented word: the L edge tone that coincides with the left edge of a Major Phrase boundary is far lower than the L edge tone coinciding with the left edge of a Minor Phrase boundary only. Based on this criterion, I compared the depth of the L edge tone of the initial word of the control cases (i.e. Word1 of the NN and the GG cases) which coincides with at least a Major Phrase boundary and that of the second word of those control cases (i.e. Word2 of the NN and the GG cases) whose boundary affiliation is our concern now. Since the initial word (Word1) is in a sentence initial position, it is already taken for granted that at least a Major Phrase boundary appears there.

It turned out that the mean F0 values of the Word1 L edge tone (L1) was 181 Hz for NN and 201 Hz for GG. On the other hand, the mean F0 values of the Word2 L edge tone (L2) were 222 Hz for NN and 251 Hz for GG. The former were 40~50 Hz lower than the latter, and the difference was statistically significant: the 95% confidence intervals of the L1 mean and those of the L2 mean did not overlap (Figure 7.12).

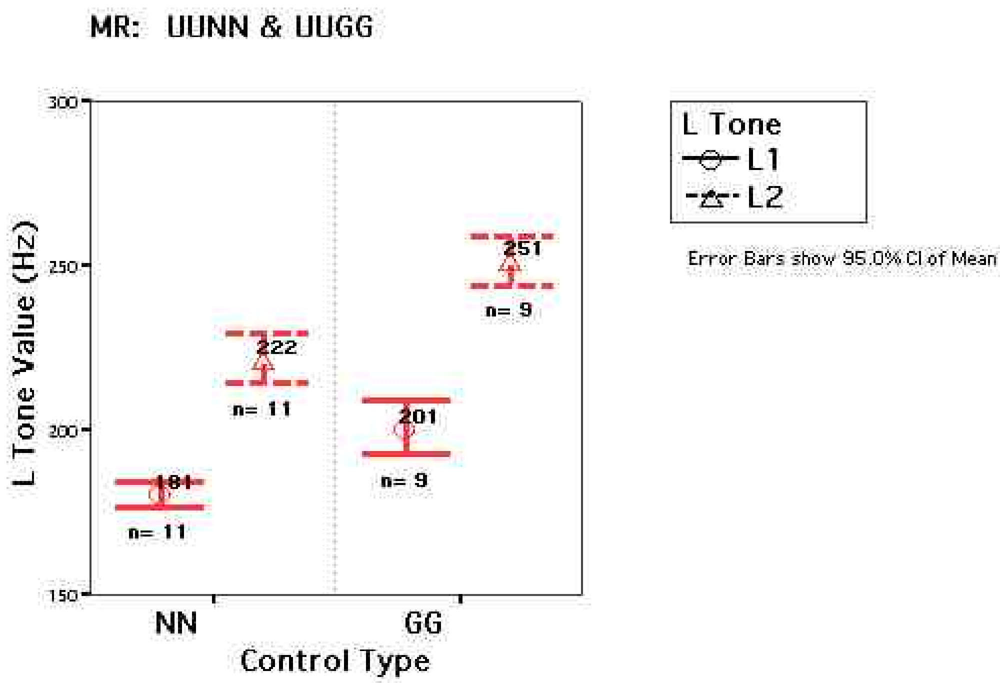


Figure 7.12. The Mean L Value of Word2 of the Control Cases (Speaker MR, UU)

The significant difference between the mean F0 values of those L tones indicates that no Major Phrase boundary is present at the left edge of Word2 in those control cases, as shown schematically in Figure 7.13.

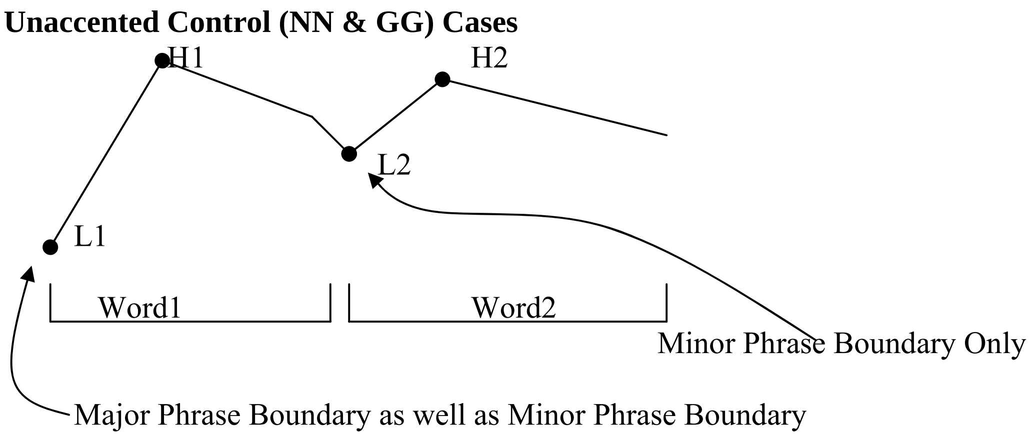


Figure 7.13. The Schematic Representation of the F0 Contour of the UU Control Cases

One may complain that the F0 of L1 should not be used for the comparison here because Word1 to which L1 belongs is not only at the Major Phrase-initial position but also at the Intonational Phrase-initial position. Because Word1 is a *sentence-initial* word, it is true that it coincides with an Intonational Phrase boundary, too. Given this, one may argue that just L2 being higher than L1 does not necessarily mean lack of a Major Phrase boundary. The F0 of the L2 boundary tone should be compared with a F0 of the L boundary tone that marks an Intonational Phrase-medial Major Phrase boundary. According to the data considered in Chapter 8 (Figure 8.04), the L edge tone of a given word that coincides with the left edge of the Intonational Phrase-*medial* Major Phrase boundary was realized at 218 Hz. This is far lower than the F0 of the L2 edge tone of the given word that we are looking at here, which is about 250 Hz (Figure 7.12). Given this, I conclude that no Major Phrase boundary is present at the left edge of the unaccented Word2 of those control cases. This conclusion is summarized in Table 7.04.

Table 7.04. The Presence/Absence of a MaP Boundary at Word2 (UU, Control Cases)

	The Presence of Major Phrase Boundary at Word2	
	UUNN	UUGG
MR	NO	NO

Because no Major Phrase boundary is present at the left edge of Word2 of those unaccented control cases, the post-FOCUS reduction of F0 excursion size between the L and the H edge tones observed in MR's unaccented cases is a non-structural effect of FOCUS.

<Accented Control Cases and Lack of a Major Phrase Boundary>

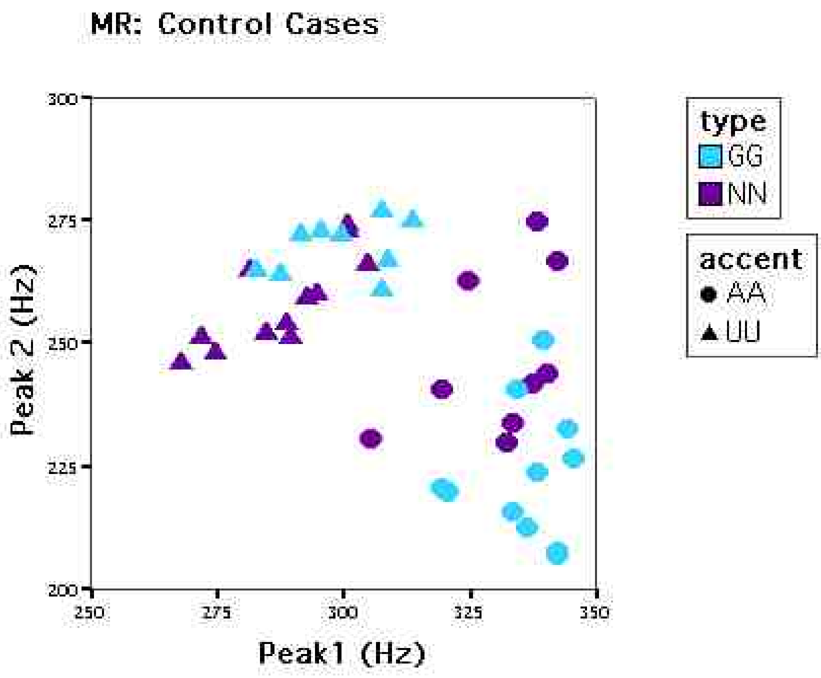
As with the unaccented cases of MR above, it is necessary to test for the presence/absence of a Major Phrase boundary at the left edge of Word2 of the accented control cases. As already introduced in Chapter 4, the diagnosis for the presence/absence of a Major Phrase boundary at the onset of a word is immediately following an accented word is the presence/absence of *catathesis*, i.e. post-accent lowering of H tone values. *Catathesis* is blocked by a Major Phrase boundary and tested by comparing the F0 peak of a word preceded by an accented word and that preceded by an unaccented word. If the former is lower than the latter, then the former is considered to have undergone *catathesis*, which is evidence for the absence of a Major Phrase boundary at its left edge.

To test for the presence of *catathesis* on Word2, the F0 peak of Word2 of the AA cases (sequence of accented words) and that of the UU cases (sequence of unaccented words) were compared.

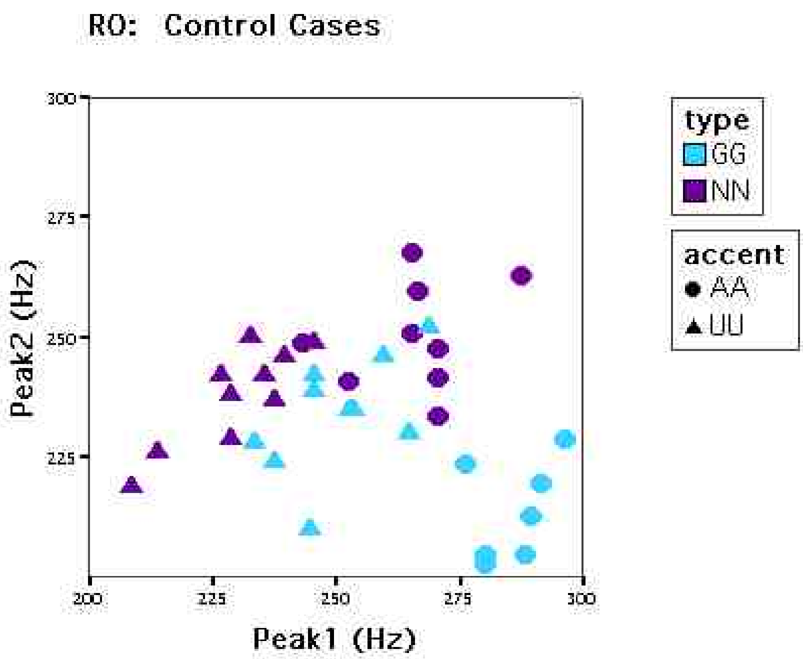
UU:	[Word1 (U)]	[Word2 (U)]	
AA:	[Word1 (A)]	[Word2 (A)]	F0 Peak Compared

Strictly speaking, however, Word2 in the AA cases and that of the UU cases are not comparable to each other because Word2 of the former is accented while Word 2 of the latter is unaccented. The F0 peak of accented words are *intrinsically* higher than that of an unaccented word because the accent H* tone usually bears a higher F0 than the H edge tone (Pierrehumbert & Beckman, 1988; Warner, 1997; my observation in Chapter 4). In unaccented words, the H edge tone is the only H tone and corresponds to its F0 peak. On the other hand, the H* accent tone is present in accented words and it is that H* tone that corresponds to the F0 peak of those accented word. Nonetheless, it is safely concluded that *catathesis* is present if even the *intrinsically higher* F0 peak of Word2 of the AA cases is lower than the *intrinsically lower* F0 peak of Word2 of the UU cases.

The F0 peak of the *unaccented* Word2 of the UU cases corresponds to the second syllable [*nyuu*]. This second syllable is where the H edge tone appears, which in turn corresponds to the F0 peak of the unaccented Word2. On the other hand, I chose the accented syllable [*rí*] for the F0 peak of the accented Word2 of the AA cases. The F0 peaks of Word2 were plotted as the y-axis of the scatter plots shown in Figure 7.14-a~c. The x-axis of those scatter plots corresponds to the F0 peak values of Word1, which I consider to represent the height of a given pitch range. If the data points representing the F0 peak of Word2 of the AA cases are lower than those of the UU given the same value of Word1 F0 peak, Word2 of the AA case has undergone *catathesis*.



a. Speaker MR

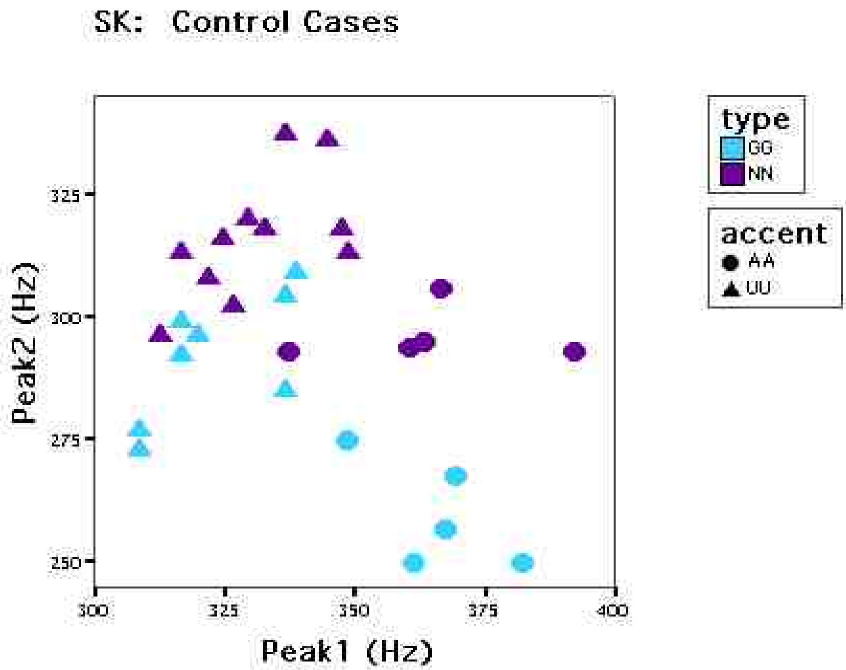


b. Speaker RO

Figure 7.14. The Relationship between Peak 1 and Peak 2 (AA vs. UU, Control Cases)

Continued next page

Figure 7.14 continued



c. Speaker SK

First, let us consider MR's scatter plot in Figure 7.14-a. It is evident from that plot that the datapoints representing the F0 peak of Word2 (i.e. Peak2) of the AAGG case were far lower than those representing Peak2 of the UUGG case. I also conclude the same for the NN cases. Though the F0 range of datapoints representing Peak 2 of the AANN case and that of the UUNN case overlapped each other, the former datapoints were relatively lower than the latter datapoints given the same F0 value of Peak1. From this, I conclude that in MR's speech Word2 of the AA cases undergoes *catathesis* regardless of their information status, given or new, i.e. there is no Major Phrase boundary at the left edge of Word2 of those two accented cases (AANN and AAGG).

Results were parallel in SK's scatter plot (Figure 7.14-c). Datapoints representing Peak2 (F0 peak of Word2) of the AA cases were consistently lower than those representing Peak2 of the UU cases regardless of their information status. Word2 of the AA cases in SK's speech also undergoes catathesis and a Major Phrase boundary is absent at the left edge of Word2 of these AA cases (AANN and AAGG).

The only exception was RO (Figure 7.14-b). Though datapoints representing Peak2 of the AAGG case were far lower than those of the UUGG case, the datapoints representing Peak2 of the AANN case were higher than those of the UUNN case. This was even true when the AA and the UU case share the same Peak1 value. This suggests that Word2 of the AANN case does not undergo *catathesis*, which in turn indicates presence of a Major Phrase boundary at its left edge. This exceptional behavior of RO's AANN case, however, is not a serious problem because Peak2 of her AAGG case still undergoes *catathesis* and lacks a Major Phrase boundary at the left edge of Word2.

Table 7.05. The Presence/Absence of a MaP Boundary at Word2 (AA, Control Cases)

	The Presence of Major Phrase Boundary at Word2	
	AANN	AAGG
MR	No	No
RO	Yes	No
SK	No	No

In summary, putting aside the only exception of RO's AANN case, a Major Phrase boundary is absent at the left edge of Word2 of all the AA control cases (i.e. AANN and AAGG). This confirms that FOCUS has the purely non-structural effect – independent of the phonological phrasing – of reducing the F0 excursion size between the L and the H edge tones of the following word.

The remaining question is whether this non-structural post-FOCUS reduction of F0 excursion size at the onset of Word2 arises from *pitch range lowering* as Pierrehumbert & Beckman (1988) suggest.

7.5.3. Result 2: F0 Values of L and H Tones

To examine whether the non-structural post-FOCUS reduction of F0 excursion size between the L and H edge tones is due to a post-FOCUS *pitch range lowering*, the F0 values of the edge tones of Word2 of the FOCUS cases were compared with those of the control cases.

According to the post-FOCUS pitch range reduction hypothesis of Pierrehumbert & Beckman (1988), post-FOCUS tones are realized in a lower compressed pitch range, which results in lowering of the F0 of those tones. Therefore, those Word2 edge tones of the FOCUS cases are expected to be consistently lower than those of the control cases.

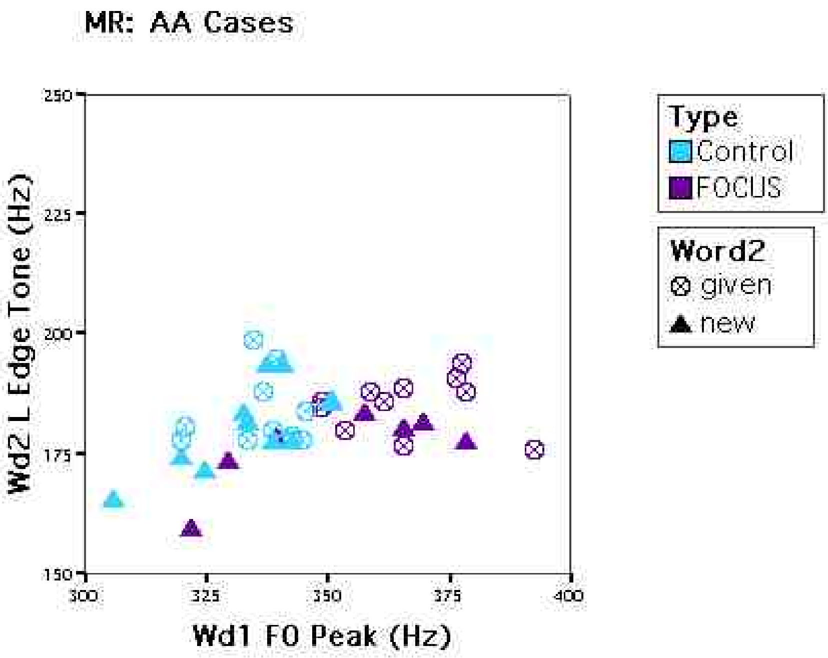
The result of the comparison between the FOCUS cases and the control cases was mixed, and I did not find strong evidence for the post-FOCUS pitch range lowering hypothesis. There was speaker variability and context variability in how to achieve post-FOCUS reduction in F0 excursion size between L and H. For example, some speakers raised the post-FOCUS L edge tone value while keeping H edge tone value intact, and others lowered the H edge tone while keeping the L edge tone value as same as that of the control case. There is no consistent pattern in terms of those edge tone values when achieving the F0 excursion size reduction.

In 7.5.3.1, I first present a comparison between the Word2 L edge tone of the accented control cases and that of the accented FOCUS cases. Comparison of unaccented cases (only MR) is presented in 7.5.3.2.

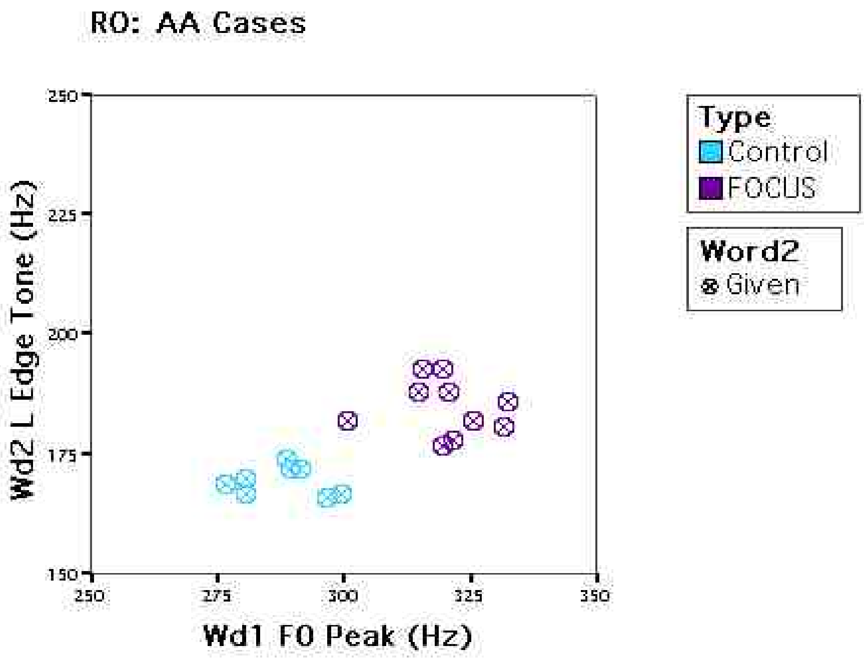
7.5.3.1. The Accented Cases

The main concern here is whether the edge tones of the FOCUS cases are realized lower than those of the control cases. Not only MR but also RO and SK's data were available for this comparison because all three speakers have a Minor Phrase boundary between Word1 and Word2 of the accented cases. However, RO's NN and FN cases were excluded because of the presence of a Major Phrase boundary at the onset of Word2 of the NN case.

Scatter plots were obtained for this comparison, and they are shown in Figure 7.15 and Figure 7.16. The F0 values of Word2 edge tones were plotted on the x-axis and those of Word1 F0 peak were plotted on the y-axis. Figure 7.15 shows variation of F0 values of Word2 L edge tone (L2) and Figure 7.16 shows that of Word2 H edge tone (H2).



a. Speaker MR

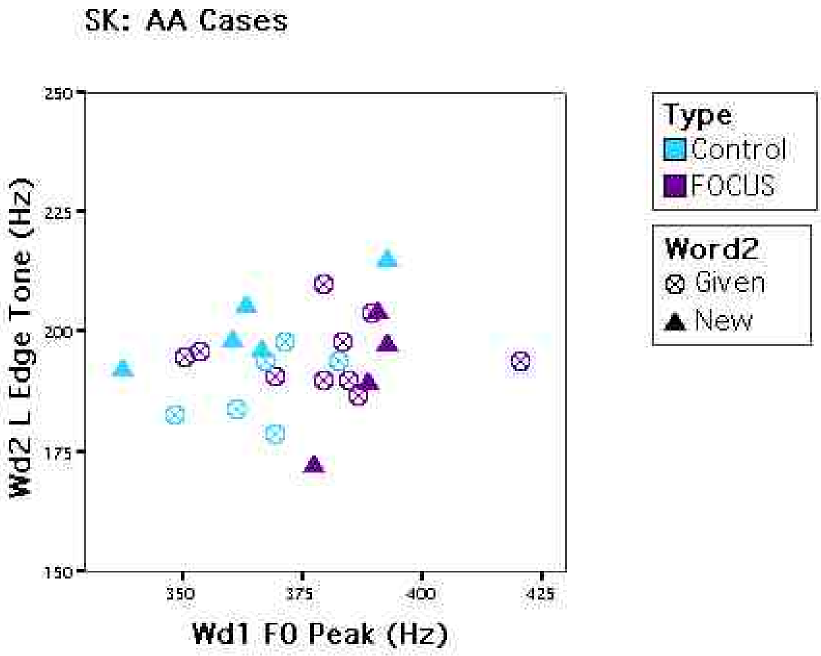


b. Speaker RO

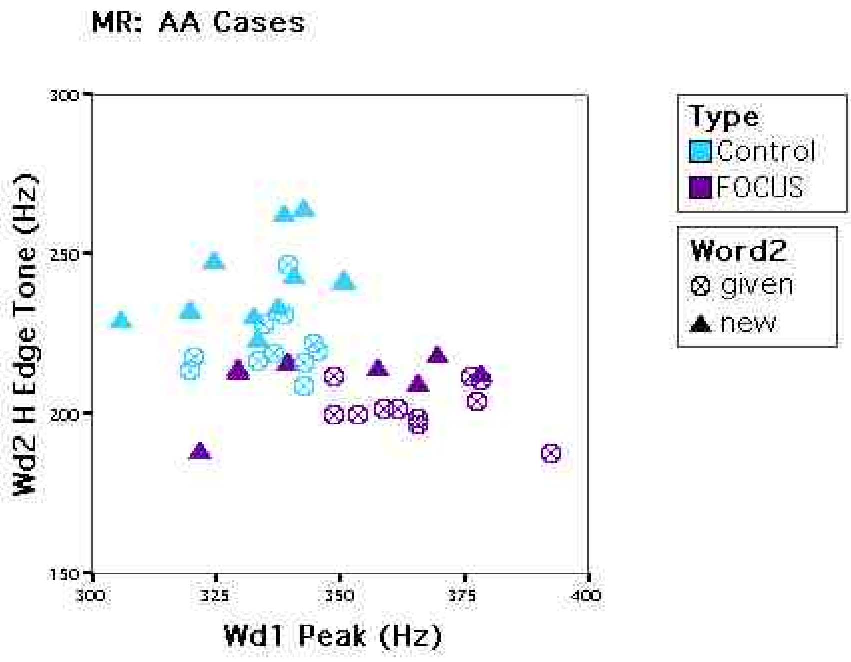
Figure 7.15. The L Edge Tone Values of Word2 (AA, Post-FOCUS vs. Control)

Continued next page

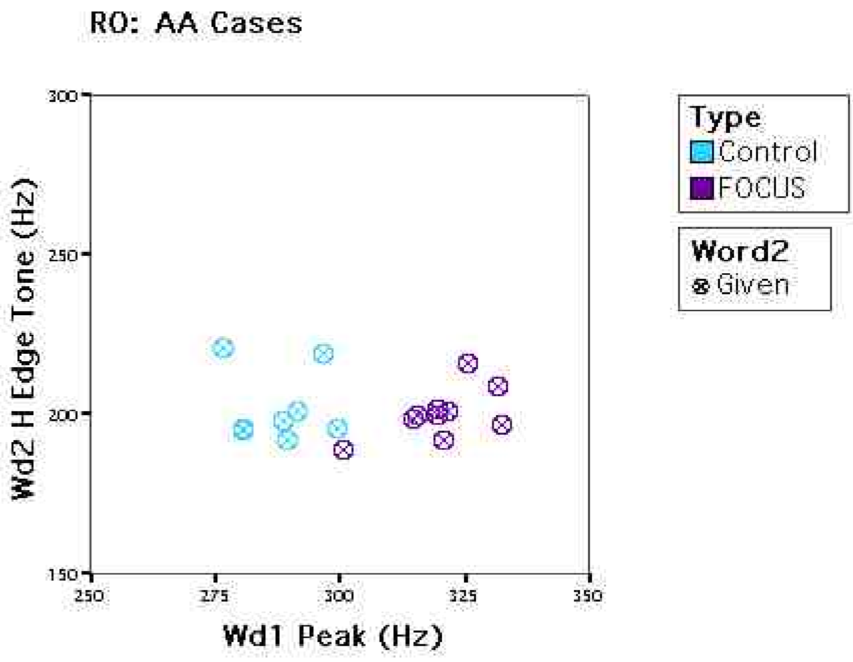
Figure 7.15 continued



c. Speaker SK



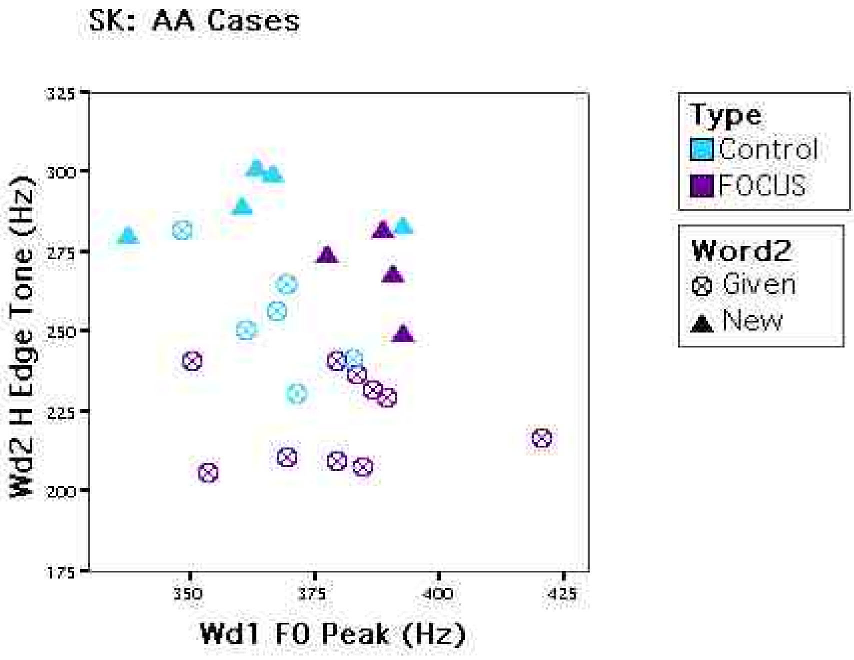
a. Speaker MR



b. Speaker RO.

