The Phonetics and Phonology of Tonal Systems

By Laura Christine Dilley

S.B., Brain and Cognitive Sciences
Massachusetts Institute of Technology, 1997

SUBMITTED TO THE HARVARD-MIT DIVISION OF HEALTH SCIENCES AND TECHNOLOGY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF

DOCTOR OF PHILOSOPHY IN SPEECH AND HEARING BIOSCIENCE AND TECHNOLOGY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2005

Signature of Author ______________________________
Harvard-MIT Division of Health Sciences and Technology
October, 2004

Certified by ______________________________
Kenneth N. Stevens, Ph.D.
Clarence J. LeBel Professor of Electrical Engineering
Harvard-MIT Division of Health Sciences and Technology
Thesis Supervisor

Accepted by ______________________________
Martha L. Gray, Ph.D.
Edwin Hood Taplin Professor of Medical and Electrical Engineering
Co-Director, Harvard-MIT Division of Health Sciences and Technology

The Phonetics and Phonology of Tonal Systems

by

Laura Christine Dilley

Submitted to the Division of Health Sciences and Technology
in Partial Fulfillment of the Requirements of the Degree of

Doctor of Philosophy

Abstract

Pitch variations are used in different languages in a variety of communicative ways, from cueing lexical item identity to conveying meaning through phrasing and accentuation. Previously, linguistic theories of intonation and tone which have sought a unifying account for tonal phenomena have not defined a clear and systematic relation between phonological representations and phonetic output in terms of acoustically observable fundamental frequency (F0) variations. This is problematic, because meaningful pitch patterns cannot be related to underlying phonological primitives in any clear way, so that it is difficult to empirically verify or falsify the theory.

This thesis addresses this problem by proposing a phonological description of intonation and tone for which there is a clear and systematic relationship between observed F0 variations and underlying phonological primitives. The theory is based on a construct called the *tone interval,* which represents an abstraction of a ratio of two fundamental frequencies. A *syntagmatic* tone interval relates two sequentially-ordered tones, while a *paradigmatic* tone interval relates a tone and a speaker-specific referent level. Tone intervals define one of three relations: a tone may be *higher*, *lower*, or the *same* level as its referent. Additional language-specific categories may be formed which restrict the pitch distance between a tone and its referent. Tones in syntagmatic tone intervals are assumed to be arranged at the phrasal level with respect to a *metrical grid*, which represents the relative prominence and timing of syllables. This permits interactions between nonsequential tones occupying metrically prominent syllables, accounting for cross-linguistic observations involving control of relative height relations on nonadjacent syllables.

Six experiments tested the predictions of tone interval theory and other phonological theories for English. Experiments 1 and 3 involved discrimination of pairs of stimuli in which the timing of an F0 extremum had been varied along a continuum with respect to segments, while Experiments 2 and 4 involved imitation of these stimuli. Experiments 5 and 6 involved imitating stimuli in which absolute F0 level had been varied along a continuum. Consistent with the tone interval theory, these results demonstrate the importance of relative pitch level for phonological representations. In particular, discrimination maxima and discreteness in production data were observed for positions in stimulus series in which either (i) the timing of an F0 extremum was varied across a vowel onset, or (ii) the F0 level of one syllable switched from higher than another syllable to lower than that syllable.

This theory provides a unifying account for a number of cross-linguistic phonetic facts, while explaining differences in the structure of tonal systems of various languages. The theory has implications for a wide array of disciplines, ranging from language acquisition to speech technology.

Thesis Supervisor: Kenneth N. Stevens
Title: Clarence J. LeBel Professor of Electrical Engineering

Acknowledgements

This work would not have been possible without the support of a great many people to whom I am extremely grateful and indebted. Foremost among them is my advisor, Ken Stevens. Over the years he has provided me with tremendous support and encouragement, and has generously permitted me to pursue my own research directions. From him I have learned to think like a scientist, and how to produce a thesis. He is a kind, generous, and intelligent man who cares a great deal for his students. I feel very lucky to have been able to learn from him.

This work could not have been created without the help of Stefanie Shattuck-Hufnagel. Stefanie inspired my interest in intonation by hiring me as an undergraduate some years ago to learn and apply the ToBI intonation system. I am deeply grateful to her for the many opportunities that she has given me, and for her tremendous support over several years. She is a wonderful person and a delight to work with.

I am lucky to have been able to work with Michael Kenstowicz, who has offered many helpful suggestions over the years, and who has patiently read many drafts. This work is much better as a result of my interactions with him. More than anyone else, Michael has helped me to understand the principles of linguistic analysis.

I am grateful to Cheryl Zoll, who taught me the first thing I ever knew about autosegmental theory. Her encouragement and helpful feedback were greatly appreciated.

Bob Ladd has generously provided feedback on many aspects of this thesis, for which I am very grateful. His papers and 1996 book have inspired me over the years to think about problems in intonational phonology. Moreover, the work he and his colleagues have done on phonetic alignment of F0 provided a great deal of substance to many theoretical questions of interest. I am extremely grateful to have had the chance to interact with him.

I am grateful to Donca Steriade, who has provided feedback on several manuscripts and other aspects of the theory. Her encouragement kept me going through some of the challenges of assembling this work.

I would like to especially thank Meredith Brown, who helped in many ways with the preparation of this thesis. She aided in running subjects, conducting analyses, putting together figures, and in various other ways picked up the ball and ran with it while I was working on my thesis. She is truly an amazing individual.

I would like to acknowledge Nelson Y.-S. Kiang for launching the interdisciplinary program in which I have been privileged to have been a student. He related to me once that he got the idea to start the program some number of years earlier, when he had met Morris Halle while a graduate student at the University of Chicago. I would also like to thank Morris Halle for taking the time to meet with me and for some interesting discussions.

I wish to thank Victor Manfredi for helpful feedback and encouragement, and for generously offering to provide a number of useful examples from the Africanist literature. I wish I could have put more of those examples in these pages.

I am grateful to Ani Patel for ongoing discussions about possible connections between music and language. Ani provided me with useful information at many points during this work.

Although they did not play a direct role in the present work, I am also very grateful to Janet Pierrehumbert and Mary Beckman. Without their many contributions to the study of intonation and other tonal systems, the present work would not have been possible.

Chao-Yang Lee is a great person and a wonderful friend. I wish to thank him for his help early in my thesis, and for his encouragement at many points along the way.

I would also like to thank the following individuals:

Catherine Best, Bashira Chowdhury, Nathan Boy, Peter Cariani, Fred Cummins, Colin Doherty, Robert Eklund, Jennifer Ford, Gregory Garretson, Ted Gibson, Seth Hall, Megan Hepler, David House, Annika Imbrie, Adria Karlsson, Lisa Lavoie, Zhiqiang Li, Melanie Matthies, Joanne Miller, John Nissenbaum, Tony Okobi, Andrew Oxenham, Joe Perkell, Carter Powers, Vivek Rao, Jason Redi, Keith

Snider, Janet Slifka, Ariel Salomon, Astrid Schepman, Chris Shera, Mark Tiede, Hubert Truckenbrodt, Virgilio Villacorta, Jennell Vick, Jimmy Vanzo, Sumudu Watugala, Duane Watson, Arlene Wint, Yi Xu, Majid Zandipour.

Also, I wish to thank my parents, Maureen and Charles, and my wonderful grandmothers, Penny McCullough and Hazel Dilley, for helping me to become the person I am today.

I am especially grateful to my husband, Devin McAuley, for supporting me at every turn during this process. Not only did he provide encouragement throughout all the ups and downs of putting this thesis together, but he aided with editing, formatting, assembling figures, and in many other ways. I would like to thank him for cheering me on and for being such a wonderful and supportive partner in every way.

Finally, I wish to thank the many individuals who have contributed to this thesis who are not named here.

Contents

Chapter 1 - Overview

This work is a beginning. It represents an attempt to apply facts about how we hear and interpret structured sound sequences, such as music, to formal problems in linguistics associated with intonation and tone in language. The result is a new framework for tonal phonology which builds on the strengths of existing frameworks while addressing outstanding theoretical problems in intonational and tonal phonology. The aims of this chapter are to motivate a new approach to the phonology of tone and intonation and to provide a roadmap to the present work.

1.1 The dual nature of tonal relations

The departure point for this work is the well-established idea that tone is inherently relative in nature. While the relativity of tone is not disputed, the issue of how to account for tonal relativity has long been a matter of controversy. This issue is complicated by the fact that tone appears to be relational in a dualistic way which has been hard to characterize in formal terms. On the one hand, tones are clearly scaled relative to a speaker's pitch range. On the other hand, tones are also scaled relative to other tones in the sequence. The present work argues that developing a representation which integrates these two aspects of tonal relations is essential for adequately characterizing tonal phenomena cross-linguistically.

Theoretical work in linguistics over the past 30 years has emphasized the view that tone is *paradigmatic* in nature, by which I mean defined relative to the speaker's pitch range, without respect to any other position in the context. Since the advent of autosegmental theory in the mid-1970's (Leben 1973, Williams 1971/76, Goldsmith 1976), tonal contours have been analyzed in terms of sequences of High (H) and Low (L) tones, implying a paradigmatic view of tone features. Under this view, H and L tones are phonological categories analogous to phonemes. Since the work of de Saussure (1966), phonemes have been depicted in terms of a network of contrasts, so that part of the "meaning" of the vowel /ɪ/ in English is its opposition to /ɛ/, as in *pin* vs. *pen*.[1] Similarly, to label Yoruba *rá* as H implies a contrast with L of *rà*. The claim that tones are analogous to phonemes has been a central assumption of autosegmental theory and work which followed, including Pierrehumbert's (1980) thesis on English intonation and the intonational transcription system known as ToBI (Tones and Break Indices; Silverman *et al.*, 1992).

This emphasis on paradigmatic tonal representations has left unclear the theoretical status of *syntagmatic* aspects of tone, which define the relative height of a tone relative to another tone in the sequence. Indeed, a number of linguists prior to the 1970's claimed that the phonological representation of tone was quintessentially syntagmatic, including Jakobson, Fant, and Halle (1952), Bolinger (1958, 1986), and Crystal (1969). There is broad consensus among linguists that theories of phonology and phonetics must include syntagmatic restrictions at some level of the representation, and a growing number of researchers have presented arguments favoring an incorporation of syntagmatic aspects of tone into phonological representations (e.g., Inkelas, Leben, and Cobler 1986, 1990; Odden 1995; Snider 1999). In order to evaluate the arguments for building syntagmatic features into the phonology, it is necessary to present evidence relevant to the representation of tonal features.

1.2 Arguments for paradigmatic tonal features in the phonology

Of the two types of tonal features under discussion, paradigmatic and syntagmatic, I will first discuss evidence supporting the existence of paradigmatic features in tonal representations. I will then turn to some arguments for syntagmatic features in tonal representations. These two aspects of tonal representations are not mutually exclusive. Indeed, the evidence will show that in order to account for cross-linguistic tonal phenomena, both paradigmatic and syntagmatic tonal features appear to be needed in the phonology.

What is the evidence that the phonology of tone includes paradigmatic features? Recall that such features entail only a comparison with the speaker's pitch range, without reference to any other tonal information in the context. A well-known argument which supports paradigmatic features is that speakers of lexical tone languages can identify words with level tones consisting of a single syllable (Gandour 1978, Fox and Qi 1990, Connell 2000). This indicates that for such cases no comparison with a surrounding context is necessary for tonal identification, suggesting that syntagmatic cues are not necessary for extracting tonal information.

It is important to note, however, that such an observation does not mean that syntagmatic characteristics play no role in tonal phonology. Indeed, there is a significant body of work suggesting that context contributes to lexical tone identification (e.g., LaVelle 1974, Leather 1983, Xu 1994, Moore and Jongman 1997, Wong and Diehl 2003). This suggests that while paradigmatic characteristics of tone are important in lexical tone languages, syntagmatic characteristics also may play a role in the representation.

While the evidence for paradigmatic features in lexical tone languages is compelling, the evidence for paradigmatic features in some other languages is less clear. One argument which has sometimes been cited in support of paradigmatic features for intonation languages, such as English, has been that certain meaning-based distinctions in such languages are based phonetically on where in a speaker's pitch range a particular contour is produced (e.g., Palmer 1922, O'Connor and Arnold 1973). However, not all descriptions of British or American English intonational phonology have made such claims; moreover, it is not clear whether such contours truly represent distinct phonological categories, or simply phonetic variants which lie along a continuum.

One line of experimentation which might be taken in support of paradigmatic features in intonation languages is that English listeners are quite good at judging the approximate location in the speaker's pitch range where an isolated vowel is produced (Honorof and Whalen, submitted). Although such data contributes important information relevant to the phonology and phonetics of tone, such results do not directly address whether paradigmatic cues give rise to categorical distinctions in English, or merely continuous phonetic variation. Indeed, a production experiment reported in Chapter 7 which uses the paradigm of Pierrehumbert and Steele (1989) suggests that pitch range cues in English fall along a continuous scale; we will return to these topics later in this work.

Taken together, these results suggest that the evidence for paradigmatic tone features in certain lexical tone languages is strong, while the evidence for paradigmatic features in other languages, such as English, is weak. One possibility is that the representations of lexical tone languages and intonation languages differ in their phonological use of paradigmatic features. This work advances the hypothesis that while many of the lexical tone languages of Africa and East Asia represent paradigmatic features, many other languages, such as English, do not represent paradigmatic features.

Indeed, there are often robust phonetic differences between what we will call *prototypical paradigmatic languages*, such as Yoruba and Cantonese, and *non-prototypical languages* such as English which have nevertheless been described in paradigmatic terms. In particular, a number of the languages of

Africa and East Asia exhibit what could be termed "frequency banding", i.e., the restriction of tones to particular frequency ranges in the speaker's pitch range (Earle 1975, Gandour 1976, Rose 1987, Connell and Ladd 1990, Connell 2000, Laniran and Clements 2003). In contrast, the scaling of intonation patterns within a speaker's pitch range in languages like English is rather free, such that individual tones are not necessarily restricted to a particular part of the speaker's pitch range. Given that African languages motivated the original claims in autosegmental theory that tone should be described in paradigmatic terms, these obvious phonetic differences between such languages and non-prototypical languages, together with evidence to be described favoring syntagmatic features, motivate re-examination of the original claims of autosegmental theory that tone is universally and exclusively paradigmatic.

Further evidence of a fundamental distinction between prototypical paradigmatic languages and nonprototypical languages comes from recent experimental evidence reported by Hallé, Chang, and Best (2004). Hallé *et al.* reported that Mandarin speakers showed quasi-categorical perception of Mandarin tones. However, French speakers did not perceive Mandarin tones in a categorical way. Because Mandarin is a prototypical paradigmatic language, these results suggest that the phonologies of such languages may differ in important ways from the phonologies of non-prototypical languages like French. Moreover, the paradigm employed by Hallé *et al.* suggests an empirical means by which languages could be tested, as a means of verifying their paradigmatic status.

Several lines of evidence therefore suggest the reasonable likelihood that prototypical paradigmatic languages have a distinct phonological representation compared to nonprototypical languages. The hypothesis advanced here is that only prototypical paradigmatic languages represent paradigmatic features, but that *all* languages represent syntagmatic features. Because the arguments favoring syntagmatic features in the phonology of tone and intonation are not as well-known, we now turn to a review of arguments supporting the existence of such features.

1.3 Arguments for syntagmatic tonal features in the phonology

There are four principal arguments supporting the existence of syntagmatic tonal features in the phonology, which will be discussed in turn in this section. The first concerns problems with theories of intonation languages which assume no syntagmatic restrictions or inadequate restrictions. The second relates to evidence for syntagmatic relations in lexical tone languages. The third concerns empirical evidence from experiments to be presented in later chapters that English speakers represent syntagmatic relations. The fourth involves evidence that other kinds of tonal sequences, such as musical melodies, are represented in terms of their syntagmatic relations. Following a discussion of these arguments, we will consider additional issues relating to parallels between music and speech.

1.3.1 Overgeneration and indeterminacy in theories lacking syntagmatic restrictions

The first argument for syntagmatic features is that theories which fail to include sufficient syntagmatic restrictions on the relative heights of adjacent tones lead to indeterminacy in phonological analyses of phonetic contours, as well as overgeneration of tonal contours from phonological sequences. Theories which have attempted to account for tonal patterns using a combination of paradigmatic tonal features plus phonetic implementation rules are at high risk of such problems. The theory of Pierrehumbert (1980), for example, attempted to account for English intonation using sequences of H and L tones, which were defined in paradigmatic terms.[2] An implicit claim of this theory, then, was that the representation of English intonation is based on the same phonological primitives as African tone languages. In Pierrehumbert's model, phonetic rules dictated both the positions of H and L tones in the pitch range, together with their relative levels with respect to one another. In this way, phonetic rules were assumed to give rise to F0 patterns.

An account of English intonation in terms of paradigmatic features plus phonetic implementation rules appears quite plausible at one level, since as noted earlier, there are some distinctions in English which have been claimed to reside in how high or low in a speaker's pitch range a contour is realized. However, it is important to note that English and other intonation languages present special problems for theories of phonology and phonetics, since the positions of H and L in a speaker's pitch range in such languages can vary quite substantially. Pierrehumbert's theory accounted for this significant variation through powerful phonetic rules which permitted the heights of tones, and consequently their relative levels with respect to one another, to vary rather freely. Unfortunately, allowing the relative levels of tones to vary without sufficient constraints effectively permits a given tonal pattern to be described in terms of a large number of different phonological sequences of tones. Conversely, a lack of sufficient restrictions on relative tone heights technically allows a phonological string to generate a large number of distinctive tonal patterns.

Pierrehumbert's description of English intonation thus leads to a significant degree of phonological indeterminacy and phonetic overgeneration, respectively. These issues are dealt with more extensively in Chapter 4, and a formal proof of the overgeneration and indeterminacy is presented. These problems can only be avoided through appropriate syntagmatic restrictions on the relative levels of tones. No theory has yet been put forward which can accomplish this, but doing so is precisely the aim of the present work. The fact that admitting syntagmatic restrictions into the phonology leads to an adequate account of the phonology and phonetics of English intonation while avoiding indeterminacy and overgeneration provides support for the claim that the phonology includes syntagmatic features.

1.3.2 Evidence from lexical tone languages

Evidence for syntagmatic features also can be found in the literature on lexical tone languages. This evidence comes both from psycholinguistic studies as well as from descriptive work on African languages. In the following section we review this evidence.

Support for syntagmatic features in the phonology of lexical tone languages comes from work in psycholinguistics showing the influence of context on tone identification. Wong and Diehl (2003) demonstrated this context-dependency through a task in which listeners judged the identity of the final word in synthetic speech versions of Cantonese sentences. Their stimuli were semantically neutral productions of the Cantonese sentence /ha6 yat1 go3 ji6 hai6 si3/ [*The next word is try*] in which the initial five-syllable context was pitch-shifted up or down while the F0 and other acoustic characteristics of the final syllable were held constant. The pitch manipulations were chosen based on observations by Chao (1947) suggesting that Cantonese Tone 3 (mid-level tone) is approximately three musical semitones lower than Tone 1 (high-level tone) and two musical semitones higher than Tone 6 (low-level tone).[3] Pitch-shifting the initial context up by two semitones led listeners to identify the final word as /si6/ [*yes*], while pitch-shifting the initial context down by three semitones caused them to identify this word as /si1/ [*teacher*].

These results show that relative height information is used by listeners to identify lexical tones in Cantonese. This presents counter-evidence to the claim of autosegmental theory that the representation of tones is exclusively paradigmatic, since under a strictly paradigmatic tonal system, we would not expect context to influence the tonal representation. The fact that context did influence word identification indicates that syntagmatic information was taken into account. If they did not, we would have expected listeners to have reported in all cases that the final syllable was /si3/ [*try*], since the acoustic material for this word and the associated paradigmatic cues did not change. This experiment therefore provides support for the claim that syntagmatic and paradigmatic tonal features can coexist in some languages.

Moreover, if the phonological representation of lexical tones includes both paradigmatic and syntagmatic features, then it should depend both on the relation of a tone to the speaker's pitch range, as well as on the relations of tones to other tones. This predicts that context should aid in identification of tones, since a listener will have more information than just paradigmatic information alone. As we noted earlier, there is indeed a large body of work that suggests that context does contribute to the identification of lexical tones. (See e.g., Xu 1994, Moore and Jongman 1997, and Wong and Diehl 2003.)

An important point to mention before proceeding further is that permitting both paradigmatic and syntagmatic features to coexist within a given language does not necessarily lead to a proliferation of tonal specifications. Indeed, in a system which permits appropriately restricted paradigmatic tonal features, the syntagmatic tonal relations are logically entailed in the paradigmatic specifications. For example, if H is higher than adjacent L in all contexts, then the syntagmatic relative height relations of a sequence of adjacent H and L tones can readily be derived. Thus, a tonal system which specifies paradigmatic features appropriately need not explicitly code for syntagmatic features, since they are readily derivable. We discuss this issue further in Section 1.5.2.

In addition to data from East Asian languages, there is evidence supporting the claim that African languages also encode syntagmatic relations. This is significant, given that the original work in autosegmental theory was based primarily on data from African languages. Here, we focus on cases in which two words are clearly specified for a single paradigmatic tonal category, yet they use syntagmatic cues of relative tone level in a contrastive manner.

Two examples of contrastive use of syntagmatic relations come from Igbo and Acatlán Mixtec, respectively.[4] In Igbo, the relative level of two syllables in words with paradigmatic High tone gives rise to a minimally contrastive word pair. The word *ama* 'street' is described as HH in Williamson (1972), while *ama* 'distinguishing mark' is described as H!H. In the former case, the two syllables are at a high level, while in the latter case, the second syllable steps down slightly. An example from Acatlán Mixtec was also reported by Pike and Wistrand (1974, p. 85). This language contrasts *?ikumida* "we incl. have," described as MHHH, with *?ikumida* "you pl. fam. have," described as MHHU, where U indicates "upstepped High". In the former case, the last two syllables are at the same level, while in the latter case, the last syllable steps up.