Figure 7.16. The H Edge Tone Values of Word2 (AA, Post-FOCUS vs. Control)
Continued next page

Figure 7.16 continued



c. Speaker SK

<The Word2 L Edge Tone>

Scatter plots in Figure 7.15 shows F0 values of the L2 edge tone (i.e. the L tone of Word2) of each of the three speakers. The data-points representing the F0 values of the L edge tones of the control cases and those of the FOCUS cases completely overlapped each other irrespective of values of the preceding Word1 F0 peak in MR and SK's figures (Figure 7.15-a, Figure 7.15-c). Only RO (Figure 7.15-b) shows a difference between the control and the FOCUS cases in their Word2 L tone F0 values: the L tone of the FOCUS cases was generally realized higher than that of the control given cases.

<The Word2 H Edge Tone>

The plots in Figure 7.16 show F0 variation of the H2 edge tone (i.e. the H edge tone of Word2). MR (Figure 7.16-a) and SK (Figure 7.16-c) shared the same pattern. In contrast, for RO (Figure 7.16-b), both the data-points of the control case and those of the FOCUS case completely overlapped each other.

In summary, absolute values of post-FOCUS tones were not necessarily lower than those of tones in the control contexts. For MR and SK, the F0 value of the post-FOCUS H2 edge tone was realized lower than that of the H2 edge tone of the control case, while the F0 value of the post-FOCUS L2 edge tone was kept as high as that of the L2 edge tone of the control case. In contrast, for RO the absolute F0 value of the post-FOCUS L2 edge tone was realized higher than that of the control counterpart, while the absolute F0 value of the post-FOCUS H2 edge tone was kept as high as that of the control counterpart. In either way, however, the F0 excursion size from the L2 edge tone to the H2 edge tone gets smaller in the post-FOCUS context than in the control context. This is graphically summarized in Figure 7.17-a.

MR & SK RO

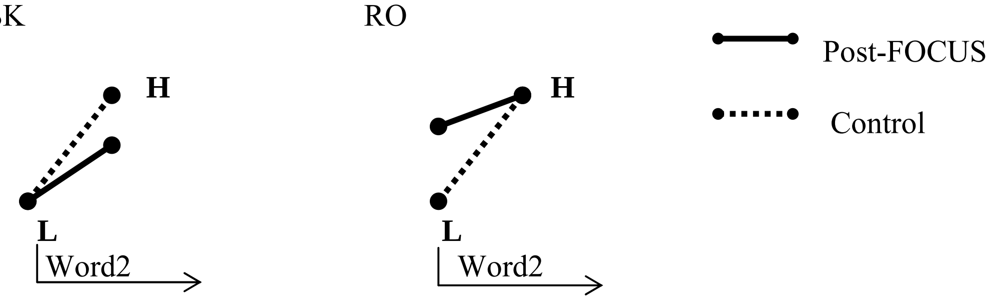


Figure 7.17. The F0 Height of Word2 Edge Tones

In this way, there was speaker variability in terms of which tonal value to be manipulated to achieve the post-FOCUS reduction of F0 excursion size between the

Word2 L and H tones. This result does not support the post-FOCUS pitch range lowering hypothesis. In the same way, no supporting evidence for the post-FOCUS pitch range lowering hypothesis was obtained from the comparison of unaccented cases which is shown in the next subsection.

7.5.3.2. The Unaccented Cases: MR

The Word2 edge tone values of unaccented cases are considered in this section. Since only MR assigned the same phonological phrase representation to the sequence of unaccented Word1 and Word2 of the FOCUS cases and that of the Control cases, only MR's data were considered. The F0 values associated with the post-FOCUS edge tones were compared with those of their control counterparts. The comparison is shown in Figure 7.18 and Figure 7.19. In Figure 7.18, the F0 value of the L2 edge tone (the L edge tone of Word2) is plotted on the y-axis and that of Peak1 (the F0 peak of the preceding Word1) is plotted on the x-axis. In Figure 7.19, the F0 values of the H2 edge tone (the H edge tone of Word2) is plotted on the y-axis.

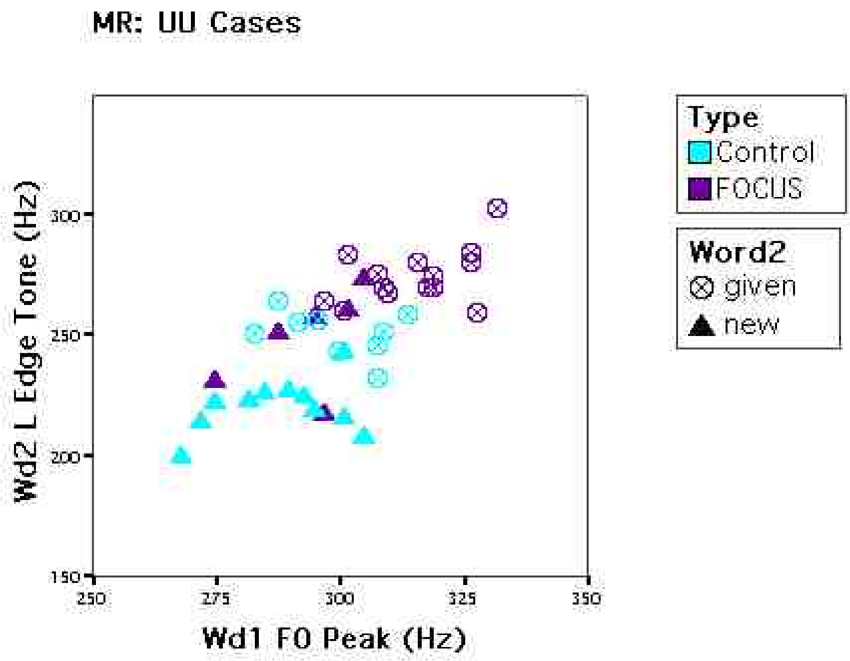


Figure 7.18. The L Edge Tone Values of Word2 (Speaker MR, AA, Post-FOCUS vs. Control)

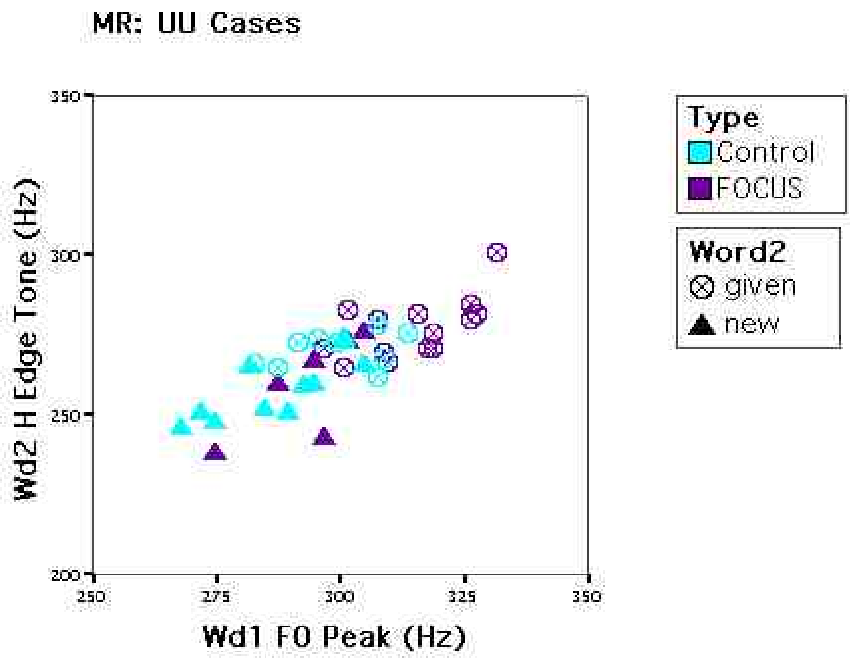


Figure 7.19. The H Edge Tone Values of Word2 (Speaker MR, AA, Post-FOCUS vs. Control)

<L Edge Tone Values>

Most of MR's L2 edge tone data points of the FOCUS shown in Figure 5.18 were realized higher than those of the control cases even when they shared the same F0 peak of the preceding words plotted on the x-axis. This is true for both the given and the new Word2.

<H Edge Tone Values>

When it comes to the H2 tone, most data points of the FOCUS cases overlapped with those of the control cases regardless of the information status of Word2, when the FOCUS and the control cases shared the same Word1 F0 peak (Figure 7.19).

<Summary>

In MR's speech, when both Word1 and Word2 were unaccented, the post-FOCUS L edge tone was realized higher than the L edge tone of the control cases while the post-FOCUS H edge tone and the H edge tone of the control cases shared the same F0. This is graphically shown in Figure 7.20.

MR's Unaccented Cases: L and H edge tones

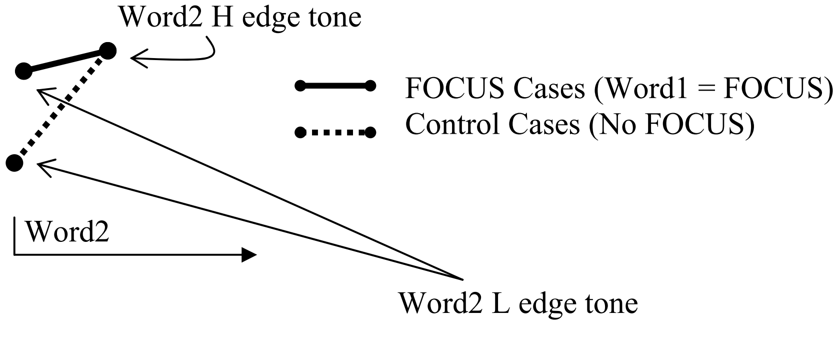


Figure 7.20. The Schematic Representation of L and H F0 Values (Speaker MR)

In summary, FOCUS had no effect of lowering the pitch range of the following Word2 when both Word1 (FOCUS) and Word2 were unaccented in MR's speech. Rather, this speaker shoved up the F0 of the post-FOCUS L edge tone, which ultimately resulted in post-FOCUS reduction of the F0 excursion size between the L and the H edge tones of Word2.

7.5.4. Conclusions of Section 7.5

I showed in this section that FOCUS had the effect of reducing the F0 excursion size between the L and H edge tones of an immediately following word. This effect is non-structural, i.e. not induced by phonological dephrasing. Such reduction takes place even when control cases and FOCUS cases share exactly the same phonological phrase formation.

Also, I asked whether this post-FOCUS F0 excursion size reduction was due to a *pitch range lowering* as Pierrehumbert & Beckman's (1988) suggested. Results obtained in this section provided little evidence for post-FOCUS pitch range lowering. There was speaker variability and context variability in how the post-FOCUS F0 excursion size reduction was achieved, which does not necessarily follow the prediction made by the post-FOCUS pitch range lowering hypothesis. The hypothesis predicts those edge tones should be realized in a lower compressed pitch range, which obviously should lower those tones. However, RO raised the L edge tone to achieve the reduction in the accented context while others lowered the H edge tone for the same purpose. In the same way, MR raised the L edge tone value to achieve the reduction in the unaccented context. From this, I conclude that the non-structural effect of FOCUS is simply to reduce the F0 excursion size between the L and H edge tones of a post-FOCUS word, and it is up to speakers how to achieve the reduction. I leave the formalization of the post-FOCUS reduction of F0 excursion size between tones for the future study.

7.6. Additional Findings: The Non-Structural Effect of Given vs. New

In Section 7.5, I found that both given and new post-FOCUS words underwent reduction of F0 excursion size between L and H edge tones. As a byproduct of this finding, we also found that the distinction between givenness and newness plays a role in determining the F0 excursion size between L and H edge tones in both the FOCUS cases and the control cases.[23] The F0 excursion size between L and H edge tones of a new word was constantly greater than that of a given word. This result is presented in Figure 7.10 and Figure 7.11. It indicates that the F0 excursion size between the L and H edge tones that marks a Minor Phrase boundary is determined by multiple factors. One of those factors is whether the Minor Phrase is preceded by a FOCUS or not, and another is whether the second Minor Phrase corresponds to a new or given word.

[23] Sugito (1996) and Venditti (2000) also investigated the effect of the given vs. new distinction on the phonetic realization of tones. Sugito found a tendency for non-downstepped F0 peaks of new items to be slightly higher than that of given items. However, no statistical analysis is provided and it is not clear whether the difference is statistically significant. Venditti (2000) showed that "salient" items in the discourse (i.e. new information) were marked with higher pitch range than already given information. I leave open for future study whether the non-structural effect of the given vs. new distinction obtained in my experiment is comparable to the "pitch range manipulation" suggested by Venditti.

7.7. Chapter Conclusion

One of the contributions of this chapter is that it has revealed both structural and non-structural aspects of FOCUS on the following XP-boundaryless word. FOCUS was responsible for deleting the following Minor Phrase boundary when both the FOCUS word and the following word were unaccented. This post-FOCUS MiP dephrasing is what Truckenbrodt's FOCUS-Prominence Theory predicts. According to the theory, there are two crucial constraints: the FOCUS-PROMINENCE (*FOC*-ΔIP) constraint and the alignment constraint between the DTE of an Intonational Phrase and the right edge of the Intonational Phrase. The former constraint demands that the FOCUS correspond to the terminal string that contains the DTE of an Intonational Phrase, and the latter constraint prefers no phonological phrase boundary between the DTE and the end of the Intonational Phrase. Given those two constraints, Truckenbrodt's theory predicts the post-FOCUS MiP dephrasing. The prediction was confirmed.

However, once the preceding FOCUS word and the following word were made accented, Minor Phrase dephrasing was not observed. I explained the lack of dephrasing in the accented context as a result of constraint interaction between the accent-MiP prominence alignment constraint and the dephrasing constraint. The dephrasing constraint requires the DTE of Intonational Phrase (i.e. FOCUS) to be as close as possible to the right edge of an Intonational Phrase, and the accent-MiP prominence constraint requires each accent to be aligned with the DTE of a Minor Phrase. The demands of these two prominence-related constraints conflict with each other, and respect of the higher-ranked accent-MiP prominence constraint leads to lack of post-FOCUS MiP dephrasing.

Though no structural effect was found in sequences of post-FOCUS accented words, there was still a non-structural effect in the accented sequences. The F0 excursion size between the L and H edge tones of the post-FOCUS word was more compressed than that of its control counterpart. Though this non-structural effect of FOCUS was present, there was little evidence for the *pitch range lowering* hypothesis proposed by Pierrehumbert & Beckman (1988). Those post-FOCUS tones were not necessarily lower than their control counterpart. For example, a speaker may raise the L tone value while keeping the H tone value intact or lower the H tone value while keeping the L tone value intact to achieve the post-FOCUS compression of F0 excursion size between those two tones.

What we did not find in this chapter is any structural effect of the new vs. given distinction on the post-FOCUS word. Both the new and given post-FOCUS words underwent dephrasing in the unaccented condition. Also, both the new and given post-FOCUS words retained their Minor Phrase boundary in the accented condition. However, I show in the following chapter that once the post-FOCUS words are made into XP, i.e. once an XP boundary appears at the left edge of the post-FOCUS words, the new vs. given distinction results in different phonological phrase formations. As already mentioned in Chapter 6, this is what the theory of FOCUS-Prominence developed by Selkirk (2000a, et seq) predicts.

CHAPTER 8

POST-FOCUS EFFECTS ACROSS XP BOUNDARIES

In the previous chapter, we examined effects of FOCUS on the following XP boundaryless word. I found in that chapter that FOCUS had both a structural and non-structural effects, i.e. the effect of deleting the following Minor Phrase boundary (the structural effect) and the effect of compressing the F0 excursion size between two edge tones that follow it (the non-structural effect). The main goal of this chapter is to investigate whether those two types of post-FOCUS effects are present even across XP boundaries.

Three questions are asked in this chapter related to the structural effect of FOCUS on the following word. One of those questions is whether there is any deletion of a Major Phrase boundary after FOCUS across an XP boundary. According to Selkirk & Tateishi (1988, 1990), a Major Phonological Phrase boundary appears at the left edge of an XP boundary. As already mentioned in Section 1.3.3. of Chapter 1, they suggested a syntax-phonology alignment constraint which demands correspondence between every XP left edge and a Major Phrase boundary, which is called ALIGN$_L$ (XP, MaP). Their study, however, is based on their observation of word sequences without FOCUS. Nagahara (1994) claimed out that those Major Phrase boundaries at the left edge of XPs disappeared once those XPs were preceded by a FOCUS. This post-FOCUS deletion of the (Major) Phonological Phrase boundary is what Truckenbrodt's (1995) and Selkirk's (2000a, 2002ab) FOCUS-Prominence theory predicts (Section 4.2.2, Chapter 4). However, Nagahara presented little empirical evidence to support his claim. Given this, it is still worthwhile asking whether the XP boundary-Major Phrase boundary correspondence is retained even in post-FOCUS part of an utterance. Lack of correspondence between those boundaries will be interpreted as the XP-MaP alignment constraint outranked by the constraints inducing post-FOCUS dephrasing, the FOCUS-related constraints proposed by Truckenbrodt (1995) and Selkirk (2000a, 2002b) introduced in Section 6.2.2. of Chapter 6. This question is investigated in Section 8.1.3.

Another question is about the effect of the information status of the post-FOCUS word with an XP boundary. Selkirk (2000a, 2000b) suggested a syntax-phonology interface constraint which demands correspondence between an XP interpreted as presentational focus (new) and the DTE of a Major Phrase, which is referred to as focus-ΔMaP. I showed in Section 6.2.3. of Chapter 6 that if observance of this focus-ΔMaP constraint is more important than observance of the post-FOCUS dephrasing constraint suggested by Truckenbrodt, then we expect a Major Phonological boundary to be present at the left edge of a post-FOCUS new (focus) XP even when MaP dephrasing takes place at the left edge of a given XP. Testing for this prediction is one of the goals of this chapter. This question is investigated in Section 8.1.3.

The other question is whether the post-FOCUS Minor Phrase dephrasing, i.e. coalescence of a FOCUS and the immediately following word into a single Minor Phrase, is blocked by an XP boundary coinciding with the left edge of that post-FOCUS word. We know from Chapter 6 that the Minor Phrase boundary at the left edge of an XP-boundaryless post-FOCUS word is deleted in an unaccented context. One of the goals of this chapter is to examine whether the same is true of the post-FOCUS word with an XP boundary. This question is investigated in Section 8.1.2.

Effects of FOCUS on the immediately following XP are considered in Section 8.1 and those on the post-FOCUS XP that are not adjacent to the preceding FOCUS are considered in Section 8.3. Phonological discussions and analyses are presented in Section 8.2.

8.1. An XP Boundary Adjacent to FOCUS

In this section, both the structural and the non-structural effect of FOCUS on the immediately following word with an XP boundary are considered. In Section 8.1.1, sentences and contexts used to investigate those questions are introduced. In Section 8.1.2 and Section 8.1.3, experimental results of both the structural and the non-structural effects are provided. Discussion and phonological analyses of the structural effects are provided in Section 8.2.

8.1.1. The Reading Materials and Contexts

In this subsection, I introduce sentences and contexts used to investigate my questions. Sentences with an XP boundary between a FOCUS (Word1) and the immediately following word (Word2) were used because our main concern is whether FOCUS has any structural and non-structural effects on an immediately following item even across an XP boundary. The description of those forms is also provided in Appendix. Those sentences were produced in four different contexts, i.e. two control contexts (NN and GG) and two FOCUS contexts (FN and FG). In the NN and the GG context, both Word1 and Word2 were interpreted as new and given, respectively. In the

FN and the FG context, Word1 was FOCUS (i.e. contrastive FOCUS) and Word2 was interpreted as new and given, respectively.

The <*Yunyuu*> Set

The set of sentences used in this section (i.e. Sentence UU/XP-Even and Sentence AA/XP-Even) are parallel to the forms used in Chapter 7 (i.e. Sentence UU/XP-odd and Sentence AA/XP-Odd). The sentences used in this chapter and those used in the previous chapter share the same lexical items. The only difference between them is their syntactic structures. While sentences used in the last chapter have XP boundaries at the left edge of odd numbered words (i.e. Word3 and Word5), those used in this chapter (i.e. UU/XP-Even and AA/XP-Even) have XP boundaries at the left edge of even numbered words (i.e. Word2 and Word4).

Table 8.01. The Two Sentence Types

UU/AA/XP-Odd (used in Chapter 5)	**UU/AA/XP-Even (used in this chapter)**
$_{XP}$[W1 W2] $_{XP}$[W3 W4] $_{XP}$[W5 …	$_{XP}$[W1] $_{XP}$[W2 W3] $_{XP}$[W4 W5] …

The actual sentence forms and a syntactic tree representation of UU/XP-Even and AA/XP-Even are shown below.

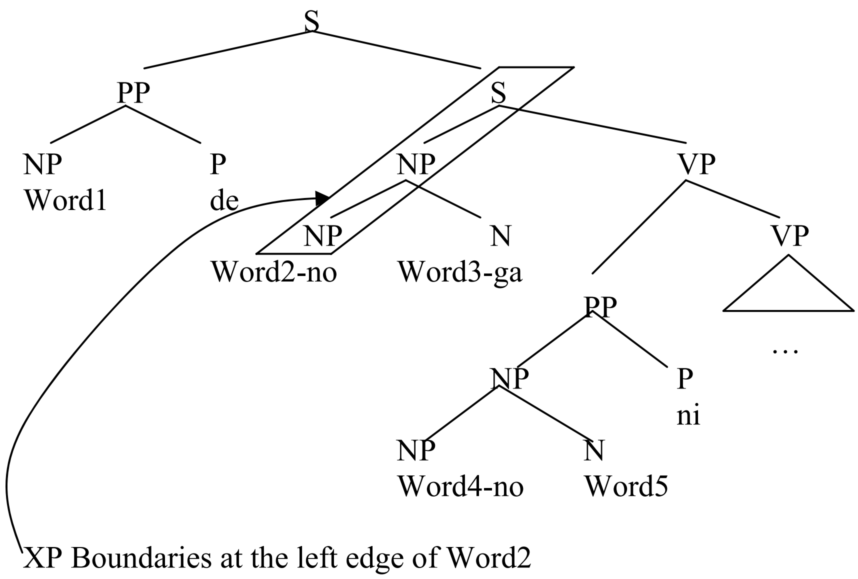
Sentence UU/XP-Even (XP Boundary between Word1 and Word2)

Wor1 Word2 Word3 Word4 Word5
[*Yokohama-de*] $_{S}$[$_{NP}$[*yunyuu-daikooya-no maneejaa-ga*] $_{VP}$[$_{NP}$[*yuujin-no moderu-ni*] …]]]
Yokohama-at importing-agency-Gen manager-Nom friend-Cop model-to …
"In Yokohama the manager of an importing agency (sold marijuana) to a model, his friend."

Sentence AA/XP-Even (XP Boundary between Word1 and Word2)

Word1 Word2 Word3 Word4 Word5 Word6
[*Háyama-de*] $_{S}$[$_{NP}$[*yunyuu-dairíten-no onnamanéejaa-ga*] $_{NP}$[*amerikájin-no uwakiáite-ni*] $_{NP}$[*mayaku-o*]]
Hayama-at importing-agency-Gen female manager-Nom American-Cop lover-to marijuana-Acc …
"In Hayama the female manager of an importing agency (sold) marijuana to an American, her lover."

(1)



As shown in the tree representation above, there are NP boundaries and a sentence boundary at the left edge of Word2. In the following, I show contexts in which those two sentences (UU/XP-Even and AA/XP-Even) were embedded.

Context "FN": Word 1 = FOC, Word 2 = new

Experimenter: *nani?*
"What?"