Syntagmatic relations can also be used contrastively at the phrasal level. For example, Inkelas, Leben, and Cobler (1986) and Inkelas and Leben (1990) report that in Hausa, relative height information is used to distinguish statements from questions. In sequences of alternating HL tonal sequences, such as HLHLHL, tonal sequences are produced with successively lower pitches on the H tones intended as statements. In contrast, the H's are at approximately the same level when such sequences are intended as questions. This indicates that a meaning-based distinction – whether an utterance is a statement or a question – depends on the relative height information among the H's in the utterance.

Odden (1995) also cites a number of examples illustrating three different syntagmatic relations among sequences of High tones in several languages. First, he notes that in Kenyang, a Mamfe Bantu language from Cameroon, High tones step down in sequence. (In this example, the site of pitch lowering is notated with ${}^!$.)

é ${}^!$béy ${}^!$mé${}^!$mwét "it hurts me"

Second, in Kimatuumbi, a Bantu language spoken in Tanzania, High tones step up in sequence. The sites of pitch raising are notated with ^.

baatí ^lyá ^kín ^dyé "they ate the birds"

Finally, in Kipare, a Bantu language from Tanzania, sequences of High tones stay at the same level. In this example, which is also from Odden, there is a step down after the first High, indicated by ${}^!$, but all tones thereafter stay at the same level.

vá${}^!$ná vékíjílá nkhúkú ndórí nkhúndú jángú "while the children eat those little red chickens of mine"

These examples collectively suggest that syntagmatic information plays a role in the representation of a second class of prototypical paradigmatic languages, namely, those of Africa. Such examples provide support for the claim that syntagmatic tonal features are universal. They also support the claim that syntagmatic relations and paradigmatic relations can coexist in the phonology of certain languages.

One possible alternative to encoding syntagmatic features in the phonology is instead to attempt to handle such data using phonetic rules for tonal scaling. Several proposals of this sort have been advanced (Pierrehumbert 1980, Liberman and Pierrehumbert 1984, Pierrehumbert and Beckman 1988, Clark 1993), which were driven by theoretical assumptions of autosegmental theory, in which syntagmatic features were assumed to be absent from the phonology. This raises the question of whether phonetic rules are capable of accounting for cross-linguistic data of the sort described above in the absence of syntagmatic relationships.

There are three principal arguments against a phonetic rule-based account of cross-linguistic distinctions of relative tone height. First, describing lexical and other semantic contrasts in terms of phonetic rules endows the phonetic component with the power to make semantic distinctions, thus endowing it with far greater functionality than has traditionally been assumed (Snider, 1999). Second, strong phonetic rules readily lead to problems with phonetic overgeneration of pitch contours and underspecification of the phonological description, as was the case with the theory of Pierrehumbert (1980). Finally, a survey of cross-linguistic data indicates that both steps down and steps up occur relative to some earlier tone. To account for both cases using phonetic rules, it would be necessary to assume that the phonetic processes behave differently in different languages. Phonetic processes might be assumed to describe universal physical or mechanistic properties, e.g. period-to-period "jitter" in F0 vibrations due to the quasi-chaotic nature of vocal fold vibration. Such physical, mechanistic processes are not expected to behave in language-specific ways. These difficulties suggest that cross-linguistic tonal data cannot be adequately accounted for using phonetic rules alone.

These examples collectively support the claim that syntagmatic information is part of the phonology in both African and East Asian languages. Together, this evidence provides support for the claim that both paradigmatic and syntagmatic phonological features are needed in order to adequately account for cross-linguistic data on tone and intonation. In the following section we provide further arguments supporting the claim that syntagmatic representations constitute part of the phonological representation of tone.

1.3.3 Evidence from perception and production studies in English

The third argument for syntagmatic tone features comes from empirical data derived from original experiments on English intonation discussed in Chapters 6 and 7. These experiments test whether the phonological representation of English intonation is basically syntagmatic, as will be claimed here, or whether it is paradigmatic, as proposed by Pierrehumbert (1980) and Beckman and Pierrehumbert (1986). In this paradigm, participants in perception and production experiments are presented with stimuli in which some acoustic aspect of F0 has been varied along a continuum. The experiments show that steps of an equivalent size along a stimulus continuum elicit responses of unequal magnitude along that continuum. Such responses can be interpreted in terms of phonological and perceptual category boundaries (Liberman, Harris, Hoffman, and Griffith 1957, Repp 1984, Kohler 1987, Pierrehumbert and Steele 1989). The patterns of resulting data fail to support the claim that English is based on paradigmatic tonal features, while simultaneously providing support for the claim that English is based on syntagmatic tonal features. In particular, subjects' responses fail to show evidence of category boundaries when tonal material is shifted through the speaker's pitch range. In contrast, subjects' responses reveal evidence of category boundaries when the relative heights of adjacent tonal material is manipulated, such that the boundaries are located precisely at positions in stimulus continua where one syllable switches from being higher than an adjacent syllable to being lower than that syllable. These empirical data thus provide additional support for the claim that syntagmatic tonal features are part of the tonal phonologies of English and other languages.

1.3.4 Importance of syntagmatic relations for musical melody

The fourth argument supporting syntagmatic tone features in language is that the representations for structurally similar tonal sequences – namely, musical melodies – are encoded in terms of their syntagmatic properties, i.e., the relations among the discrete tones or notes. This is precisely the reason that we can recognize that the melodic sequence G-G-A-G-C-B and the sequence C-C-D-C-F-E both correspond to the beginning of the tune "Happy Birthday," even though they are played in different musical keys. Indeed, there is a large body of experimental work showing that melodies are represented in terms of the relations among the notes (White 1960, Dowling and Fujitani 1971, Attneave and Olson 1971, Dowling and Harwood 1986, Handel 1989). Burns (1999: 218) writes:

> As is overwhelmingly evident from both everyday musical experience, from music theory, and as has been shown in formal experiments... melodic information in music is mediated by the frequency ratio relationships among tones (i.e., the musical intervals) not by their absolute pitches.

Experiments on the representation of musical melody have further emphasized the importance of the patterns of ups and downs of the notes, as separate from the specific sizes of the intervals (Dowling and Fujitani 1971, Dowling and Harwood 1986). The pattern of ups and downs of the notes, of course, is also a syntagmatic property, one which likely provides a more direct analogy with the scaling of tones in intonation languages.[5] Because musical melodies are structurally similar to tonal patterns in language, it is reasonable that linguistic tonal patterns are encoded in terms of syntagmatic relations as well.

We have presented four lines of argument supporting the existence of syntagmatic features in tonal phonology. Given this evidence favoring syntagmatic tonal features, how can such features be reconciled with existing frameworks? We briefly consider this issue in the next section.

1.4 Theoretical frameworks for tonal features

Given that both paradigmatic and syntagmatic features appear to be warranted in phonological representations for tone, what sort of theoretical framework might accommodate both types? The most obvious approach is to consider whether autosegmental theory might be modified to incorporate syntagmatic features, given that theory's many strengths. Moreover, we would like to avoid many of the pitfalls of previous theories which posited syntagmatic features by avoiding a return to "suprasegmental" treatments of tone (cf. Jakobson, Fant, and Halle 1952), which have been critiqued by Leben (1973) and others. In the following we consider some issues related to incorporating syntagmatic tonal features into autosegmental theory.

In order to justify the theoretical approach that we have taken, it is worth considering some of the strengths, as well as the weaknesses of autosegmental theory. An important observation is that this theory made many distinct, separable claims on a variety of distinct issues. Probably the most influential and successful of these claims was its assumption that tonal features are separate and quasi-autonomous from segmental features. This autonomy was represented in the theory by placing tones on a different level of representation, or *tier*, from segments. Doing so provided a way of accounting for certain classical problems in phonology, such as the observation that tones could "persist" even when segments did not. This particular phenomenon is termed *tonal stability*; it describes the process by which a tone on a deleted segment shows up on a different segment.

Moreover, autosegmental theory was attractive because it presented a solution to another longstanding puzzle. This concerned the issue of why tone features could sometimes appear to be "segmental," so that they were realized on a given segment, and at other times "suprasegmental," so that they were realized over a string of segments. By assuming that tonal features were paradigmatic in nature but that they could "spread" over multiple segmental positions, autosegmental theory presented an attractive solution for the representation of this duality.

However, as we have seen, the claim that tones are exactly like segments, and therefore exclusively paradigmatic, is not commensurate with the evidence we described earlier in this chapter, which indicates support for syntagmatic features as well. The obvious question then to ask is, how might syntagmatic features be incorporated into an autosegmental framework? Standard autosegmental theory, for its part, cannot account for the sorts of evidence presented in this introduction. One reason for this is that descriptive work in linguistics has shown evidence of three relative height relations between tones (Odden 1995); however, standard autosegmental theory acknowledges only a binary-valued mechanism for encoding relative height relations, which is obviously insufficient to account for a three-way distinction.[6, 7]

Several authors have proposed incorporating syntagmatic tonal features into the phonology (Inkelas, Leben, and Cobler 1986, Inkelas and Leben 1990, Snider 1999).[8] These proposals build on an approach to modifying autosegmental theory known as *register tier theory* which permits two autosegmental tonal tiers, one which encodes the tone type, and the other which encodes the tone register (Yip 1980, 1993, Pulleyblank 1986, Inkelas, Leben, and Cobler 1986). Among this work, Snider (1999) provides the most explicit account for the integration of syntagmatic features into the phonology. In Snider's proposal tones are represented in terms of paradigmatic "tonal" features, as well as syntagmatic "register" features; each tone is assumed to possess both a tonal feature and a register feature.

While incorporating syntagmatic features into the phonology seems to be a step in the right direction, three issues prevented us from adopting Snider's approach. First, the theory assumes that the relative heights of some tones are controlled by a parameter which takes a gradient range of values called the tone-register ratio. Such a gradiently-valued parameter seems inappropriate to capture the relative height relations of tones, which seem quite categorical in their natures. Second, the tone-register ratio is assumed to vary on a language-specific basis, so that it is not possible to predict what the phonetic relative heights of certain tones will be for any given language without knowing both the tonal features and the tone-register ratio. Third and finally, the sorts of restricted register phenomena which motivated register tier theory in the first place do not occur in English and other intonation languages, which permit relatively freer scaling of tones. As a result, it was not clear how English or other non-prototypical languages could be appropriately described in terms of tone registers.

Finally, we consider another theory which has proposed syntagmatic features, namely that of Clark (1978). Clark proposed a description of Igbo using a syntagmatic interpretation of lexical tone features. The fact that many aspects of Igbo tonology can be adequately described in terms of syntagmatic features is quite significant. However, evidence was presented earlier suggesting that paradigmatic features are also needed in tonal representations, which Clark's theory did not permit.

Given these considerations, it seemed necessary to pursue a different approach to integrating syntagmatic and paradigmatic features than had been attempted before. The most reasonable approach was to build on the strengths of autosegmental theory, while at the same time avoiding some of its weak points. In this way, it seemed possible to develop a stronger theory overall which built on the vast array of phonetic data and descriptive work which has come to light in the past 30 years. The following section provides background on the proposed theoretical approach, which is inspired by work in music theory.

1.5 A new approach: Integrating paradigmatic and syntagmatic features

The theoretic approach to tonal description which is proposed here is known as *tone interval theory*. Because music theory provided the conceptual foundation for this theoretic approach, we will present some background on musical systems in this section. In this way, the parallels between tone interval theory, which is proposed in Chapters 2 and 3, and music theoretic constructs will be made clearer.[9]

Skeptics will undoubtedly question the validity of comparisons between musical melody and tonal patterns in speech. Indeed, there are a number of differences between these two types of sequences which at first glance would appear to categorially distinguish the two. This section therefore addresses such concerns by showing that apparent differences between music and speech break down, revealing underlying similarities between the two domains. In the following we first defend the analogy between the melody aspects of speech and music. We then discuss background on music theory relevant to the proposals to be made. Finally, we discuss some differences among musical melodies which seem to parallel differences among tonal systems.

1.5.1 A critical examination of parallels between music and speech

Section 1.3.4 discussed the fact that tonal patterns in music are encoded primarily in terms of the relations among the notes in sequence. Such evidence tends to support the idea that tonal sequences in other domains, such as language, are also represented in terms of such syntagmatic relations. However, the analogy between musical melodies and speech could be disputed on the grounds that these two types of tonal sequences bear many important differences. We will see that when we consider music and speech stimuli more generally, apparent differences between music and speech break down, revealing underlying similarities between both domains.

The first dimension along which music and speech seem to differ is the fact that musical melodies consist of a sequence of intervals of fixed sizes, while the intervals between F0 patterns across speech syllables are not fixed. While this is generally true for the melodies of Western music, it is not true when we look across musical traditions and cultures. For example, melodies in Australian aboriginal music are described in terms of a pattern of ups and downs in frequency; these melodies are not comprised of fixed intervals (U. Will, personal communication). Moreover, in Javanese music the tunings of the notes are quite wide, so that the frequency ratio realizing the change from one note to another can vary substantially (Perlman and Krumhansl 1996). These examples illustrate that like speech, musical melodies can also consist of a rather freely varying sequence of frequency ratios.

Moreover, tonal patterns in speech can apparently also realize constrained frequency ratios, as indicated by examples of apparent restrictions on interval sizes of tonal excursions. For example, we noted earlier that Chao's (1947) description of Cantonese identified Tone 3 as being approximately three musical semitones lower than Tone 1 and two musical semitones higher than Tone 6. In addition, Thorsen's (1980) work on Danish intonation indicates that steps down of different sizes distinguish different meanings. Yet another example is the English calling contour, which involves an interval of about three semitones (Liberman 1975). These examples illustrate that tonal patterns in both music and speech may or may not consist of intervals with restricted sizes. Thus, music and speech apparently do not differ along this dimension.

A second dimension along which music and speech might be claimed to differ is the fact that musical melodies consist of sequences of discrete, sustained pitches, while F0 in speech varies continuously. While this is often true of musical melodies, it is not always the case. For example, vocal music involves continuously varying F0 to approximate a sequence of discrete pitches. When the singing is even moderately fast, the F0 curve associated with the string of sung tones takes on a dynamic character which can begin to look like an F0 curve in speech. Moreover, psychophysical experiments have shown that melodies can be heard out of continuously varying fundamental frequency curves (Demany and McAnally 1993).

Moreover, F0 contours in speech are perceived in terms of a sequence of discrete pitches (House 1990, d'Alessandro and Mertens 1995), indicating that the discreteness of pitches alone does not differentiate music and speech. In general, dynamic F0 curves interact with spectral cues to make some portions of the F0 curve sound more salient and others less salient (e.g., House 1990, d'Alessandro and Mertens 1995, d'Imperio and House 1997). This differential salience engenders something like a sequence of discrete pitches in speech, although listeners are generally not aware of it. These observations from speech agree with the auditory perception literature, which suggests that interruptions of dynamic frequency trajectories through amplitude reduction, noise, and so on cause certain pitches to be "heard out" (Brady, House, and Stevens 1961, Dannenbring 1976, Nabelek, Nabelek, Hirsh 1970, 1973, Ciocca and Bregman 1987, d'Alessandro, Rosset, and Rossi 1998). In other words, both speech and music can be described in terms of a sequence of discrete pitches.

A third dimension which might appear to categorically distinguish music and speech concerns the fact that musical melodies have a tonal center. The tonal center corresponds to a "most important" or referent note for a musical scale. What we see on closer examination is that once again, music and speech do not differ so dramatically. First, musical melodies need not have a tonal center; such is the case for so-called "atonal melodies" (Dowling and Fujitani 1971, Handel 1989). Second, it appears that tonal patterns in speech apparently can exhibit a "referent pitch" analogous to the tonal center in music. For example, phonological analyses of lexical tone languages have sometimes described one tone as a referent (e.g., Wang 1967, Hyman 1993).

In sum, apparent differences between tonal patterns in music and in speech break down when we look across tonal systems in music and language. Other possible differences between music and speech will be considered later. Having thus defended the analogy between music and speech, we turn to an examination of aspects of music theory which have inspired the tone interval approach to linguistic description.

1.5.2 Background in music theory

We begin by describing the most basic analogies between tone interval theory and music theory. What is needed to account for the linguistic data is a way of describing both paradigmatic and syntagmatic constructs in a unified framework; music readily permits both types of constructs. In music, syntagmatic relations code for the relative heights and intervals between notes in a sequence. Moreover, paradigmatic relations code for the "scale structure" of musical melodies, representing the relative height and interval between each note and a referent pitch.

If we assume that the tonal systems of music and of speech are analogous, then an important issue concerns the relative distributions of syntagmatic and paradigmatic structures. Of the two types of constructions, syntagmatic relations in musical melodies are the more basic; they are common to all musical sequences. In contrast, paradigmatic constructs are observed in only a subset of musical melodies. Similarly, the claim made here for language is that all linguistic tonal systems represent syntagmatic features among the tones. In contrast, a subset of linguistic systems are claimed to represent paradigmatic features.

A further point is that paradigmatic constructs in music logically entail syntagmatic relations. This is because each note in the scale system is defined with respect to a common reference note. Thus, when the notes are strung together, their mutual reference to a common note means that their relative heights can be derived. Similarly, the claim advanced here is that paradigmatic constructions in language logically entail syntagmatic relations as well.

Having defined the most basic aspects of analogy to be claimed between musical and linguistic systems, we turn to presentation of more background on music theory. Although most of us, when we think of the term *music* call to mind Western tonal music, the term as used here refers to a broad range of musical systems across cultures. Each distinct musical genre from a given culture or group has its own rules and structure. The diversity of tonal systems across musical cultures suggests that comparison to the diversity of linguistic tonal systems may be fruitful. Indeed, the goals of music theory and linguistics are much the same: to model the intuitions of informed listeners who have experience – though not necessarily expertise – with the language or music of a culture.

Musical scales across cultures are usually constructed by dividing an octave into a certain number of scale steps. An octave is defined as an interval consisting of two pitches which have a frequency ratio of 2 to 1. Thus, an octave separates a note of frequency 200 Hz and a note of frequency 400 Hz. Across cultures, there is little consistency in how the octave is divided up into interval scale steps. In standard ("equal tempered") Western musical scales, the scale steps are based on an interval known as the semitone, which corresponds to the twelfth root of 2, or approximately 1.0594. Thus, the 12 steps in a Western scale generated from a base note of 200 Hz have frequencies of about 212, 224, 238, 252, 267, 283, 300, 317, 336, 356, 378 Hz, and 400 Hz.[10] Many of the notes in this scale have frequencies which are close to small, whole-number ratios, such as 3:2, 4:3, 5:3, etc.; notes with frequencies closely approximating these ratios correspond to the harmonically important, "consonant" notes in the scale.

Musical systems across cultures divide the scale very differently. On the one hand, Western music divides the octave into 12 scale steps, with narrow frequency ranges corresponding to each of the notes. In contrast, other musical systems divide the octave into more or fewer steps. Across musical cultures, there is a preponderance of scale systems which are limited to five or seven tones, including those of Southeast Asian, Javanese and Balinese traditions (Malm 1967, Wade 1979, Tenzer 1991, Burns 1999). In such scale systems, a much wider range of frequencies may correspond to a given musical note than in Western music; such is the case in the Javanese scale system (McPhee 1966, Perlman and Krumhansl 1996).[11] In general, musical systems which have a richly-developed instrumental tradition, such as Western or Indian musical systems, divide the scale more finely than those which do not (U. Will, personal communication).

There are important structural relations among the notes of the Western scale, or any scale for that matter. One property which is apparently common to nearly all scale systems is the notion of a "most important note", i.e., a tonal center which all the other notes of the scale are referred to. The tonal center in Western music is termed the *tonic*, which is defined as the first note in a major or minor scale. Major and minor scales correspond to well-defined subsets of the 12-note scale consisting of seven notes each; most of these notes in turn are "consonant" notes deriving approximately from small, whole-number ratios. Handel (1989: 335-6) describes the relations of all the notes of the Western scale to the tonic in the following way:

> Within a musical context, each tone has a direction or balance. Beginning with the major scale, we can analyze the quality of each tone. If we start with the tonic note, we "hear" the second, third, and fourth notes as pointing back to the tonic. There is a tension: a tendency to return to the tonic. The fifth note is a pivot… it can fall back to the tonic or it can bounce ahead to the eighth note, an octave above the tonic. The sixth note can move to the fifth note or move toward the seventh note and from there to the eighth note, an octave above… Movement to the tonic represents balance.

This passage illustrates that all the notes in a scale "point toward" the tonal center or tonic. In this way, the tonic is analogous to a paradigmatic referent. When a listener "knows" that a particular note is the tonic, any other note in the scale can be played and a listener will be able to extract the structural "meaning" of the note in the context of the musical scale. This suggests an analogy between melodies possessing a tonic note and paradigmatic tonal features in language.

A second observation about musical sequences is that a tonic referent note restricts the remaining scale notes to a particular region of the frequency space. This recalls linguistic observations concerning prototypical paradigmatic systems, in which each of the lexical tones occupies a restricted region or range of the frequency space. Indeed, linguistic descriptions of such systems have sometimes described one of the lexical tones as a "referent" to the other tones (Wang 1967, Hyman 1993).

An important point is that not all melodic sequences evoke a sense of a tonal center or tonic. In standard Western music, identification of the tonic is usually made possible by the fact that most melodies are comprised entirely or predominantly of only the seven notes of a major or minor scale, out of 12. When a melody is restricted to such a seven-note subset, the sense of a tonic is strong. Indeed, it is possible for Western listeners to determine the relationships among the notes and how they relate to a tonic, even if they are musically untrained (Attneave and Olson 1971, Bartlett and Dowling 1980). On the other hand, the notes of a melody may also be drawn more generally from the 12 notes of the Western scale. When this happens, listeners cannot use their implicit knowledge of major and minor scale structure to determine which note corresponds to the tonic. In other words, some melodies evoke a sense of a tonic, while others do not.