Speaker: *Tookyoo-ya Oosaka-no yoona hanzai-no tahatsusuru tóshi-dewanaku,*
Tokyo-and Osaka like high crime rate city-Not,

tian-no yóikotode sirareru, ano **[*YOKOHAMA-DE*]$_{\mathbf{Wd1}}$**
safety-Gen good known, that YOKOHAMA-AT,

[*yunyuu-daikooya-no*]$_{\mathbf{Wd2}}$ *maneejaa-ga*
Gen-agency-Gen manager-Nom

yuujin-no moderu-ni mayaku-o uttarasíiyo.
friend-Copula model-to marijuana-Acc sold-I heard.

"I've heard that in YOKOHAMA, the city known to be safe unlike those cities like Osaka or Tokyo, which are notorious for their high crime rate, a manager of an importing agency sold marijuana to a model who is a friend of his/hers."

Context "FG": Word 1 = FOC, Word 2 = given

Experimenter: *Aóyama-de yunyuu-daikooya-no maneejaa-ga yuujin-no moderu-ni marifana-o uttá-to kiitaga, hontookánee?*
"I've heard that in Aoyama a manager of an importing agency sold marijuana to a model who is a friend of his/hers. But is that true?"

Speaker: *chigaimásu. Aóyama-dewa-naku* **[*YOKOHAMA-DE*]$_{\mathbf{Wd1}}$**
"No" Aoyama-Copula-Not YOKOHAMA-AT

[*yunyuu-daikooya-no*]$_{\mathbf{Wd2}}$ *maneejaa-wa*
importing-agency-Gen manager-Topic

yuujin-no moderu-ni mayaku-o uttá-ndesu.
friend-Gen model-to marijuana-Acc sold-Copula.

"In YOKOHAMA but not in Aoyama, a manager of an importing agency sold marijuana to a model who is a friend of his/hers.

Context "NN": Word 1 = new, Word 2 = new

Speaker: *chotto kiiteyo.*
"Hey, just listen to me."

Experimenter: *nani?*
"What?"

Speaker: **[*Yokohama-de*]$_{Wd1}$** **[*yunyuu-daikooya-no*]$_{Wd2}$** *maneejaa-ga*
Yokohama-at importing-agency-Gen manager-Nom

yuujin -no *moderu-ni* *mayaku-o* *uttarashíiyo.*
friend-Cop model-To marijuana-Acc sold-I heard.

"I've heard that in Yokohama the manager of an importing agency sold marijuana to a model who is his/her friend.

Context "GG": Word1 = given, Word2 = given

Experimenter: *Yokohama-de yunyuu-daikooya-no juugyóoin-ga yuujin-no moderu-ni mayaku-o uttarashíi-ga, hontookane?*
"Is it true that in Yokohama the employee of an importing agency sold marijuana to a model who is his/her friend?

Speaker: *iie, chigaimásu.*
"No, that's not correct."

[*Yokohama-de*]$_{Wd1}$ **[*yunyuu-daikooya-no*]$_{Wd2}$** *MANEEJAA-GA*

Yokohama-at importing-agency-Gen manager-Nom

yuujin -no *moderu-ni* *mayaku-o* *uttándesu.*
friend-Cop model-To marijuana-Acc sold-I heard.

"The MANAGER of an importing agency sold marijuana to a model, his/her friend, in Yokohama."

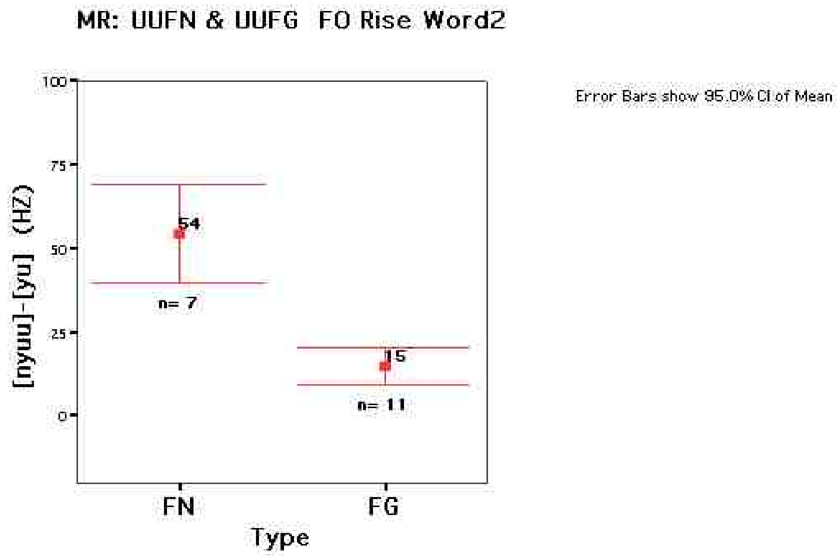
8.1.2. The Lack of MiP Dephrasing

The main question asked in this section is whether post-FOCUS Minor Phrase dephrasing takes place even across a syntactic phrase boundary, i.e. even when there is an XP boundary coinciding with the left edge of Word2. I found the reliable presence of at least a Minor Phrase boundary at the left edge of the post-FOCUS XP regardless of its information status and its accent condition. In other words, we found that the XP boundary blocked Minor Phrase dephrasing.

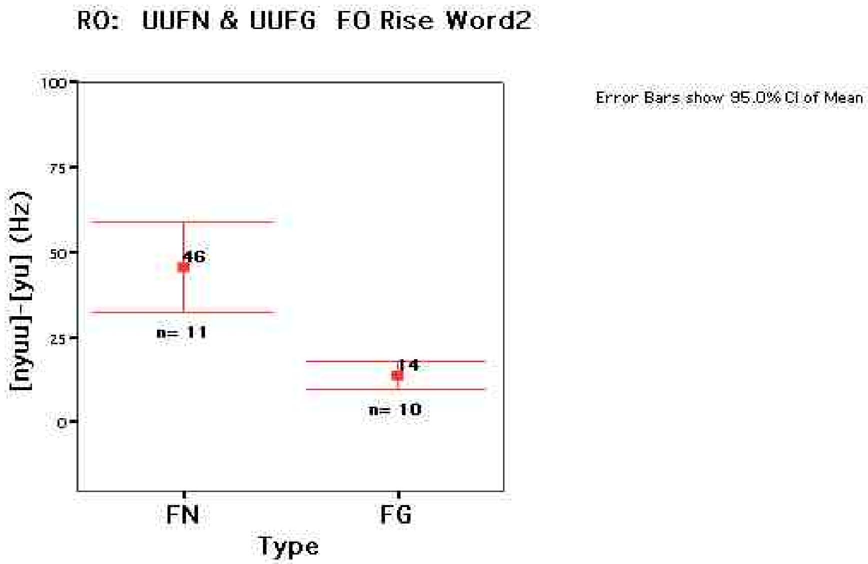
8.1.2.1. Results

F0 excursion size from the first ([yu]) to the second syllable ([nyuu]) of Word2 was obtained for both FN and the FG cases and both accent conditions (UU and AA). As for the sequence of unaccented words, all the data from all three speakers were used. As for the sequence of accented words, I excluded three tokens of the eight AAFN tokens of RO because those did not sound natural enough.

If the F0 rise from the initial syllable to the second syllable is greater than zero, then we conclude that there is at least a Minor Phrase boundary at the left edge of Word2. Error bar plots in Figure 8.01 in Figure 8.02 show mean values and 95% confidence intervals of the F0 excursion size between the first to the second syllable of Word2 of the unaccented cases and those of the accented cases, respectively.



a. Speaker MR

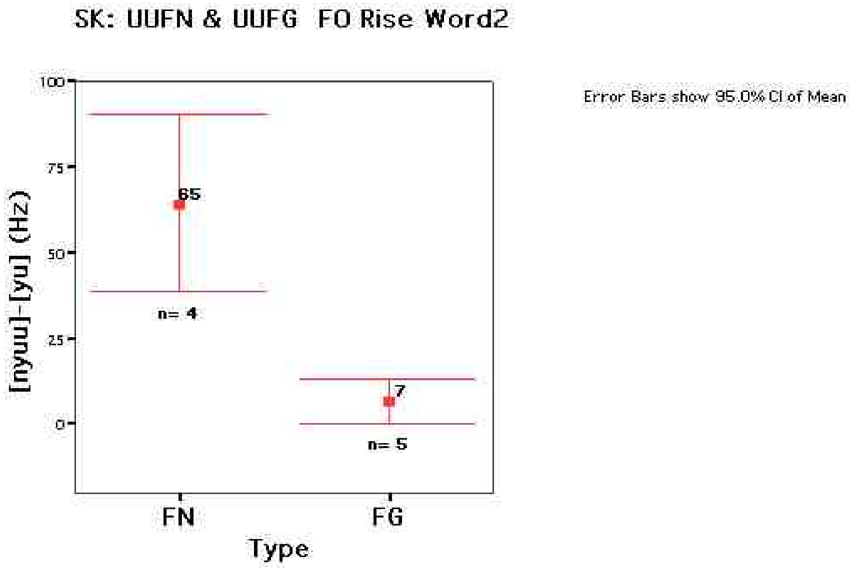


b. Speaker RO

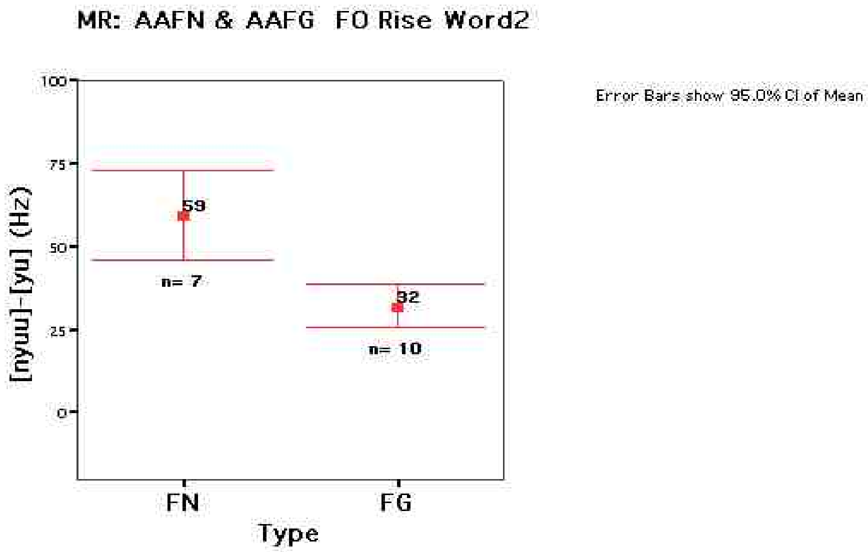
Figure 8.01. The Mean F0 Difference between the First and the Second Syllable of Word2 of the Post-FOCUS Cases (UU)

Continued next page

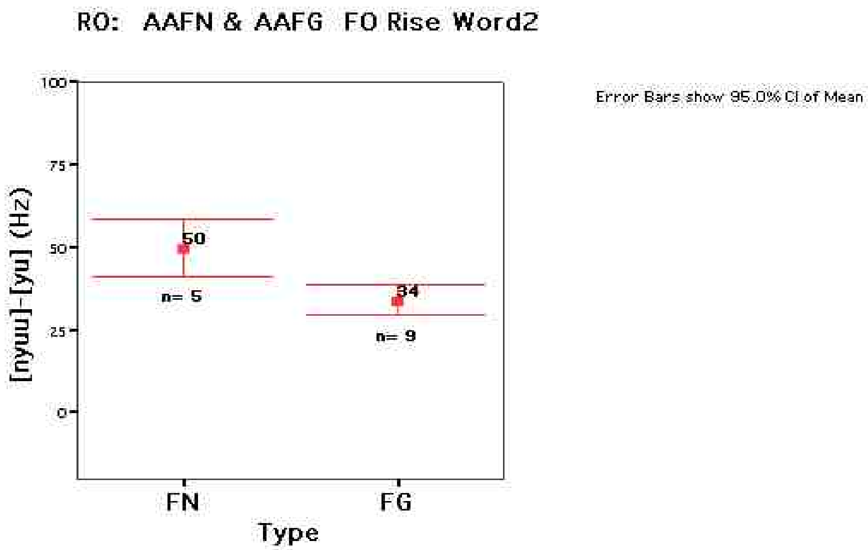
Figure 8.01 continued



c. Speaker SK



a. Speaker MR

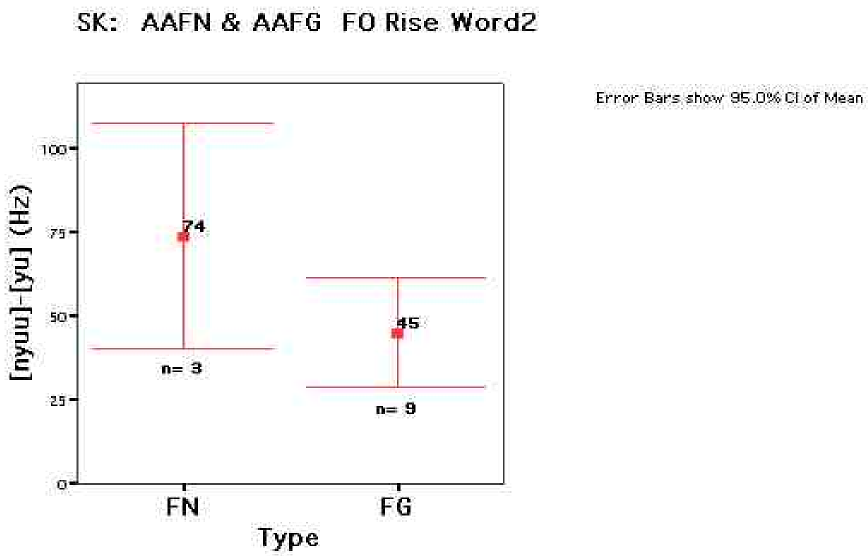


b. Speaker RO

Figure 8.02. The Mean F0 Difference between the First and the Second Syllable of Word2 of the Post-FOCUS Cases (AA)

Continued next page

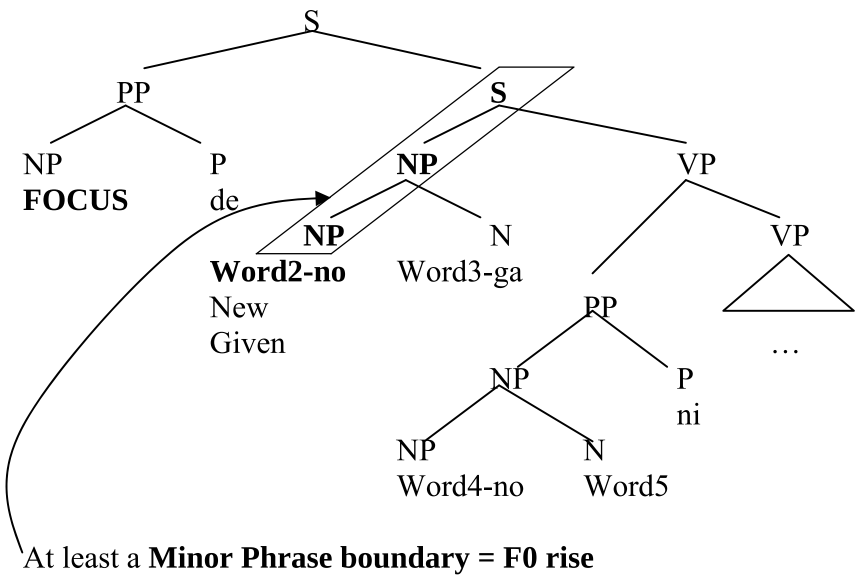
Figure 8.02 continued



c. Speaker SK

What is commonly shared by these three speakers is that the mean values of F0 excursion size between the first and the second syllable of Word2 were all above zero and their confidence intervals did not overlap with zero regardless of the information status (i.e. given vs. new) of the post-FOCUS Word2 and accent conditions. This is good evidence for the L and H edge tones, i.e. a Minor Phrase boundary, at the left edge of the post-FOCUS Word2.

(2)



The presence of a Minor Phrase boundary in the accented cases is already expected by ALIGN$_{L}$ (H*+L, ΔMiP), the constraint that demands at most one accent to appear in a Minor Phrase, which was already introduced in Section 7.3.3, Chapter 7. However, it is still necessary to account for the presence of a Minor Phrase boundary at the left edge of the unaccented Word2. In Section 8.2 of this chapter, I propose an XP-MiP alignment constraint, and this constraint is responsible for the reliable presence of a Minor Phrase boundary at the left edge of the post-FOCUS unaccented Word2 with an XP boundary.

8.1.3. The Post-FOCUS Major Phrase Boundary

That post-FOCUS Minor Phrase dephrasing was blocked by an XP boundary at the left edge of Word2 does not necessarily mean absence of any post-FOCUS dephrasing across an XP boundary. It is still possible that FOCUS has an effect of

Major Phrase dephrasing, i.e. deletion of a Major Phrase boundary at the left edge of a post-FOCUS XP. Remember that the left edge of an XP boundary is the default location for a Major Phrase boundary in a neutral context without any FOCUS (Selkirk & Tateishi, 1988, 1990). The XP-MaP alignment constraint in (3) is at work.

(3) ALIGN$_L$ (XP, MaP)

The left edge of every XP must coincide with the left edge of a Major Phrase.

One of the questions is whether this alignment constraint is observed even in the post-FOCUS context, i.e. whether the Major Phrase boundary at the left edge of an XP boundary is *retained* even when it is preceded by a FOCUS.

The other question is whether the difference in the information status of the post-FOCUS XP leads to any difference in the Major Phrase formation of the post-FOCUS item. Selkirk (2002a) proposed a focus (new) XP-MaP Prominence constraint which requires each focus XP to correspond to the head of a Major Phrase. Following her proposal, I showed in Section 6.2.3 of Chapter 6 that post-FOCUS Major Phrase dephrasing is blocked if this focus XP-MaP Prominence constraint is at work. One of the goals of this section (Section 8.1.3) is to test the prediction made by the focus XP-MaP constraint.

I found the Major Phrase dephrasing effect when the post-FOCUS word (XP) was interpreted as given. I interpret this result as evidence for the post-FOCUS dephrasing constraint outranking the XP-MaP constraint. However, when the post-FOCUS XP was interpreted as new (focus), the Major Phrase boundary was kept intact at the left edge of the post-FOCUS word. This outcome supports Selkirk's focus XP-MaP Prominence constraint and is evidence for the constraint outranking the post-FOCUS dephrasing constraint.

8.1.3.1. Presence of a Major Phrase Boundary in Control Cases

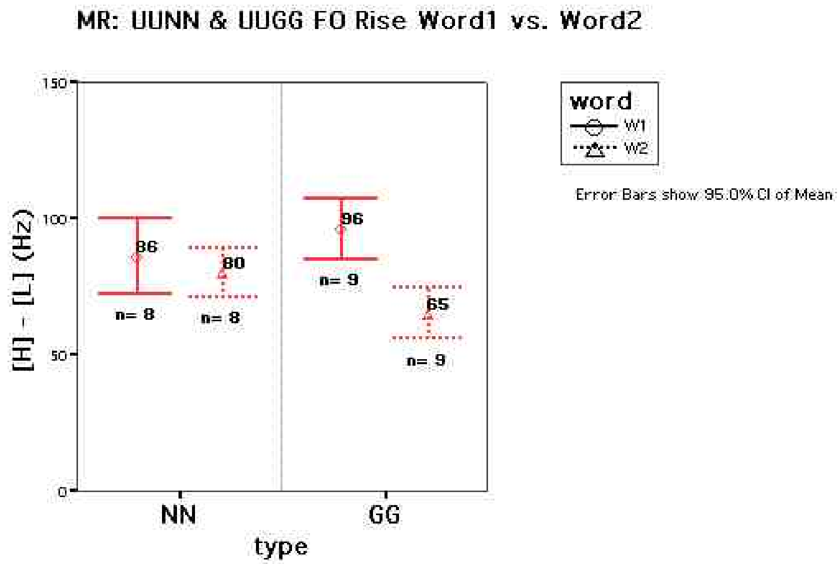
To start with, it is necessary to confirm the presence of a Major Phrase boundary at the left XP boundary of Word2 in the control contexts, i.e. NN and GG. Only if the presence of that higher order boundary is confirmed in those control contexts can we ask whether FOCUS has any effect of deleting a Major Phrase boundary at the left edge of a post-FOCUS Word2 of the FOCUS cases, i.e. FN and FG. I show in this subsection that both of the control cases (i.e. NN and GG) have a Major Phrase boundary at the left edge of the post-FOCUS Word2 with an XP boundary. Based on this observation, I will ask in Section 8.1.3.2. whether that Major Phrase boundary is *retained* even when Word1 is made into a FOCUS.

8.1.3.1.2. The Unaccented Control Cases

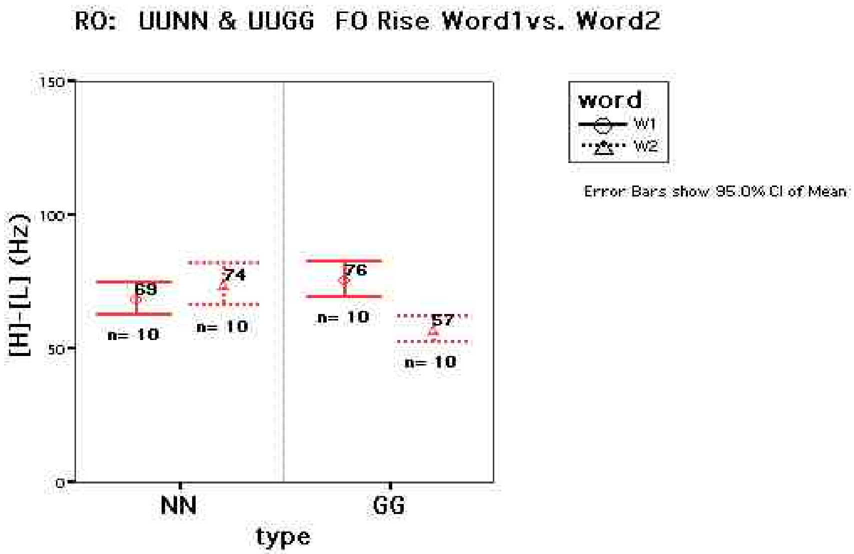
The UU control cases (i.e. control cases with unaccented words) are considered in this subsection. The presence or absence of a Major Phrase boundary at the onset of a word preceded by an unaccented word is judged by the F0 excursion size from a L edge tone to a H edge tone. According to Pierrehumbert & Beckman (1988), the greater the phrase boundary is, the lower the L edge tone is that marks the phrase edge. In other words, we expect greater F0 excursion size from the L to the H edge tone when a higher order phrase boundary is present.[24] I adopt this as criterion for the presence and absence of a Major Phrase boundary at the onset of a word preceded by an unaccented word. One possible way to apply this criterion is to compare the F0 excursion size from L to H of Word1 to that of Word2. I will conclude that there is a Major Phrase boundary at the onset of Word1 if the F0 excursion size from L to H of Word2 is as large as that of Word1. This argument comes from the fact that the left edge of Word1 is a sentence-initial position, which coincides with at least a Major Phrase boundary.

In order to carry out the comparison, the F0 difference between the first syllable (L) and the second syllable (H) was obtained for both Word1 and Word2. Mean values of the excursion size as well as their 95% confidence intervals are summarized in Figure 8.03.

[24] Shinya (2002) systematically shows correlation between the F0 excursion size between the L and the H edge tones and phrasal boundary strength.



a. Speaker MR

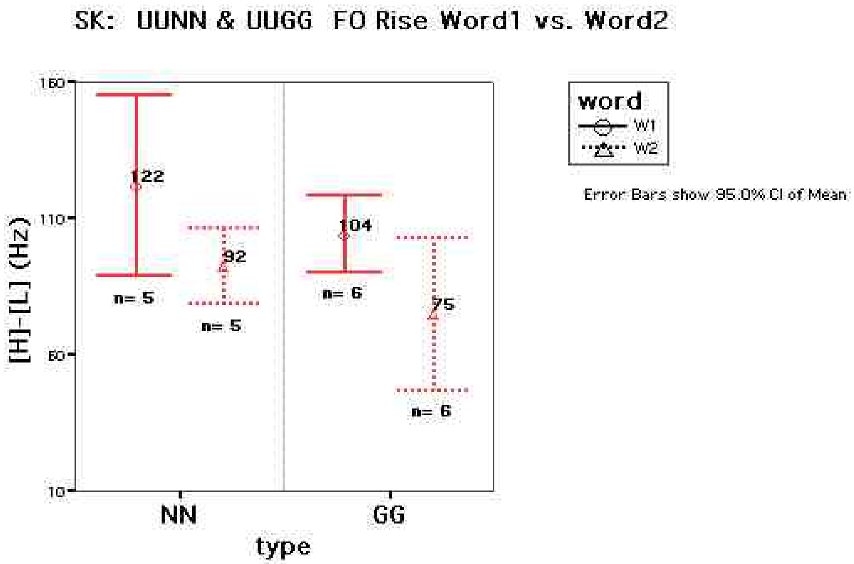


b. Speaker RO

Figure 8.03. The Mean F0 Difference between L and H of Word2 of the Control Cases (UU)

Continued next page

Figure 8.03 continued



c. Speaker SK

<The UUNN Case>

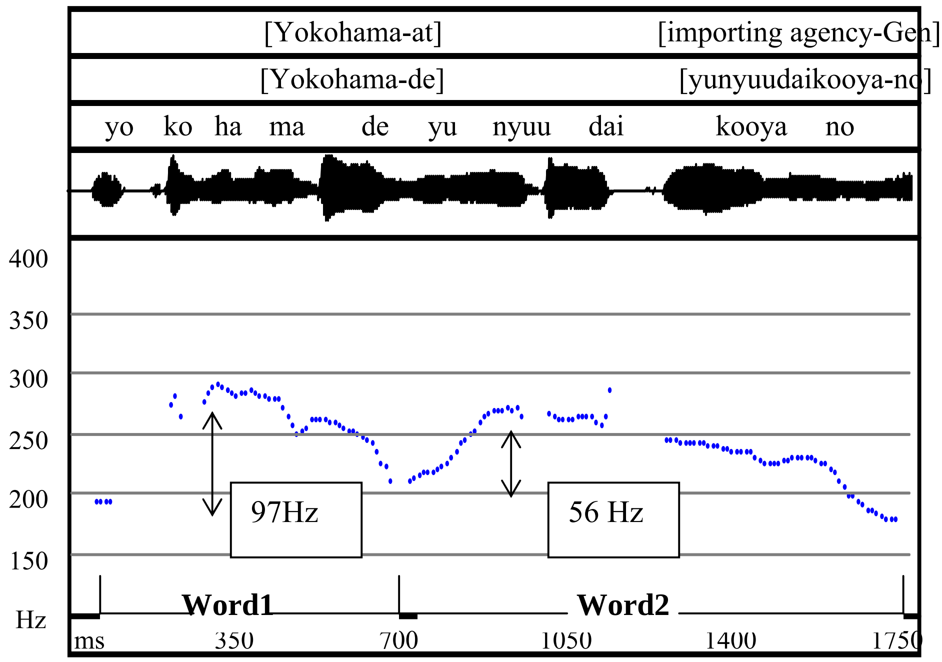
As shown in Figure 8.03, the F0 excursion size of Word1 and that of Word2 were not statistically different for the NN cases in all three speakers' speech. For example, the mean value of Word1 and that of Word2 of the NN case were 86 Hz and 80 Hz respectively in MR's figure (Figure 8.03-a), and their confidence intervals substantially overlapped each other. A 2-tailed paired t-test result also confirmed that there was no significant difference between those two means: $p = 0.43$. The same was true for RO (Figure 8.03-b): the mean value of Word1 and that of Word2 of the NN case were 69 Hz and 74 Hz respectively and their confidence intervals overlapped each other. A t-test result (2-tailed, paired) also confirmed that the difference between them was not significant: $p = 0.2$. When it comes to SK, the mean value of Word1 and that of Word2 of the NN case was not close to each other: the former was 122 Hz while the latter was 92 Hz. This 30 Hz difference, however, was only marginally significant according to a t-test (2-tailed, paired): $p = 0.056$. In this way, there was no strong evidence to confirm that the F0 excursion size of Word1 and that of Word2 of the NN case are different, and I conclude that there is a Major Phrase boundary at the left edge of Word2 of the NN case based on this result.

<The UUGG Case>

When it comes to the GG case, things were not that straightforward. The F0 excursion size at the onset of Word2 was significantly smaller than that of Word1 in MR and RO's speech (see Figure 8.03-a,b). For example, as shown in MR's speech (Figure 8.03-a), the F0 excursion size mean of Word1 was 96 Hz while that of Word2 was 65 Hz. This 30 Hz difference turned out to be statistically significant because there was no overlap between their 95% confidence intervals. The same thing was true of RO (Figure 8.03-b). The F0 excursion size mean of Word1 was 76 Hz while that of Word2 was 57 Hz, and this 20 Hz difference was significant because their confidence intervals did not overlap. SK, too, had a parallel difference between the F0 excursion size mean of Word1 (104 Hz) and that of Word2 (75 Hz), while that 30 Hz difference was only marginally significant: $p = 0.089$ (2-tailed paired t-test). Nonetheless, it is too early to give any conclusion about the presence or absence of a Major Phrase boundary at the left edge of Word2 of the GG case just based on this result. It may be the case that the F0 excursion size of the given Word2 of the GG case was phonetically reduced because of its givenness while that of the given Word1 "stays" as large as that of the new Word1 of the NN case because of an “utterance initial" effect, in spite of its givenness.

Impressionistically speaking, I still hear a higher order boundary at the left edge of the given Word2 of the GG cases in all three speakers' speech. Especially because those LH edge tones of Word1 and those of Word2 are far apart (i.e. three syllables intervening between them), it is almost impossible to detect any F0 excursion size difference between Word1 and Word2 of the GG case. For example, the following F0 contour of Word1 and Word2 of the GG case obtained from MR's speech has about a 40 Hz difference between the F0 excursion size of Word1 and Word2. Nonetheless, it is clear to the author, a native speaker of Tokyo Japanese, that there is a phrase boundary higher than a simple Minor Phrase boundary, i.e. a Major Phrase boundary, at the left edge of Word2.

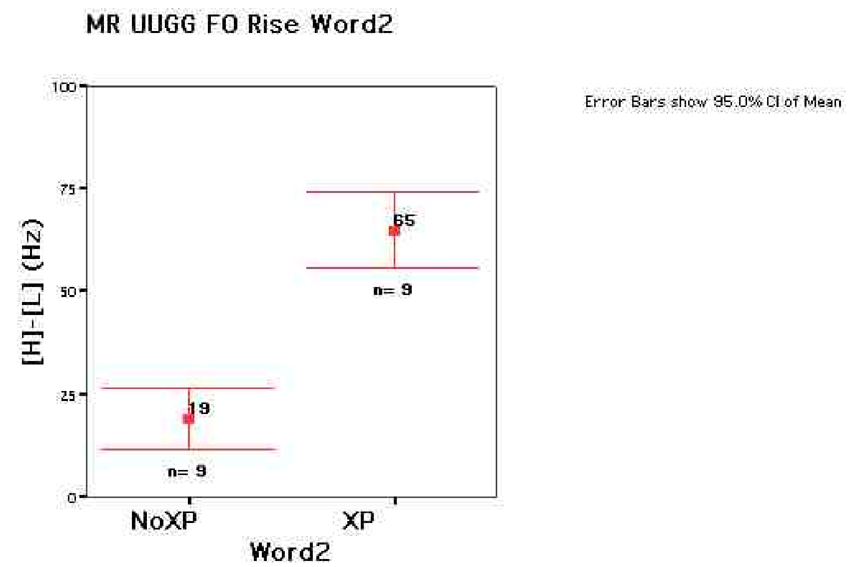
MR 6-77-3



H = 294 Hz H = 274 Hz
L = 197 Hz L = 218 Hz

Figure 8.04. The Example F0 Contour of Word1 and Word2 of the GG Case (Speaker MR)

To test this impressionistic judgment on a Major Phrase boundary at the left edge of Word2 of the GG case with an XP boundary, another test was carried out. In Chapter 7, it was already concluded that there was no Major Phrase boundary at the left edge of a given Word2 of the GG case when no XP boundary was present at its left edge. If there is a Major Phrase boundary at the left edge of the given Word2 with an XP boundary, then we expect that the F0 excursion size of Word2 with an XP boundary should be greater than that of Word2 without an XP boundary. Error bar plots in Figure 8.05 show the comparison between the F0 excursion size of Word2 with an XP boundary and that of Word2 without such a boundary.

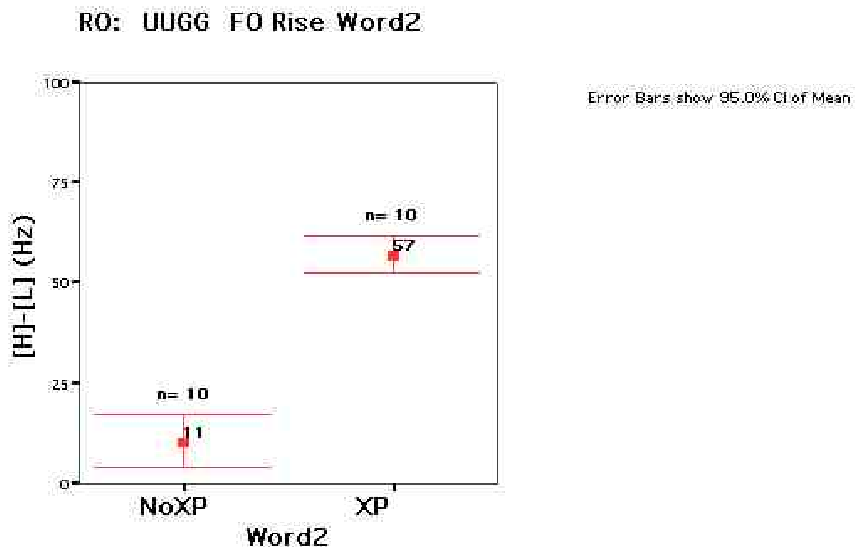


a. Speaker MR

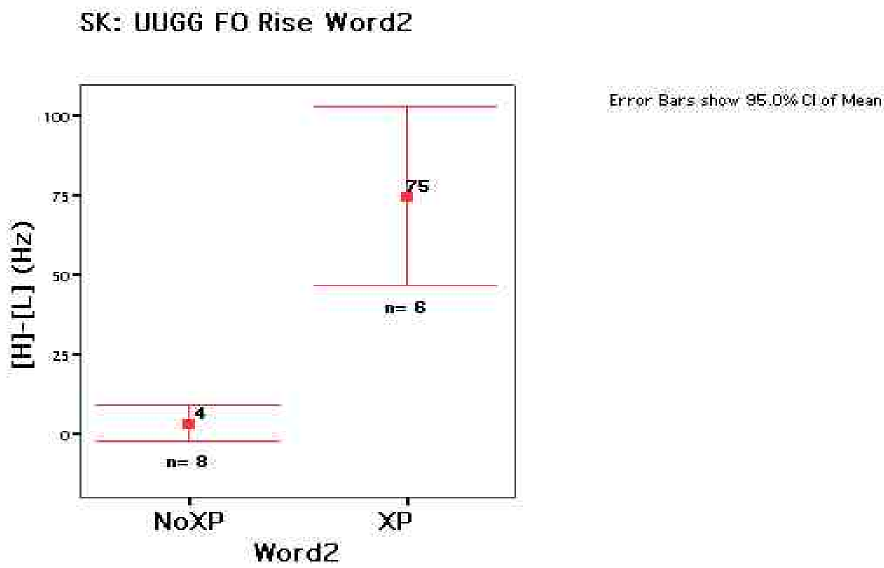
Figure 8.05. The Mean F0 Difference between L and H of Word2 (No-XP vs. XP, UUGG)

Continued next page

Figure 8.05 continued



b. Speaker RO



c. Speaker SK

From these figures, it is obvious that the mean of F0 excursion size of Word2 with an XP boundary was significantly greater than that of Word2 without it. The mean of the former was 46 Hz greater than the latter in MR and RO's speech and 71 Hz greater in SK's speech. Also, their 95% confidence intervals did not overlap at all. This is a good indication that the GG case also has a Major Phrase boundary at the left edge of Word2 with an XP boundary.

In summary, I conclude that there is a Major Phrase boundary at the left edge of Word2 with an XP boundary when it is produced in control contexts, i.e. NN and GG.

(4) **The UU Control Cases**

[New(U)$_{\text{Word1}}$]	$_{\text{XP}}$[New(U)$_{\text{Word2}}$ …	[Given(U)$_{\text{Word1}}$]	$_{\text{XP}}$[Given(U)$_{\text{Word2}}$ …
($_{\text{MaP}}$)	($_{\text{MaP}}$	($_{\text{MaP}}$)	($_{\text{MaP}}$
($_{\text{MiP}}$)	($_{\text{MiP}}$	($_{\text{MiP}}$)	($_{\text{MiP}}$

8.1.3.1.2. The Accented Control Cases

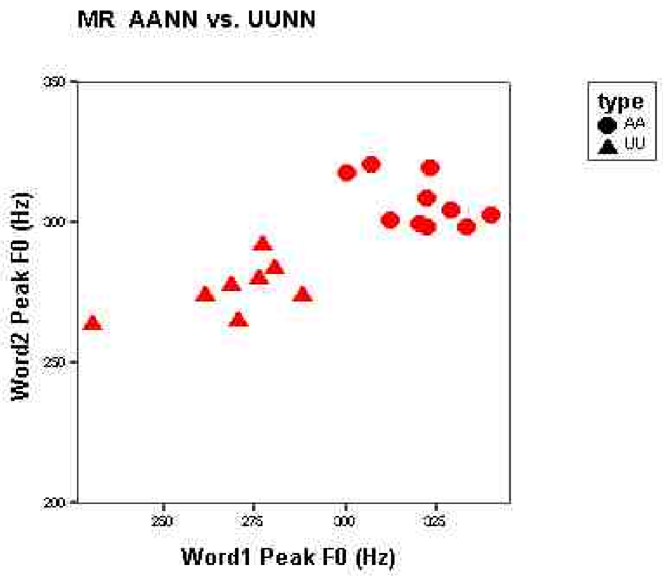
The AA control cases (i.e. control cases with accented words) are considered in this subsection. For this purpose, we examined whether Word2 of those control cases underwent *catathesis*. The domain of *catathesis* is a Major Phrase, and its propagation is blocked by a new Major Phrase boundary. If *catathesis* is blocked at the left edge of Word2 of the control cases, it will be concluded that a Major Phrase boundary is present at its left edge.

The presence or absence of *catathesis* was tested by comparing the F0 peak value of Word2 of the AA cases (i.e. sequence of accented words) and that of the UU cases (i.e. sequence of unaccented words). If the F0 peak of Word2 of the AA case is lower than that of Word2 of the UU case when both of those two cases share the same

F0 value of Word1, then it is concluded that Word2 of the AA case undergoes *catathesis*: no Major Phrase boundary at the left edge of the accented Word2. However, if the peak F0 value of Word2 of AA is realized as high as that of UU when they share the same F0 peak of Word1, or the former is higher than the latter, then it is interpreted as evidence for lack of *catathesis*, i.e. evidence for presence of a Major Phrase boundary, at the left edge of the accented Word2.

<The AANN Case>

Scatter plots in Figure 6.06 show F0 peak values of Word2 and those of Word1 of the AANN and the UUNN case. F0 peak values of Word1 are on the horizontal axis and those of Word2 are on the vertical axis. Data points represented by circles are those of the AANN case and those represented by triangles are those of the UUNN case.

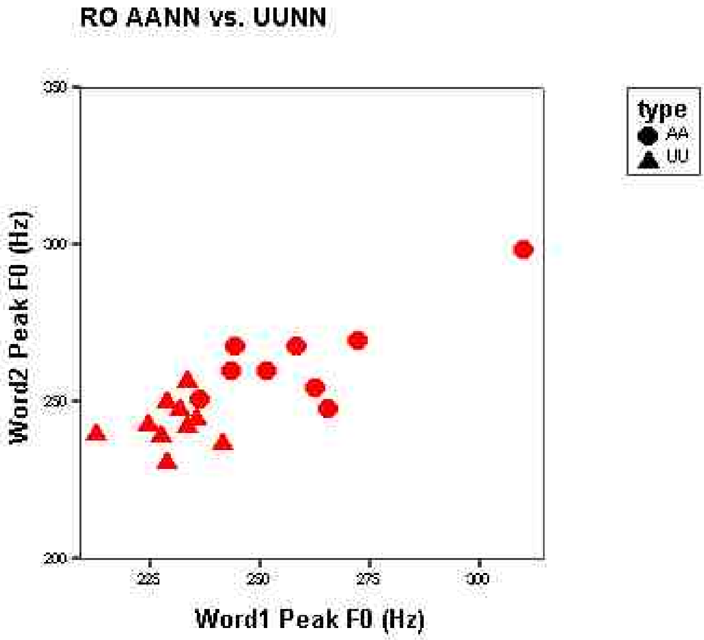


a. Speaker MR

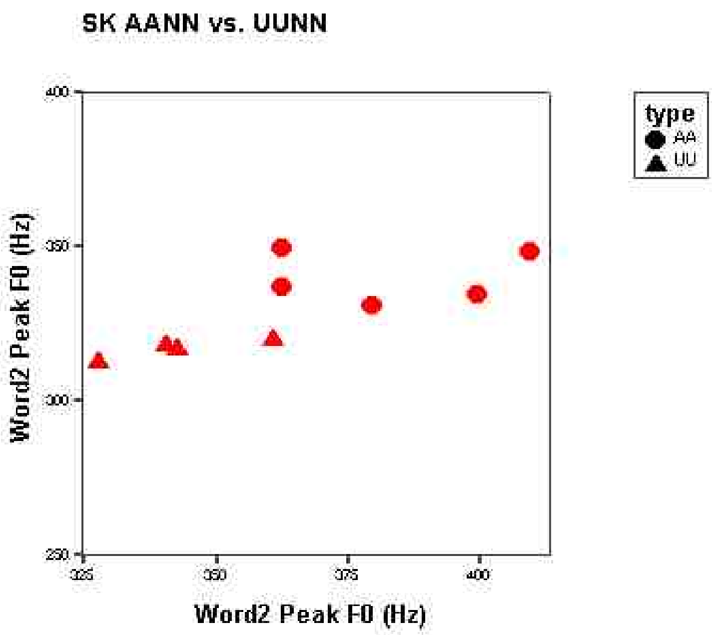
Figure 8.06. F0 Peak Values of Word2 (AA vs. UU, The NN Cases)

Continued next page

Figure 8.06 continued



b. Speaker RO

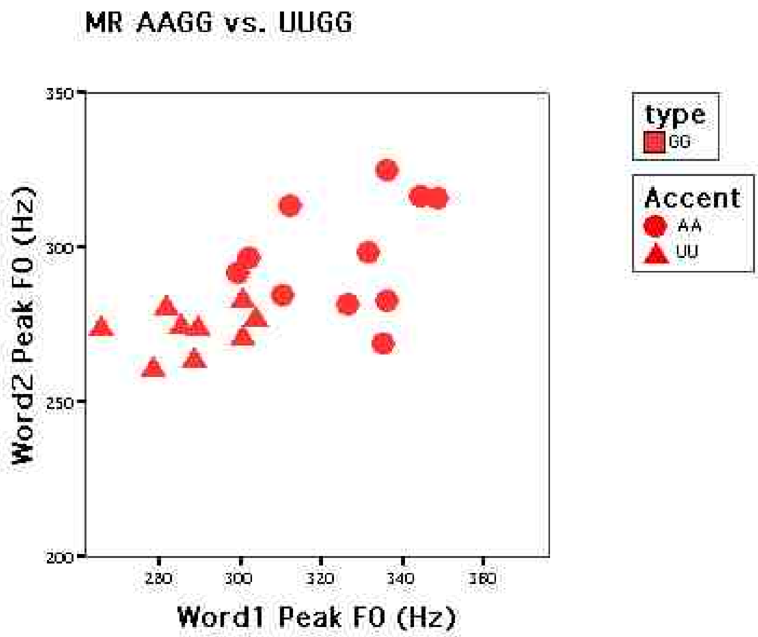


c. Speaker SK

In these plots, the data points representing the peak F0 of Word2 of the AANN case (circles) are either higher than or as high as those representing the peak F0 of Word2 of the UUNN case (triangles). Even when both AANN and UUNN shared the same peak F0 of Word1 (horizontal axis), AA's peak F0 of Word2 was slightly higher than UU's peak F0 of Word2. Given this observation, it is evident that Word2 of the AANN case did not undergo *catathesis*, and I conclude that there is a Major Phrase boundary at the left edge of Word2 of the AANN case.

<The AAGG Case>

The same result was observed in the AAGG case. Scatter plots in Figure 8.07 show F0 peak values of Word2 and those of Word1 of the AAGG and the UUGG cases.

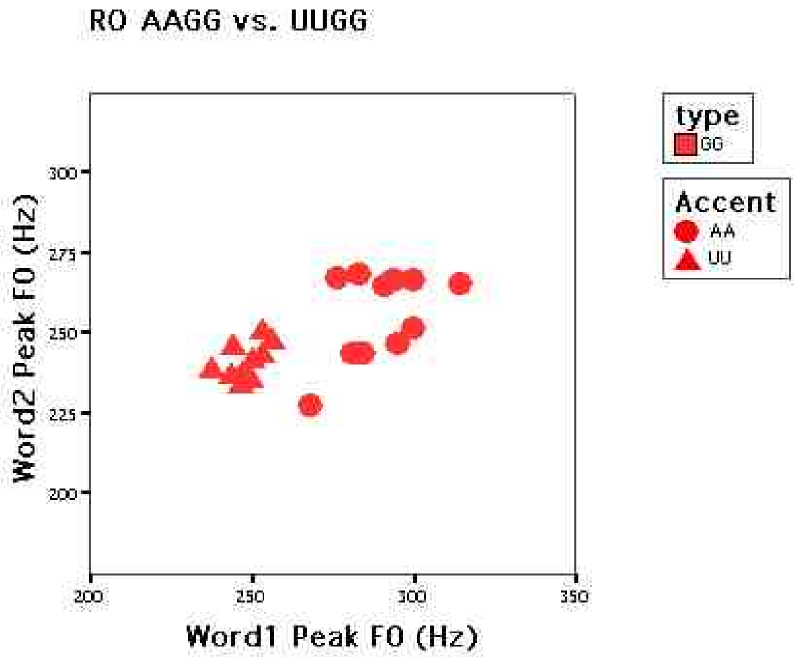


a. Speaker MR

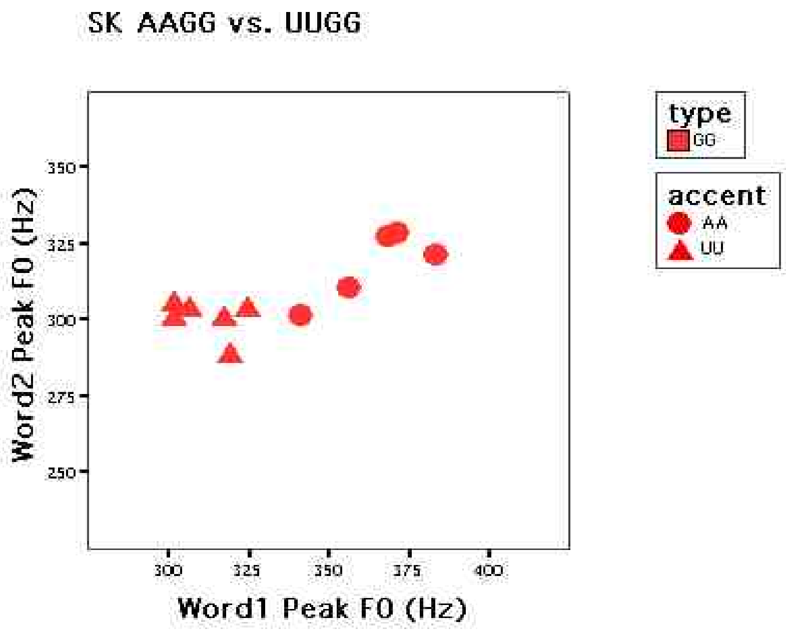
Figure 8.07. F0 Peak Values of Word2 (AA vs. UU, The GG Cases)

Continued next page

Figure 8.07 continued



b. Speaker RO



c. Speaker SK

Data points representing Word2 F0 peak of the AAGG case were generally realized higher than or as high as those of the UUGG case. When the AAGG and the UUGG case shared the same peak F0 of Word1, they shared the same peak F0 value of Word2: the F0 peak of Word2 of the AAGG case was not lower than that of Word2 of the UUGG case. This leads to a conclusion that Word2 of the AAGG case did not undergo *catathesis* and there is a Major Phrase boundary at the left edge of Word2 of the AAGG case.

(5) **The AA Control Cases**

$[\text{New(A)}_{\text{Word1}}]\ _{\text{XP}}[\text{New(A)}_{\text{Word2}} \ldots$	$[\text{Given(A)}_{\text{Word1}}]\ _{\text{XP}}[\text{Given(A)}_{\text{Word2}} \ldots$
$(_{\text{MaP}} \quad) \quad (_{\text{MaP}}$	$(_{\text{MaP}} \quad) \quad (_{\text{MaP}}$
$(_{\text{MiP}} \quad) \quad (_{\text{MiP}}$	$(_{\text{MiP}} \quad) \quad (_{\text{MiP}}$

8.1.3.1.3. Summary of 8.1.3.1.

In summary, both the accented control cases (i.e. AANN and AAGG) and the unaccented control cases (i.e. UUNN and UUGG) have a Major Phrase boundary at the left edge of Word2 with an XP boundary. This result confirms Selkirk & Tateishi's (1988, 1991) XP-Major Phrase boundary alignment constraint, $\text{A}\textsc{lign}_{\text{L}}$ (XP, MaP), which demands alignment between the left edge of every XP and the left edge of a Major Phrase. In Section 8.1.3.2, I ask whether the demand of this alignment constraint is respected even when the XP boundary is immediately preceded by FOCUS.

8.1.3.2. The Post-FOCUS MaP Dephrasing and Given vs. New Distinction

It is already inferred from Figure 8.01 and Figure 8.02 that there may be a Major Phrase boundary at the left edge of the post-FOCUS new Word2 of the FN case while such a boundary may be absent at the left edge of the post-FOCUS given Word2 of the FG case. This inference comes from the fact that the Word2 F0 excursion size of the FN case was far greater than that of the FG. For example, the mean value of the F0 rise from the first syllable to the second syllable of Word2 of the UUFN case of MR was 54 Hz while that of her UUFG case was only 15 Hz. This 40 Hz difference between those two FOCUS cases sounds too large to be attributed to a non-structural effect of a given vs. new distinction on Word2, and my speculation is that a Major Phrase boundary is present at the left edge of Word2 of the FN case while it is absent in the FG case. The main interest of this section is to verify this speculation.

For this purpose, I compared the F0 excursion size difference between the GG case and the NN case and that between the FG case and the FN case. If the FG case really deletes a Major Phrase boundary at the left edge of Word2 while the FN case retains it, then we expect that the F0 excursion size difference between the FG and the FN case should be significantly greater than that between the NN and the GG case. It is because both of those control cases (i.e. NN and GG) equally have a Major Phrase boundary at the left edge of Word2, and they are both expected to be associated with relatively large F0 excursion size. On the other hand, the FOCUS cases have a phonological contrast between the FG case and the FN case: the FG case lacks a Major Phrase boundary but the FN case retains it.

Though the control NN and GG cases share the same representation, we observed in the previous section that the former was associated with slightly larger F0 excursion size than the latter, because of the non-structural effect of new vs. given distinction. On the other hand, we expect not only such a non-structural difference of the new vs. given but also a structural difference (i.e. presence and absence of a Major Phrase boundary) in the FOCUS cases. Therefore, we expect the F0 excursion size difference between the FN and the FG to be greater than that between the NN and the GG case. In other words, we expect an interaction between the "given vs. new" factor and the “control vs. FOCUS" factor. The graphical representation in Figure 6.08 summarizes this expectation.