Differences among melodies in whether or not there is a tonic again suggest a parallel with language. Earlier we distinguished between prototypical paradigmatic languages, including Cantonese and Yoruba, which exhibited something like a referent pitch as well as banding of frequencies in the pitch range, from non-prototypical languages, such as English, which exhibited neither of these characteristics. Our claim is that the distinction between these two language groups parallels the distinction in music between melodies which have a tonic referent and those which lack one. Melodies with a tonic are analogous to linguistic tonal patterns exhibiting a referent pitch and frequency banding; such sequences are claimed here to represent paradigmatic tonal features (where syntagmatic tonal features are logically entailed). In contrast, melodies lacking a tonic are analogous to linguistic tonal patterns lacking a referent pitch or frequency banding; such sequences are claimed here to represent on syntagmatic tonal features.

Supporting evidence for a typological distinction of this sort comes from evidence discussed earlier by Hallé, Chang, and Best (2004). They showed that Mandarin lexical tones were perceived quasi-categorically by Mandarin listeners, but not by French listeners. Mandarin should probably be classified as a prototypical paradigmatic language. In contrast, French would be a non-prototypical language. Since French speakers lack experience with Mandarin tones, they did not show evidence of these categories.

These results mirror with findings from music showing that musical intervals give rise to categorical perception (CP). CP for musical intervals has been demonstrated for both musicians and nonmusicians (Siegel and Siegel 1977a, b; Burns and Ward 1978; Zatorre and Halpern 1979; Howard, Rosen, and Broad 1992). The fact that nonmusicians also exhibit CP for musical intervals is significant, because it suggests that such categories may arise largely through passive exposure to the structure of a particular musical grammar within one's culture. Similarly, the linguistic influence of language-specific input structures on categories is indicated by the differential performance of French and Mandarin listeners in Hallé *et al.*'s study. These observations indicate that for tonal patterns in both music and speech, the categories of representation depend on the structure of the input.

Another aspect of music theory which will prove relevant to tone interval theory concerns the fact that musical melody is integrated with the overall metrical structure. The metrical structure of music can be described in terms of structures not unlike metrical grids in linguistic theory (e.g., Liberman and Prince 1977, Selkirk 1984, Halle and Vergnaud 1987, Hayes 1995). The importance of rhythm for the representation of melody is well documented. (See e.g., Lerdahl and Jackendoff 1983, Jones and Ralston 1991, Jones 1993.) Theoretical frameworks for the interaction of melody and rhythm have been proposed by Lerdahl and Jackendoff (1983), Narmour (1991), and others. In speech, interactions between tonal patterns and metrical structure have now been documented in a variety of languages with very distinctive tonal patterns (e.g., Liberman 1975, Rice 1987, Inkelas and Zec 1988, Manfredi 1993, Hayes 1995, Bickmore 1995, Zec 1999, de Lacy 2002).

Finally, in both music and speech, long-distance syntagmatic relations of relative height can hold between nonadjacent notes or pitches. Consider the fact that variations on a simple melodic theme often involve the insertion of "less-important" notes between the "important" notes of the melodic theme, such that the melody is realized through non-adjacent syntagmatic relations of relative height between the important notes. In spite of the fact that the melody is carried by temporally nonadjacent notes, listeners nevertheless recognize the melody (Lerdahl and Jackendoff 1983). That syntagmatic relations of relative height can hold between nonadjacent pitches is well-established in the music and auditory perception literature (van Noorden 1977, Bregman 1990, Lerdahl and Jackendoff 1983, Jones and Boltz 1989, Narmour 1991, 1999). Similarly, syntagmatic relations of relative height hold between nonadjacent tones in speech as well (Pierrehumbert 1980, Liberman and Pierrehumbert 1984, Inkelas, Leben and Cobler 1986, Ladd 1988). It is significant that in both music and speech, such nonadjacency relations occur between notes which are in metrically strong positions.

This brief overview of music theory is intended to be suggestive of the kinds of analogies that I have drawn on in the proposals of tone interval theory. The following section discusses the structure of the thesis, as well as some outstanding theoretical issues which the framework will address.

1.6 Structure of the present work

This work is organized as follows. Chapters 2 and 3 provide the groundwork for a theory of tone and intonation based on tone intervals. Chapter 2 focuses on redefining the nature of tonal features as dually syntagmatic and paradigmatic by introducing the tone interval as a phonological construct. In this way the present theory provides a new explanation for why the tonal features of languages alternately appear to be "segmental" and "suprasegmental". Establishing a solid theoretical basis for syntagmatic tonal features in particular permits an account in Chapter 4 for a range of phonetic data from intonation languages concerning the presence and alignment of certain F0 turning points, such as peaks and valleys, which find no clear account under current theories. In addition, syntagmatic tonal features will be shown in Chapter 5 to permit an account of English intonation which is both more phonetically transparent and simpler than earlier theories.

Chapter 3 elaborates on the theory by describing how tone intervals are arranged at the phrasal level. We will propose that tones universally associate with structures known as *metrical grids* (e.g., Liberman 1975; Liberman and Prince 1977; Hayes 1980, 1995; Selkirk 1984; Halle and Vernaud 1987). In this way, the present theory fills a gap in the literature by assuming a common explanation for cross-linguistic observations of interactions between tone and meter. Several principles of association between tones and metrical grid structures will be proposed in the course of the chapter. These proposals will permit an account in Chapter 5 of a host of phenomena in English and other intonation languages, including the phonetic alignment patterns of bitonal pitch accents reported by Arvaniti, Ladd, and colleagues, as well as long-distance interactions between accents demonstrated in English by Pierrehumbert (1980), Liberman and Pierrehumbert (1984), and Ladd (1988).

Next, Chapters 4 and 5 provide motivation for the tone interval framework by showing how it accounts for a range of phonetic facts which have no account under existing theories. Chapter 4 prepares for a demonstration of some problems with existing theories by first focusing on epistemological issues related to how theories of phonology and phonetics may be evaluated, and what sorts of facts such theories should seek to account for. A body of phonetic work is reviewed showing consistent alignment of F0 turning points, such as maxima and minima, and it is argued that any theory of the phonology and phonetics of intonation languages should provide some account of this alignment data. We then evaluate the ability of different theories to explain this data, as well as the complexity of those accounts. A formal proof is presented showing that the theory of Pierrehumbert (1980) overgenerates possible phonetic contours and underdetermines the phonological description, since it does not include appropriate syntagmatic restrictions on adjacent tone pairs. In contrast, the tone interval theory, which permits syntagmatic primitives directly in the phonology, is found to be both more descriptively adequate as well as simpler than theories based on paradigmatic primitives plus phonetic implementation rules.

Next, Chapter 5 shows how the tone interval framework permits a description of English intonational patterns using just six simple primitives. We demonstrate that tone interval theory provides a solid theoretical basis for previous descriptive work in linguistics which was based on unfounded assumptions about the relationship between phonology and phonetics. In addition, it is shown that a simple change of notation readily permits previous descriptions in terms of H's and L's to be expressed in terms of tone intervals.

Chapter 5 also demonstrates that the tone interval framework permits a phonological account for a number of phonetic facts which are at odds with the claims and predictions of previous frameworks. For example, we show how the tone interval framework provides for a more flexible, abstract definition of a "starred tone." This new definition of "starredness" not only accounts for phonetic observations in English, but also in Greek, a language for which Arvaniti, Ladd, and Mennen (2000) recently demonstrated the inadequacy of current conceptions of starred tones. Moreover, an account is provided for findings that the two tones in bitonal pitch accents align with respect to segments rather than with respect to each other by a fixed temporal interval (Arvaniti *et al.*, 1998; Ladd *et al.*, 1999, 2000; Atterer and Ladd 2004; Dilley *et al.*, to appear). In addition, we will provide a more consistent definition of the relation between phonetics and phonology for F0 "dips" between high accents, thereby providing a means of accommodating the findings of Ladd and Schepman (2003) concerning the alignment of such points. Finally, we show how the association with the metrical grid permits a cleaner analysis of English data reported by Pierrehumbert (1980) and Liberman and Pierrehumbert (1984). We will show that this data provides not just qualitative, but quantitative support for the proposals of tone interval theory.

Chapters 6 and 7 present the results of several perception and production experiments which test the claims and predictions of the paradigmatic theory of English (Pierrehumbert 1980, Beckman and Pierrehumbert 1986, Pierrehumbert and Beckman 1988, Beckman and Ayers-Elam 1997), as well as the syntagmatic account of English presented in Chapter 5. Chapter 6 discusses a series of paired perception and production experiments which investigate the validity of assumptions of the paradigmatic theory of English about the mapping from phonetic characteristics to phonological categories. In these experiments, F0 maxima or minima are shifted through SW, WS, or SWW syllable sequences in order to determine whether listeners interpret the differences in a categorical way. The results provide support for some of the assumptions of the paradigmatic theory about the mapping from phonology to phonetics, such as the distinctiveness of H* vs. H+L* and L* vs. L+H*. However, the results also provide evidence against certain other assumptions, such as the claim that high accents with an F0 peak on a stressed syllable versus after that syllable are mapped to a single category, H*.

Chapter 7 tests the claims and predictions of the paradigmatic theory of English against those of the syntagmatic theory of English presented in Chapter 5. The first of two production experiments uses stimuli in which the F0 level of a syllable or syllables has been shifted through the pitch range. The results show no support for several phonological contrasts which are claimed to be distinguished on the basis of paradigmatic F0 level (H* vs. L+H*, H* vs. L*+H, and %H vs. no tone). In contrast, support is obtained for the syntagmatic theory, in that categorical effects in production are observed at positions along stimulus continua in which a syllable changes from higher than an adjacent syllable to lower than that syllable, consistent with a change from one syntagmatic category to another. The second production experiment directly tests the predictions of the syntagmatic theory using an imitation task. The F0 levels of two adjacent syllables are shifted along a continuum, so that their levels relative to one another changes halfway through each continuum. The results show categorical effects in production at positions in stimulus continua in which the level of one syllable changes from being higher than to lower than the following syllable, consistent with a change from one syntagmatic category to another. The experimental design rules out an account of the data by paradigmatic theories.

Finally, Chapter 8 summarizes the main arguments of each chapter and the contributions of the thesis. Moreover, the chapter charts some directions for future research. At this point, we will turn to a discussion of the basic properties of tone intervals in the upcoming chapter.

Chapter 2 – Introduction to tone intervals

2.1 Some theoretical considerations and goals of the chapter

A widely known and puzzling fact has been that tones sometimes behave like segments, and at other times unlike segments. Previously, the accepted explanation for this fact had been that tonal features are exactly like segmental features. According to this explanation, the dualistic behavior of tones arises due to cross-linguistic differences in patterns of phonological association with segments, together with distinctive kinds of phonetic processes (Leben 1973, Goldsmith 1976, Pierrehumbert 1980).

The present work offers a different explanation for the behavior of tones as alternately segmental and suprasegmental. This dualistic behavior is claimed here to arise from the fact that tonal features are both syntagmatic and paradigmatic, which is made possible by phonological structures known as *tone intervals*. The goal of Chapter 2 is to introduce tone intervals and their properties. In particular, we will show how these constructs permit tonal features to define either paradigmatic or syntagmatic relations. Moreover, we will show that "tonal" properties of language derive from the properties of tone intervals, rather than of tones. Indeed, tones are viewed in this theory as featureless "timing markers" which are associated with particular positions in the metrical structures of utterances. In contrast, we will show that tone intervals are the structures which give rise to phonological categories of "tonal" representation.

It is important to point out that although tone interval theory admits syntagmatic features into the phonological description, it does not advocate a return to the views of Jakobson, Fant, and Halle (1952). Jakobson *et al.* posited a fundamental distinction between "intrinsic" features, which were definable on independent acoustic properties, from "prosodic" features, which were definable only in relational terms. The former category included segmental features, while the latter included all tonal features. The critique of Leben (1973) showed, among other things, that tonal features in many cases exhibit paradigmatic properties. Such observations provide evidence against the distinction posited by Jakobson *et al.*, and an account based strictly on paradigmatic features for tone was posited. The present theory offers a different account by permitting tonal features to be either paradigmatic or syntagmatic. Because tone is assumed in this way to have a dual nature, the present theory cannot be said to express the same position as that of distinctive feature theory.

Having defined the major goals of this chapter, we turn to the task of defining the nature of tone intervals. This is accomplished in several sections. First, Section 2.2 defines tone intervals formally and describes their basic properties. Next, Section 2.3 describes the basic inventory of possible distinctions for tone intervals. Section 2.4 discusses some of the properties of tone intervals which derive from their status as quasi-mathematical constructs. Section 2.5 shows how tone intervals define language-universal phonological categories. Next, Section 2.6 defines some additional properties of tone intervals deriving from their quasi-mathematical status. Finally, Section 2.7 summarizes the major points of the chapter.

2.2 Definition and basic properties of tone intervals

In this section we propose a basic phonological construct, known as the *tone interval*. Here, a tone interval is defined both conceptually and formally. It will be shown that tone intervals can define either paradigmatic or syntagmatic tonal relations in the phonology.

Simply put, tone intervals are phonological constructs which define a relationship between a tone and a referent tonal entity. Tone intervals are thus functional constructs: they serve the function of relating a tone to a referent tonal entity in a particular way. In so doing, tone intervals endow tones with particular relational tonal properties. A tone interval I is defined in its general form as follows:

$$I = \frac{T}{r} \tag{2.1}$$

The expression above formally describes a relationship of reference, in which a tone, T, is related to a referent tonal entity, r, thereby defining tonal features in a way described in this chapter. By themselves, tones have no features, nor any tonal properties. Rather, tones mark particular positions in the metrical structure of the utterance associated with segments or syllables. Those marked positions then instantiate relative tonal relations when tones are joined into tone intervals, as shown later in Chapter 3.

Depending on the nature of the referent tonal entity r, a tone interval can define a syntagmatic or a paradigmatic tonal relationship. On the one hand, r may correspond to another tone, so that the tone interval defines a syntagmatic relation, as in (2.2). On the other hand, the referent tonal entity may correspond to a referent tonal level called the *tonic*, symbolized μ, so that the tone interval defines a paradigmatic relation, as in (2.3). The representations given above in expressions (2.1) through (2.3) indicate phonological abstractions of phonetic reality. In particular, these expressions describe a relation between an abstract, phonological tone, and an abstract referent.

The borrowing of the term *tonic* from music theory intentionally invokes an analogy with melodic systems which define a tonic referent pitch or "tonal center". Just as the tonic in music represents a "normalizing tonal level" for a particular musical scale, the tonic as used here represents a normalizing level within the abstract tonal space.[12] More specifically, the tonic has both a phonetic as well as a phonological interpretation. Phonetically, the tonic corresponds to a speaker-specific referent pitch. Phonologically, the tonic corresponds to an abstract referent level. Later, we will claim that tonal systems which represent paradigmatic relations specify one paradigmatic tonal category to be at the level of the tonic. The implication of this claim is that the tonic referent level is realized phonetically whenever an instance of the corresponding tonal category is produced.[13,14]

$$I = \frac{T}{T_r} \text{ (syntagmatic form)} \tag{2.2}$$

$$I = \frac{T}{\mu} \text{ (paradigmatic form)} \tag{2.3}$$

By realizing abstractions of frequency ratios, tone intervals codify a relation of relative height between a tone and a referent. By *relative height*, we mean whether a tone is higher than, lower than, or at the same level as the referent. In certain tonal systems, tone intervals may also codify the distance between a tone and a referent in the abstract tonal space in a language-specific manner. How this is accomplished is the topic of Section 2.5.

Consider how an F0 contour arises from the phonological representation as specified under this formulation. In this theory, tone intervals phonologically encode tonal information, rather than tones. This tonal representation is expressly relational, in that it captures a relationship of relative height between a tone and a referent. The function of tones in this theory is to mark positions in the metrical structure of an utterance corresponding to particular segments or syllables. By extension, tones are understood to inherently lack "tonal" content. Tones ultimately gain tonal specifications by joining into tone intervals along with some referent, which may either be another tone in the sequence or the tonic level. Tone intervals thus effectively endow tones with their tonal properties, which are viewed as intrinsically involving a relation between a tone and a referent. Phonological specification for the tonal features of tone intervals involves associating one of three primitive, relational tone features with each tone interval, as described in Section 2.3; this feature specifies the relative height of a tone with respect to its referent. The distance between the tone and the referent is determined either by language-specific phonological categories or by the phonetics. At the phonetic implementation stage, the tones in this abstract representation are then translated into a set of target pitches at positions of underlying tones. These target pitches are then connected up via monotonic pitch interpolation functions. Finally, the vocal folds are enlisted to produce an F0 contour which acoustically instantiates the fully-specified pitch pattern.

It will be noted that tone intervals resemble the form of a mathematical fraction, which is a type of ratio. This analogy is deliberate, since, as we will see, tone interval constructs function phonologically much like ratios in mathematics. One way in which this structure is advantageous is that it permits us to make a rather direct connection with categories of representation in music. Recall that in music, the representational categories are understood as abstractions of frequency ratios; musical intervals constitute true categories, in the sense that they give rise to categorical perception (Siegel and Siegel 1977a, b; Burns and Ward 1978; Zatorre and Halpern 1979; Howard, Rosen, and Broad 1992). Similarly, the phonological categories in the present theory are understood as abstractions of frequency ratios.

There are a number of other phonological properties of tone intervals which follow from the analogy with mathematical ratios. First, consider the fact that in mathematics, ratios entail a process of comparison through division of one entity by another. Similarly, tone interval constructs embody an explicit comparison between the "value" of a tone T and the "value" of a referent r, where T and r represent phonological abstractions of phonetic reality.

Another property concerns limits on the phonological values of tone intervals. As discussed earlier, tone intervals represent abstractions of ratios of frequency values. Because T and r are phonetically realized in terms of positive, real quantities given in units of Hz, tone intervals, too, are restricted to positive, real quantities. The specific values of tone intervals as given in a particular language are then further limited by the phonological structure of the language in question as well as the extrema of the speaker's pitch range.

Now that some of the properties of tone intervals have been defined, we can consider in more detail how these constructs represent phonological categories. Here an analogy with music is once again helpful. In music, the categories are embodied in the musical notes in a particular melody and/or musical scale. Each such category is associated with restrictions on the values of frequency ratios. Similarly, the present theory proposes that the tonal categories are embodied in restrictions on values of tone interval ratios. These tone intervals can assume a syntagmatic form, as shown earlier in (2.2), or a paradigmatic form, as in (2.3). These two types of constructs give rise to syntagmatic or paradigmatic phonological categories, respectively. The claim made here is that all linguistic tonal systems represent syntagmatic constructs, while only a subset of systems represent paradigmatic constructs.

An important point concerns the difference between abstract phonological constructs and acoustic-phonetic observations. Given that the phonological categories correspond to ranges of tone interval values, the measurable phonetic values of a tone and a referent realizing this abstract phonological relation should bear close correspondence to the underlying range of tone interval values but need not match it exactly. An analogous example comes from music. In music theory, an octave corresponds to an idealized frequency ratio of 2:1. An opera singer who produces an octave starting with an A of 440 Hz might end with an A note of 880.5 Hz. While this does not precisely match the idealized 2:1 ratio, it nevertheless is a close approximation, and we might expect that perceptually, the pitch of the 880.5 Hz tone is a good example of an A for a listener. This serves to illustrate the difference between an *abstract* frequency ratio of the sort which is assumed here to characterize tonal categories in music and in language, from a *realized* frequency ratio of the sort expected to be produced acoustically. Now that we have more carefully defined this distinction, we turn to a more detailed discussion of the tonal features of tone intervals in the following section.

2.3 Features of tone intervals

In the previous section we presented the basic claim that the phonological representation of tone in language is based on constructs known as tone intervals, which relate a tone and a referent. In this section we discuss the three primitive phonological relationships which may obtain between a tone and a referent: a tone may be higher than, lower than, or at the same level as its referent. These three basic relationships serve two important functions. First, they phonologically define a particular "spatial arrangement" between a tone and its referent in the abstract tonal space. Second, each of these three basic relations is associated with a distinct and restricted range of tone interval values. *A restricted range of tone interval values defines a phonological category in this system.* In this regard, we draw an explicit analogy between the representations of musical categories, which are based on ranges of frequency ratios, and the representations of linguistic tonal categories, which are based on ranges of tone interval values.

As mentioned above, there are three primitive relative height relations which may be specified for a given tone interval: a tone may be higher than, lower than, or at the same level as a referent. These three possibilities derive from the selection for a given tone interval construct of a set of two binary-valued relational features: [±same] and [±higher]. These features are inherently relational; they define a relative position of the tone and the referent with respect to one another in the abstract tonal space. The binary features are arranged in the feature geometry shown in Figure 2.1. It can be seen that the feature [±higher] is a subsidiary of the specification [-same]. Thus, a specification for [±higher] is given only if the value of [±same] is specified in the negative. We note that the distinction between [+same] and [-same] mirrors the distinction between "same" and "different," which is so important in perception generally.

An important consequence of this geometry is that one of exactly three basic relative height relations define the relationship between a tone and a referent. First, a tone can be at the same level as a referent, so that the tone interval is [+same]. Second, a tone can be higher than its referent, so that the tone interval is [-same, +higher]. Finally, a tone may be lower than its referent, so that its specification is [-same, -higher]. In the following, we briefly illustrate how these features specify the relative height relation between a tone and a referent for syntagmatic and paradigmatic tone intervals, respectively. We will then examine the implications of these relative height restrictions for ranges of tone interval values.