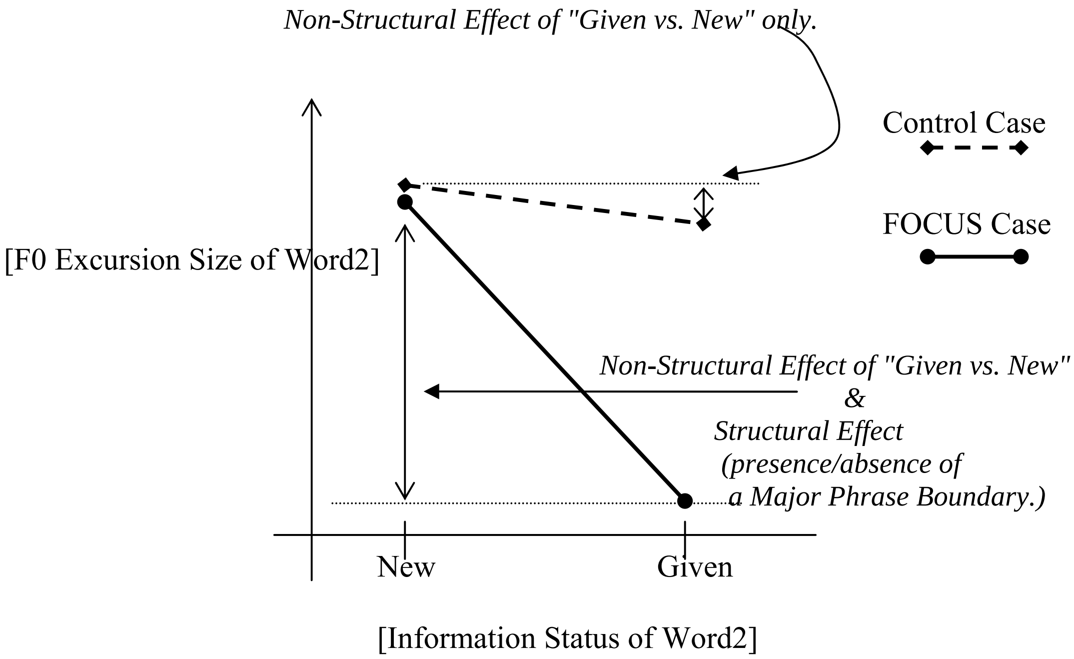


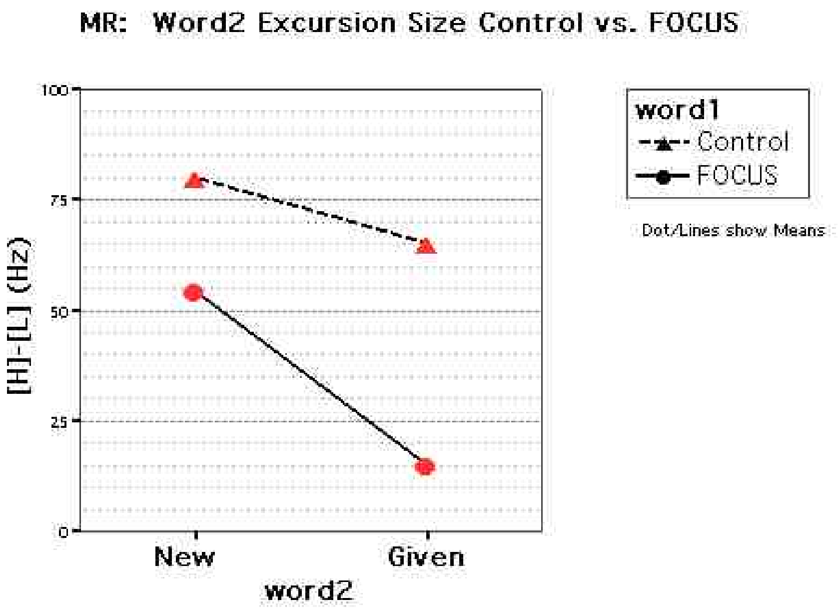
Figure 8.08. The Expected Interaction between the Given vs. New Factor and the Control vs. Post-FOCUS Factor

8.1.3.2.1. Results: The UU Cases

All three speakers conformed to the prediction above. The results are shown in Figure 8.09. Plots in Figure 8.09 summarize the three speakers’ F0 excursion difference between the control cases and the FOCUS cases as well as that between the given and the new Word2. There was only a subtle difference between the F0 excursion size of

Word2 of the NN case and that of the GG case, but there was a substantial difference between that of Word2 of the FN case and that of the FG case. For example, in MR's speech there was only about a 15 Hz difference between the NN and the GG case, while the F0 excursion size of the FN case was about 40 Hz greater than that of the FG case. An ANOVA analysis was carried out to see whether the interaction between the "post-FOCUS vs. control factor" and the "new vs. given factor" was statistically significant. Also, Scheffe's multiple comparison test was carried out to examine whether the differences between the two control cases (NN and GG) and between the two FOCUS cases (FN and FG) were significant.

According to the ANOVA results, the interaction between the given vs. new factor and the control vs. FOCUS factor was significant for MR ($F(1,31) = 9.471$, $*p = 0.004$) and SK ($F(1,16) = 6.309$, $*p = 0.023$) and marginally significant for RO ($F(1,37) = 3.782$, $p = 0.059$). According to Scheffe's multiple comparison test, the mean difference between the control cases (i.e. the NN and the GG case) was not significant for MR and SK (MR, mean difference = 15.13, SE = 5.6, $p = 0.084$; SK, mean difference = 17.57, SE = 10.71, $p = 0.46$), while that between the FN and the FG case was significant (MR, mean difference = 39.43, SE = 5.57, $*p < 0.001$; SK, mean difference = 57.70, SE = 11.86, $*p = 0.002$). RO also shared the same tendency: the difference between the control cases (NN and GG) was smaller than the difference between the FOCUS cases (FN and FG), though those differences were both significant (RO's control case, mean difference = 16.80, SE = 5.49, $*p = 0.037$; RO's FOCUS case, mean difference = 31.73, SE = 5.36, $*p < 0.001$).



		N
TYPE	UUFG	11
	UUFN	7
	UUGG	9
	UUNN	8

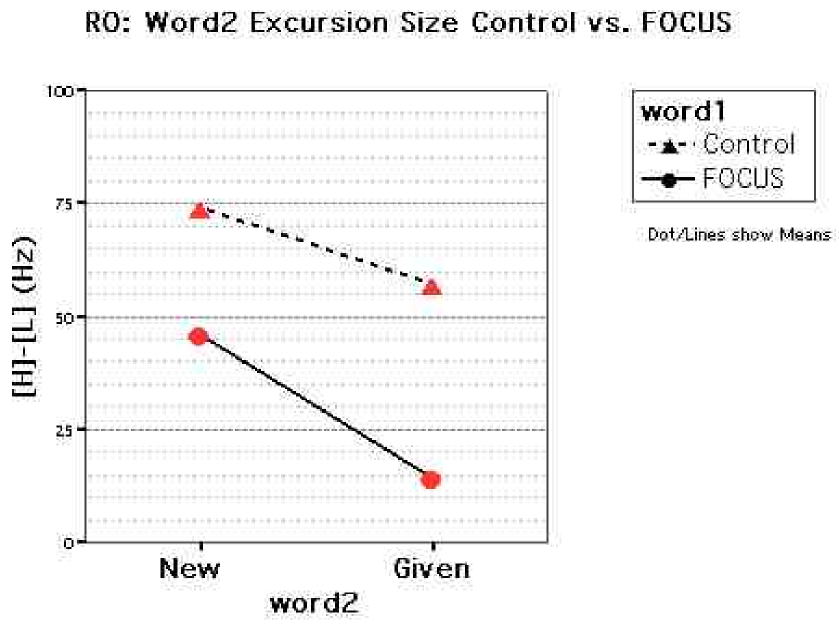
	Mean	Std. Error	95% Confidence Interval	
TYPE			Lower Bound	Upper Bound
UUFG	15.000	3.474	7.915	22.085
UUFN	54.429	4.354	45.548	63.310
UUGG	65.000	3.840	57.168	72.832
UUNN	80.125	4.073	71.818	88.432

a. Speaker MR

Figure 8.09. The Mean F0 Difference between L and H of Word2 (New vs. Given, Control vs. Post-FOCUS, UU)

Continued next page

Figure 8.09 continued



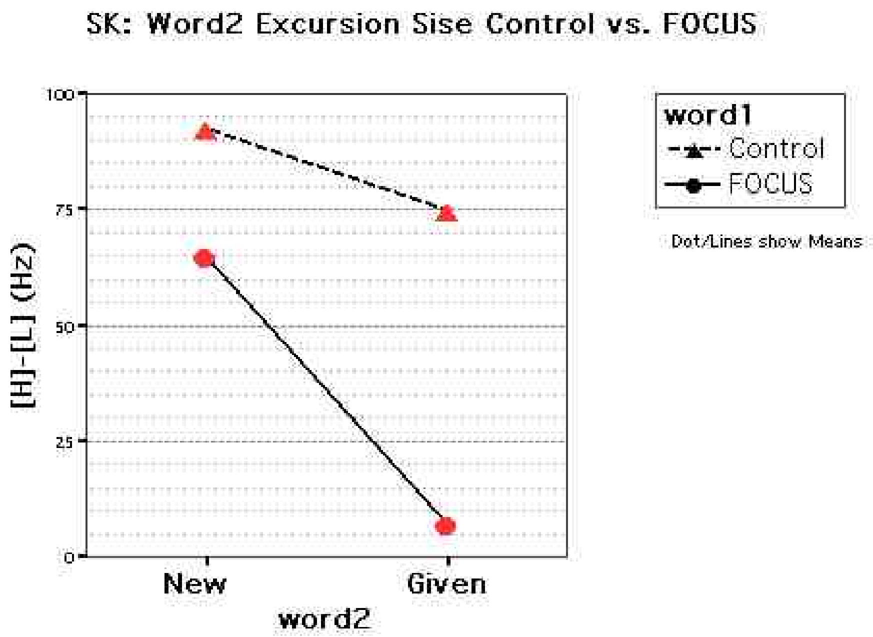
		N
TYPE	UUFG	10
	UUFN	11
	UUGG	10
	UUNN	10

	Mean	Std. Error	95% Confidence Interval	
TYPE			Lower Bound	Upper Bound
UUFG	14.000	3.882	6.134	21.866
UUFN	45.727	3.701	38.228	53.227
UUGG	57.100	3.882	49.234	64.966
UUNN	73.900	3.882	66.034	81.766

b. Speaker RO

Continued next page

Figure 8.09 continued



		N
TYPE	UUFG	5
	UUFN	4
	UUGG	6
	UUNN	5

	Mean	Std. Error	95% Confidence Interval	
TYPE			Lower Bound	Upper Bound
UUFG	6.800	7.907	-9.962	23.562
UUFN	64.500	8.840	45.759	83.241
UUGG	74.833	7.218	59.531	90.135
UUNN	92.400	7.907	75.638	109.162

c. Speaker SK

Given these results, I conclude that there is a Major Phrase boundary at the left edge of the post-FOCUS new Word2 of the UUFN case while there is only a Minor Phrase boundary at the left edge of the given Word2 of the UUFG. This conclusion is summarized in (6).

(6) **Unaccented Control Cases**

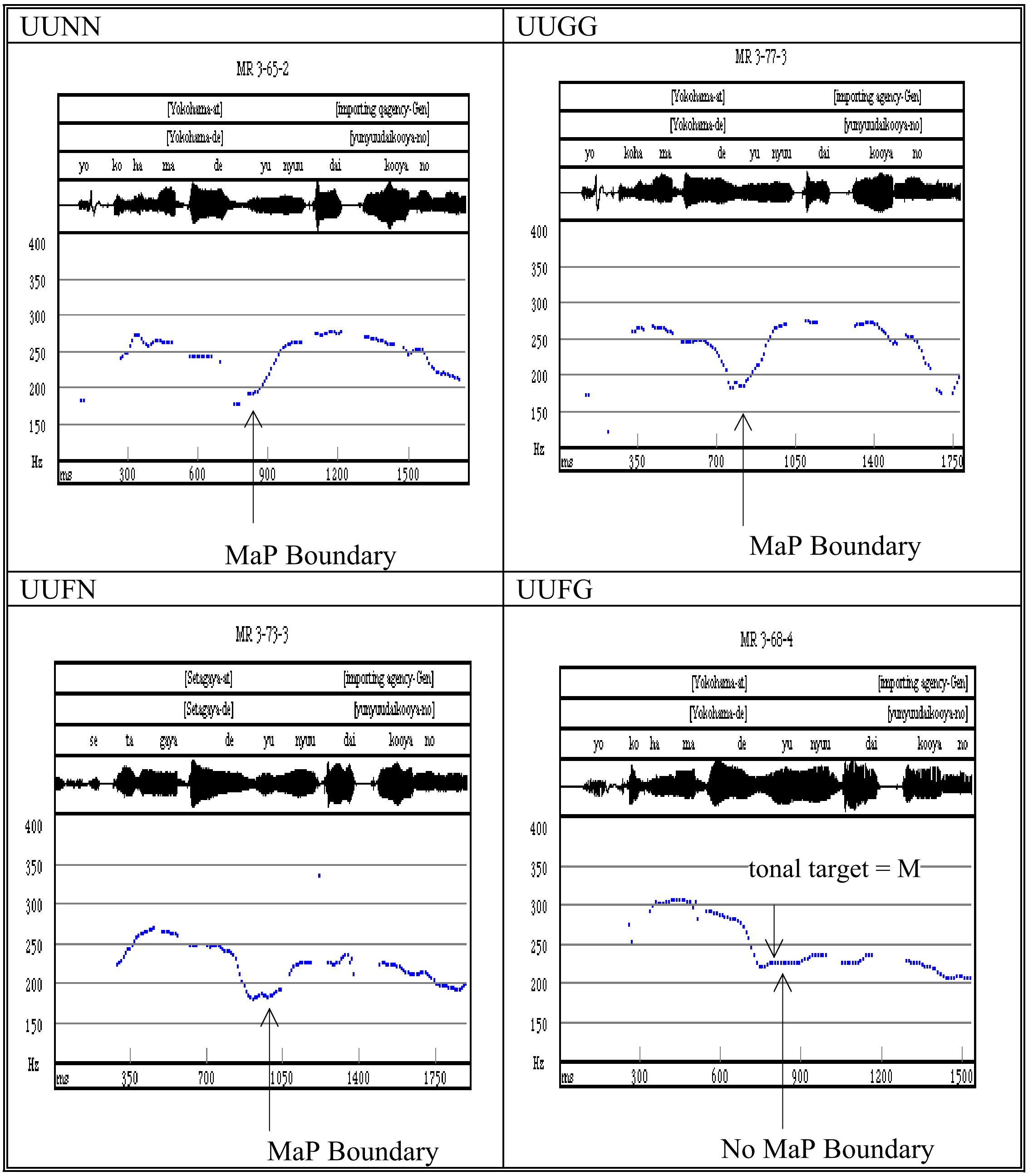
[New(U)$_{\text{Word1}}$]	$_{\text{XP}}$[New(U)$_{\text{Word2}}$ …	[Given(U)$_{\text{Word1}}$]	$_{\text{XP}}$[Given(U)$_{\text{Word2}}$ …
($_{\text{MaP}}$)	($_{\text{MaP}}$	($_{\text{MaP}}$)	($_{\text{MaP}}$
($_{\text{MiP}}$)	($_{\text{MiP}}$	($_{\text{MiP}}$)	($_{\text{MiP}}$

⇕ ⇕

Unaccented FOCUS Cases

[FOC(U)$_{\text{Word1}}$]	$_{\text{XP}}$[New(U)$_{\text{Word2}}$ …	[FOC(U)$_{\text{Word1}}$]	$_{\text{XP}}$[Given(U)$_{\text{Word2}}$ …
($_{\text{MaP}}$)	($_{\text{MaP}}$	($_{\text{MaP}}$ ---MaP Dephrasing ---	
($_{\text{MiP}}$)	($_{\text{MiP}}$	($_{\text{MiP}}$)	($_{\text{MiP}}$

Example F0 contours of those unaccented control cases and FOCUS cases are shown in Figure 8.10.

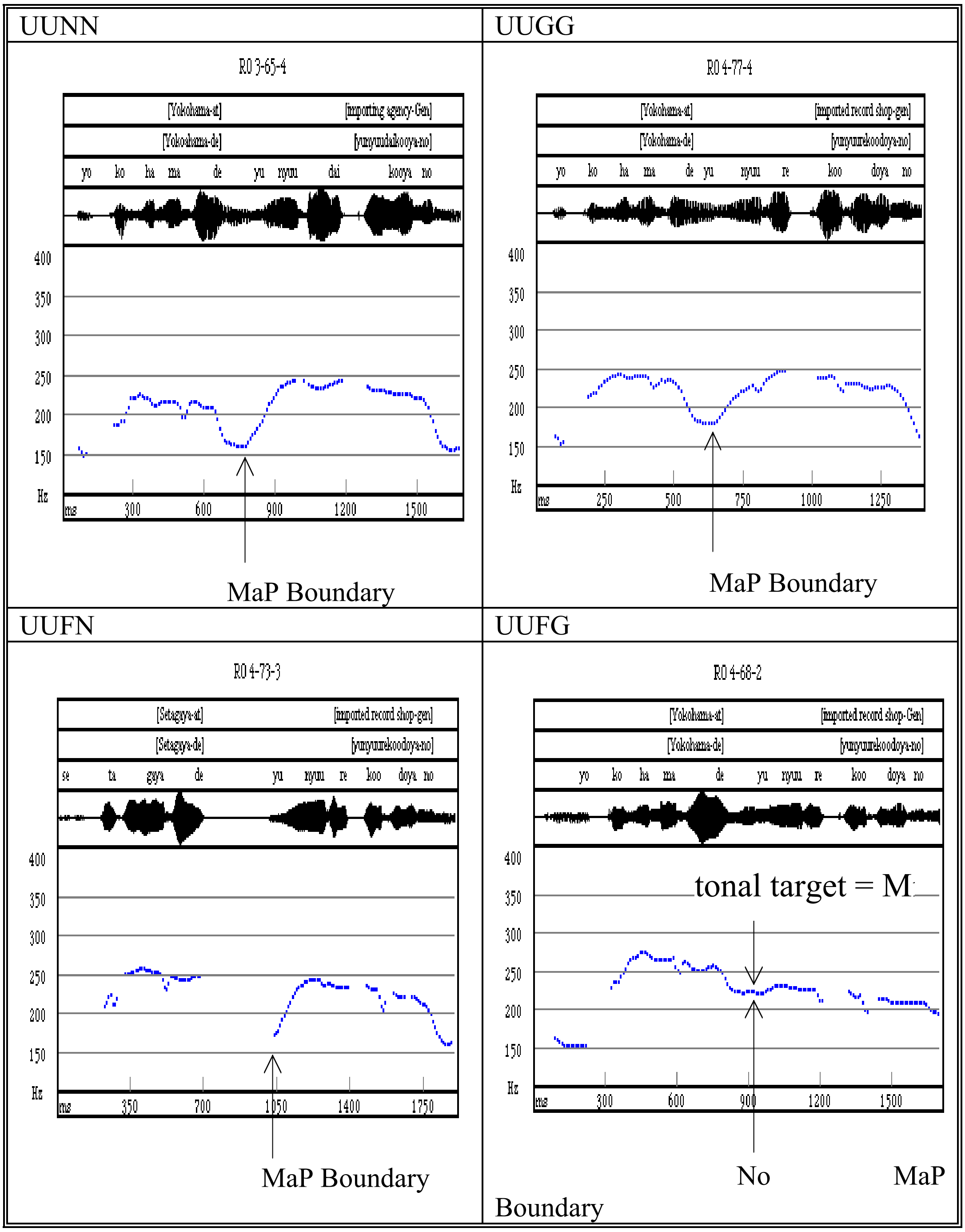


a. Speaker MR

Figure 8.10. Example F0 Contours of Word1 and Word2 of the UU Cases

Continued next page

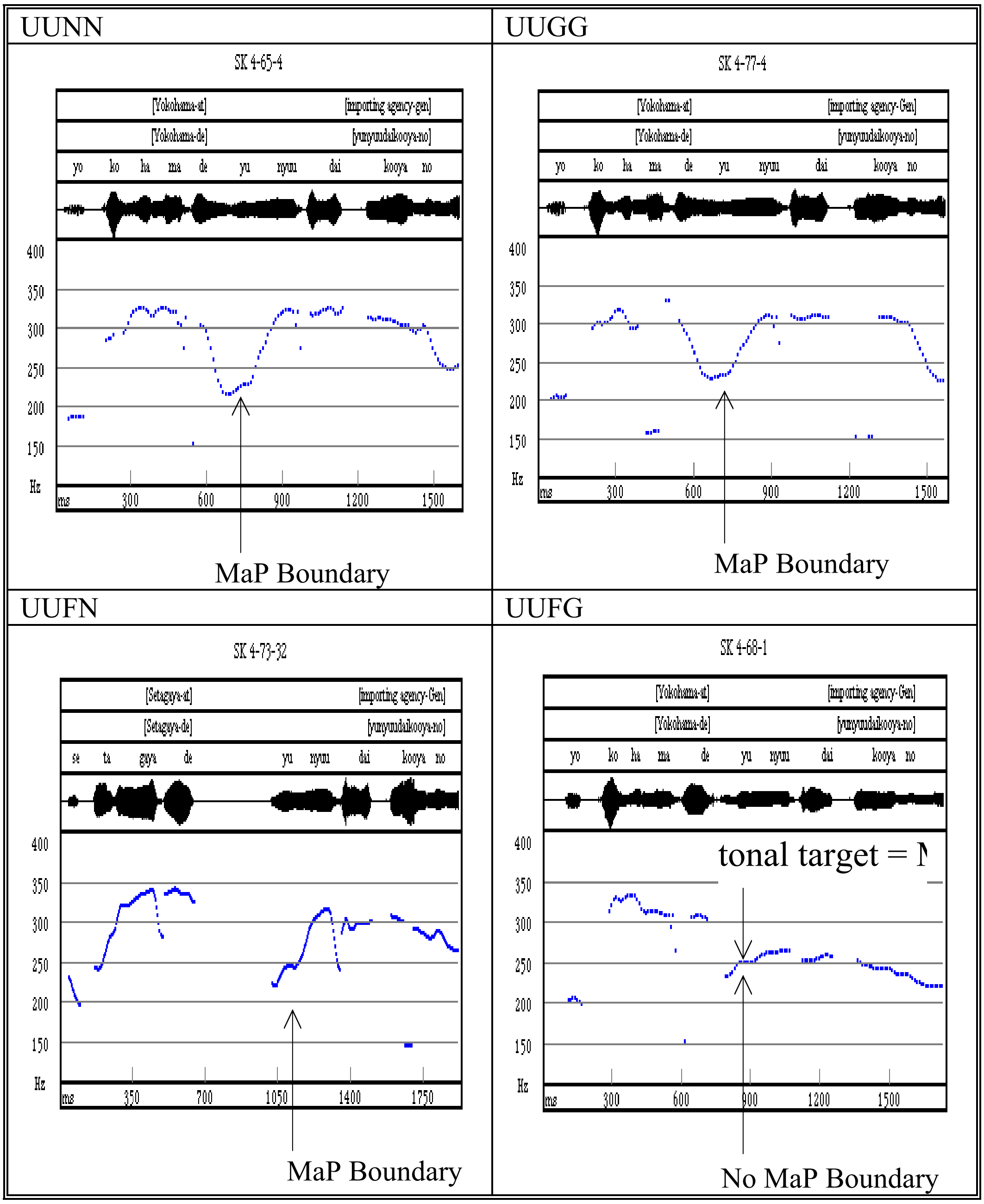
Figure 8.10 continued



b. Speaker RO

Continued next page

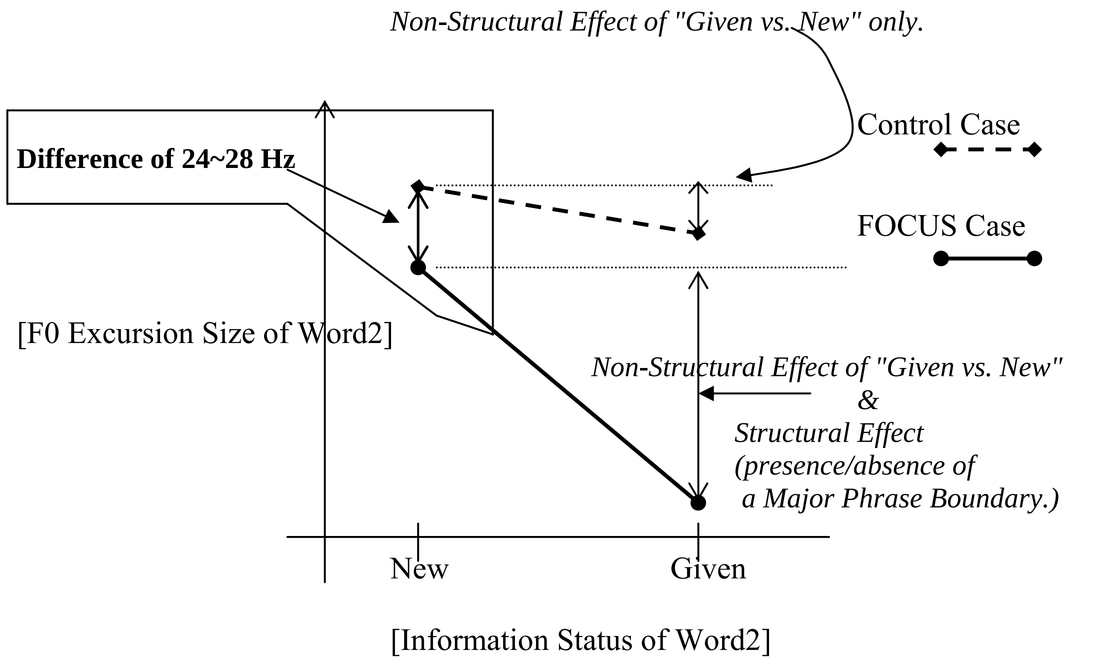
Figure 8.10 continued



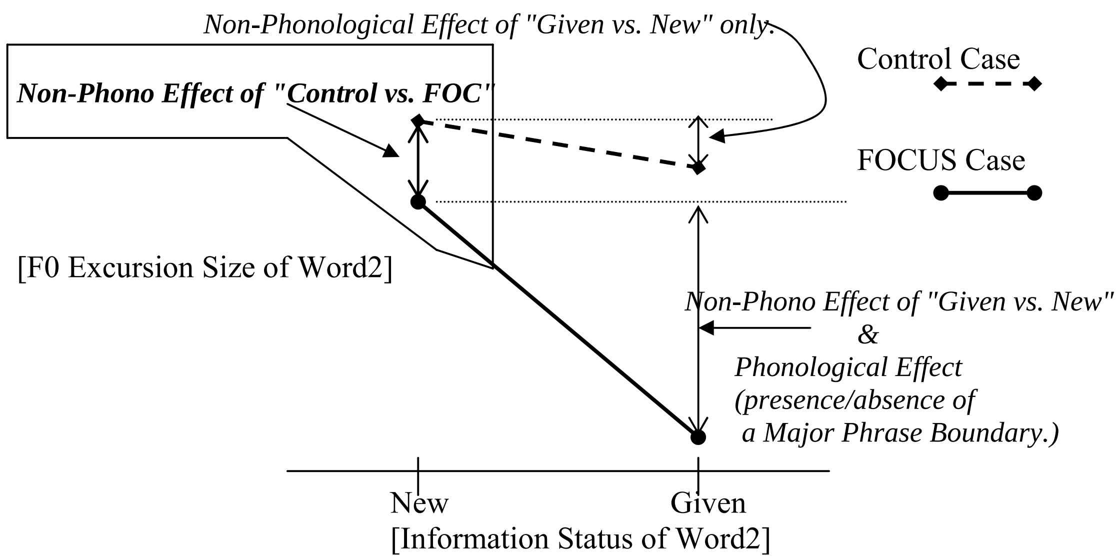
c. Speaker SK

There is one outcome that is not part of the picture shown in Figure 8.08: the F0 excursion size of the FN case was 24~28 Hz lower than that of the control NN case as graphically summarized in Figure 8.11-a.

I explain this F0 excursion size difference between the FN and the control NN case as an additional post-FOCUS non-structural effect. The non-structural effect of FOCUS on the F0 realization of the following word is not a new finding in this chapter. We already saw in Chapter 7 that FOCUS has the effect of compressing or reducing the F0 excursion size between post-FOCUS LH edge tones without changing the phonological phrase organization. I suggest that the F0 excursion size between the FN and the NN case observed in this section be parallel to the non-structural effect found in Chapter 7. That is, both the FN case and the NN case have a Major Phrase boundary at the left edge of Word2. However, the F0 excursion size difference between the L and the H edge tone of the post-FOCUS Word2 of the FN case is more compressed/reduced than that of its control counterpart (i.e. the NN case) because of the non-structural post-FOCUS effect. This is graphically summarized in Figure 8.11-b.



a. The 24~28 Hz Difference between the FN case and the NN Case



b. The Non-Phonological Effect of the Control vs. the Post-FOCUS Factor

Figure 8.11. The Actual Outcome of the Interaction between the New vs. Given Factor and the Control vs. Post-FOCUS Factor

8.1.3.2.2. Results: The AA Cases

In this subsection, I show results obtained from the sequence of accented words (AA cases). The results of the AA cases are consistent with those of the UU cases.

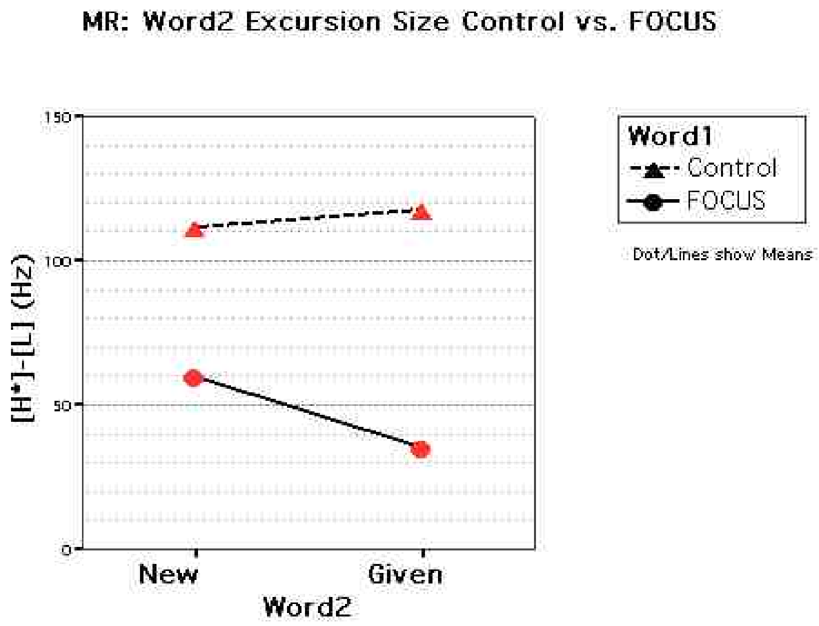
For the AA cases, I took the difference between the L edge tone (first syllable [yu]) and the H* pitch accent (fourth syllable [rí]) as the F0 excursion size relevant to this analysis. The relative height of the pitch accent H* of Word2 (i.e. the F0 peak of Word2) more correctly reflects the “prominence” of that word.

An ANOVA analysis was carried out to examine the interaction between the post-FOCUS vs. control factor and the given vs. new factor. Scheffe’s multiple comparison was used to examine whether the F0 excursion mean differences between the two control conditions (NN and GG) and between the two FOCUS conditions (FN and FG) are significant. According to the ANOVA analysis, MR and RO had a significant interaction between the control vs. FOCUS factor and the given vs. new factor (MR: $F(1,34) = 8.485$, $*p = 0.006$; RO: $F(1,30) = 5.063$, $*p = 0.032$)[25]. According to Scheffe’s multiple comparison test, the F0 excursion size mean difference between the control cases (NN and GG) was not significant (MR, mean difference = 5.67, SE = 6.81, $p = 0.874$; RO, mean difference = 12.11, SE = 5.06, $p = 0.15$), while that between the FOCUS cases (FN and FG) was significant (MR, mean difference = 24.23, SE = 7.68, $*p = 0.031$; RO, mean difference = 30.27, SE = 6.28, $*p = 0.001$).

When it comes to SK (Figure 8.12-c), she had no significant interaction between those two factors ($F(1,18) = 2.098$, $p = 0.165$). Nonetheless, we found the same relation

[25] As already mentioned in Section 1.2, I excluded three tokens out of the eight AAFN tokens of RO because they sounded unnatural. She assigned too low F0 values to post-FOCUS "new" words in those tokens, which were even lower than those of post-FOCUS given words.

between the F0 excursion size difference of the FOCUS cases and that of the control cases. The difference between the AAFN case and the AAFG case of SK was 48 Hz (i.e. the AAFG case is 48 Hz smaller than the AAFN case) while the difference between the AANN and the AAGG case was only 28 Hz (i.e. the AAGG case is 28 Hz lower than the AANN case). The difference between the FOCUS cases was 20 Hz greater than that of the control cases. Also, according to Scheffe's multiple comparison test, the F0 excursion size mean difference between the control cases (NN and GG) was only marginally significant (SK's control case, mean difference = 28.00, SE = 9.50, p = 0.064), while the mean difference between the FOCUS cases (FN and FG) was significant (SK's FOCUS case, mean difference = 48.00, SE = 10.02, *p = 0.002). In this way, SK still had a contrast between the control case and the FOCUS case with respect to the F0 excursion size difference. This is consistent with the other speakers.



		N
TYPE	AAFG	10
	AAFN	7
	AAGG	11
	AANN	10

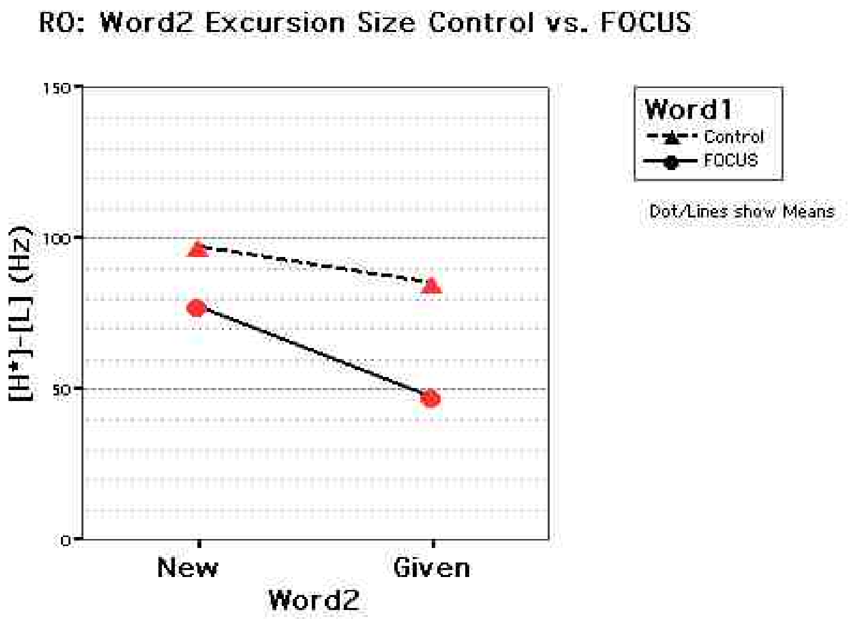
	Mean	Std. Error	95% Confidence Interval	
TYPE			Lower Bound	Upper Bound
AAFG	35.200	4.929	25.183	45.217
AAFN	59.429	5.891	47.456	71.401
AAGG	117.273	4.699	107.722	126.823
AANN	111.600	4.929	101.583	121.617

a. Speaker MR

Figure 8.12. The Mean F0 Difference between L and H* of Word2 (New vs. Given, Control vs. Post-FOCUS AA)

Continued next page

Figure 8.12 continued



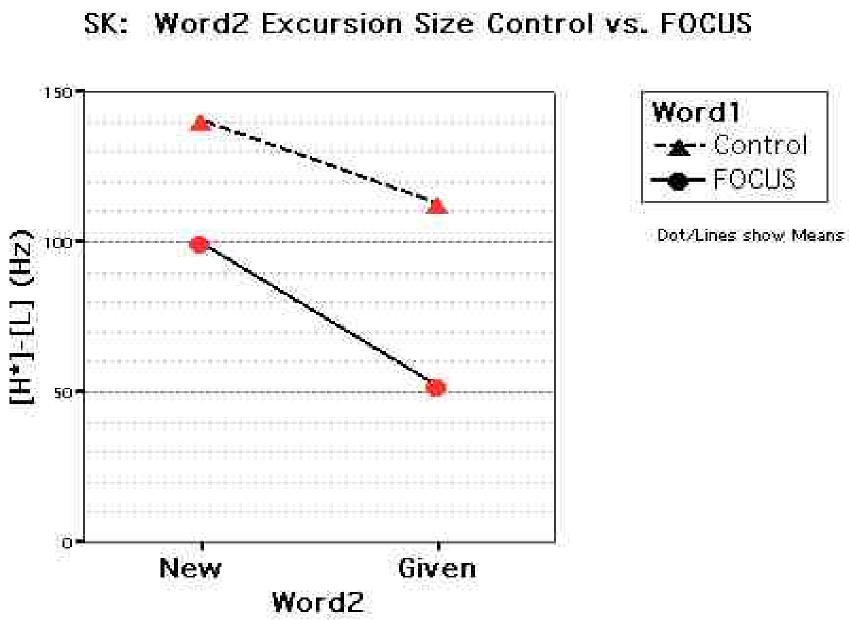
		N
TYPE	AAFG	9
	AAFN	5
	AAGG	11
	AANN	9

	Mean	Std. Error	95% Confidence Interval	
TYPE			Lower Bound	Upper Bound
AAFG	47.333	3.755	39.666	55.001
AAFN	77.600	5.037	67.313	87.887
AAGG	85.000	3.396	78.064	91.936
AANN	97.111	3.755	89.443	104.779

b. Speaker RO

Continued next page

Figure 8.12 continued



		N
TYPE	AAFG	9
	AAFN	3
	AAGG	5
	AANN	5

	Mean	Std. Error	95% Confidence Interval	
TYPE			Lower Bound	Upper Bound
AAFG	51.667	5.008	41.145	62.188
AAFN	99.667	8.674	81.443	117.891
AAGG	112.400	6.719	98.284	126.516
AANN	140.400	6.719	126.284	154.516

c. Speaker SK

Given these results, I conclude that there is a structural difference between the AAFN case and the AAFG case, while no such difference is present between the AANN case and that of the AAGG case. In other words, the AAFG case lacks a Major Phrase boundary at the left edge of Word2; the AAFN case preserves the boundary. This is summarized in (7).

(7) **Accented Control Cases**

<table>
<tr><td>[New(A)$_{\text{Word1}}$] $_{\text{XP}}$[New(A)$_{\text{Word2}}$ …</td><td>[Given(A)$_{\text{Word1}}$]$_{\text{XP}}$[Given(A)$_{\text{Word2}}$ …</td></tr>
<tr><td>($_{\text{MaP}}$) ($_{\text{MaP}}$</td><td>($_{\text{MaP}}$) ($_{\text{MaP}}$</td></tr>
<tr><td>($_{\text{MiP}}$) ($_{\text{MiP}}$</td><td>($_{\text{MiP}}$) ($_{\text{MiP}}$</td></tr>
<tr><td>$\Updownarrow$</td><td>$\Updownarrow$</td></tr>
</table>

Accented FOCUS Cases

<table>
<tr><td>[FOC(A)$_{\text{Word1}}$]$_{\text{XP}}$[New$_{\text{Word2}}$ …</td><td>[FOC(A)$_{\text{Word1}}$] $_{\text{XP}}$[Given(A)$_{\text{Word2}}$ …</td></tr>
<tr><td>($_{\text{MaP}}$) ($_{\text{MaP}}$</td><td>($_{\text{MaP}}$ ---MaP Dephrasing ---</td></tr>
<tr><td>($_{\text{MiP}}$) ($_{\text{MiP}}$</td><td>($_{\text{MiP}}$) ($_{\text{MiP}}$</td></tr>
</table>

In addition, there is again an additional non-structural effect of FOCUS. In spite of a Major Phrase boundary at the left edge of Word2 of both the AAFN case and the AANN case, the F0 excursion size of the post-FOCUS Word2 of the AAFN case was consistently smaller than that of Word2 of the AANN case (i.e. the control case). For example, in MR's data the former was about 50 Hz smaller than the latter (Figure 8.12-a). I again interpret the difference between those two cases as the non-structural effect of FOCUS reducing/compressing the F0 movement associated with post-FOCUS words without changing the phonological phrase formation.

Example F0 contours of those four accented forms of MR (AANN, AAGG, AAFN and AAFG) are presented in Figure 8.13.

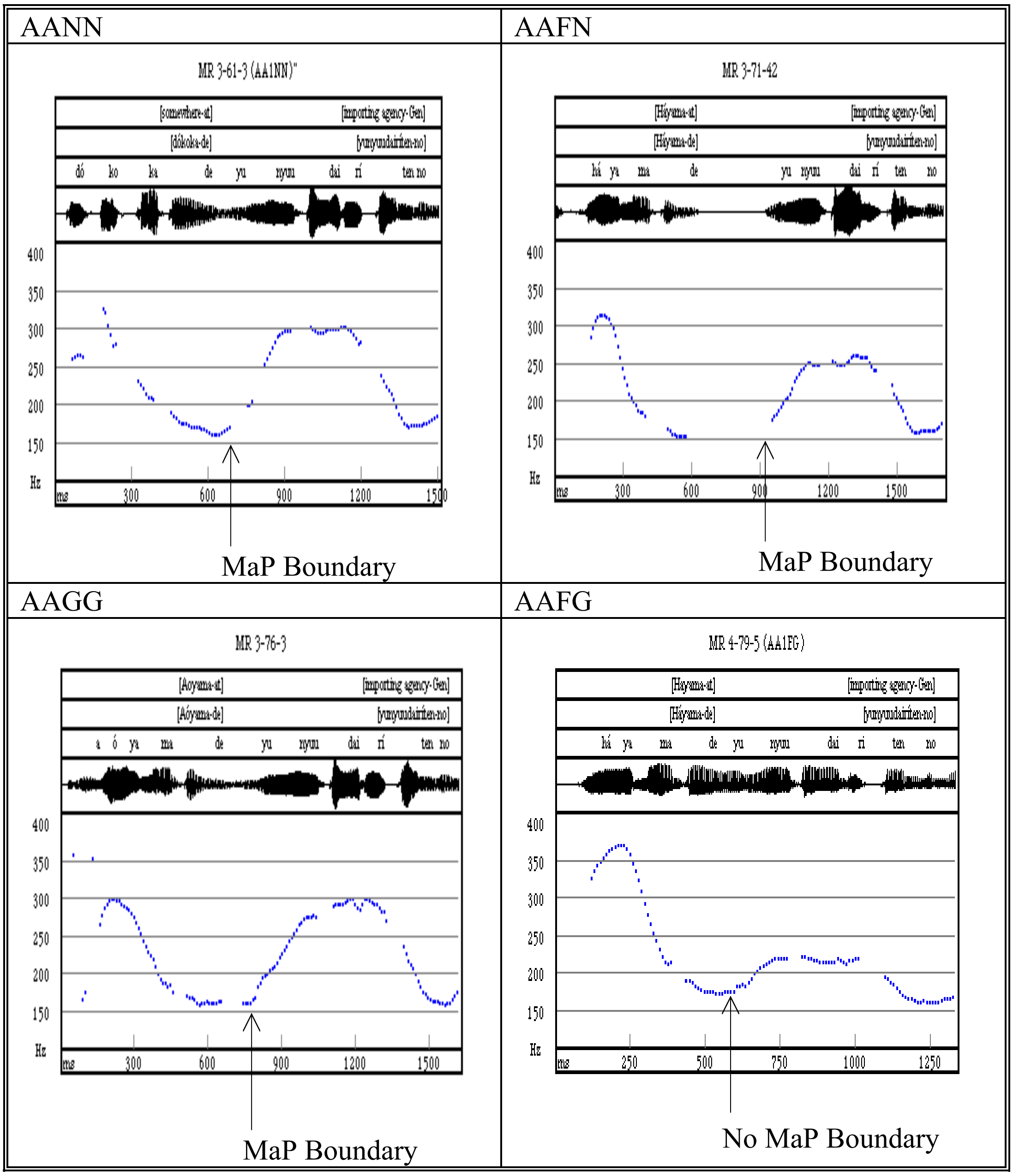


Figure 8.13. An Example F0 Contour of Word1 and Word2 of the AA Cases (Speaker MR)

8.2. Discussions

In the last section of this chapter, we found that FOCUS has both a structural and the non-structural effect on the immediately following word even across an XP boundary. FOCUS deleted the Major Phrase boundary of the following given word regardless of accent conditions (the structural effect). At the same time, even when no post-FOCUS dephrasing took place, F0 movement associated with the post-FOCUS word was still compressed (the non-structural effect). In this section, I focus on the structural effect and provide discussions and analyses of the effect.

8.2.1. The Post-FOCUS MaP Dephrasing

One of the major findings in this chapter is that FOCUS deletes the Major Phrase boundary of the post-FOCUS given word in spite of the presence of an XP boundary at the left edge of that word. This result suggests that the XP-MaP boundary alignment constraint proposed by Selkirk & Tateishi (1988, 1991), which is again shown in (8), is outranked by constraints that induce post-FOCUS dephrasing, i.e. FOCUS-ΔIP in (9) and ALIGN$_R$(Δ_{IP}, IP) in (10).[26]

(8) ALIGN$_L$ (XP, MaP)

Every XP left edge coincides with a Major Phrase left edge.

[26] Detailed discussions on those FOCUS-related constraints that play a crucial role in the post-FOCUS dephrasing are provided in Section 2.2. of Chapter 4.

(9) FOCUS-ΔIP (FOCUS-PROMIENCE)

The FOCUS-marked constituent in the morpho-syntactic representation should correspond to a string of phonological representation which contains the highest prominence (DTE,Δ) of an Intonational Phrase.

(10) ALIGN$_R$(Δ_{IP}, IP)

The DTE (Δ) of an Intonational Phrase must coincide with the right edge of an Intonational Phrase.

(11) FOCUS-ΔIP, ALIGN$_R$(Δ_{IP}, IP) >> ALIGN$_L$ (XP, MaP)

Given the constraint ranking in (11), the output form that best satisfies the FOCUS-related constraints in (9) and (10) is chosen as the grammatical representation. As a result, no Major Phrase boundary is present after FOCUS even when the post-FOCUS words coincide with an XP boundary at their left edges. This is shown in Tableau 8.01.

Tableau 8.01.

	[Wd1$_{FOCUS}$] $_{XP}$[Wd2] $_{XP}$[Wd3]	FOCUS-ΔIP	ALIGN$_R$ (Δ_{IP}, IP)	ALIGN$_L$ (XP, MaP)
a.	$_{IP}$ ( ) $_{MaP}$(______) $_{MaP}$( ) $_{MaP}$( ) Wd1 Wd2 Wd3 Δ_{IP}		*!* (2MaPs)	
b. ❍	$_{IP}$( ) $_{MaP}$(________________________) Wd1 Wd2 Wd3 Δ_{IP}			**

Heads of prosodic constituents are underlined.

8.2.2. The Given vs. New Contrast

Another significant finding is the contrast between the phonological phrase formation of the post-FOCUS given word and that of the post-FOCUS new word. The post-FOCUS given word underwent Major Phrase dephrasing while the new word blocked the dephrasing. This result supports the additional Focus-Prominence interface constraint proposed by Selkirk (2000b), which was introduced in Section 6.2.3 of Chapter 6. The constraint demands correspondence between a new (presentational focus) XP and the Major Phrase prominence (the DTE of MaP). This focus-MaP Prominence constraint is again shown in (12).

(12) focus(XP)-ΔMaP Selkirk (2002b)

The terminal string of a focus-marked XP (i.e. new XP) in the input syntactic representation must correspond to a terminal string in the output phonological representation which contains the DTE (Δ) of a prosodic constituent Major Phrase.

The focus-MaP Prominence constraint in (12) outranks the constraint in (10) that calls for no phonological phrase boundary after FOCUS, i.e. $\text{ALIGN}_R(\Delta_{IP}, IP)$. This constraint ranking and the result of the ranking are shown in Tableau 8.02.

Tableau 8.02.

$_{XP}$[Wd1(*FOC*)]$_{XP}$ $_{XP}$[Wd2(*foc*)]$_{XP}$	FOCUS-Δ_{IP}	focus(XP)-Δ_{MaP}	ALIGN$_R$ (Δ_{IP}, IP)
a. ❍ $_{IP}$ ( ) $_{MaP}$( ) $_{MaP}$( ) $_{MiP}$( ) $_{MiP}$( ) Wd1 Wd2 Δ_{IP} Δ_{MaP}			* (1 MaP) * (1 MiP)
b. $_{IP}$ ( ) $_{MaP}$( ) $_{MiP}$( ) $_{MiP}$( ) Wd1 Wd2 Δ_{IP}		*!	* (1 MiP)
c. $_{IP}$ ( ) $_{MaP}$( ) $_{MiP}$( ) $_{MiP}$( ) Wd1 Wd2 $\Delta_{IP(MaP)}$	*!		

Since absence of Major Phrase boundary at the left edge of the post-FOCUS new word with an XP boundary results in the lack of correspondence between the word and the DTE(Δ) of a Major Phrase, the optimal output form must have a Major Phrase boundary at the left edge of the post-FOCUS XP. The presence of the Major Phrase boundary leads to violation of the dephrasing constraint, ALIGN$_R$ (Δ_{IP}, IP). However, this violation is not fatal because ALIGN$_R$ (Δ_{IP}, IP) is less important than the focus XP-MaP prominence constraint.

8.2.3. The Lack of MiP Dephrasing

Finally, let us return to the first observation made in Section 8.1.3 that there was a Minor Phrase boundary at the left edge of both new and given post-FOCUS XPs. The post-FOCUS given word with an XP boundary underwent Major Phrase dephrasing, while its Minor Phrase boundary was always kept intact regardless of the information status and accent conditions of the post-FOCUS word.

For accented post-FOCUS words, it is already expected that there is a Minor Phrase boundary at its left edge regardless of whether it is given or new. This boundary is there because of the accent-MaP prominence constraint introduced in Section 7.3.3 of Chapter 7, ALIGN$_L$(H*+L, ΔMiP), is highly ranked. This constraint requires each accented word to coincide with the head of a Minor Phrase, which ultimately calls for a Minor Phrase boundary at the left edge of each accented word.

For an unaccented new post-FOCUS XP, the presence of a Minor Phrase boundary was also predicted by a constraint that we have already seen in the previous section, i.e. focus XP-ΔMaP. The focus XP-ΔMaP constraint requires the presence of a Major Phrase boundary at the left edge of the new post-FOCUS XP, and the presence of a Major Phrase boundary entails presence of a Minor Phrase boundary. [27]

An unaccented given post-FOCUS XP, however, needs an additional constraint to guarantee the presence of a Minor Phrase boundary at its left edge. I propose that in addition to the standard XP-MaP alignment constraint there is another XP-targeting alignment constraint which calls for a Minor Phrase boundary, i.e. ALIGN$_L$(XP, MiP).

(13) ALIGN$_L$ (XP, MiP)

Every XP left edge coincides with some Minor Phrase left edge.

[27] Here, I assume that EXHAUSTIVITY, one of the prosodic structure well-formedness constraints proposed by Selkirk (1995) is undominated, and the presence of a phonological phrase boundary of Level i must coincide with a phonological phrase boundary of Level i-1.

EXHAUSTIVITY (Selkirk 1995)
No phonological constituent of C_i dominates a phonological constituent of C_j, $j < i-1$.

This alignment constraint, then, outranks the FOCUS-*Dephrasing* constraint, i.e. ALIGN$_R$ (Δ_{IP}, IP), which requires deletion of all the post-FOCUS phonological phrase boundaries. This is shown in Tableau 8.03.

Tableau 8.03.

$_{XP}$[Word1(*FOC*)]$_{XP}$ – Acc $_{XP}$[Word2(given)]$_{XP}$ –Acc	FOCUS-Δ_{IP}	ALIGN$_L$ (XP, MiP)	ALIGN$_R$ (Δ_{IP}, IP)
b. ❍ $_{IP}$( ) $_{MaP}$( ) $_{MiP}$( ) $_{MiP}$( ) Wd1 Wd2 Δ_{IP}			*
c. $_{IP}$( ) $_{MaP}$( ) $_{MiP}$( ) Wd1 Wd2 Δ_{IP}		*!	

I assume that FOCUS-Δ_{IP} is undominated.
Heads of Intonational Phrase are underlined.

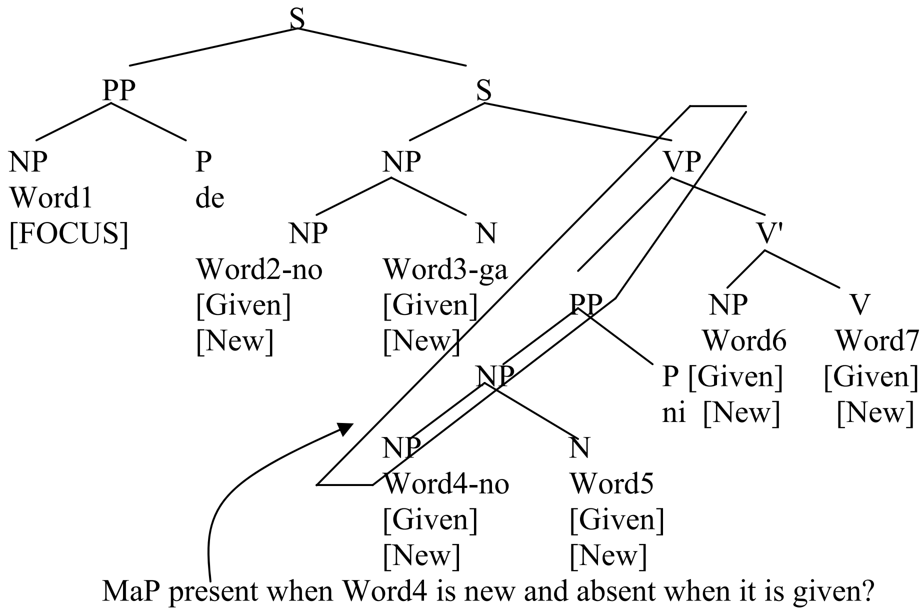
8.3. The Post-FOCUS Effect on a Given XP Further Away from FOCUS

In this section, we ask whether the presence and absence of a Major Phrase boundary at the left edge of a post-FOCUS new and given XPs, respectively, is even true in a later part of a sentence where those XPs are not adjacent to the preceding FOCUS. Sentence AA/XP-Even was again used for this purpose. There is an XP boundary at the left edge of not only Word2 but also Word4 in that sentence and we examine whether a Major Phrase boundary appear at the left edge of Word4 when it is interpreted as given and preceded by a FOCUS Word1.