Recall that either a syntagmatic or paradigmatic relation can be defined in a tone interval, depending on whether referent is another tone or a tonic referent, respectively. Consider first for a syntagmatic tone interval how the choice of relative height relation affects the abstract configuration of two temporally ordered tones, T_1 T_2. Suppose T_1, and T_2 participate in the following syntagmatic tone interval construct:

$$I_{1,2} = \frac{T_2}{T_1} \tag{2.4}$$

The fact that we are dealing with a syntagmatic tone interval is indicated by the presence of a tone in the denominator of the expression. We will refer to tone in the numerator as the *referring tone* and the tone in the denominator as the *referent.* By convention, the first subscripted element for a syntagmatic tone interval I – namely, the numeral "1" – denotes the subscript of the referent, and the second subscripted element – namely, the numeral "2" – denotes the subscript of the referring tone.

In general, any syntagmatic tone interval has a "direction". By this we mean that either an upcoming tone or a preceding tone can serve as the referent for a particular referring tone. Thus, we must somehow specify which tone of the two comes first in time. We will take the convention of left-to-right directionality, such that the earlier tone in time serves as the referent unless otherwise stated. Then for an ordered pair of tones T_1 T_2, T_1 is the referent by default.[15]

The selection of a primitive relative height relation for a syntagmatic tone interval affects the spatial arrangement of two tones in the abstract tonal space. As mentioned above, the second tone may be higher than, lower than, or at the same level as the earlier tone. When these tones are subsequently supplied to the phonetic implementation module, the tones are connected up through monotonic interpolation functions. In this way, pairing the relations *higher*, *lower*, or *same* with the syntagmatic tone interval above in (2.4) defined by the tonal sequence T_1 T_2 generates a rising, falling, or level contour, respectively, at the phonetic output.

A note of caution is that the relational features *higher*, *lower*, and *same* alone do not specify the distance between a tone and its referent. We must still define mechanisms in this framework for distinguishing categorical contrasts like "small fall" and "large fall", consistent with the existence of attested distinctions like H$^{!}$H, which corresponds phonetically to a small fall, versus HL, which corresponds phonetically to a comparatively larger fall. How these distinctions are made is discussed in Section 2.6.

Now we turn to consideration of paradigmatic tone intervals. How does the choice of a primitive relative height relation affect the abstract configuration of a tone with respect to a tonic referent? Consider the expression for a paradigmatic tone interval relating a tone T_0 to the tonic μ which is repeated from above:

$$I_{\mu,0} = \frac{T_0}{\mu} \tag{2.5}$$

The fact that we are dealing with a paradigmatic tone interval in (2.5) is indicated by the presence of the tonic, μ, in the denominator. By convention, the first subscripted element in a paradigmatic tone interval I is the referent μ, while the second subscripted element denotes the subscript of the referring tone.

The manner in which relational features specify the abstract spatial configuration between the tone and the tonic referent is quite straightforward. A specification of *higher*, *lower*, or *same* for the paradigmatic tone interval requires that T be higher than μ, lower than μ, or at the same level as μ, respectively. This phonological relative height relation is subsequently passed to the phonetic implementation module, which interprets μ with respect to a speaker's own pitch range. T is then assigned some pitch by the phonetic module which is higher than, lower than, or at approximately the same level as μ, as appropriate.

There are two additional points worthy of note. First, a tone T might not be produced in isolation; rather, it could form part of a sequence of tones. These other tones might or might not belong to a paradigmatic tone interval. Regardless of whether they do or not, they must ultimately be connected up by phonetic interpolation functions to yield an F0 contour. In Chapter 3 we will describe a principle which requires that any sequence of tones, whether specified for paradigmatic tonal features or not, is required to join into syntagmatic tone interval constructs at the phonological level. This makes sense, because stringing together a sequence of paradigmatically-specified tones naturally yields a configuration of relative height relations. For tones belonging to paradigmatic tone intervals, the syntagmatic relations may be derived from the spatial arrangement of tones in the tonal space dictated by the paradigmatic specification of relative height, or they may be derived by rule. Thus, the tones in a paradigmatic tone interval will also belong to at least one syntagmatic tone interval, so long as the tone is not produced in isolation. These syntagmatic relations are then readily interpreted in terms of phonetic interpolation functions.

A second point worthy of note is that the present framework so far only generates a relative height relation between μ and T. We have not described how the theory permits the capture of differential distances between μ and T. This is dealt with in Section 2.5.2, where we formalize a notion of the tonal distance between a tone and its referent. For paradigmatic tone intervals, this formal notion of distance permits, among other things, an account for systems with multiple tone levels.

We have just described how the three basic relative height relations associated with syntagmatic and paradigmatic tone intervals affect the spatial configuration of a tone and its referent. We now turn to the issue of how these three basic relative height relations lead to restrictions on tone interval values. We will see that each of these relations has a natural interpretation in terms of a distinct pattern of restrictions on tone interval values; in this theory, restrictions on ranges of tone interval values *are* the phonological categories of the system.

2.4 Forming phonological categories: How features restrict tone interval values

Earlier we described tone intervals as having properties of mathematical fractions or ratios. In this section we will see that when a tone interval is paired with a particular relational feature, the mathematical properties of tone intervals give rise to emergent, categorical behavior in tone intervals. In particular, we will see that the choice of relational feature leads to restrictions on ranges of tone interval values. Recall that musical categories are represented through restricted ranges of frequency values. The analogy between tonal categories in music and in language thus leads to the insight that pairing a relational feature with a syntagmatic or paradigmatic tone interval engenders the formation of phonological categories in this system. In the following we explore how the pairing of a relational feature with a tone interval engenders categorical behavior on the part of tone intervals.

First, we will consider how the pairing of a relative height relation with a tone interval correspondingly restricts the range of tone interval values. Consider the fact that each of the relational features has an interpretation in terms of the relative magnitudes of the numerator and denominator of a tone interval "fraction". Ultimately, we will see that these relative magnitudes have an interpretation in terms of tone interval values. Note that from here onward, we will preferentially refer to abstract notions like *magnitude, value,* or *quantity* when describing entities in the numerator and denominator, rather than referring to numbers. Such descriptors are useful, because they are amenable conceptually to processes involving a comparison at an abstract level.

Suppose we consider the implications of particular relative height relations on the relative magnitudes of tones and referents in more detail. A generalized tone interval is defined as $I = T/r$. Each relative height relation defines a relative magnitude for the numerator and denominator; thus, for example, the relation *higher* can be interpreted as meaning that the value in the numerator is greater than the value in the denominator, such that $T > r$. Similarly, *lower* means that the value in the numerator is less than the value in the denominator, so that $T < r$. Finally, *same* means that the values in the numerator and denominator are equal, so that $T = r$.[16]

The relative magnitudes of a numerator and denominator have a natural interpretation in terms of ranges of tone interval values, where these restricted tone interval ranges comprise the phonological categories of the system. Given that the relation higher implies that $T > r$, dividing T by r yields a range of tone interval values greater than 1, or unity. In other words, pairing the relation *higher* with a tone interval restricts its range of values to $I > 1$. Similarly, *lower* entails that $T < r$, so that dividing T by r yields a range of tone interval values less than unity, $I < 1$.[17] Finally, same implies that $T = r$, such that dividing T by r restricts the tone interval value to $I = 1$. This is summarized in Table 2.1.

Tone interval relation	Range of tone interval values
higher	$I > 1$
lower	$I < 1$
same	$I = 1$

Table 2.1: Correspondence between relational features and tone interval ranges.

Table 2.1 makes clear that the relational features described early in this chapter have a formal interpretation in terms of restrictions on tone interval ranges. These relational features phonologically restrict the relative position of a tone with respect to a referent. In the same way, each feature corresponds to a distinctive range of values for tone interval ratios. Expressions like $I > 1$, $I < 1$ and $I = 1$ will prove very important for understanding tonal phenomena within the context of the present framework, as will be shown in upcoming discussion.

It is worth noting that the mathematical operators which relate the features *higher*, *lower*, and *same* to restrictions on tone interval values – namely, greater than (">"), less than ("<"), and equals ("=") – are quite categorical in their nature. The distinctiveness of mathematical categories can be illustrated by considering that if two quantities are equal, then one cannot be greater than the other. Similarly, if one quantity is less than the other, then the two cannot be equal, and so forth. These examples illustrate that mathematical operators of this sort indeed have the properties of distinct categories.

So far we have been discussing only the generalized form of tone intervals. However, tone intervals can take either syntagmatic or paradigmatic forms. In the following, we first briefly elaborate on the previous derivation, showing how restrictions on the ranges of syntagmatic and paradigmatic tone intervals give rise to syntagmatic and paradigmatic phonological categories, respectively. Of the two kinds of categories – syntagmatic and paradigmatic – we claim that syntagmatic categories are universally attested in linguistic tonal systems. Next, we will consider what sorts of phonetic outputs may be generated by a given phonological specification. Finally, we describe another way of conceptualizing the referent, which is as a normalizing level in the tonal space.

For syntagmatic tone intervals, three phonological categories are generated by pairing one of three relative height relations with a syntagmatic tone interval construct. We claim that the resulting syntagmatic phonological categories are common to all linguistic tonal systems. For example, pairing a syntagmatic tone interval defined by the tone sequence T_1 T_2 with the relations *higher*, *lower*, or *same* generates three phonological categories: $I_{1,2} > 1$, $I_{1,2} < 1$ and $I_{1,2} = 1$, respectively. Moreover, pairing a paradigmatic tone interval defined by T_0 and μ with *higher*, *lower* and *same* also yields three phonological categories: $I_{\mu,0} > 1$, $I_{\mu,0} < 1$ and $I_{\mu,0} = 1$, respectively.

It is significant that when either a syntagmatic or paradigmatic tone interval is associated with the relation *higher* or *lower*, multiple values of I will satisfy the relation. In contrast, associating the relation *same* with a tone interval means that only a single value of I will satisfy the relation: $I = 1$. In phonetic terms, this means that when the underlying relations are $I > 1$ or $I < 1$, the pitch distance between a tone and a referent can vary. However, when the underlying relation is $I = 1$, the pitch distance between a tone and a referent must be approximately zero phonetically, so that the tone occurs at about the same level as the referent. It is noteworthy that this difference in the behavior of *higher* and *lower*, on the one hand, versus *same* on the other, is readily described by the fact that the former two relations are associated with the feature [-same], while the latter is associated with the feature [+same].

These differences in phonetic behavior can be further illustrated by considering, for a generalized tone interval $I_{r,T}$, the range of values of a tone T that will satisfy a given relation when we hold the referent r fixed at some level, for each of the three features. If the relation *higher* is specified, so that $I_{r,T} > 1$, then T can take on a range of values, so long as it is higher than r. Thus, r "sets the floor" for T, given the relation *higher*; in this way, r effectively "normalizes" the scaling of T in the tonal space. Some phonetic outputs that might result from such a relation are illustrated in Figures 2.2(a) and (b) in the case of syntagmatic and paradigmatic tone interval constructs, respectively.

The situation is quite comparable in the case of the relation *lower*, which gives the relation $I_{r,T} < 1$. Once again, T can again take a range of values, so long as it is lower than r. In this way r "sets the ceiling" for T, given the relation *lower*, thereby normalizing the scaling of T in the tonal space. Some of the phonetic outputs that might result from such a relation are illustrated in Figures 2.2(c) and (d) in the case of syntagmatic and paradigmatic tone intervals, respectively.

Finally, consider the case in which the relation *same* is specified for a tone interval, yielding the relation $I_{r,T} = 1$. In this case, T can only take a single value, namely $T = r$. The phonetic outcome of this is that only a level contour can be generated for a syntagmatic tone interval, as in Figure 2.2(e), while a tone must always be located on the tonic level in the case of a paradigmatic tone interval, as in Figure 2.2 (f).

There is one additional factor which limits the phonetic scaling of tones. Each speaker's voice is associated with certain physiological limitations on the highest and lowest pitch levels that can be produced; such limits play a role in the phonetic scaling of tones. Assuming that each tone T gives rise to a pitch p, then the following relation holds: $l < p < h$. In this relation, l and h are some absolute low and high pitch limits that can be produced by an individual speaker's voice. Then the relation $l < p < h$ effectively limits the scaling of tones phonetically such that their associated pitches cannot exceed the specified range.

We will pause here briefly to summarize this section. We proposed three primitive relations of relative height: *higher*, *lower*, and *same*. We also showed that these relations can be interpreted as statements about the relative magnitudes of the numerator and the denominator in tone interval expressions. As a result, the relations give rise to restrictions on the ranges of syntagmatic and paradigmatic tone intervals. *Higher* implies that $I_{r,T} > 1$, *lower* implies that $I_{r,T} < 1$, and finally *same* implies that $I_{r,T} = 1$. In this way, categorical behavior emerges from a function which can take a continuous range of values. Such restrictions on ranges of tone interval values are understood in the present framework to correspond to phonological categories. In the following section we describe how additional, language-specific phonological categories can be formed through more stringent restrictions on tone interval values.

2.5 The representation of language-specific phonological categories

The last section defined three basic relations of relative height between a tone and a referent: *higher*, *lower*, and *same*. When paired with syntagmatic or paradigmatic tone intervals, these relations generated phonological categories corresponding to restrictions on syntagmatic or paradigmatic tone interval values. In this section we discuss how additional, language-specific phonological categories are defined through the specification of further restrictions on tone interval values. These language-specific phonological categories are analogous to intervallic categories in musical melodies and scale systems, which are defined on a culture-specific basis across musical systems.

Simply put, language-specific phonological categories are created by subdividing the portions of the tonal space which are associated with the relations *higher* and *lower*, for paradigmatic or syntagmatic tone intervals. Recall that a wide range of phonetic values could satisfy the relations *higher* or *lower*, which are given as $I > 1$ and $I < 1$while only a single value would satisfy the relation *same* ($I = 1$). Thus, only *higher* and *lower* can be further subdivided into more restricted regions of tone interval values, in order to give rise to further phonological categories.

Specifying further subdivisions of the tonal space associated with *higher* or *lower* effectively places language-specific limits on the permitted intervallic distances between a tone and a referent. To see why limiting the distance between a tone and a referent could be a productive approach, consider that certain tonal distinctions appear to be made on the basis of tonal distance. For example, some languages distinguish H$^{!}$H, which corresponds to a small fall, from HL, which corresponds to a comparatively larger fall.[18] Moreover, the size of a "small step down" apparently varies across languages; for example, H$^{!}$H has been used to describe a pitch drop of about 1 semitone in Mende (Goldsmith 1976), as compared to a pitch drop of about 3 semitones in English (Liberman 1975). Such observations support the idea that the size of an interval can be restricted in language-specific ways. Restricting the distance between a tone and a referent also permits the system to distinguish multiple tone levels in languages with register tone. For example, the distinction between High and Extra High tones in a language possessing both might be characterized as involving different restrictions on distances between each tone and the tonic.

The key to defining additional phonological categories in this system is to codify the notion of a *cutoff value*. Cutoff values divide the tonal space up into phonologically distinct regions. The relations *higher*, *lower*, and *same* effectively share a cutoff value of unity. These relations defined restrictions on tone interval ranges corresponding to three distinct regions or levels: $I > 1$, $I < 1$ and $I = 1$. In each case, the value at unity separates the regions of the tonal space corresponding to distinct categories. In the same way, language-specific cutoff values can be defined which serve to subdivide the regions associated with the relations *higher* and *lower*. When a cutoff value is specified for a syntagmatic tone interval, it phonetically restricts the size of a rising or falling contour. When a cutoff value is specified for a paradigmatic tone interval, it phonetically restricts the distance between a tone and the tonic.

Language-specific phonological categories are created by taking the mathematical intersection of two regions of the tonal space. One of these regions is defined by *higher* or *lower* ($I > 1$ or $I < 1$, respectively). The other region is defined with respect to a language-specific cutoff value. This second region can corresponds to the portion of the tonal space which is greater than the cutoff value, less than the cutoff value, or equal to cutoff value. In the following we illustrate how additional phonological categories are defined in the cases of syntagmatic and paradigmatic tone intervals, respectively.

2.5.1 Syntagmatic phonological categories

We will consider first how syntagmatic tone intervals can give rise to additional phonological categories. Suppose we wish to distinguish a contour which falls a short distance, such as H$^{!}$H, from a contour which falls a larger distance, such as HL. This distinction can be captured in terms of two distinct phonological categories defined by different restrictions on the syntagmatic tone interval space, a result which we demonstrate below.

As we discussed earlier, additional phonological categories are defined through the intersection of the regions defined by *higher* or *lower* with a region which lies above, at, or below a language-specific cutoff value. Because we wish to describe a difference between two falling contours (H$^{!}$H and HL), we know that the appropriate relation is *lower*, which describes a tone being lower than the preceding tone. Moreover, distinguishing between a "smaller distance" and a "larger distance" then entails determining the intersection between $I_{1,2} < 1$ and another region of the tonal space defined with respect to the appropriate cutoff value.

In order to define an additional phonological category, we define a cutoff tone interval level δ_{-}. The "negative" subscript indicates that this cutoff value is less than unity.[19] Given this cutoff value, there are three regions that can be defined with respect to it: the region lying above the cutoff value ($I_{1,2} > \delta_{-}$), at the cutoff value ($I_{1,2} = \delta_{-}$), or below the cutoff value ($I_{1,2} < \delta_{-}$). A phonological category is defined by the intersection of $I_{1,2} < 1$ with one of these three regions. In the following, we consider these three cases.

First, a phonological category can be formed through the intersection of the region lying *below unity* ($I_{1,2} < 1$) with the region lying *above the cutoff value* ($I_{1,2} > \delta_{-}$). This intersection is given by the expression in (2.6), which in turn can be rewritten as in (2.7):

$$(I_{1,2} < 1) \cap (I_{1,2} > \delta_{-}) \tag{2.6}$$

$$1 > I_{1,2} > \delta_{-} \tag{2.7}$$

Second, a phonological category can be formed through the intersection of the region lying *below unity* ($I_{1,2} < 1$) with the level *at the cutoff value* ($I_{1,2} = \delta_{-}$). This intersection is given in (2.8), which can be rewritten as in (2.9):

$$(I_{1,2} < 1) \cap (I_{1,2} = \delta_{-}) \tag{2.8}$$

$$1 > I_{1,2} = \delta_{-} \tag{2.9}$$

Third, a category can be formed through the intersection of the region lying *below unity* ($I_{1,2} < 1$) with the region lying *above the cutoff value* ($I_{1,2} > \delta_-$). This intersection is given in (2.10) and can be rewritten as in (2.11):

$$(I_{1,2} < 1) \cap (I_{1,2} < \delta_-) \tag{2.10}$$

$$1 > \delta_- > I_{1,2} \tag{2.11}$$

These three cases permit us to describe the difference between a small fall, as in H$^{!}$H, and a larger fall, as in HL. In general, this descriptive approach distinguishes between comparatively smaller and comparatively larger pitch excursions. Thus, cutoff values which are closer to 1 distinguish small falls from larger falls, while cutoff values which are closer to 0 distinguish large falls from very large falls. Thus, terms like "small fall" or "large fall" in the following are intended in a comparative sense, rather than an absolute sense. Then the expression in (2.7) defines a phonological category which corresponds phonetically to a *small fall of variable size* – relatively speaking – such that the distance between T_1 and T_2 is restricted phonologically to intervals lying between unity and δ_-. This is shown in Figure 2.3(a). Second, the expression in (2.9) defines a phonological category which corresponds to a *small fall of consistent size*, such that the distance between T_1 and T_2 is restricted to an interval equal to δ_-. This is shown in Figure 2.3(b). Finally, the expression in (2.11) defines a phonological category which corresponds to a *larger fall of variable size*. The associated tone interval range for this category is shown in Figure 2.3(c). This expression indicates that the distance between T_1 and T_2 is restricted to intervals which are smaller than δ_-. In general, categories defined by relations of equality, such as (2.9), are characterized by phonetic consistency, while categories defined by relations of inequality, such as (2.7) and (2.11), are characterized by comparative phonetic variability. We will refer to a phonological category defined by a relation of equality, as in (2.9), an *intervallic level* and a phonological category defined by a relation of inequality, as in (2.7), an *intervallic range.*

The expressions just derived capture distinctions like H$^{!}$H versus HL. In particular, expressions (2.7) and (2.9) adequately describe phonological variants of a small fall. Either of these might be selected by a given language to represent a small fall phonetically. In contrast, the expression in (2.11) corresponds to the category of a comparatively larger fall.

Given that the present approach generates two separate representations for a small fall, as in (2.7) and (2.9), it is interesting that these cases appear to be attested in English and Danish, respectively. Anecdotally, the size of the steps down in the English calling contour (e.g., *Anna! Dinner's ready!*) can vary from about 1 to 4 semitones but does not exceed this range (Liberman 1975, Ladd 1996). On the other hand, Danish appears to attest specific step sizes, where the size of the step permits listeners to distinguish among different meanings (Thorsen, 1980). We therefore hypothesize that English represents the calling contour through an expression of the form in (2.7), while Danish represents steps down through expressions of the form in (2.9).

Having illustrated how falls of different sizes can be specified in the present framework, we now turn to the issue of specifying rises of different sizes. The derivation is quite parallel to the case of falls. Distinctions in the size of a rise are made by taking the intersection of the region corresponding to the relation *higher* with one of three possible regions that can be defined with respect to the cutoff δ_+; here, the "positive" subscript indicates that this cutoff value is greater than unity. We will consider each of these cases in turn.

First, a phonological category can be formed through the intersection of the region lying *above unity* ($I_{1,2} > 1$) with the region lying *below the cutoff value* ($I_{1,2} < \delta_+$). The intersection is given by the expression in (2.12). Second, a category can be formed through the intersection of the region lying *above unity* ($I_{1,2} > 1$) with the level *at the cutoff value* ($I_{1,2} = \delta_+$). The intersection of these two regions is shown in (2.13). Finally, a category can be formed through the intersection of the region lying *above unity* ($I_{1,2} >$

1) with the region lying *above the cutoff value* ($I_{1,2} > \delta_+$). The intersection of these regions is shown in (2.14).

$$1 < I_{1,2} < \delta_+ \qquad (2.12)$$

$$1 < I_{1,2} = \delta_+ \qquad (2.13)$$

$$1 < \delta_+ < I_{1,2} \qquad (2.14)$$

These three cases distinguish comparatively smaller and comparatively larger rises. Figure 2.3(d) illustrates (2.12) by showing a *small rise of variable size* in which the distance between T_1 and T_2 is restricted to intervals lying between unity and δ_-. Figure 2.3(e) illustrates (2.13) by depicting a *small rise of consistent size* in which the distance between T_1 and T_2 is restricted to an interval equal to δ_-. Finally, Figure 2.3(f) illustrates (2.14) by showing a *larger rise of variable size* in which the distance between T_1 and T_2 is restricted to intervals which are greater than δ_-.