Sentence AA/XP-Even **(XP boundary at the left edge of Word4)**

Word1 Word2 Word3 Word4 Word5 Word6
[*Háyama-de*] $_{S}$[$_{NP}$[*yunyuu-dairíten-no onnamanéejaa-ga*] $_{NP}$[*amerikájin-no uwakiáite-ni*] $_{NP}$[*mayaku-o*]]
Hayama-at importing-agency-Gen female manager-Nom American-Cop lover-to marijuana-Acc …
"In Hayama the female manager of an importing agency (sold) marijuana to an American, her lover."

(14)



To test for the presence and absence of a Major Phrase boundary at the left edge of Word4 in the FN and the FG context (i.e in the context where Word1 is a FOCUS and Word4 is interpreted as "new" or given respectively), the F0 excursion size of Word4 of those FOCUS cases was obtained. Then, an ANOVA test analogous to those adopted in Section 8.1.3 of this chapter was carried out. In the ANOVA analysis, the given vs. new factor and the control vs. post-FOCUS factor are predictors and F0 excursion size of the target word was the dependent variable. If there is an interaction between the two predictors, i.e. if the F0 excursion size difference between the FG case and the FN case is greater than that between the GG and the NN case as shown in

Figure 8.14, then we can conclude that there is no Major Phrase boundary at the left edge of the post-FOCUS given Word4.

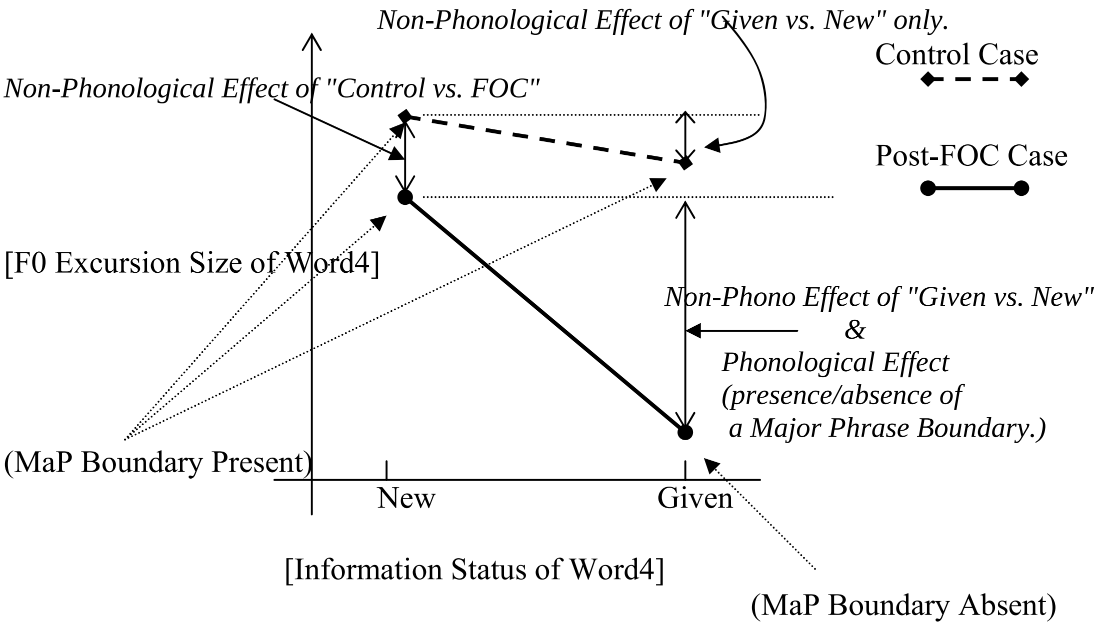
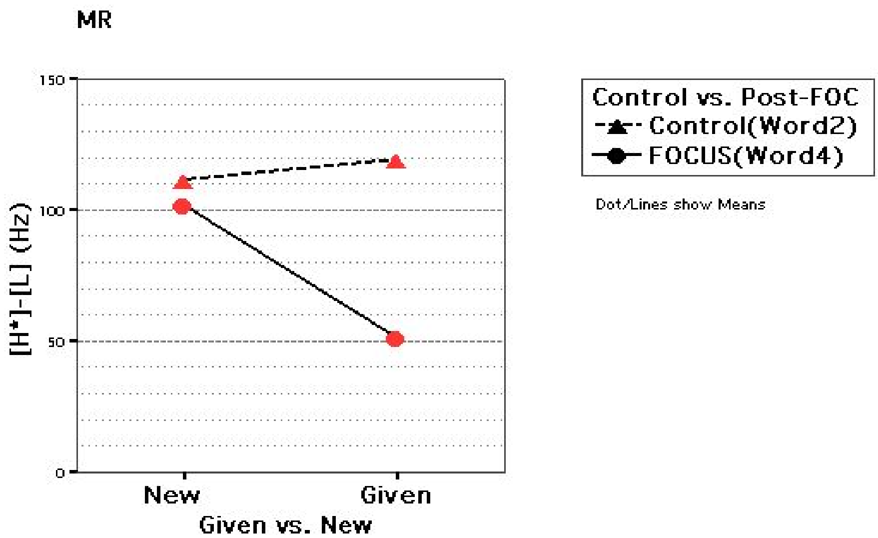


Figure 8.14. The Expected Interaction between the Given vs. New Factor and the Control vs. Post-FOCUS Factor

There is one problem in carrying out this comparison, however. The GG case control case that we have is a sentence in which both Word1 and Word2 are interpreted as given followed by a FOCUS Word3 and given Word4 (see the GG context, Section 8.1.1). That is, the given Word4 of the GG control case is preceded by a FOCUS Word3, and we cannot use the F0 excursion size of that given Word4 as an appropriate control case for this comparison. Facing this problem, a compromise was made: I decided to compare the F0 excursion size of Word2 of the GG and the NN control case with that of Word4 of the FG and the FN case. Since Word2 also coincides with an XP and we already know that Word2 has a Major Phrase boundary in those control cases, it could be still used as a control for the comparison that we are going to carry out.

<Results>

There was an interaction between the control vs. FOCUS factor and the given vs. new factor (MR: $F(1,33) = 28.881$, $^*p < 0.001$; RO: $F(1,33) = 5.839$, $^*p = 0.021$; SK: $F(1,18) = 30.572$, $^*p < 0.001$). Line plots in Figure 8.15 graphically show this interaction.



		N
TYPE	AAFG	10
	AAFN	7
	AAGG	10
	AANN	10

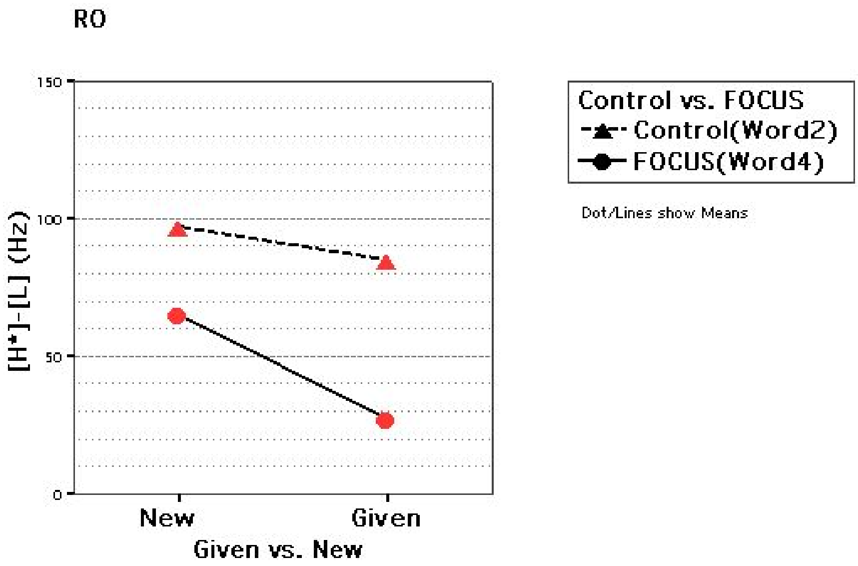
	Mean	Std. Error	95% Confidence Interval	
TYPE			Lower Bound	Upper Bound
AAFG	51.300	5.128	40.866	61.734
AAFN	102.000	6.130	89.529	114.471
AAGG	118.900	5.128	108.466	129.334
AANN	111.600	5.128	101.166	122.034

a. Speaker MR

Figure 8.15. The Mean F0 Difference between L and H* of Word4 (New vs. Given, Control vs. Post-FOCUS AA)

Continued next page

Figure 8.15 continued



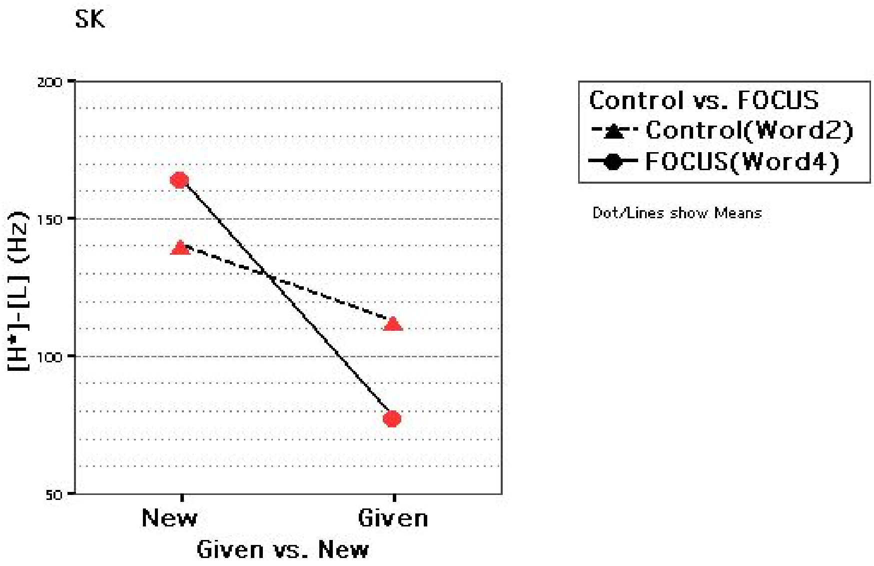
		N
TYPE	AAFG	9
	AAFN	8
	AAGG	11
	AANN	9

	Mean	Std. Error	95% Confidence Interval	
TYPE			Lower Bound	Upper Bound
AAFG	27.222	5.401	16.234	38.211
AAFN	65.250	5.729	53.595	76.905
AAGG	85.000	4.885	75.060	94.940
AANN	97.111	5.401	86.123	108.100

b. Speaker RO

Continued next page

Figure 8.15 continued



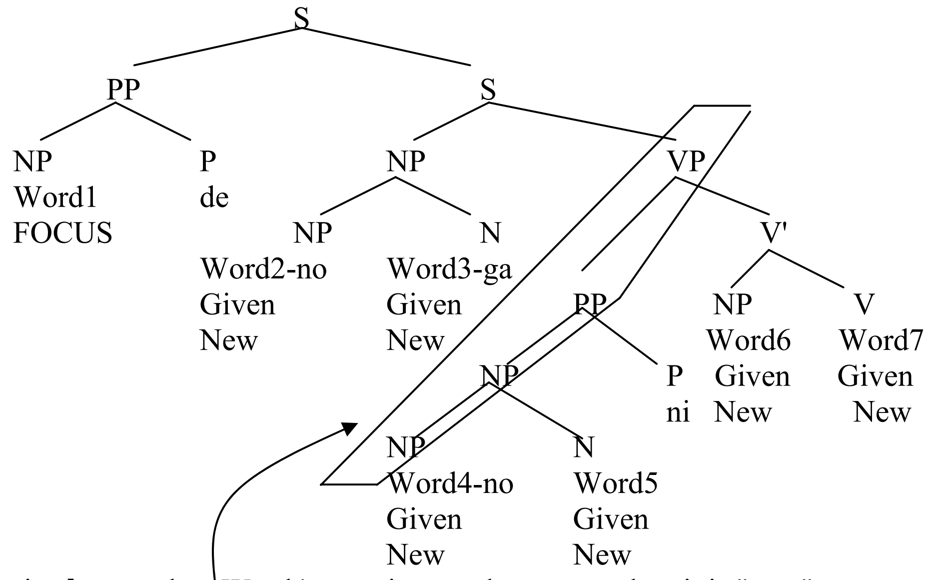
		N
TYPE	AAFG	9
	AAFN	3
	AAGG	5
	AANN	5

	Mean	Std. Error	95% Confidence Interval	
TYPE			Lower Bound	Upper Bound
AAFG	77.333	3.871	69.201	85.465
AAFN	164.333	6.704	150.248	178.418
AAGG	112.400	5.193	101.490	123.310
AANN	140.400	5.193	129.490	151.310

c. Speaker SK.

It is evident from those plots that the F0 excursion size difference between the post-FOCUS given and the post-FOCUS new Word4 was far greater than that between the control given and the control new Word2. At the same time, the F0 excursion size of the post-FOCUS given Word4 was far smaller than that of post-FOCUS new Word4. I take this result as evidence for lack of a Major Phrase boundary at the left edge of the post-FOCUS given Word4 while that boundary is present at the post-FOCUS new Word4. This conclusion is summarized in (15).

(15)



MaP Boundary is **absent** when Word4 was given and **present** when it is "new".

<Discussion>

In summary, we obtained the same results as those obtained in Section 1.3 of this chapter. That is, no Major Phrase boundary appeared at the left edge of the given post-FOCUS XP while one appeared at the left edge of the new post-FOCUS XP. This confirms the FOCUS-related constraints that induce post-FOCUS dephrasing, i.e. FOCUS-ΔIP and ALIGN$_R$(Δ_{IP}, IP), and the focus XP-MaP Prominence constraint that demands correspondence between a new XP and the DTE of a Major Phrase, i.e. focusXP-ΔMaP.

8.4. Chapter Conclusion

We have considered in this chapter the effects of FOCUS on a post-FOCUS part of an utterance across XP boundaries. We found both structural and non-structural effects of FOCUS there. In what follows, let us review and discuss some of the core findings made in this chapter.

Unlike the structural effect of FOCUS within an XP that we saw in Chapter 7, there was no post-FOCUS Minor Phrase dephrasing, i.e. no coalescence of the following word and a FOCUS into a single Minor Phrase. However, FOCUS still had a structural effect of Major Phrase dephrasing across XP boundaries, i.e. deletion of Major Phrase boundaries in a post-FOCUS position. Unlike the post-FOCUS Minor Phrase dephrasing of Chapter 7 which equally took place for both given and new post-FOCUS words, this post-FOCUS Major Phrase dephrasing was not automatic: it only took place when post-FOCUS XPs were given. It was blocked when post-FOCUS XPs were interpreted as new. This is not simply explained by a constraint ranking in which a constraint calling for post-FOCUS Major Phrase dephrasing outranks an XP-MaP alignment constraint because such ranking predicts that both given and new post-FOCUS XPs should undergo Major Phrase dephrasing.

I solved this problem by adopting Selkirk's (2000a) idea that the phonological grammar refers to a syntactic feature [focus] (i.e. newness of a morpho-syntactic constituent) as well as [FOCUS] (i.e. Contrastive Focus). According to Selkirk, there is a constraint that requires a Major Phrase boundary at the left edge of each [focus] XP, i.e. new XP. I proposed that this focus XP-MaP constraint outranks the post-FOCUS Major Phrase dephrasing constraint while the standard XP-MaP alignment constraint is dominated by the dephrasing constraint. This constraint ranking results in the presence of a Major Phrase boundary at the left edge of the new post-FOCUS XP while the absence of the boundary at the left edge of the given post-FOCUS XP.

Another contribution of this chapter as well as the last chapter (Chapter 7) is the finding that information structure (such as FOCUS, new vs. given distinction) have not only structural but also non-structural effects. Information structure not only determines the phonological representation via syntactic representation but also directly affects the phonetic representation/interpretation. We found that a post-FOCUS new XP was associated with smaller F0 excursion size than a new XP preceded by no FOCUS in spite of the fact that both of them equally had a Major Phrase boundary at their left edge. In the same way, even in a control context without FOCUS, a given item was associated with a smaller F0 excursion size than a new item despite the fact that they equally had the same phonological phrase boundary at their left edge. This is only explained if information structure is capable of directly influencing the phonetic representation/interpretation without affecting the phonological representation.

CHAPTER 9

CONCLUDING REMARKS

I provided in this dissertation empirical and theoretical analyses of downtrends and reduction of F0 movement focusing on several factors. One of them is the passage of time, which affects how low the base line slope of a pitch range reaches (Chapter 3). I showed that there is not only a global declination slope which unfolds across phonological phrase boundaries but also a more local tone-bound declination slope that is associated with each tone. This local tone-bound declination slope is a new finding.

Another factor is the phonological representation of a sentence, which includes hierarchical organization of phonological constituents and tonal configurations. The hierarchical structure of phonological constituents is partially determined by the input syntactic representation such as the presence/absence of XP boundaries, Focus features ([FOCUS] and [focus]), and the lexically provided pitch accent H*+L (Chapter 6, Chapter 7, Chapter 8). FOCUS corresponds to the most prominent mora of an Intonational Phrase, and the correspondence between FOCUS and prominence together with additional prosodic requirements leads to deletion of post-FOCUS phonological phrase boundaries. The lack of phonological phrase boundaries, then, leads to reduction of F0 movements. For example, the lack of a Minor Phrase boundary leads to absence of edge tones and the absence of edge tones results in a flat F0 movement. Also, the lack of higher phrase boundaries such as a Major Phrase boundary results in "weaker" values of edge tones (i.e. higher F0 for L tones and lower F0 for H tones) and smaller F0 excursion size between those tones.

The other factor is the semantic/pragmatic interpretation of the sentence such as new and given information and contrastive FOCUS (Chapters 7 and 8). I showed in those chapters that even when phonological representation is kept constant for post-FOCUS items and non-post-FOCUS items, the left edge of the post-FOCUS phonological phrase was marked with smaller F0 excursion size between the L and H edge tones than that of the non-post-FOCUS phonological phrase. Also, we found that a phonological phrase associated with given information was marked with smaller F0 excursion size between those edge tones than that associated with new information.

In addition to those factors, the interaction between two neighboring tones was also considered as one of the factors that induces a downtrend in Japanese, i.e. the post-pitch accent downstep known as *catathesis* (Chapter 4). In that chapter, I proposed a local tone-by-tone scaling model: the F0 value of each tone is a function of the F0 value of the immediately preceding tone as well as the F0 of the base line (for L tones) or the top line (for H tones) of a pitch range. Those tone-by-tone scaling rules are expressed as a function that contains a coefficient which varies between 0 and 1. This coefficient value determines the "salience" of each tonal value. For H tones, the greater the coefficient value is, the more salient (the higher) the H tone F0 is. For L tones, however, the smaller the coefficient value is, the more salient (the lower) the L tone is. This coefficient value, or the saliency of each tone, is determined by factors such as the kind of the target tone whether it is part of a pitch accent (more salient) or an edge tone (less salient). Its value is also determined by phrase boundary strength, phrasal prominence, and information status of words with which the target tone is associated. Though detailed formalization of the relationship between the coefficient values (the tonal saliency) and those higher phrasal and information factors was left open for our future studies, I consider this proposal to be the first step for understanding how those phonological and semantic factors should be interpreted by the phonetics and reflected on the surface F0 values of tones.

In the following figure, I summarize how those factors considered in this dissertation contribute to the surface downtrends and F0 excursion size reduction.

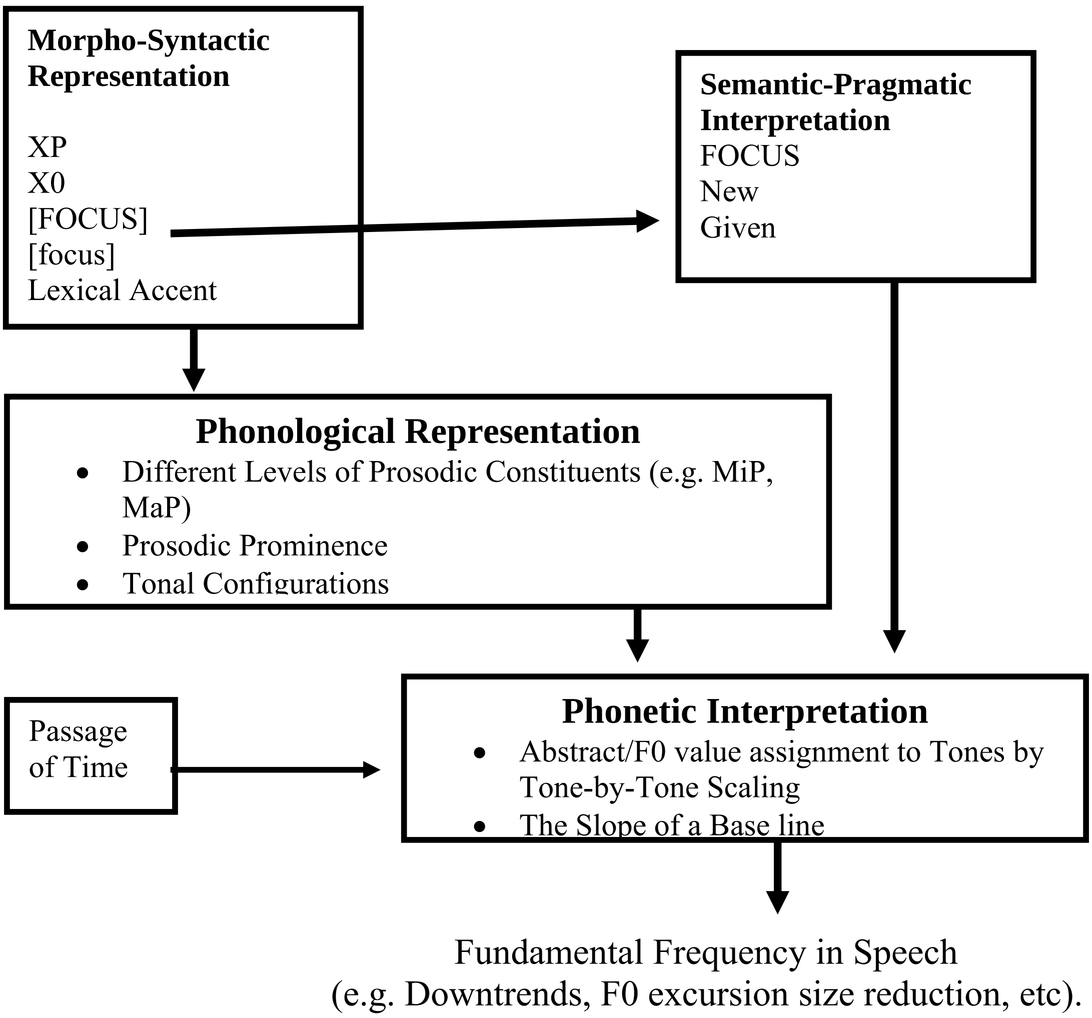


Figure 9.01. Factors Influencing the Fundamental Frequency in Speech

APPENDIX

DATASETS

Dataset I: The <*Umaya*> Set

The <*umaya*> set is a paradigm of phrases designed to show the effect of time on tonal scaling and used in Chapter 3. The target sequence of words in those sentences consists of two words, each of which is supposed to form its own Minor Phrase. The syllable number of the initial word of the sequence was varied to see the effect of the duration change on the F0 value of tones associated with the second word.

(a)	*yama-no*	*umaya-no*	…
	mountain-Gen	barn-Gen	…
(b)	*yamamura-no*	*umaya-no*	…
	(village name)-Gen	barn-Gen	…
(c)	*yamanakamura-no*	*umaya-no*	…
	(village name)-Gen	barn-Gen	…
(d)	*yamanakagawamura-no*	*umaya-no*	…
	(village name)-Gen	barn-Gen	…
(e)	*minamiyamanakagawamura-no*	*umaya-no*	…
	(village name)-Gen	barn-gen	…

Each of those expressions was embedded in the sentence shown below.

chikágoro,	___(target form)___	*yaneura-no/amádo-no*	*kaischiku-o meguri,*
recently,	___(target form)___	attic-Gen/storm window-Gen	reform-Acc about,
juumin-no	*íken-ga*	*wareteiru-yóodesu.*	
residents-Gen	opinions-Nom	disagrees-it seems.	

"It seems that the residents recently disagree with each other regarding the repair of the attic/storm window of __ (target form)__."

Dataset II: The <*Maronmónaka*> Set

The <*maronmónaka*> set is a paradigm of compound nouns designed to show the effect of duration between two neighboring H tones (the H edge tone and the H* accent tone) on the F0 of the H* accent tone. They appear in Chapter 3. Compound nouns were ideal for this investigation because it is possible to vary the syllable number between a H- edge tone and a H* accent tone just by adding a word with different length between the initial and the final member of the same compound noun.

(a) *maronmónaka-o*
chestnut monaka-Acc

(b) *maroniromónaka-o*
chestnut-colored monaka-Acc

(c) *maronkuriimumónaka-o*
cream-cream monaka-Acc

(d) *maronaisukuriimumónaka-o*
chestnut-ice cream monaka-Acc

Those words were embedded in the following sentence.

áruhi	*Taro-ga*	*(adjective forms)*	*ringo-o*	*bóribori*	*tábeteimasita.*
one day	Taro-Nom		apple-Acc	crunching on	

suruto	*_____(target form)_____-o*	*kago ippai-ni mótta*	*onnánoko-ga*	*toorikakarimásita.*
then	______________________-Acc	full of basket	girl-Nom	passed by.

"One day, Taro was crunching on (---) apples. Then, a girl with a basket full of ___(target form)___ passed by him."