Two points are worthy of note before we proceed further. First, it is possible for more than one cutoff value to be specified in the phonology. This raises the question of how many values can be specified. We propose that perception and production constraints effectively limit the number of distinct cutoffs for tone interval values which can be specified phonologically. We will return to this issue shortly. Second, phonological categories in this framework are always defined through the intersection of two overlapping regions of the tonal space. If two regions are non-overlapping, their intersection will correspond to the "empty set." This suggests that phonological categories can never be formed through the intersection of two non-overlapping regions, and they also can never be defined by taking the union of two regions. Because ***higher*** and ***lower*** define primitive regions of the tonal space, one consequence is that all phonological categories will correspond to subdivisions of $I > 1$ or $I < 1$. Thus a language might distinguish phonologically among rises or among falls, but we predict no possibility that a language would group rises and falls into the same category. Having thus dealt with syntagmatic categories, we consider in the next section how paradigmatic tone interval constructs can give rise to additional phonological categories.

2.5.2 Paradigmatic phonological categories

Additional phonological categories can also be defined in terms of restrictions on paradigmatic tone intervals. For example, suppose a language distinguishes between two lexical tones which lie above the tonic level; these tones might be termed High and Extra High.[20] Such a distinction can be represented in the present framework by assuming that each lexical category is associated with a distinct range of tone intervals. In particular, the paradigmatic tone interval range corresponding to Extra High would lie further above the tonic than the paradigmatic tone interval range corresponding to High.

Because both High and Extra High are expected to be realized with a pitch which is above the tonic level, we can infer that phonological categories are defined through the intersection of a paradigmatic tone interval specified for the relation ***higher*** with a region defined with respect to a cutoff value. Suppose that we examine different possible phonological categories that might be specified for the j^{th} tone in a sequence, T_j, with respect to a cutoff tone interval level Δ_+. Note that the subscript for Δ_+ indicates that the cutoff is greater than unity.[21] Given this cutoff value, additional phonological categories are formed by taking the intersection of the region defined by higher ($I_{\mu j} > 1$) with the region above the cutoff value ($I_{\mu j} > \Delta_+$), at the cutoff value ($I_{\mu j} = \Delta_+$), or below the cutoff value ($I_{\mu j} < \Delta_+$). We consider each of these cases below.

First, a phonological category can be formed through the intersection of the region lying *above the tonic* ($I_{\mu j} > 1$) with the region *below the cutoff value* ($I_{\mu j} < \Delta_+$), as in (2.16). Second, we can form a category through the intersection of the region *above the tonic* ($I_{\mu j} > 1$) with the level *at the cutoff value*

($I_{\mu j} = \Delta_+$), as in (2.17). Third, we can form a category through the intersection of the region lying *above the tonic* ($I_{\mu j} > 1$) with the region *above the cutoff value* ($I_{\mu j} > \Delta_+$), as in (2.18).

$$1 < I_{\mu j} < \Delta_+ \tag{2.16}$$

$$1 < I_{\mu j} = \Delta_+ \tag{2.17}$$

$$1 < \Delta_+ < I_{\mu j} \tag{2.18}$$

The expression in (2.16) corresponds to a phonological category in which tones phonetically occur in a region of the speaker's pitch range that is above the tonic but below a cutoff Δ_+, as shown in Figure 2.4(a). This case corresponds to a tone which is a *small, variable distance above the reference level.* Next, the expression in (2.17) corresponds to a phonological category in which tones consistently show an approximate value of Δ_+, as shown in Figure 2.4(b). This case therefore corresponds to a tone which is a *small, consistent distance above the reference level.* Finally, the expression in (2.18) corresponds to a phonological category in which tones fall in a region of the speaker's pitch range which is above the cutoff value of Δ_+, as shown in Figure 2.4(c). This corresponds to a tone which is a *larger, variable distance above the reference level.*

How does this help us distinguish two tones higher than the tonic, like High versus Extra High? We have derived expressions for three possible phonological categories which lay above the tonic, where the tonic corresponds to the level where $I = 1$. Then High and Extra High might be equated with two of these categories, such that the category for Extra High corresponded to a higher region in the tonal space than High. For example High and Extra High could correspond to (2.16) and (2.17), respectively, or (2.17) and (2.18), or (2.16) and (2.18). Without more work on the phonetics characteristic of these tonal categories, it is impossible to know how such a theory would be applied to various descriptive categories. However, the expressions in (2.16) and (2.18) correspond to intervallic ranges, so the observed phonetic distance from the tonic to the tone is expected to vary within limits; in contrast, the expression in (2.17) corresponds to an intervallic level, so that the phonetic distance from the tonic to the tone should be consistent. Phonetic studies would thus aid in determining how descriptive lexical tone categories would be captured in this framework.

We can extend this account to phonological categories which lay below the tonic. Such representations could distinguish different low tones, e.g., Low and Extra Low. Here, three phonological categories may be formed by determining the intersection of the region $I_{\mu j} < 1$ with the regions lying above some cutoff level Δ_-, or at Δ_-, or below Δ_-. We quickly review these cases in turn.

First, a category can be formed through the intersection of the region lying *below the tonic* ($I_{\mu j} < 1$) with the region *above the cutoff value* ($I_{\mu j} > \Delta_-$), as in (2.19). A second category can be formed through the intersection of the region *below the tonic* ($I_{\mu j} < 1$) with the level *at the cutoff value* ($I_{\mu j} = \Delta_-$), as in (2.20). Yet a third category can be formed from the intersection of the region lying *below the tonic* ($I_{\mu j} < 1$) with the region *below the cutoff value* ($I_{\mu j} < \Delta_-$), as in (2.21).

$$1 < I_{\mu j} < \Delta_- \tag{2.19}$$

$$1 < I_{\mu j} = \Delta_- \tag{2.20}$$

$$1 < \Delta_- < I_{\mu j} \tag{2.21}$$

These three possibilities are illustrated in Figure 2.4(d)-(f). Figure 2.4(d) shows the region bounded by 1 and Δ_-. This case therefore corresponds to a tone which is a *small, variable distance below the reference level.* Figure 2.4(e) illustrates the expression corresponding to the level at Δ_-. This case corresponds to a tone which is a *small, consistent distance below the reference level.* Finally, Figure 2.4(f) illustrates the expression for the region below Δ_-. This case corresponds to a tone which is a *larger, variable distance below the reference level.*

It is clear that more phonetic studies will be needed in order to determine how the categories posited in the present framework match up with descriptive categories like "High" and "Extra High". In many cases, earlier descriptive phonological analyses have dispensed with phonetic "details" which would be useful and relevant to determining the phonological analysis under the current approach. Before discussing how the correspondence between traditional descriptive categories such as High and Low and tone interval categories can be assessed, it is necessary to address the more general question of how to describe languages in general. We turn to this issue presently.

An important and central question raised by the proposed descriptive approach is whether a language should be described in terms of paradigmatic categories or not. In Chapter 1 we claimed that all languages represent syntagmatic categories of relative height, which were defined in Section 2.X to be *higher*, *lower*, and *same*, while only a subset of languages represent paradigmatic categories. It is therefore important to define how we can distinguish languages which represent paradigmatic tonal categories, as opposed to those which don't.

We propose that several criteria can be used to assess whether a given language should be described in terms of paradigmatic categories. First, languages for which the tonal categories are realized in relatively fixed positions in a speaker's pitch range across utterances are likely to represent tone in a paradigmatic way.[22] Second, the distances among the tones should be relatively fixed, such that it is possible to characterize the inventory in terms of frequency ratios. Third, it should be possible to identify at least some of the tonal contrasts in isolation (cf. Connell 2000). Fourth, the frequency ratios among the candidate paradigmatic categories should be relatively fixed within and across speakers. Fifth, and finally, languages which genuinely represent paradigmatic are expected to entail a perceptual warping of the tonal space which can be assessed through appropriate perceptual experiments (cf. Hallé, Chang, and Best 2004).

Assuming that a language meets most or all of the above criteria for a paradigmatic system, how can its phonology be described in the present framework? That is, how can we determine the phonological description of paradigmatic tonal categories in terms of the present system? There are three considerations which should significantly aid in this endeavor, which we turn to now.

The first is that, of the paradigmatic lexical tone categories for a given language, we claim that one of these categories will always be equated with the tonic. In other words, the level of some tonal category C, where $T_j \in C$, will be set to the level of the tonic in the phonology. Then $I_{\mu j} = 1$. This claim has implications for the relation between the phonology and the phonetics. In phonological terms, one of the paradigmatic tonal categories will always be given by $I_{\mu j} = 1$. In phonetic terms, whenever category C is realized, then there is an explicit normalizing value against which other tones can be gauged directly.[23]

Second, the present theory proposes a rather straightforward mapping between phonetic values and the phonological description. We assume that perceived pitch is the phonetic parameter which is controlled by the speaker and interpreted by the listener with respect to the phonology. Because F0 is highly correlated with pitch (Moore 1997, Rausch and Plomp 1999), it should be possible to characterize the phonological categories in a rather straightforward way through measurements of F0 ratios. Thus, any phonetically explicit phonological description of a language should include not just F0 values, but theoretically relevant ratios of those values. Given that one tonal category is always equated with the tonic, as discussed above, the remaining tonal categories may be described with reference to that category.

These considerations should be useful in determining the phonological representations of the categories of a language. A related issue concerns how many distinctive "tone levels" a language is permitted to have under the present theory. Systems with five distinctive tone levels have been reported, and it is generally agreed that to achieve descriptive adequacy, a theory of tone must predict more than four levels.[24] The answer is that the present system has no built-in limitations on the number of tone levels that may occur. However, we think that perception and production constraints probably limit the number of levels that can exist in a given linguistic system; the net result of these constraints is that tonal systems apparently cannot sustain more than about five levels. Earlier, we likened paradigmatic tonal systems to musical scale systems. We take credence of the fact scale systems across cultures do not typically represent more than five or seven separate scale notes (Malm 1967, Wade 1997, Burns 1999). We leave the issue of how the number of tonal categories comes to be limited for future work.

On a final note, the present framework has implications for the analysis of contour tones. On the basis of evidence from some East Asian languages, Yip (1980) and others have argued that rising and falling tonal movements cannot be decomposed into a sequence of level tones, and hence that certain rising and falling patterns should be treated as unitary contour tones. The most compelling evidence for the existence of contour tones has been the fact that in many cases, the endpoints of rises and falls do not have the same pitch as level tones in the language, suggesting that it would be incorrect to analyze rises and falls in terms of level tones. However, the present framework casts a different light on this evidence by admitting the possibility that syntagmatic and paradigmatic relations may be specified separately. That is, a contour tone could be described by a combination of two tones, one of which was specified paradigmatically while the other was not. Then the fact that the endpoints of rises or falls did not match the pitches of level tones, as has been observed in some languages (Yip 1980, Xu 1998), would readily be explained by distinct paradigmatic and/or syntagmatic restrictions on discrete tones, as opposed to unitary contour tones.

2.6 Some additional properties of tone intervals

In the following, we briefly consider some additional properties of tone intervals. The first property (the "Reciprocal Property") derives from the fact that tone intervals behave like fractions, so that it is possible to take the reciprocal of such constructs. The second property (the "Multiplicative Property") derives from the fact that tone intervals are multiplicative, just as fractions are. Each of these properties will be useful when we consider how tone intervals can account for "long-distance" interactions between nonadjacent tones in English and other languages. The foundation for an account of such long-distance effects is given by these properties, together with theoretical groundwork which is laid in Chapter 3. Later, in Chapter 5, we provide some examples of how these mathematical properties are useful in providing a phonological description of English data which were originally described in a phonetic way by Pierrehumbert (1980) and Liberman and Pierrehumbert (1984). Pierrehumbert and Liberman demonstrated that the phonetic heights of two nonadjacent tones on stressed syllables are raised or lowered in systematic and highly coordinated fashion which maintains a fixed relative height relationship. Ladd (1988, 1990, 1993) has argued that such systematic control is indicative of a phonological relationship. The mathematical properties to be described help to lay the foundation for a phonological account of such long-distance effects based on syntagmatic tonal features. We therefore turn to a discussion of these two properties now.

2.6.1 The Reciprocal Property

One property of tone intervals which will aid in building a formal account of long-distance interactions in Chapter 3 the fact derives from that such constructs behave like mathematical fractions. Thus, like other fractional ratios, tone intervals can be "flipped upside down" under reciprocal operations, so as to reverse the places of the numerator and the denominator. The effect of such operations is to change the tonal properties of tone intervals in well-defined ways.

We can define the reciprocal operation in terms of a formal linguistic statement, the Reciprocal Property, which is given as follows.

Reciprocal Property: For any tone interval $I_{r,T}$, there exists a reciprocal tone interval, $(I_{r,T})^{R} = I_{T,r}$.

The Reciprocal Property corresponds to a general property of tone interval constructs and as such cannot be considered a rule. A raised letter "R" is used to indicate the reciprocal operation. To see how the Reciprocal Property works, suppose we have a tone interval $I_{1,2}$, which is given as in (2.22):

$$I_{1,2} = \frac{T_2}{T_1} \tag{2.22}$$

Then the reciprocal tone interval of $I_{1,2}$ is $(I_{1,2})^{\mathrm{R}} = I_{2,1}$. This is shown below.

$$I_{2,1} = \frac{T_1}{T_2} \tag{2.23}$$

While one effect of the Reciprocal Property is to reverse the numerator and denominator, a concomitant effect is to reverse the referring tone-referent relationship. Thus, in $I_{1,2}$, T_2 is the referring tone and T_1 is the referent, while in the reciprocal tone interval $I_{2,1}$, T_1 becomes the referring tone and T_2 is the referent.

What are the effects of taking the reciprocal on the features of a tone interval? In general, taking the reciprocal corresponds to taking the opposite value of [±higher] while leaving the value of [±same] unchanged.[25] This means that under reciprocal operations, tone intervals which are *higher* become *lower*, and vice versa, while tone intervals which are *same* remain unchanged. For syntagmatic tone intervals, this means that rises become falls, and vice versa, while level contours remain unchanged. For paradigmatic tone intervals, this means that a tone which is higher than the tonic suddenly becomes lower than the tonic, and vice versa; while a tone which is at the tonic remains unchanged.

The effects of taking a reciprocal on tone interval features can be summarized as a corollary of the Reciprocal Property. We refer to this corollary as the Relative Feature Complementarity Corollary.

Relative Feature Complementarity (Corollary): For tone interval $I_{r,T}$ associated with referring entity X, referent Y, and a relational feature specification [α same, (β higher)], there is a reciprocal tone interval $I_{T,r} = (I_{r,T})^{\mathrm{R}}$ associated with referring entity Y, referent X, and a relational feature specification [α same, (-β higher)].

Relative Feature Complementarity simply codifies the fact that the features of a tone interval and of its reciprocal are related in a regular way. In particular, the reciprocal of a given tone interval will have the same value of [±same] compared with the original tone interval. However, the reciprocal will have an opposite specification of [±higher] compared to the original tone interval.

What sorts of attested linguistic phenomena can this reciprocal process explain? One example comes from the Wu dialect of Chinese, in which rising contours become falling contours, and vice versa (Chen 2000). Such processes have been cumbersome to describe in autosegmental terms, which has treated such phenomena as the outcome of metathesis operations (so that LH changed to HL, or the other way around). A straightforward account is permitted under tone interval theory, in which the operation of "taking the reciprocal" arises as a natural property of the construct itself, so that a rule-based account is unnecessary. Indeed, we can now understand this behavior in terms of a reciprocal operation on the feature [±higher]. Suppose the relation of T_2 to T_1 is [+higher], so that the sequence generates a rise. Then the reciprocal of this relation will be [-higher], so that the sequence generates a fall. In other words, we can understand such phonological processes not as the outcome of metathesis *per se*, but as the outcome of a reciprocal operation.

Finally, we can consider some ways in which the existence of reciprocal operations yields greater insight into the nature of syntagmatic systems in general, in two ways. Consider first the fact that the same syntagmatic relation between two tones can be expressed either in forward-time or in reverse-time.

The Reciprocal Property shows why this should be so. That is, when the numerator and denominator of a tone interval are flipped under a reciprocal operation, it reverses the referring tone-referent relationship. For syntagmatic tone intervals, the referring and referent tones always corresponds to a time-ordered pair; thus, "flipping" a tone interval under a reciprocal operation time-reverses the syntagmatic relation.

Moreover, the Reciprocal Property explains a puzzling fact about relational tonal features in general. This stems from the fact that there are two technically equivalent ways of expressing a relation of relative height between a tone and a referent which appear to be polar opposites on the surface. A contour which exhibits a rise from T_1 to T_2, for example, may be accounted for in one of two ways in terms of relational features. On the one hand, we can say that T_2 is *higher* than T_1. On the other hand, we can say that T_1 is *lower* than T_2. However, we have just described the same physical state of affairs in terms of two opposite-valued relations – *higher* vs. *lower*! This seems like a paradox; how can the same situation be described in terms of opposite relational values? This observation is explained in a straightforward way by the Reciprocal Property. According to the Reciprocal Property, the statement "T_2 is higher than T_1" is formally equivalent to the statement "T_1 is lower than T_2", under a reciprocal transform. Then the value of [+higher] is reversed under a reciprocal transform, consistent with the observation that the contour can be described either in terms of the relation *higher*, or in terms of the relation *lower*. The fact that we can account for an apparent paradox – namely, that the same physical state of affairs can be described in two equivalent but seemingly opposite ways – provides one argument for basing a relational theory on tone intervals, rather than just tones. It is unclear how the tone-only theory could explain the fact that two tones may be associated with opposite-valued, but clearly situationally equivalent, features. In contrast, appealing to expressly relational tone interval constructs, as well as reciprocal operations, provides a natural explanation for this puzzling phenomenon. It is not clear how a theory based solely on tones plus relational features could capture these facts. Having elucidated one important property of tone intervals, we turn to the description of a second significant property.

2.6.2 The Multiplicative Property

Another property of tone intervals which will be useful in describing long-distance interactions between tones relates to the fact that, like mathematical fractions, tone intervals are multiplicative. However, tone intervals cannot be multiplied arbitrarily; only adjacent tone intervals can be multiplied. In Chapter 3 we will define a general notion of adjacency in terms of a structure known as a *metrical grid*, which is a linguistic device for representing timing and relative prominence (Liberman and Prince 1977; Hayes 1980, 1995; Selkirk 1984; Halle and Vergnaud 1987). For the moment, however, we can present an example of how a sequence of adjacent tones T_1 T_2 T_3 gives rise to two consecutive tone intervals which may be multiplied. Suppose that we define two tone intervals $I_{1,2}$ and $I_{2,3}$ which respectively relate T_1 to T_2, and T_2 to T_3, as follows:

$$I_{1,2} = \frac{T_2}{T_1} \tag{2.24}$$

$$I_{2,3} = \frac{T_3}{T_2} \tag{2.25}$$

Then multiplying these tone intervals gives the following expression:

$$I_{1,2} \cdot I_{2,3} = \frac{T_2}{T_1} \cdot \frac{T_3}{T_2} \tag{2.26}$$

Just as for mathematical fractions, like "terms" can be cancelled from the numerator and denominator. Then on the right hand side of the expression, T_2 can be cancelled. We are left with the following:

$$I_{1,2} \cdot I_{2,3} = \frac{T_3}{T_1} \tag{2.31}$$

Note that the right hand side of the expression now has the form of a T_I with referring tone T_3 and referent tone T_1. This is equivalent to $I_{1,3}$, which we substitute into the above expression to give the desired result.

$$I_{1,3} = I_{1,2} \cdot I_{2,3} \tag{2.32}$$

The multiplicative property of tone intervals ultimately permits a relation between nonadjacent tones to be derived from a sequence of adjacent tones. In Chapter 3, we derive a formal phonological description of relative height relations between nonadjacent tones which will permit us to account for data from English in Chapter 5. This formal account makes use of the Multiplicative Property of tone intervals, as well as the Reciprocal Property and the metrical grid.

2.7 Summary

The primary goal of this chapter was to redefine the features of tonal systems so as to permit either syntagmatic or paradigmatic relations to be defined. This was accomplished by introducing a phonological construct known as a tone interval, which relates a tone to a referent. A tone interval captures a syntagmatic relation of reference when the referent entity is another tone, and a paradigmatic relation of reference when the referent is the speaker's tonic level. We also introduced two binary-valued relational features: [±same] and [±higher]. Because of the feature geometry of these relations, they give rise to three primitive relations: *higher*, *lower*, and *same*. These relations describe the relative height relation between a tone and its referent for a given tone interval.

It was shown that the relative height relations (*higher*, *lower*, and *same*) have a straightforward interpretation in terms of ranges of tone interval values. This is because relations of relative height can be interpreted in terms of the relative magnitudes of the referring tone and referent. These patterns of relative magnitude, in turn, have an interpretation in terms of ranges of values for tone interval ratios. Then *higher* was shown to correspond to a range of tone interval values such that $I > 1$, while *lower* and *same* were shown to correspond to $I < 1$ and $I = 1$, respectively.

Phonological categories were defined as restrictions on ranges of values for paradigmatic or syntagmatic tone intervals. We claimed that the tonal relations $I > 1$, $I < 1$, and $I = 1$ for syntagmatic tone intervals are universally attested in all linguistic tonal systems. Moreover, we showed how language-specific phonological categories could be constructed by taking the intersection of two regions of tone interval space. The first region corresponded to a paradigmatic or syntagmatic tone interval which was specified to be *higher* or *lower*, while the second region corresponded to a region which was above, below, or at the same level as some language-specific cutoff value.

Finally, we illustrated that tone intervals behave in several ways like mathematical ratios or fractions. For example, we showed that it is possible to take the reciprocal of a tone interval. The effect of this is to alter the features of the resulting tone interval relative to the original, so that a tone interval which was *higher* becomes *lower*, and vice versa, while a tone interval which was *same* remains unaffected. This property helps to provide a more cogent explanation for why rises change to falls, and vice versa, in some languages. Finally, we described how tone intervals can be multiplied, so that a sequence of adjacent tones gives rise to a relationship between nonadjacent tones.

Having thus redefined tonal features, we turn in the next chapter to two issues: the role of tones in the phonology, and the organization of tones and tone intervals at the phrasal level.

Chapter 3 – Organization of tone intervals at the phrasal level

The last chapter showed how tone intervals permit a redefinition of tone features as dually syntagmatic and paradigmatic, thereby providing a new explanation for why tone sometimes appears to be "segmental" and at other times "suprasegmental". In this chapter, we build the account by discussing how tone intervals are organized at the phrasal level. The theoretical groundwork which is laid in this chapter will ultimately permit us to provide a new description of English intonation in Chapter 5. It will be shown that this account avoids problems of phonological overgeneration and phonetic indeterminacy that were associated with paradigmatic descriptions of English, such as Pierrehumbert (1980), while also providing a description which is better in line with empirical data collected over the past two decades. Moreover, we will show that this account is simpler than previous descriptions, in that it assumes only six singleton primitives and no additional rules for phonetic implementation.

The present chapter also fills a gap in the literature by proposing that *tones universally associate with respect to metrical structures*. In this way, the chapter presents a theoretical framework capable of accommodating a range of cross-linguistic facts concerning interactions between tones and metrical structure. The inspiration for these proposals comes from music, which is a domain where interactions between tones and metrical structures are not only well-attested, but well-understood. We will show that by virtue of their associations with hierarchical metrical structures, nonadjacent tones can participate in tone intervals over long distances. This not only presents a new way of representing "stepping" phenomena like downstep and upstep, but also nested relative height relations across intonational phrase boundaries of the sort reported in English by Pierrehumbert (1980), Liberman and Pierrehumbert (1984), and Ladd (1988), as we show in Chapter 5. Once again, the inspiration for these proposals comes from music, where such long-distance interactions have been extensively studied.

This chapter is organized as follows. Section 3.1 introduces the topic of how tone intervals and tones are organized at the phrasal level by developing an account for some simple intonation contours in English. Along the way we will discuss fundamental issues concerning how tones associate with metrical structures, and we will propose some principles for their coordination. Next, Section 3.2 elaborates on the topic of how tones associate with respect to metrical structures. Finally, Section 3.3 describes the principles by which the association of tones with hierarchical metrical grid structures permits the syntagmatic interactions between nonconsecutive tones. In Chapter 5 we will show how the theoretical groundwork laid in this chapter permits a straightforward, phonetically transparent phonological account of intonation data from English.

3.1 Developing an account for some simple examples

A key issue to be dealt with in any theory of tone is how the tonal representation is organized at the levels of the word and the phrase. Because our focus in this work is on intonation languages, it is necessary to limit our attention largely to the issue of tonal representation at the phrasal level. In contrast, we will have little more to say on the topic of the paradigmatic representation of tone at the lexical level, which is a topic that must necessarily be left for future work.

To introduce the topic of how syntagmatic tone intervals and their associated tones are organized at the phrasal level, we will show how we can derive a tone interval description for a simple English intonation pattern. This will permit us to introduce some of the principles which in this theory govern how the tones of tone intervals associate with metrical structures of utterances. In the following, we first review some of the properties of tones. Next, we will describe how the tones in tone intervals come to be associated with particular positions in metrical structures. This will permit us to introduce some of the principles which influence how tones are arranged with respect to metrical structures. We will also derive a useful variant notation for tone intervals which will permit us to more readily see how their tones come to find places within a metrical structure. Finally, we will describe briefly how the final structure comes to be realized in terms of an F0 contour; this topic will be discussed in more depth in Chapter 4, when we consider the issue of how to properly evaluate theories of the phonetics and phonology of intonation systems.

In Chapter 2, we identified a dual function for tones. First, they serve as constituents which enter into a tone interval construct, thereby participating in a tonal relationship of reference. We noted that tone intervals, not tones, give rise to the phonological categories of this system. Second, we noted that tones also serve as a kind of "pointer" to a timing position in a metrical structure. In this chapter, it will become clear how these two notions of tones are related.

We will focus in this section on how we can describe the utterance shown in Figure 3.1. This simple contour shows a rise to a prominent, accented syllable, followed by a fall; such a pattern is common for declarative statements. We assume that this contour arises through the association of tones with particular positions in the utterance, where these tones are arranged into tone intervals designating a particular set of relative height relations among the tones.

In order to describe the utterance in Figure 3.1 in terms of tone intervals, we will need to consider several issues. First, we need to elucidate the principles by which tones come to be associated with the metrical structures of utterances. Moreover, we must deduce what relative height relations are specified for tone pairs residing in the tone intervals that give rise to the tonal pattern in Figure 3.1.

We assume that tones universally align with timing positions in utterances by referencing structures known as *metrical grids*. A metrical grid is a linguistic device for representing the relative prominence and timing of syllables. In a metrical grid representation, the timing and relative prominence pattern of a sequence of syllables is represented in terms of rows and columns of X's. Each column of X's represents a single timing position, and every syllable in a metrical grid is associated with one or more X's on the lowest row of the grid. The number of X's in a column for a single syllable indicates the relative prominence of that syllable with respect to other syllables. Timing positions associated with metrically stressed or prominent elements are represented by a column of height 2 or higher, while timing positions which are associated with metrically weak elements are represented by a column of height 1. A metrical grid for the utterance in Figure 3.1 is shown in Figure 3.2.

We will refer to each column of X's in a metrical grid as a *timing position* (or a *timing slot*), which we interpret as an abstract phonological quantity of time which is associated with a segmental or syllabic position in a word or a phrase. Moreover, we will refer to the segmental or syllabic position which the grid column in turn is associated with as the *tone-bearing unit.*[26, 27] We will indicate an association between a tone and a timing slot by drawing a line from one to the other; here, association is defined as temporal co-occurrence. For example, the association of a tone *T* with a timing slot would be indicated in the following way:

$$\begin{array}{c} \mathrm{X} \\ | \\ T \end{array}$$

Later on, when we consider the relationship between phonology and phonetics, it will be important to clarify further how tone intervals give rise to tonal patterns. In this regard, it is worth noting that tones play an important role in the eventual phonetic realization of the phonological structure. In particular, the fact that tones associate with particular timing positions in metrical grids means that they indirectly "point to" a particular segmental or syllabic position – that is, a tone-bearing unit. This is because each metrical grid column is itself associated with one of two types of metrical primitives, i.e. a mora or a syllable, where this appears to be a difference among languages (Hayes 1995). In this way, tones have the effect of designating the corresponding tone-bearing unit to realize the relations codified in tone intervals. In other words, by associating with particular timing slots, tones indirectly mark the associated tone-bearing unit to fulfill tone interval constituency relations of relative height.[28]

An important factor which affects how tones are associated with metrical grids concerns the type of tone in question. We assume that there are two types of tones – *starred tones* and *unstarred tones.* These two tone types have distinct requirements regarding how they can associate with respect to the metrical structure of an utterance. Following the conventions of Goldsmith (1976) and Pierrehumbert (1980), *starred tones* are required to associate with metrically prominent positions. Such tones are marked with an asterisk: T^*. On the other hand, *unstarred tones* must associate with metrically nonprominent positions. These tones lack an asterisk: T.

Now that we have described the basics of how tones associate with timing positions in a metrical grid, we can consider how we might describe the contour in Figure 3.1 in terms of tones and tone intervals. This contour shows a simple rising-falling pattern, which we can describe as arising from a sequence of three tones: one at the beginning, one on the stressed syllable, and one at the end. We will describe these by the tone sequence T_1* T_2* T_3. Given that we have now distinguished between starred and unstarred tones, we claim that the representation for this contour is therefore given by a sequence of two tone intervals, as in (3.1) and (3.2). As we will show, these two tone intervals are sufficient to describe the tonal properties of the entire phrase in Figure 3.1.

$$I_{1,2} = \frac{T_2{}^*}{T_1{}^*} \qquad (3.1)$$

$$I_{2,3} = \frac{T_3}{T_2{}^*} \qquad (3.2)$$

Here, we have used a notation for tone intervals which is identical to that used in Chapter 2, except for the fact that that we have distinguished starred and unstarred tones. In general, we assume that the phonological representation of syntagmatic tone intervals codes not only for the relative heights of tones, one with respect to one another, but also whether the corresponding tones comprising those tone intervals are starred or unstarred.

The tone interval pair in (3.1) and (3.2) simply names the tones participating in $I_{1,2}$ and $I_{2,3}$; no relative height relations are indicated by this representation. Examining the contour in Figure 3.1, we can see that the first pair of tones must ultimately generate the rise, while the second pair of tones must generate the fall. We can therefore infer that the tone interval values specified at the level of the phonology are given as in (3.3) and (3.4):

$$I_{1,2} > 1 \tag{3.3}$$

$$I_{2,3} < 1 \tag{3.4}$$

The notation shown in (3.3) and (3.4) should again be familiar from Chapter 2; it represents the relative height relations between T_1*, T_2*, and T_3. However, there is one disadvantage to this representation, in that it does not permit us to simultaneously represent both the relation of relative height coded in the tone interval, as well as the starred or unstarred status of the tones themselves. Therefore, we will occasionally utilize a different notation for tone intervals which will be especially useful in visualizing how the tones in tone intervals come to be aligned with particular timing positions at the phrasal level. We can represent the relative height relations in tone intervals expressed in (3.3) and (3.4) in a way that makes explicit their connection with particular tones, as in (3.5) and (3.6), respectively:

$$\begin{array}{|cc|} T_1^* & T_2^* \\ \multicolumn{2}{|c|}{I > 1} \\ \hline \end{array} \tag{3.5}$$

$$\begin{array}{|cc|} T_2^* & T_3 \\ \multicolumn{2}{|c|}{I < 1} \\ \hline \end{array} \tag{3.6}$$

The next step in describing the translation from the phonological representation to the phonetic contour in Figure 3.1 is to describe how the tones comprising syntagmatic tone intervals come to be associated with particular timing positions. In intonation languages, tones are associated with timing positions via reference to two kinds of properties: *metrical properties* and *positional properties*. We will first focus on how metrical properties affect the placement of tones in a phrasal structure.

Our assertion above was that the tonal contour in Figure 3.1 could be described by three tones, which we claimed were located at the beginning, on the stressed syllable, and at the end of the utterance. However, we would like to derive some principles by which such a description might follow naturally. One such principle concerns the association of tones with the edges of utterances. In general, we will assume that each and every utterance always has a tone at its beginning and a tone at its end. We codify this assumption as the Minimal Tone Principle, which is given below.

Minimal Tone Principle. The first and last timing positions of an utterance must be associated with a tone.

The Minimal Tone Principle is essentially a statement about the well-formedness of the tonal representation at the phrasal level. It states that at the phrasal level every utterance must be associated minimally with one initial tone and one final tone. These tones can be thought of as "boundary tones," although there is nothing special about them other than the fact that they are initial or final in their domains. Depending on the phonology of individual languages, initial and final tones may be supplied lexically, or they may be supplied at the phrasal level.

The Minimal Tone Principle derives from a claim about the possible relations between the phonetic F0 contour and phonological targets. In particular, it ensures that every portion of a tonal contour arises either through direct phonetic instantiation of a phonological target, or through interpolation between phonological targets. Indeed, the only way that we can ensure that F0 contour associated with every timing position in an utterance arises through one of these two mechanisms is to require that the initial and final positions in an utterance are specified for tone. This principle resolves an ambiguity in the theory of Pierrehumbert (1980) regarding how the initial position of a contour is instantiated phonetically.[29, 30]

The effect of the Minimal Tone Principle with respect to Figure 3.1 is to require that the timing positions associated with *Show* and *-ney* must be associated with tones. At this point, we now have enough information to understand how tones must be associated with timing positions in the phrase for our example contour in Figure 3.1. In particular, tones T_1* and T_3 must be aligned with the first and last timing positions in the utterance, while T_2* must be aligned with the remaining metrically prominent position in the utterance, on *mo-*. This pattern of association is shown in Figure 3.3. Because the expressions in (3.5) and (3.6) mutually refer to T_2*, this tone is indicated only once in the figure. However, its dual constituency with respect to each of the tone intervals is indicated by the fact that two separate tone intervals connect with it.

We assume that the geometry of the tonal representation is such that a tone associates with every X in a particular grid column. The net effect of this pattern of association between tones and timing slots is to permit tones to be represented on higher grid levels. We defer discussion of this topic until a fuller treatment of metrical grids is possible in Section 3.2. For now, we simply note that each tone associates with all X's in its respective column. As a result, tone T_2* shows a single vertical line stemming upward and connecting with an X on grid level 2, as shown in Figure 3.4.

We are nearly finished with our discussion of the example in Figure 3.1. The last step is to discuss how the phonological representation gives rise to a phonetic F0 contour, a topic which will be a major focus of Chapter 4. The first step in this process is that the tone-bearing units which are marked by tones are assigned target pitches which fulfill the relative height relations specified in the associated tone intervals. In this way, target pitches are implemented on *Show*, *mo-*, and *-ney* in conjunction with the underlying tones T_1, T_2* and T_3. *Show* has a higher pitch than *mo-*, and *-ney* has a lower pitch than *mo-*, as required by the associated tone intervals. Next, these target pitches are connected up by monotonic pitch interpolation functions.[31] We note that the emphasis here is on pitch, rather than F0, since we assume here that the primary phonetic correlates of the phonological representation are *perceptual* in nature. An F0 contour readily follows from a representation based on pitch, since F0 is indeed the primary acoustic correlate of pitch. The end result is the contour seen in Figure 3.1.

One important distinction between the present theory and earlier work concerns the nature of the tonal inventory. Here, we assume that there are only two types of tones: starred tones and unstarred tones. As a result, the number of categories is significantly smaller in the present theory than in previous theories. This may raise questions for some readers concerning the status of phrasal constituents in the present theory, such as the intonational phrase. Consistent with earlier work, we assume that tones and tone intervals are organized into intonational phrases.[32] Phrase-level phenomena which were previously attributed to distinct tonal types, such as "boundary tones" (Liberman 1975; Pierrehumbert 1980), are dealt with in the present theory in two ways. First, we will propose in Chapter 5 that tone intervals at phrase edges exhibit distinctive phonetic, as well as phonological, properties. In that chapter we will also present a hypothesis that universal principles of auditory perception restrict possible sequences of tone intervals and how those tone intervals may be realized phonetically. In our view, such principles have the potential to explain a number of phrase-level tonal phenomena.[33] Second, we assume that there are interactions between intonational phrase boundaries and the metrical grid, so that durational lengthening sometimes attributed to the intonational boundaries *per se* instead results from metrical adjustments to the grid at phrasal boundaries (e.g., Selkirk 1984). For the moment, however, we will simply introduce the notation "%", which indicates the edge of an intonational phrase constituent; this symbol will be indicated alongside tonal information to indicate a well-formed intonational phrase constituent. These issues are taken up again in Chapter 5.

Suppose now that we consider how to describe the tonal pattern in Figure 3.6. This contour shows an overall falling-rising pattern, consistent with a typical interrogative pattern in English. We note that the F0 level at the end of the utterance is high, but well with the range of the rest of the utterance. The falling-rising pattern is consistent with a sequence of three tones comprising two tone intervals, just as for the contour in Figure 3.1. Moreover, the Minimal Tone Principle dictates that the first and last timing slots are necessarily associated with tones. By inference, then, the positions of the tones are the same as in the contour in Figure 3.1. Moreover, the edges of intonational phrase boundaries are indicated.

This suggests that the only difference between the phonological representation for the contour in Figure 3.1 and the representation for the contour in Figure 3.6 is in the sequential identities of the tone intervals. Rather than a sequence up-down (i.e., *higher-lower*) as in Figure 3.4, the representation for the contour in Figure 3.6 involves a sequence down-up (i.e., *lower-higher*). Thus, the representation for the contour in Figure 3.6 is given in Figure 3.7. These examples were intended as a brief overview of the kinds of phonological descriptions which the present framework provides. We will return to the description of English intonational patterns in Chapter 5.

3.2 Interactions between tones and metrical grids

In the examples of the previous section, we briefly reviewed the sequence of steps by which a pitch contour is constructed from a tone interval representation for an English utterance. This permitted us to illustrate a few things about the association of tones with timing positions. However, there are some additional issues regarding the nature of associations between tones and timing slots that require further discussion, which we consider in this section. We will first address some issues related to what sorts of patterns of association between tones and timing slots are permitted or restricted in this theory. Next, we will discuss how our proposal that tones associate with respect to metrical grids fits in with earlier work. Finally, we will propose a principle by which unstarred and starred tones are associated with metrical grids in a coordinated fashion.

3.2.1 Restrictions on the association of tones with timing slots

In this section, we consider two issues regarding the nature of associations between tones and timing slots. One issue concerns our implicit assumption that some timing slots are unspecified for tone at the surface phonological level. This phenomenon is termed *surface underspecification*, and we assume that such underspecification is freely permitted in many languages. The claim that timing positions may be underspecified for tone is justified by phonetic experiments of Pierrehumbert and Beckman (1988). They showed that for Japanese H*+L accents, the slope of the F0 contour between the H and L tonal targets is increasingly shallow, the more syllables that intervene between the targets. This is consistent with the claim that these intervening syllables are not specified for tone.

A second issue concerns restrictions on the possible geometries by which tones and timing slots may be associated. We assume that possible patterns of association between tones and timing slots are described by the Principle of Tonal Association (PTA), which is stated below.

Principle of Tonal Association (PTA). Every tone must associate with exactly one timing slot.

There are two central implications of the PTA. The first claim of the PTA is that *tones do not perseverate in time.* In other words, tones do not "spread" across multiple segments. The second claim of the PTA is that *all tones associate with timing positions.* The implication is that there are no phonetically unrealized "floating" tones. We clarify these points below.

The PTA essentially codifies possible configurations between tones and timing slots. In particular, the PTA permits only the configurations like those in Figures 3.5(a) and 3.5(b) to exist between tones and timing slots. These two structures indicate one-to-one and many-to-one relations, respectively.[34]

A number of possible configurations which had been permitted under earlier theories are ruled out by the PTA. In particular, the patterns of association shown in Figures 3.5(c) and 3.5(d) are not permitted by the PTA; these indicate perseveration of a given tone across multiple timing positions. Ruling out perseveration of tones requires that all tones be localized to a particular syllable (or sub-syllabic unit). Moreover, the patterns of association shown in Figures 3.5(e) and (f) are also ruled out by the PTA; these indicate association of a tone with something other than a timing position. For example, Figure 3.5(e) indicates a "floating" tone which is not associated with any timing position.35 Moreover, Figure 3.5(f) indicates a tone which is associated not with a timing position, but rather with another tone; this structure was implicit in the bitonal accents proposed by Pierrehumbert (1980). Having clarified the possible geometries by which tones can associate with timing slots, we are ready for a fuller discussion of how tones and their corresponding tone intervals are associated with metrical grids.

3.2.2 Evidence for interactions between tones and metrical grids

A central claim of tone interval theory is that tones universally associate with timing slots via metrical grids. This section elaborates on the representation of tone at the phrasal level. In particular, we will explore the hierarchical structure of metrical grids and the implications of this structure for the representation of tones and tone intervals.

Metrical grid structures are actually one of several devices which have been proposed for representing the timing and relative prominence of utterances. These structures have their roots in a rich body of work since the mid-1970's collectively known as metrical stress theory (e.g., Liberman 1975; Liberman and Prince 1977; Hayes 1980, 1995; Selkirk 1984; Halle and Vergnaud 1987). Of the linguistic devices which have been proposed for representing relative prominence and timing, there are a number which also indicate constituency, such as bracketed grids and metrical trees. In what follows, we have chosen to leave out overt constituency relations, instead utilizing pure grid structures. While English clearly does indeed represent metrical constituents, such as feet, there are two reasons why we have left brackets out of our grids. First, leaving out constituency relations appears to avoid a set of problems in tune-to-text matching which were encountered by Liberman (1975), as discussed in Chapter 1 of Pierrehumbert (1980). Second, it is apparently possible to account for intonational phenomena in English without the use of bracketed grids. It may be that certain facts about tonal alignment (i.e., association) are more elegantly captured using bracketed grids; we leave this issue for future work. However, the lack of brackets here should not be taken as denial of the need for brackets in grids in general. Indeed, many linguistic phenomena appear to require constituency relations for explanatory adequacy (e.g., Halle and Vergnaud 1987, Manfredi, 1993, Hayes 1995). Moreover, constituency relations are attested across languages. We see it as an open question, however, whether the tonal representation necessarily must make reference to constituency relations in all languages.

In the three decades since metrical stress theory was first proposed, a substantial body of evidence has accumulated concerning the interaction of metrical systems and tonal systems. Metrical systems have been shown to interact with tonal systems in a wide variety of languages, including lexical tone, pitch accent, and intonational systems (e.g., Kisseberth 1984, Rice 1987, Inkelas and Zec 1988, Manfredi 1993, Kenstowicz 1994, Duanmu 1999, Kenstowicz and Sohn 2001, deLacy 2002). This burgeoning body of evidence about the interaction between tonal and metrical systems has not been placed within a comprehensive framework, in spite of recent work which has focused on the interaction between tone systems and metrical systems (e.g., Purnell, 1997).