Dataset III: The <*Ánnasan*> Set

The <*ánnasan*> set is a paradigm of phrases designed to show the effect of duration between the H*+L pitch accent and the following L edge tone on the F0 value of the L tone (Chapter 3). Initial accented person name nouns are used as Word1. Each of those name nouns were different in terms of the number of morae/syllables between their initial accented mora of Word1 and the beginning of the following Word2, where the L edge tone appears.

(a)	*án(na)-san-no*	*omiaiáite…*
	Anna-Ms.-Gen	date
(b)	*márinamu-san-no*	*omiaiáite…*
	Marinam-Ms.-Gen	date
(c)	*ándoryuu-san-no*	*omiaiáite…*
	Andrew-Mr.-Gen	date

Each of those words was embedded in the following sentence template.

konokatá-wa ____*(target forms)*_____*-no* *amerikájin-de* *Jan-san-to iimásu.*
this person-Top ____________________-Copula American Jan Mr/Ms. Jan called.
"This person is an American who is __________________, and called Ms/Ms. Jan.

Dataset IV: The <*Omiaiáite*> Set

The <*omiaiáite*> set is a paradigm designed to investigate post-accent downtrend (Chapter 4), and consists of two forms. One of them is from the previous dataset, i.e. the <*ánnasan*> set, consisting of two accented words; the other consists of the initial unaccented word and the following accented word. They were also embedded in the sentence template shown in 4.2.3.

(a) **AA**
Án(na)-san-no omiaiáite-no ---
Ms. An(na)-Gen date-Copula ---
"(an American) who is a date of Ms. Anna"

(b) **UA**
Manaeda-no omiaiáite-no ---
Ms/Mr. Manaeda-Gen date-Copula ---
"(an American) who is a date of Ms/Mr. Manaeda"

Dataset V: The <*Maronkéeki*> Set

The <*maronkéeki*> set is also a paradigm designed to investigate post-accent downtrend (Chapter 4), especially to test the model proposed and predictions made in that chapter. The set consists of four forms. The location and presence/absence of accent varies across those forms.

(a) **UAA**
nijukko-iri-no *yuuháimu-no* *maronkéeki*
twenty-pieces-all-in-one box-Copula Juheim-Gen chestnut cake
"twenty pieces of chestnut cake of Juheim in a single box"

(b) **AAA**
nihyakúen-no *yuuháimu-no* *maronkéeki*
two hundred yen Copula Juheim-Gen chestnut cake
" chestnut cake of Juheim of two hundred yen"

(c) **AUA**
nihyakúen-no *imuraya-no* *maronkéeki*
two hundred yen-Copula Imuraya-Gen chestnut cake
"chestnut cake of Imuraya of two hundred yen"

(d) **UUA**

nijukko-iri-no *imuraya-no* *maronkéeki*
twenty-pieces-all-in-one box-Copula Imuraya-Gen chestnut cake
"twenty pieces of chestnut cake of Imuraya in a single box

Both were embedded in the following sentence.

asoko-ni áru ____(target)___ -o hitóhako kudasái
"please give me a box of ___________ which is right over there"

Dataset VI: The <*Yonjúuen*> Set

This dataset consists of two forms in (a) and (b). They are both numeral expressions which denote the price in Japanese yen, and used in Chapter 4.

(a)	*sanmán*	*yonjúuen*
	thirty thousand	forty yen
(b)	*sánbyaku*	*yonjúuen*
	three hundred	forty yen

Both of those forms were embedded in a template:

asoko-ni áru ___(target form)____-no sukáafu-o tóttekudasái

"please pass me _______ over there."

Dataset VII: The <*Yunyuu*> Set

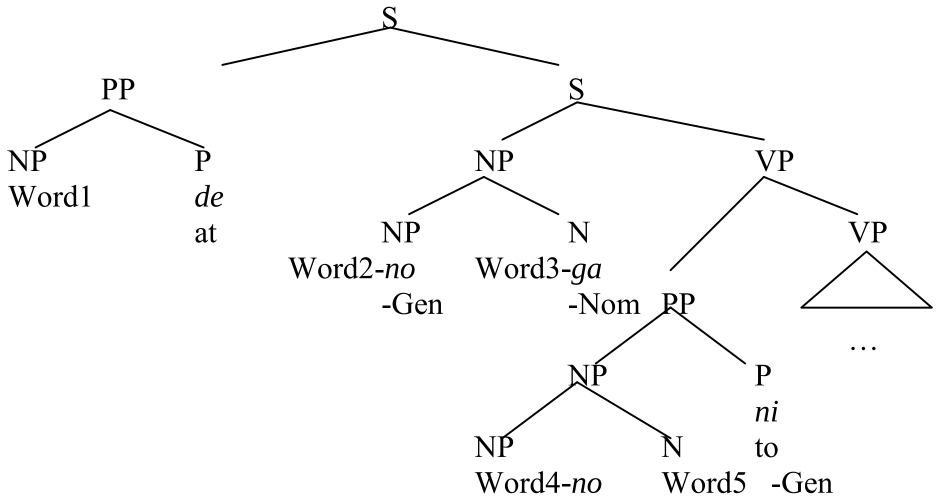
The <*yunyuu*> set is designed to investigate the effect of FOCUS on the F0 scaling of post-FOCUS words, and the effect of "given" vs. "new" distinction on F0 scaling (Chapter 7 and Chapter 8). Sentences in the set varied in terms of their (i) information structure (context), (ii) accent conditions and (iii) syntactic structures. All of those sentences were put into a dialogue to give the right information structure to them. There are four information structures (contexts): (A) all neutral (new) context referred to as NNNN; (B) FOCUS and post-FOCUS neutral (new) items referred to as FNNN; (C) FOCUS and post-FOCUS given items referred to as FGGG; (D) two given items followed by FOCUS and given items referred to as GGFG. There are mainly two accent conditions: all accented, referred to as AA, and all unaccented referred to as UU. There are two syntactic structures but with the same lexical items in the same order. Each word is numbered such as Word1, Word2, and one of the syntactic structures has an XP boundary at the left edge of odd-numbered words (XP-Odd). The other syntactic structure has an XP boundary at the left edge of Word1 and that of the even-numbered words (XP-Even).

Lexical Items and Accent Conditions

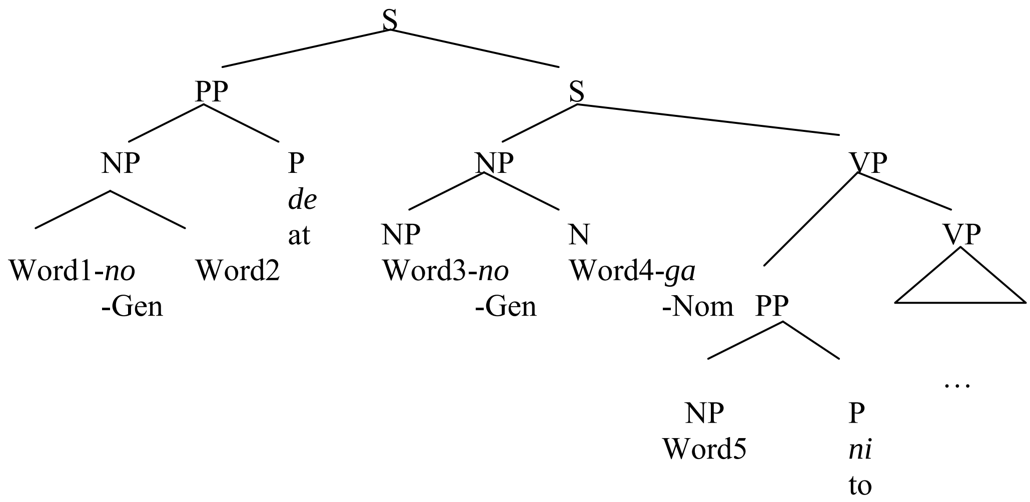
	Word1	Word2	Word3	Word4	Word5
AA:	*dókoka/Háyama/Aóyama*	*yunyuudairíten*	*onnamanéejaa*	*amerikájin*	*uwakiite* ---
	"somewhere/Hayama/Aoyama	importing agency	female manager	American	lover ---
UU:	*Yokohama/Setagaya*	*yunyuudaikooya*	*maneejaa*	*yuujin*	*joyuu* ---
	"Yokohama/Setagaya	importing agency	manager	friend	actress ---

Syntactic Conditions

XP-Odd



XP-Even



Context FNNN

Experimenter: *nani?*
"What?"

Speaker: *Tookyoo-ya Oosaka-no yoona hanzai-no tahatsusuru tóshi-dewanaku,*
Tokyo-and Osaka like high crime rate city-Not,

tian-no yóikotode sirareru, ano **[*YOKOHAMA-DE*]$_{\mathbf{Wd1}}$**
safety-Gen good known, that YOKOHAMA-AT,

[*yunyuu-daikooya-no*]$_{\mathbf{Wd2}}$ *maneejaa-ga*
Gen-agency-Gen manager-Nom

yuujin-no moderu-ni mayaku-o uttarasíiyo.
friend-Copula model-to marijuana-Acc sold-I heard.

"I've heard that in YOKOHAMA, the city known to be safe unlike those cities like Osaka or Tokyo, which are notorious for their high crime rate, a manager of an importing agency sold marijuana to a model who is a friend of his/hers."

Context FGGG

Experimenter: *Aóyama-de yunyuu-daikooya-no maneejaa-ga yuujin-no moderu-ni*
marifana-o uttá-to kiitaga, hontookánee?
"I've heard that in Aoyama a manager of an importing agency sold marijuana to a model who is a friend of his/hers. But is it true?"

Speaker: *chigaimásu. Aóyama-dewa-naku* **[*YOKOHAMA-DE*]$_{\mathbf{Wd1}}$**
"No" Aoyama-Copula-Not YOKOHAMA-AT

[*yunyuu-daikooya-no*]$_{\mathbf{Wd2}}$ *maneejaa-wa*
importing-agency-Gen manager-Topic

yuujin-no moderu-ni mayaku-o uttá-ndesu.
friend-Gen model-to marijuana-Acc sold-Copula.

"In YOKOHAMA but not in Aoyama, a manager of an importing agency sold marijuana to a model who is a friend of his/hers.

Context NNNN

Speaker: *chotto kiiteyo.*
"Hey, just listen to me."

Experimenter: *nani?*
"What?"

Speaker: **[*Yokohama-de*]$_{\mathbf{Wd1}}$** **[*yunyuu-daikooya-no*]$_{\mathbf{Wd2}}$** *maneejaa-ga*
Yokohama-at importing-agency-Gen manager-Nom

yuujin -no *moderu-ni* *mayaku-o* *uttarashíiyo.*
friend-Cop model-To marijuana-Acc sold-I heard.

"I've heard that in Yokohama the manager of an importing agency sold marijuana to a model who is his/her friend.

Context GGFG

Experimenter: *Yokohama-de yunyuu-daikooya-no juugyóoin-ga yuujin-no moderu-ni mayaku-o uttarashíi-ga, hontookane?*
"Is it true that in Yokohama the employee of an importing agency sold marijuana to a model who is his/her friend?

Speaker: *iie, chigaimásu.*
"No, that's not correct."

[*Yokohama-de*]$_{\mathbf{Wd1}}$ **[*yunyuu-daikooya-no*]$_{\mathbf{Wd2}}$** *MANEEJAA-GA*

Yokohama-at importing-agency-Gen manager-Nom

yuujin -no *moderu-ni* *mayaku-o* *uttá-ndesu.*
friend-Cop model-To marijuana-Acc sold-Copula.

"The MANAGER of an importing agency sold marijuana to a model, his/her friend, in Yokohama."

The second word of both UU (and AA) consists of a compound noun *yunyuudaikooya (yunyuudairíten)* "importing agency". The compound word was intentionally adopted to examine whether presence of a preceding FOCUS (Word1 as a FOCUS) has any effect of "deleting" a Minor Phrase boundary at the left edge of Word2 when there is no XP boundary between those two words. For this purpose, the presence of a Minor Phrase boundary had to be guaranteed at the left edge of Word2 of the neutral case (NNNN) and the GGFG case, which were treated as "control" cases. As mentioned in the introductory chapter (Chapter 7), a sequence of unaccented words are likely to put into a single Minor Phrase even in a neutral context when no XP boundary is present between those two words and both of those two words are relatively short, consisting of a single stem. To avoid such unwanted "dephrasing" in the neutral context (and in the GGFG context), a relatively long compound *yunyuudaikooya* was used as Word2. Also, to make forms as parallel as possible, a compound *yunyuudairíten* was also adopted for the AA case (accented case).

BIBLIOGRAPHY

Banno, M. (1999). Towards a discourse model of prosodic structure: a prosodic and syntactic analysis of declination in Japanese. Ph.D dissertation. University of California, Santa Barbara.

Beckman, M. E. and J. Pierrehumbert (1986). "Intonational Structure in English and Japanese." Phonology 3: 225-309.

Beckman, M. E. (1995). "Local Shapes and Global Trends". Proceedings of ICPhS 95. Stockholm. 2.

Berg, R. v. d., C. Gussenhoven and T. Rietveld (1992). "Downstep in Dutch: Implications for a Model". Papers in Laboratory Phonology, II: Gesture, Segment, Prosody. G. J. Docherty and R. D. Ladd (eds.) Cambridge, UK, Cambridge University Press: 335-367.

Botinis, A. (1998). "Intonation in Greek". Intonation Systems. D. Hirst and A. D. Cristo, Cambridge University Press.

Bruce, G. (1977). Swedish Word Accents in Sentence Perspective. Lund, Gleerup.

Cho, T. and P. Keating (2001). "Articulatory and Acoustic Studies on Domain-Initial Strengthening in Korean." Journal of Phonetics 29: 155-190.

Cho, Y.-M. Y. (1990). "Syntax and Phrasing in Korean". The Phonology-Syntax Connection. S. Inkelas and D. Zec (eds.) Chicago, CSLI: 47-62.

Clark, M. M. (1990). The Tonal System of Igbo. Dordrecht, Foris Publications.

Clements, G. N. and K. C. Ford (1979). "Kikuyu Tone Shift and Its Synchronic Consequences." Linguistic Inquiry 10: 179-210.

Collier, R. (1975). "Physiological Correlates of Intonation Patterns." Journal of the Acoustical Society of America 58: 249-255.

Collier, R. (1985). "F0 Declination: The Control of Its Setting, Resetting, and Slope". Laryngeal Function in Phonation and Representation. T. Baer, C. Sasaki and K. S. Harris (eds.) Boston, A College-Hill Publication. 39-40: 137-170.

Collier, R. and C. E. Gelfer (1984). "Physiological explanations of F0 declination". Proceedings of the Xth International Congress of Phonetics Sciences. M. P. R. v. d. Droecke and A. Cohen (eds.) Dordrecht, Foris Publications.

Collier, R. and J. t'Hart (1975). "A Grammar of Pitch Movements in Dutch Intonation." IPO Annual Reports (6): 17-21.

Cristo, A. D. and L. Jankowski (1999). "Prosodic Organization and Phrasing after Focus in French." Proceedings from ICPhS 99: 1565-1568.

Dilly, L. and S. Shattuck-Hufnagel (1996). "Glottalization of Word-Initial Vowels as a Function of Prosodic Structure." Journal of Phonetics 24: 423-444.

Fougeron, C. (1997). "Articulatory Strengthening at Edges of Prosodic Domains." Journal of the Acoustical Society of America 101(6): 3728-3740.

Fougeron, C. (2001). "Articulatory Properties of Initial Segments in Several Prosodic Constituents in French." Journal of Phonetics 29: 109-136.

Fougeron, C. and S.-A. Jun (1998). "Rate Effects on French Intonation: Prosodic Organization and Phonetic Realization." Journal of Phonetics 26: 45-69.

Gårding, E. (1987). "Speech Act and Tonal Pattern in Standard Chinese." Phonetica 44.

Gårding, E. (1993). "Focal Domains and Their Manifestations in Some Swedish Dialects." Nordic Prosody VI: 65-75.

Gelfer, C. R., K. S. Harris and T. Baer (1985). "Controlled Variables in Sentence Intonation". Laryngeal Function in Phonation and Representation. T. Baer, C. Sasaki and K. S. Harris (eds.) Boston, A College-Hill Publication: 422-435.

Gelfer, C. R., K. S. Harris, R. Collier and T. Baer (1983). "Is Declination Activity Controlled?". Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control. I. R. Titze and R. C. Scherer (eds.) Denver, The Denver Center for the Performing Acts.

Ghini, M. (1993). "Phi-Formation in Italian: A new Proposal." Toronto Working Papers in Linguistics 12(2): 41-78.

Goldsmith, J. (1976). Autosegmental Phonology. Ph.D dissertation. MIT, published by Garland Press, 1979, Cambridge.

Grønnum (Thorsen), N. (1980). "Intonation Contours and Stress Group Patterns in Declarative Sentences of Varying Length in Asc Danish." ARIPUC 14.

Grønnum (Thorsen), N. (1989). "Stress Group Patterns, Sentence Accents and Sentence Intonation in Southern Jutland (Sønderborg and Tønder) - with a View to German." ARIPUC 23: 1-85.

Grønnum (Thorsen), N. (1995). "Superposition and Subordination in Intonation: A Non-Linear Approach". Proceedings of ICPhS 1995. Stockholm. 2: 124-131.

Gussenhoven, C. (1984). On the Grammar and Semantics of Sentence Accents. Dordrecht, Foris Publications.

Gussenhoven, C. and T. Rietveld (1988). "Fundamental Frequency Declination in Dutch: Testing Three Hypotheses." Journal of Phonetics 16(3): 355-369.

Haraguchi, S. (1977). The Tone Structure in Japanese: an Autosegmental Theory of Tonology. Tokyo, Kaitakusha.

Hayes, B. (1995). Metrical Stress Theory. Chicago, The University of Chicago Press.

Hayes, B. and A. Lahiri (1991). "Bengali Intonational Phonology." Natural Language and Linguistic Theory 9: 47-96.

Homma, Y. (1973). "An Acoustic Study of Japanese Vowels." Study of Sounds 16: 347-368.

Jackendoff, R. (1972). Semantic Interpretation in Generative Grammar. Cambridge, MA, MIT Press.

Jun, S.-A. (1993). The Phonetic and Phonology of Korean Prosody. Ph.D dissertation. Ohio State University.

Jun, S.-A. (1998). "The Accentual Phrase in the Korean Prosodic Hierarchy." Phonology 15: 189-226.

Jun, S.-A. and C. Fougeron (2000). "A Phonological Model of French Intonation". Intonation: Analysis, Modeling and Technology. A. Botinis (ed.) Dordrecht, Kluwer Academic Publishers.

Jun, S.-A. and M. Oh (1996). "A Prosodic Analysis of Three Types of Wh-Phrases in Korean." Language and Speech 39(1): 37-61.

Kenstowicz, M. and H.-S. Sohn (1997). "Phrasing and Focus in Northern Kyungsang Korean". PF: Papers at the Interface, MITWPL. B. Bruening, Y. Kang and M. McGinnis (eds.), Department of Linguistics, MIT. 30: 25-47.

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

Kubozono, H. (1995). "Constraint Interaction in Japanese Phonology: Evidence from Compound Accent". Phonology at Santa Cruz, Department of Linguistics, University of California at Santa Cruz. 4: 21-38.

Kubozono, H., J. Ito and A. Mester (1995). "On-in Koozoo kara mita Go to Ku no Kyookai". Bunpoo to Onsei. Tokyo, Kuroshio Publishers.

Lacy, P. d. (1999). Tone and Prominence. University of Massachusetts, Amherst.

Ladd, R. D. (1980). The Structure of Intonational Meaning: Evidence from English. Bloomington, Indiana University Press.

Ladd, R. D. (1990). "Metrical Representation of Pitch Register". Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech. J. Kingston and M. E. Beckman (eds.) Cambridge, Cambridge University Press: 35-57.

Ladd, R. D. (1995). "Linear and Overlay Descriptions: an Autosegmental-Metrical Middle Way". Proceedings of ICPhS 1995. Stockholm. 2: 116-123.

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

Leben, W. (1973). Suprasegmental Phonology. Ph.D dissertation. MIT.

Leben, W. (1976). "The Tones in English Intonation." Linguistic Analysis 2: 69-107.

Le Gac, D. (2002). "Tonal Alternations and Prosodic Structure in Somali". Speech Prosody 2002 Proceedings. B. Bel and I. Marlien (eds.) Aix-en-Provence: 455-458.

Liberman, M. (1975). The Intonational System of English. Cambridge, MIT Press.

Liberman, M. and A. Prince (1975). "On Stress and Linguistic Rhythm." Linguistic Inquiry 8: 249-336.

Liberman, M., M. Schltz, S. Hong and V. Okeke (1993). "The Phonetic Interpretation of Tone in Igbo." Phonetica 50: 147-160.

Lindblom, B. (1963). "Spectrographic Study of Vowel Reduction." Journal of the Acoustical Society of America 35(11): 1773-1781.

Maeda, S. (1976). A Characterization of American English Intonation. Ph.D dissertation. MIT.

McCarthy, J. and A. Prince (1993). "Generalized Alignment". The Yearbook of Morphology 1993. G. Gooij and J. v. Maarle (eds.), Kluwer Academic Publishers.

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

Möbius, B. (1995). "Components of a Quantitative Model of German Intonation". Proceedings of ICPhS 1995. Stockholm. 2: 108-115.

Nagahara, H. (1994). Phonological Phrasing in Japanese. Ph.D dissertation. University of California, Los Angeles.

Nespor, M. and I. Vogel (1986). Prosodic Phonology, Dordrecht, Foris Publications.

Pierrehumbert, J. and M. Beckman (1988). Japanese Tone Structure. Cambridge, MIT Press.

Pierrehumbert, J. B. (1980). The Phonology and Phonetics of English Intonation. Ph.D dissertation. MIT, Cambridge.

Poser, W. J. (1984). The Phonetics and Phonology of Tone and Intonation in Japanese. Ph.D dissertation. MIT, Cambridge.

Prince, A. and P. Smolensky (1993). Optimality Theory, Rutgers University.

Pulleyblank, D. (1986). Tone in Lexical Phonology. Dordrecht, Reidel.

Rooth, M. (1992). "A Theory of Focus Interpretation." Natural Language Semantics 7: 141-177.

Selkirk, E. (1984). Phonology and Syntax: The Relation between Sound and Structure. Cambridge, MA, MIT Press.

Selkirk, E. (1986). "On Derived Domain in Sentence Phonology." Phonology 3: 371-405.

Selkirk, E. (1995a). "The Prosodic Structure of Function Words". Papers in Optimality Theory. J. Beckman, L. W. Dickey and S. Urbanczyk (eds.) Amherst, GLSA Publishers: 439-470.

Selkirk, E. (1995b). "Sentence Prosody: Intonation, Stress, and Phrasing". The Handbook of Phonological Theory. J. Goldsmith (ed.), Blackwell.

Selkirk, E. (2000a). "The Interaction of Constraints on Prosodic Phrasing". Prosody: Theory and Experiment. M. Horne (ed.) Dordrecht, Kluwer Publishing.

Selkirk, E. (2000b). Focus Types and Tone. Paper presented at the First North American Phonology Conference. Concordia University, Montreal.

Selkirk, E. (2001a). "The Syntax-Phonology Interfaces". <u>IESBS</u>. Section 3.9.

Selkirk, E. (2001b). "Testing the Prosodic Structure Hypothesis." <u>Journal of the Phonetic Society of Japan</u> 5(1): 53-56.

Selkirk, E. (2002a). "Contrastive FOCUS vs. Presentational Focus: Presentational Evidence from Right Node Raising in English". <u>Speech Prosody 2002: Proceedings of the First International Speech Prosody Conference</u>. B. Bel and I. Marlien (eds.) Aix-en-Provence, Laboratorie Parole et Language, University of Provence: 643-646.

Selkirk, E. (2002b). "Bengali Intonation Revisited: An Optimality Theoretic Analysis in which FOCUS Stress Prominence Drives FOCUS Phrasing". <u>Papers in Optimality Theory II (UMOP 2002)</u>. A. Coetzee and P. D. Lacy (eds.) Amherst, GLSA.

Selkirk, E. and T. Shen (1990). "Prosodic Domains in Shanghai Chinese". <u>The Phonology-Syntax Connection</u>. S. Inkelas and D. Zec (eds.) Chicago, CSLI: 313-338.

Selkirk, E. and K. Tateishi (1988). <u>Constraints on Minor Phrase Formation in Japanese</u>. Proceedings of the 24th Annual Meeting of the Chicago Linguistics Society, Chicago, Chicago Linguistic Society.

Selkirk, E. and K. Tateishi (1991). "Syntax and Downstep in Japanese". <u>Interdisciplinary Approaches to Language: Essays in Honor of S-Y Kuroda</u>. C. Geogopoulos and R. Ishihara (eds.) Dordrecht, Kluwer.

Shinya, T. (2002). <u>Intonational Asymmetry between Argument and Adjunct in Japanese</u>. Manuscript. University of Massachusetts.

Smith, J. L. (1997). <u>Noun Faithfulness: On the Privileged Behavior of Nouns in Phonology</u>. Manuscript. University of Massachusetts at Amherst.

Sugito, M. (1996). "Rodoku ni okeru "shinjoho" no hyogen (Expression of "New Information" in Read Speech)". <u>Nihonjin no eigo</u>. Tokyo, Izumi Shoin: 221-234.

t'Hart, J. and A. Cohen (1993). "Intonation by Rule: a Perceptual Quest." <u>Journal of Phonetics</u> 1(4): 309-327.

Truckenbrodt, H. (1995). <u>Phonological Phrases: Their Relation to Syntax, Focus and Prominence</u>. Ph.D dissertation. MIT, Cambridge.

Uechi, A. (1988). <u>An Interface Approach to Topic/Focus Structure</u>. Ph.D dissertation. University of British Columbia.

Venditti, J. (1995). <u>Japanese ToBI Labeling Guidelines</u>. Ohio State University.

Venditti, J. (2000). Discourse Structure and Attentional Salience Effects on Japanese Intonation. Ph.D dissertation. Ohio State University.

Vogel, I. and I. Kenesei (1990). "Syntax and Semantics in Phonology". The Phonology-Syntax Connection. S. Inkelas and D. Zec (eds.) Chicago, CSLI: 365-378.

Warner, N. (1997). "Japanese Final-Accented and Unaccented Phrases." Journal of Phonetics 25: 43-60.

Whalen, D. H. and A. G. Levitt (1995). "The Universality of F0 of Vowels." Journal of Phonetics 23: 349-366.

Xu, Y. (1997). "Contextual Tonal Variations in Mandarin." Journal of Phonetics 25: 61-83.

Xu, Y. (1999). "Effects of Tone and Focus on the Formation and Alignment of F0 Contours." Journal of Phonetics 27: 55-105.