The present work fills this gap by proposing that *tonal representations are universally derived through association with metrical grid structures*. For reasons stemming from production and perception, it makes sense to us that the tonal and metrical structures of utterances would be integrated. From the standpoint of production, there is substantial evidence that tones are temporally coordinated with syllables and segments, as discussed in Chapter 4. Such coordination requires that abstract timing structures be generated to coordinate laryngeal and supralaryngeal gestures. From the standpoint of perception, there is good evidence from experiments in psychology that the perception of tone and meter are interrelated (e.g., Jones and Ralston 1991). These arguments, in addition to cross-linguistic evidence of the interaction of tonal and metrical systems, appear to warrant a comprehensive proposal of tone and meter.

In what follows, we will assume that metrical grid structures are computed before tonal rules apply.[36] In general, the metrical structure at the phrasal level is by and large derived from the metrical properties associated with individual words (Liberman and Prince 1977, Selkirk 1984, Halle and Vergnaud 1987, Hayes 1980, 1995). For those unfamiliar with metrical stress theory, a useful discussion of how metrical structures are derived, together with a review of the older literature, is given in Hogg and McCully (1987).[37]

We have already claimed in Section 3.1 that tones associate with timing positions in metrical grids; as we discussed, timing positions are indicated by a sequence of X's on the lowest row of the grid structure. That tones align with timing positions is consistent with phonetic evidence regarding tonal alignment, which will be reviewed in Chapter 4. In particular, tones have been shown to align with respect to segments and/or syllables; such elements can be represented as timing positions in a grid in a straightforward way. The number of timing positions associated with a syllable depends on its segmental composition, lexical stress properties, and its position in a phrase (e.g., Broselow, 1995). Stressed syllables are typically longer than unstressed syllables; hence, they are typically represented by two timing positions at the lowest grid level, compared with one for unstressed syllables. In general, we assume that phonetic timing properties reflect the structure of an underlying grid structure in a straightforward way. (See also Broselow, Chen, and Huffman 1997.) However, many of the details of the correspondence between grid structures, timing positions, and phonetic length remain to be worked out.[38]

There is good evidence that the manner in which tones align with timing positions reflects the influence of metrical prominence. In English, tunes are temporally expanded or compressed so that the turning points (i.e., the approximate phonetic exponents of tones) are associated with respect to metrically prominent syllables in utterances (e.g., Halliday 1967, Liberman 1975; Ladd 1996). Moreover, there are well-attested restrictions on the alignment of tones with metrical prominence patterns in many African languages; see Odden (1995) for a review. Goldsmith (1976) proposed to model the characteristic tonal alignment patterns by designating tones as "starred" or "unstarred". Pierrehumbert (1980) adopted this notation of the star for her description of English; in more recent work, starred and unstarred tones have been interpreted in terms of metrical prominence, such that starred tones must align with metrically prominent syllables. It is worth pointing out, however, that Pierrehumbert defined the "star" notation in a dual way which has led to difficulties in applying the description to other languages (Arvaniti, Ladd, and Mennen 1998, 2000). In Chapter 4, we will clarify the definition of the "star" notation in a way that removes the ambiguity pointed out by Arvaniti *et al.* (1998, 2000). In the following section we consider another way in which metrical grids influence the patterns of association of tones.

3.2.3 More on the association of tones with timing slots

We have discussed several factors which affect the association of tones with metrical grid structures. In this section, we will discuss two more such factors, which affect starred and unstarred tones, respectively. We will begin by addressing a principle which seems to govern permitted patterns of association between unstarred tones and metrical grids. This principle will be important in Chapter 5 when we discuss intonational phenomena in English described by Pierrehumbert (1980) as complex bitonal pitch accents. By framing tonal alignment in terms of metrical structures, we will be able to propose a means for reconciling phonological theory with a significant body of phonetic evidence concerning the alignment of tones in these bitonal pitch accents. In particular, this principle permits a means of accounting for empirical data suggesting that the two tones in bitonal pitch accents are not separated by a fixed temporal interval, as originally claimed by Pierrehumbert (1980).

It appears that unstarred tones may apparently associate only with certain metrically weak timing positions which are *next to a prominent position occupied by a starred tone.* This restriction is formulated as the Metrical Prominence Association Rule (MPAR), which is given below.

Metrical Prominence Association Rule. Unstarred tones not located at phrasal edges must associate with weak positions adjacent to prominent positions associated with a starred tone.

The MPAR is motivated by facts about alignment in intonation languages, as reviewed in Chapter 4. The MPAR suggests that the structure in Figure 3.8(a) represents a well-formed association pattern for an unstarred tone, because it associates with a weak position which is adjacent to a prominent position that is linked to a starred tone. However, the structure in Figure 3.8(b) is ill-formed under the MPAR. This is because an unstarred tone occupies a position adjacent to a prominent position, but that position does not have a starred tone.

We note that the MPAR exempts unstarred tones located at phrasal edges from aligning next to a "starred position". One reason for the exemption is that the Minimal Tone Principle takes precedence over the MPAR, so that unstarred tones associated with the initial and final timing slots in a phrase need not be adjacent to "starred positions." Another reason for the exemption is that the behavior of unstarred tones at phrase edges appears to be governed by varied and language-specific factors which are not easily characterized. A full description of the principles of unstarred tone alignment in phrase-final positions will therefore not be undertaken here.[39]

There is one additional fact about unstarred tones which should be mentioned. That is, we propose that unstarred tones are designated in the phonology to be attracted either to a *leftward* "starred position" or to a *rightward* "starred position." We will indicate unstarred tones which are leftward attractees with a "+" sign to the left of an unstarred tone, *+T*. Moreover, we will indicate unstarred tones which are rightward attractees with a "+" sign to the right of an unstarred tone, *T+*. Then a sequence *T* +T, T** would align with a metrical grid structure as in Figure 3.9(a), while a sequence *T* T+ T** would align with a metrical grid structure as in Figure 3.9(b).[40]

Next, we turn our attention to starred tones. As discussed earlier, these tones must associate with grid columns of height 2 or higher. However, we have not said anything about their distributions with respect to metrical grids. In this regard, we take note of some observations by Liberman (1975) and Pierrehumbert (1980) concerning the distribution of accentual tones across metrical structures. In particular, Pierrehumbert (1980: 37) characterizes the interaction between metrical strength and accentuation this way:

> If a foot has a pitch accent, any foot of equal or stronger metrical strength in the phrase also has a pitch accent[,] except that there are no pitch accents after the nuclear stress of the phrase.

Here, a "pitch accent" is interpreted to mean a "starred tone". We will adopt this general insight concerning the interaction of starred tones and metrical grids, which will be stated as a formal principle below. However, we think that the exceptional clause is not necessary, and that the restriction on pitch accentuation after the nuclear stress can be explained in terms of other factors, a topic to which we will return in Section 5.4.2. We therefore reformulate Liberman's and Pierrehumbert's insights in terms of the Metrical Strength Rule, which is given below.

Metrical Strength Rule. If a grid column is associated with a starred tone, any grid column of equal or stronger metrical strength in the phrase also has a starred tone.

Having now discussed factors which affect the association of starred and unstarred tones with metrical grids, we are ready to explore how the hierarchical structure of metrical grids affects the kinds of tonal structures permitted in the phonology.

3.3 Hierarchical structuring of tone intervals

The preceding principles dealt chiefly with how tones associate with respect to the "horizontal dimension" of metrical grids, which relates to utterance timing. We turn our attention to how tones associate with respect to the "vertical dimension" of grids, which reflects the relative prominence patterns of syllables. With regard to this "vertical dimension," our goal in this section is to describe how hierarchical relations come to be specified between tones which are not sequential in time. What we will see is that tones which are not adjacent on lower grid rows may come to be adjacent on higher grid rows. When this happens, they give rise to a relative height relation of some sort, as specified in a syntagmatic tone interval. However, the relative height relations specified on different levels are not at all independent of one another; rather, they are highly interdependent. Describing these interdependencies among grid rows is dealt with in the appendices to this thesis.

Recall that in Section 3.1 we claimed that the vertical dimension of grids affects the tonal representation in that *tones spread upward to associate with every X in a column.* The fact that tones spread upward is assumed to be a general aspect of the "geometry" of tone. Ultimately, the result of the upward association of tone is that tones are permitted to interact with one another across long distances. It turns out that the adjacency of tones on higher grid rows permits long-distance interactions between tones. To explain the concept of tone adjacency as well as how tones are permitted to interact, we turn to an example.

Consider the metrical grid shown in Figure 3.10. Tones are associated with timing positions in the grid structure on the lowest row. Moreover, in this figure an association line links the starred tones to every position upward in a grid column. We will consider how tones are organized with respect to the horizontal and vertical dimensions of the grid in turn.

With regard to the horizontal structure of the grid, we note that the association of tones with timing positions is consistent with the principles discussed earlier. In particular, the MPAR is respected, in that each (non-phrasal) unstarred tone is associated with a weak timing position that is adjacent to prominent position occupied by a starred tone. Finally, the first and last timing positions of the utterance are associated with tones, as required by the Minimal Tone Principle.

With regard to the vertical structure of the grid, we note again that an association line links the starred tones to every position upward in a grid column. By virtue of the association line drawn from each tone to X's on each upper grid row, the representation indicates that these tones "have a presence" on higher rows of the grid. Because of their presence on a higher grid row, the representation indicates that certain tones which are not sequentially organized in time on the lowest grid row may nevertheless come into *adjacent, contiguous positions* on upper grid rows. From this point onward, we will say that two tones are *adjacent* on some grid row if they are linked to X's on that row which directly precede or follow one another. Moreover, we will say that two tones are in *sequential* positions if they directly precede or directly follow one another *in time*, i.e., they are adjacent on the lowest grid row.

The importance of tone adjacency on different grid rows will become clear shortly. However, it is important to first observe the patterns of adjacency which emerge through the upward spread of tone. We first distinguish the rows in the metrical grid by assigning these rows a number in order from bottom to top. Consider the fact that on the lowest row of the grid, row 1, certain tones are adjacent while others are nonadjacent. Then T_1 is adjacent to T_2*, T_2* is adjacent to T_3, etc. However, T_2* is not adjacent to T_4*, since these tones are separated by T_3; nor is T_4* adjacent to T_6*. However, T_2*, T_4* and T_6* spread upward to all of the X's in their respective columns, causing them to be "represented" on higher grid rows. The effect of this upward spreading is that these tones come to be in adjacent positions on higher rows, despite their nonadjacency on lower rows. Given that tones T_2*, T_4* and T_6* are represented on grid row 2 by virtue of having spread upward, then T_2* is adjacent to T_4*, and T_4* is adjacent to T_6*. Finally, T_2* and T_6* are associated with higher grid columns than T_4*, so T_2* and T_6* spread upward to the third grid row. Then on grid row 3, T_2* and T_6* are in adjacent positions. Finally, note that T_6* is associated with the highest column of all. However, on that grid row it is not adjacent to any other tones.

What is the significance of the fact that tones come into adjacent positions on higher grid rows? We assume that *tones which are adjacent on a given row of the grid are required to join into a syntagmatic tone interval.* In other words, tones on higher grid rows give rise to some relation of relative height that can hold across long distances. This is captured in the Syntagmatic Adjacency Restriction, which is given below.

Syntagmatic Adjacency Restriction: Every pair of tones [T_n T_{n+1}] which is adjacent on some row of the metrical grid gives rise to a syntagmatic tone interval $I_{n,n+1}$.

The Syntagmatic Adjacency Restriction simply requires that every pair of tones which is adjacent on some row of the metrical grid must participate in a relation of relative height. As we have seen already, syntagmatic tone intervals control the relative height relations between tones. Then the result of the Syntagmatic Adjacency Restriction is that relations of relative height hold between tone pairs at every level of the grid. Moreover, this restriction implies that relative height relations can hold between tones which are not contiguous in time and which are separated by many syllables.

The claim that tones may interact over long distances is well-justified by phonetic data from several languages, including English, Japanese, and Swedish. We will discuss some of this evidence in Chapter 5. The issue of how to account for long-distance interactions between tones has been an active area of interest among phonologists for the past two decades. (See Liberman and Pierrehumbert 1984; Ladd 1990, 1993; van den Berg *et al.* 1992.) In Chapter 5 we will show how the present proposal can account for such long-distance interactions.

We just discussed the fact that that tones which come into adjacent positions must give rise to relative height relations in the form of syntagmatic tone intervals, which are generated under the Syntagmatic Adjacency Restriction. However, we have not yet specified how these syntagmatic tone intervals come to be specified for relative height relations. This is because there are important interdependencies in the relative height relations specified at different grid levels, which we must understand before going into detail about relative height specifications on different levels. We now turn to an important question which lays the groundwork for discussion of how the syntagmatic categories come to be specified. In particular, how are the tones in a metrical grid matched up with syntagmatic tone intervals and tone interval values?

The answer is that tones are matched up with their corresponding syntagmatic tone intervals via an auxiliary structure known as a *tone interval matrix.* A tone interval matrix is organized into rows and columns in a way that mirrors the structure of a metrical grid. The tone interval matrix encodes an entry for each syntagmatic tone interval corresponding to each adjacent pair of tones represented on all rows of the grid. Ultimately, this matrix encodes the sequence of relative height relations shared by pairs of tones at all levels of the grid. In this section we will be concerned with a simple representational issue: how pairs of tones at different grid rows match up with entries in the tone interval matrix. That is, we simply want to state how the system keeps track of which tones go with which tone intervals.

Figure 3.11 shows the tone interval matrix associated with the metrical grid structure in Figure 3.10. Each row of the tone interval matrix encodes a sequence of syntagmatic tone intervals (and associated values) associated with each pair of adjacent tones on the corresponding row of the metrical grid. As a result, if there are N tones represented on a given grid row, there are N-1 syntagmatic tone intervals represented on the corresponding row of the tone interval matrix.

Consider how tones on different grid rows give rise to syntagmatic tone intervals in a tone interval matrix. Tones T_1 and T_2* are adjacent on row 1 of the grid; these tones then correspond to the initial entry $I_{1,2}$ on row 1 of the tone interval matrix. Every pair of adjacent tones on this grid row is associated with a syntagmatic tone interval of the tone interval matrix in this way. Then tones T_2* and T_3 are associated with tone interval $I_{2,3}$ of the tone interval matrix, tones T_3 and T_4* are associated with tone interval $I_{3,4}$, etc.

Pairs of tones on higher rows of the grid correspond to syntagmatic tone intervals on higher rows of the tone interval matrix. There are three tones represented on row two of the grid: T_2*, T_4*, and T_6*. Then T_2* and T_4* are associated with tone interval $I_{2,4}$ on row two of the tone interval matrix; similarly, T_4* and T_6* are associated with tone interval $I_{4,6}$ on row two of the tone interval matrix. On row three of the grid, there is only one pair of tones: T_2* and T_6*. Then these tones are associated with tone interval $I_{2,6}$ on row three of the tone interval matrix. Finally, row four of the grid contains a singleton tone, T_6*, but no tone pairs; hence, there are no entries on row four of the tone interval matrix. As a result, only three rows of the tone interval matrix have entries.

We will take the convention that the tone interval matrix is written underneath a metrical grid. Then a given syntagmatic tone interval will be drawn underneath its associated referring tone. For example, in $I_{1,2}$, T_2 is the referring tone, so $I_{1,2}$ is aligned vertically with T_2. In this way the representation makes clear how many syntagmatic tone intervals a given tone participates in. It is noteworthy that T_6 participates in three tone intervals, derived from its adjacency to three different tones on rows one through participates in three tone intervals, derived from its adjacency to three different tones on rows one through three of the metrical grid.

It turns out that there are significant restrictions on the sorts of relative height relations which can coexist in the tone interval matrix simultaneously. We have worked out these restrictions in detail, and they are provided in Appendix A.1-A.2. Among other things, an understanding of these restrictions leads to a new treatment of downstep, upstep, and other tonal "register" phenomena. However, it is not necessary to delve deeper into these points here, since by now we have laid the basic foundation for an account of interactions between nonconsecutive tones. In Chapter 5 we will show how the hierarchical relations permitted in the theory permits us to phonologically account for long-distance interactions between accented tones demonstrated in experiments by Pierrehumbert (1980) and Liberman and Pierrehumbert (1984).

3.4 Summary

This chapter discussed how tone intervals are organized at the phrasal level. It was proposed that tones universally associate with metrical grid structures. This filled a gap in the literature which for the first time permits a unified treatment of interaction between tone and metrical structures cross-linguistically. Moreover, doing so laid the groundwork for an account of English intonation to be presented in Chapter 5. We will briefly review the major points of this chapter.

We initially reviewed the functions of tones in this theory. First, tones were noted to inherently lack "tonal" features, but instead to serve as constituents that entered into tone interval constructs. Second, tones associate with timing positions in the metrical grid. Significantly, by associating with grid columns, tones have the secondary effect of "pointing" to a particular segment or syllable, thereby designating it to be a "tone-bearing unit".

We then discussed several factors which affected how tones may associate with metrical grids. An intrinsic factor affecting tonal association was the type of tone. Starred tones associate with grid columns of height 2 or higher, while unstarred tones associate with grid columns of height 1.

Next, the Minimal Tone Principle requires that the first and last timing positions of an utterance must be associated with a tone. This ensures that every portion of a tonal contour arises either through direct phonetic instantiation of a phonological target or through interpolation between phonological targets.

Another principle we discussed was the Principle of Tonal Association (PTA), which restricted the possible geometries by which tones and timing slots may be associated. The PTA requires that every tone must associate with exactly one timing slot. The effect of this principle is to eliminate a range of structures which previously had been permitted in phonological theory.

Two factors were then discussed which affect the association of unstarred and starred tones. First, the Metrical Prominence Association Rule requires unstarred tones not at phrasal edges to associate with weak positions that are "next to" a starred tone. This principle will be important in accounting for the behavior of "bitonal pitch accents" in Chapter 5. Second, the Metrical Strength Rule ensures that if a grid column is associated with a starred tone, any grid column of equal or stronger metrical strength in the phrase also has a starred tone.

We noted that tones spread upward to associate with every X in a grid column, causing starred tones to be "represented" on higher grid rows. The Syntagmatic Adjacency Restriction stipulates that every pair of tones which is adjacent on some row of the metrical grid must give rise to a syntagmatic tone interval, thereby making it possible for accented syllables to participate in relative height relations with one another over long distances. This will make possible a cleaner account of data presented by Pierrehumbert (1980) and Liberman and Pierrehumbert (1984), as we will shown in Chapter 5.

We have now laid the groundwork for a new treatment of English intonation. It will be shown in Chapter 5 that this theory presents a simpler and more phonetically transparent account of English intonational phenomena than previous accounts. Moreover, it will be demonstrated that this account builds on earlier work by providing a solid theoretical basis for assumptions used in descriptive linguistics which previously had been unfounded. Finally, we will show that the account a reconciliation of phonological theory with phonetic evidence from a wide variety of languages.

In order to demonstrate the advantages of the account of English to be proposed, it is necessary first to consider several more issues in Chapter 4. First, we must discuss in more detail the relationship between phonology and phonetics as it is viewed in this theory. Second, we must evaluate the present theory alongside other theories of phonetics and phonology using a reasonable set of criteria, so that we may evaluate the theories on equal footing. We will show that with respect to intonational phenomena, the theory of phonetics and phonology presented here, which admits syntagmatic phonological primitives, is both more descriptively adequate and simpler than theories based on paradigmatic primitives. We will demonstrate that this is true even when we consider versions of such theories which have not yet been proposed, but which might be.

Chapter 4 – Evaluating theories of phonetics and phonology for intonation languages

4.1 Objectives of the present chapter

The present chapter has three primary objectives. The first is to add substance to the claim made in Chapter 1 that theories of the phonology and phonetics of intonation languages must include syntagmatic restrictions on tones. We will show that failing to include such restrictions leads to overgeneration of phonetic contours from the phonological representation, as well as underdetermination of the phonological representation from pitch contours. The second objective of the present chapter is to evaluate several such theories of phonology and phonetics, including the tone interval theory described in Chapters 2 and 3. This is accomplished by establishing a set of criteria for theoretical adequacy and then determining how each of the theories performs with respect to those criteria. A third objective is to illustrate how the tone interval theory can provide an account for a large body of outstanding phonetic data and empirical findings which cannot be accounted for under existing theories. In particular, a number of studies have now shown that F0 turning points, such as F0 peaks and valleys, are consistently aligned by speakers with the segmental string, and that differences in the type of turning point and in its temporal alignment are interpreted as meaningful by listeners (e.g., Bruce 1977; Kohler 1987; Xu 1997; Arvaniti, Ladd, and Mennen 1998; Ladd, Mennen, and Schepman 2000; Grice, Ladd, and Arvaniti 2000).

This chapter is organized as follows. First, in Section 4.2 we consider some criteria for judging a theory of phonology and phonetics of tonal systems. These criteria are subsequently used to evaluate theories of phonology and phonetics for intonation languages in particular. Next, Section 4.3 considers the question of which phonetic facts a theory of the phonology and phonetics of intonation should strive to account for. A body of work is reviewed which shows that the types and temporal alignment characteristics of F0 turning points, such as F0 peaks and valleys, are significant for the phonological representation. It is argued that any descriptively adequate theory of phonology and phonetics of tone should provide some account of this body of phonetic data. Section 4.4 then identifies a minimal set of restrictions that must be in place in any theory, in order for it to adequately account for this intonational data. Moreover, it is argued that intonation languages present special problems for a theory of phonology and phonetics which are not encountered in many lexical tone languages. Section 4.5 then evaluates the proposed tone interval theory with respect to the criteria described in Section 4.2. It is shown that the theory achieves descriptive adequacy for the F0 data discussed in Section 4.3, and that it performs well with respect to the evaluation criteria. Finally, Section 4.6 evaluates other theories against these criteria, including that of Pierrehumbert (1980). Finally, Section 4.7 provides a summary of the major points of the chapter.

4.2 Some criteria for evaluating theories of phonetics and phonology

One goal of the present chapter is to evaluate theories of the phonology and phonetics of tonal systems with respect to one another, in order to more clearly assess the relative strengths and weaknesses of such theories. In order to carry out such an evaluation process, it is necessary to determine which criteria are to be used for the evaluation. It is generally agreed that the following two criteria are important for any theory which seeks to account for a set of observations or facts.

4.2.1 Capacity to explain significant data and findings

The first goal of any theory should be to provide an explanation for data and empirical findings. A central goal of a theory of phonology and phonetics, then, should be to provide an explanation for phonetic and phonological facts which have linguistic significance. In order to do so, the theory should in general posit a well-defined association between a set of phonetic observations and a clear and unambiguous set of underlying phonological primitives. Moreover, the theory should ideally provide some insight or rationale for why the phonetic observations and associated phonological representations take the forms that they do.

The ability of a theory to associate a set of observations with a set of underlying primitives is dependent to a large extent on the *clarity of its definition of input/output relations* for the phonetics and phonology. Indeed, a central focus of linguistic research is to clearly define the relation between the underlying structure and the observable phonetic output (Goldsmith 1995). The converse of this is that phonetic data should be clearly interpretable in terms of a set of well-defined underlying primitives.[41]

What constitutes a clear definition of the input/output relations for the phonetics and phonology? In our view, a clear definition has the following complementary form. On the one hand, a set of phonological primitives should give rise to a clear-cut and readily predictable set of phonetic attributes. On the other hand, observed phonetic data should be readily interpretable in terms of a small set of well-defined phonological primitives. Since this should be true of any theory of phonology and phonetics, it should also be true of a theory which focuses on the phonology and phonetics of tonal phenomena.

An implication of these insights is that in general, a theory which does not clearly define how a set of phonological primitives relates to a set of phonetic observations is liable to be descriptively insufficient. Moreover, a model which does not provide rationale for why the phonetic observations and phonological representations take the forms that they do is less desirable than a model which does provide such insight. Thus, in evaluating different theories, we will be interested in the extent to which the data are explained by each theory, as well as the degree to which each theory provides insight into the phonetic and phonological forms. The ability of a theory to provide a phonological explanation for phonetic data is closely related to how clearly the theory defines the relationship between the underlying phonology and the observed surface phonetics. Thus, in evaluating different theories of phonology and phonetics in this chapter, it will be especially important to ascertain the clarity of the definition of the input-output relation put forward by each theory.

4.2.2 Simplicity

Another important criterion by which a theory of phonetics and phonology should be judged is *simplicity*. A collective ideal which guides both theory-making and research efforts in science is expressed in a philosophical principle known as *Occam's Razor*, which espouses simplicity in scientific explanation. This principle suggests that, all other things being equal, the theory which posits the simplest explanation is more likely to be correct. This principle would lead us to favor a theory which affords the fewest parameters and postulates as the most likely explanation for a set of phenomena.

There is a second sense of the term "simplicity" which can be used to evaluate the "goodness" of a theory, particularly as it regards a theory of phonetics and phonology. Suppose we have two theories which are equal in all respects, except for the following difference. Suppose that Theory A posits that certain parameters are necessary for understanding a spoken message, but a listener can never obtain direct phonetic evidence for them. In contrast, suppose that Theory B posits an explanation in which all of the necessary parameters are predicted to be recoverable from the speech signal. All else being equal, Theory B is the more desirable of the two theories, since it provides the simpler explanation. By this logic, we should prefer Theory B to Theory A.

The issue of phonetic realization is especially important when we consider the need for eventually developing accounts of real-world processes, such as the linguistic acquisition of tonal systems. Clearly, the phonetic information necessary for tonal acquisition *is* present in the speech signal, since children *do* successfully acquire tonal systems. Thus, real-world observations also tend to support theories which assume that the necessary phonetic information is directly recoverable from the speech signal, as opposed to theories which do not assume this. Thus, a simpler theory wins not only on philosophical, but also practical grounds.

We have thus defined two desiderata for any theory of phonology and phonetics. First, a theory should provide an explanation for phonetic and phonological facts which have linguistic significance. This is necessarily accomplished by defining a clear input/output relation, so that the phonological input gives rise to a well-defined set of phonetic outputs and vice versa. Second, a theory should be simple, positing as few parameters and postulates as possible. Moreover, a theory of phonology and phonetics should not assume that parameters necessary to recovering the phonological representation cannot be heard, especially in the face of evidence to the contrary. Having defined some criteria by which a theory of phonology and phonetics might be evaluated, we turn to the question of which phonetic facts such a theory should seek to explain.

4.3 Some critical phonetic data: The presence and alignment of F0 turning points

Given that theories of phonology and phonetics should try to explain linguistically significant phonetic facts, one strategy for evaluating such theories is to identify a set of phonetic data with crucial linguistic significance and evaluate how well different theories are able to account for this data. This leads us to an important question: what phonetic data should a phonological theory of tone and intonation strive to account for? Evidence from approximately three decades of phonetic studies has now demonstrated the significance of *F0 turning points* for the phonological representation of tonal systems across languages. F0 turning points are defined here in a preliminary way as F0 maxima, minima and corners, that is, positions where a level contour turns into a fall or a rise, or vice versa; a more specific definition will be given below.

This section has two objectives. The first is to review evidence of the significance of F0 turning points for the linguistic representation. We will argue that because of the linguistic significance of this data, any theory of the phonology and phonetics should provide some account for it. The second goal is to raise the issue of what sorts of phonological primitives might best account for F0 turning point data. In this regard, we will describe how phonetic studies of F0 turning points have helped to resolve a longstanding theoretical debate in theories of tone and intonation concerning whether the phonological primitives are dynamic rises and falls, or tones. Having reviewed this evidence, we will then move on in Section 4.4 to discussing why it is particularly difficult to account for F0 turning point data in intonation languages and why such languages present special problems for theories of phonetics and phonology. Finally, in Sections 4.5-4.6 we will consider how well various theories meet the challenges of accounting for this data.

Before embarking on a review of phonetic data concerning F0 turning points, we need to define more clearly what is meant by this term. In this chapter we distinguish two complementary usages of the term *F0 turning points*. In the context of reporting phonetic studies of intonation contours, the term refers to the fine details of timing of observable F0 maxima, minima, etc. Indeed, anyone who has examined F0 contours in detail knows that there are many small obtrusions in such curves due to interactions with segments, the chaotic nature of vocal fold mechanics, etc. Variability in the shapes of pitch contours in regions of interest can present methodological challenges in phonetic studies of F0 (Knight 2004). Nevertheless, there is also a general trend which is apparent in the overall direction of curvature. The gross shape of this F0 curvature tends to be rather invariant across repetitions of a given intonation pattern by a speaker. This suggests that some aspect of the gross contour shape is determined by the phonological representation, whatever the form of that representation. A standard assumption of phonetic studies on F0 is that measurements of fine-grained F0 timing characteristics will yield insight into phonologically significant aspects of F0 curves, so long as a sufficiently large number of measurements are made. Part of this chapter deals with the issue of predicting the types and alignment characteristics F0 turning points from phonological theory. In the context of discussions of the relation between phonetics and phonology in Sections 4.4-4.6, we will therefore be concerned with how well a given theory is expected to predict *the significant F0 turning points*, by which we mean the gross attributes of F0 contour shape, when obtrusions due to segmental influences and chaotic vocal fold fluctuations are ignored. These "significant" turning points are expected to match well fine-grained measurements of F0 maxima, minima, and corners averaged across many measurements. In the context of discussion on prediction of F0 curve attributes from phonological theory, the term F0 turning points refers to the peaks, valleys, and corners of this time-averaged curve.

There are several lines of evidence which have demonstrated the importance of F0 turning points for the phonological representation of tonal patterns. First, experiments using synthetic speech have demonstrated that manipulating the timing of an F0 turning point affects the meanings that listeners hear for utterances (Purcell 1976; Bruce 1977; Kohler 1987; Nagano-Madsen and Eriksson, 1989; Gårding and Eriksson 1989; House 1990; D'Imperio and House 1997; D'Imperio 2000). For example, in lexical pitch accent languages like Swedish and Japanese, shifting the timing of an F0 peak through segmental material causes listeners to hear different words (Bruce 1977, Gårding and Eriksson 1989, Nagano-Madsen and Eriksson 1989, House 1990). In intonation languages like German and Italian, shifting the timing of an F0 maximum or minimum causes listeners to infer different meanings (Kohler 1987, D'Imperio and House 1997, D'Imperio 2000). For example, a listener might hear an utterance alternately as given or as new information in the discourse (Kohler 1987).

Another line of evidence that F0 turning points are significant for the linguistic representation concerns the fact that listeners interpret continuous variation in F0 turning point alignment in terms of discrete categories (Kohler 1987; Pierrehumbert and Steele, 1989; House 1990; d'Imperio and House 1997; d'Imperio 2000). By *alignment* we mean the temporal position of an F0 turning point with respect to the segmental string. This suggests that the timing of F0 turning points is interpreted by listeners in a categorical way, which is consistent with phonological distinctions.

A third line of evidence comes from production studies, which have shown consistent timing of F0 turning points with respect to segments or syllables across a wide variety of languages. These languages include Swedish (Bruce 1977), Mandarin (Xu 1998, 1999, 2001; Li 2003), German (Atterer and Ladd 2004), Dutch (Ladd, Mennen, and Schepman 2000), English (Ashby 1978; Silverman and Pierrehumbert 1990; Ladd, Faulkner, Faulkner and Schepman 1999; Ladd and Schepman 2003; Dilley, Ladd, and Schepman, forthcoming), Japanese (Nagano-Madsen and Eriksson 1989; Ishihara 2003), Greek (Arvaniti, Ladd, and Mennen 1998), Spanish (Prieto, van Santen, and Hirschberg 1990), Serbo-Croatian (Smiljanić 2004) and a number of others.[42] In these studies, speakers produce F0 turning points so that they are aligned consistently with segmental boundaries; variation in their timing by a speaker is typically on the order of 10-30 ms. In general, phonetic consistency is usually indicative of phonological control. Thus, this body of data provides further evidence for the significance of F0 turning points with respect to the linguistic representation.

We can conclude from these results that a theory of the phonology and phonetics of tonal systems should strive to provide an account for linguistically significant data concerning F0 turning points. That is, we would like such a theory to predict the general forms of phonetic contours, given some phonological representation. Moreover, we would like to be able to interpret F0 turning point data in terms of a well-defined set of phonological primitives. This raises the question of exactly *which* aspects of F0 data should be described under such a theory.

Consideration of the evidence suggests that two aspects of F0 data in particular should be defined through clear input-output relations in a theory of phonology and phonetics. The first aspect of F0 data to be accounted for concerns which *type* of F0 turning point will arise from the phonological representation – that is, whether it is a peak, a valley, or a corner. That distinct F0 turning point types arise from distinct phonological representations is a central tenet of phonological theories of tonal systems (Liberman 1975, Pierrehumbert 1980, Ladd 1996, Gussenhoven 2004). Moreover, the notion that different turning point types have distinct phonological representations is supported by empirical work showing that F0 maxima and F0 minima behave in categorically distinctive ways (Gussenhoven and Rietveld 2000).

The second aspect of F0 data which a theory must account for is the *temporal alignment* of F0 turning points with respect to segments or syllables. To what degree of accuracy should a theory seek to account for the temporal alignment of such points? In our view, a theory which is concerned with describing those aspects of the phonetics which are linguistically significant need not account for small degrees of variation which do not give rise to categorical distinctions. Standard deviations in the timing of F0 maxima of about 10-30 ms are typical in production experiments (e.g., Arvaniti, Ladd, and Mennen 1998). Moreover, variation in F0 turning point timing across certain segmental positions does not elicit categorical effects (Kohler 1987, Pierrehumbert and Steele 1989, House 1990, d'Imperio and House 1997). Such within-category variation does not presumably contribute significant information to the linguistic message, and thus need not be modeled in a theory which seeks to describe only the linguistically significant aspects of the data.

Assuming that a theory should seek to account for F0 turning point data, what sorts of phonological primitives provide the best account? It turns out that phonetic work on F0 turning points has played a significant role over the last decade in empirically evaluating two types of theories, which will be termed here *tonal target theories* and *dynamic rise-fall theories*.[43] In the following, we will describe these two types of theories and review the relevant evidence, which clearly supports tonal target as phonological primitives.

The two types of theories can be described as follows. According to tonal target theories, the phonological primitives are tones or targets which are associated with a given position in the segmental string, such as a segment or syllables (Pike 1945, Liberman 1975, Pierrehumbert 1980). Targets are then connected up under phonetic interpolation processes to yield an F0 contour at output. Crucially, these targets do not exceed the duration of a syllable. In contrast, dynamic rise-fall theories, the phonological primitives are rising or falling pitch movements which associate with particular prominent syllables (e.g., t'Hart, Collier and Cohen 1990; Palmer 1922; Halliday 1967). These primitive rising or falling movements may extend beyond a syllable in duration, so that an F0 contour results from concatenations of such primitive elements.

The primacy of tonal target theories over dynamic rise-fall theories has now been demonstrated through significant body of work showing that F0 turning points are consistently aligned with respect to the segmental string (e.g., Xu 1998, 2001; Ladd *et al.* 1999, 2000; Arvaniti *et al.* 1998, Ladd and Schepman 2003, Dilley, Ladd, and Schepman, forthcoming). Moreover, F0 turning points have been shown to be consistently aligned with the segmental string in studies of both English and Mandarin, even when speaking rate is varied (Xu 1998, Ladd *et al.*, 1999). These results are readily interpretable with respect to tonal target theories, in which tonal primitives align with segmental positions; such theories predict that tones, like F0 turning points, will remain consistently aligned with segments or syllables, even under phonetic pressures such as changes in speaking rate. In contrast, results of this sort cannot be readily be accounted for in terms of dynamic rise-fall theories, in which the primitives are rises and falls. This is because in such theories, F0 turning points are treated merely as positions where concatenation processes cause a rise to meet a fall, or vice versa; such an explanation does not predict F0 turning points to necessarily be aligned consistently with segments or syllables. Thus, we can conclude that F0 turning point data is best modeled in terms of theories which assume that the primitives are discrete tonal targets whose duration does not extend beyond a syllable.

In summary, we have identified a body of data which must be accounted for in any theory of the phonology and phonetics of tonal systems. In order to account for this data, it is necessary for the phonology to predict both the *type* of turning point, as well as the *timing* of this turning point within a phonologically significant range of values. Conversely, distinct types and timing characteristics for F0 turning points should be interpretable in terms of a clear-cut set of phonologically distinctive categories. We also showed that such data is best accounted for by a theory which assumes that the primitives are discrete tonal targets. In the following section we discuss some challenges for any theory in accounting for such data.

4.4 Some challenges in accounting for F0 data in intonation languages

In the last section we described a critical set of phonetic data which any theory of the phonology and phonetics of tone and intonation needs to account for. This section focuses on challenges and obstacles to accounting for such data specifically in *intonation languages*. It will be argued that intonation languages are unlike other languages, in that they present special challenges for a theory of phonology and phonetics. In particular, intonation languages permit a given tonal contour to be scaled in very different positions in the speaker's pitch range, unlike lexical tone languages such as Yoruba, which tend to restrict tonal patterns to particular regions of a speaker's pitch range. We will show that the key to meeting the challenges of modeling F0 data is to include syntagmatic restrictions on the relative heights of tones in sequence in the theoretical definitions of input-output relations. Indeed, it will be shown that a theory which does not include sufficient syntagmatic restrictions on the relative heights of tones in sequence in its input-output relations is unable to account for F0 data. Thus, an understanding of the special problems posed in modeling data from intonation languages and how to overcome them will provide support for a major contention of this thesis, namely, that syntagmatic restrictions are needed at the level of the phonology in order to adequately account for cross-linguistic tonal data. A major goal for this section, then, is to come to terms with some of the challenges which intonation languages present for theories of phonology and phonetics, and the resulting need for syntagmatic restrictions. A better understanding of these challenges and the necessity of syntagmatic relations will then allow us to better evaluate theories of phonology and phonetics.

Why might accounting for tonal patterns in intonation languages be challenging for a theory of phonology and phonetics? Consider first the fact that intonation patterns have typically been described in terms of discrete *tones* in theoretical work in phonology over the last 30 years or so. We saw in the last section that such a descriptive approach is consistent with phonetic evidence, which supported tonal target theories over dynamic rise-fall theories. F0 turning points correspond to significant points in intonational contours, and they have typically been described in terms of specific types of tones. For example, an F0 maximum has typically been assumed to correspond to an underlying H tone, while an F0 minimum has typically been described in terms of an underlying L tone.

The challenging part of accounting for F0 turning points in terms of primitive tones like H and L stems from the fact that across utterances, the levels of tones vary not only with respect to the speaker's pitch range, but also with respect to one another. For example, consider a rising-falling contour, which can be modeled by a LHL tonal sequence with the H tone corresponding to the F0 peak. If this contour is produced in the upper part of the speaker's pitch range, the H corresponds to a relative high pitch. In contrast, if the contour is produced in the lower part of the speaker's pitch range, the H corresponds to a relatively low pitch. It might even be the case that a pitch which corresponded to a H tone in one context was lower than a pitch which corresponded to a L tone in another context. The important observation is that no matter where in the pitch range the contour is produced, the pitch levels of the L and H tones must be coordinated with respect to one another so that the entire LHL sequence always generates a rising-falling contour.

This suggests that a theory of the phonology and phonetics of intonation languages must carefully balance two kinds of factors in defining input-output relations, in order to adequately describe intonational data. On the one hand, the input-output relations must allow for the fact that "vertical scaling factors" permit both H and L tones to take on a wide range of values within a speaker's pitch range. On the other hand, the input-output relations must adhere to "horizontal scaling factors" so as to restrict the levels of the L and H tones with respect to one another so that in a sequence of tones like LHL, the H always corresponds to an F0 peak.

This raises some important issues. Given that two kinds of factors affect the scaling of H and L tones, how do these factors interact with one another? Careful consideration suggests that these two factors must be coordinated in some way. For example, in order for a H tone in a LHL sequence to correspond to an F0 peak, neither L tone can go above the H tone. If L tones were permitted to go higher than adjacent H, then a H tone in a LHL sequence could either give rise either to an F0 peak, or an F0 valley. This would result in serious problems; for example, such a theory would be indeterminate with respect to whether an F0 valley should be described as L or as H. The input-output relations would be confusing, and ultimately, such a theory of phonetics and phonology would lose its explanatory power.

Obviously, the level of indeterminacy in such a theory would lead us to reject it as insufficient both on phenomenological and logical grounds. This leads us to two critical questions. First, what are the restrictions on input-output relations in a theory of phonology and phonetics that will permit us to say with certainty that a given phonological string predicts a particular sequence of F0 turning points? Conversely, what are the restrictions that will permit us to say that a given sequence of F0 turning points necessarily arose from a given phonological string?

In order to understand the crucial restrictions on input-output relations, consider what constraints must be in place in order for us to describe a simple rising-falling-rising F0 contour in terms of a sequence $L_1H_2L_3H_4$. Describing the contour in this way implies that the local F0 maximum corresponds to H_2 and the following local minimum corresponds to L_3. Then what set of restrictions must be in place on tonal scaling factors, such that H_2 will necessarily correspond to an F0 maximum and L_3 will necessarily correspond to an F0 minimum?

The answer is that in order for H_2 to have a locally high pitch, L_1 and L_3 must each have locally lower pitches. Similarly, L_3 must be lower in pitch than both H_2 and H_4. Considered reflection reveals that *in order for a sequence of tones to give rise to a particular sequence of F0 turning points, the input-output relations must constrain the relative heights of every pair of adjacent tones in the sequence.* If the relative height of even one pair of adjacent tones is not constrained, then it will not be possible to fully predict what sorts of F0 turning points will be observed at the output, nor how such a sequence should be described in terms of input tones. If a theory somehow permits the relative heights of *two* consecutive tone pairs to vary freely, then we will not be able to predict *anything* about the shape of the F0 curve that would arise from those tones. Such a theory would logically fail to account for data concerning F0 turning points by virtue of the utter lack of clarity of its input-output definitions.

These observations suggest a productive way of evaluating the input-output relations for theories of phonetics and phonology for intonation systems. In particular, the crucial issue for the success of such a theory is whether it defines restrictions on the relative height relations of every pair of tones in the sequence *at some level of the theory.* If so, then such a theory can plausibly account for the F0 turning point data discussed in Section 4.3. If not, then such a theory logically fails to be able to account for the F0 turning point data discussed earlier.

There is a further conclusion that we can draw from these considerations which is perhaps more significant. We have shown that syntagmatic restrictions on the relative levels of tones are required at some level of a theory of phonology and phonetics for intonation languages, in order for the theory to define clear input-output relations. The fact that syntagmatic relations are needed in order to account for crucial aspects of the speech signal is an argument that such relations are part of the representation of tone and intonation, which is a major contention of this work.

Now that we have a better understanding of the necessary requirements that a theory of phonology and phonetics must meet in order to adequately describe intonation data, we can turn to the matter of evaluation. We will first evaluate the proposed tone interval theory with respect to the considerations described in this section. We will subsequently consider how several other theories fare with respect to these considerations, including the theories of Pierrehumbert (1980) and Pierrehumbert and Beckman (1988).

4.5 An evaluation of tone interval theory

Sections 4.2-4.4 marshaled several arguments relevant to evaluating theories of the phonology and phonetics of intonation languages. First, we identified some criteria by which the explanatory capacity and simplicity of theories of phonology and phonetics can be gauged. Second, we identified a body of phonetic data which a theory of phonology and phonetics must account for concerning the alignment and types of F0 turning points. Third, we considered some special problems associated with accounting for the scaling of tones in intonation languages.

Having considered these issues, we are now well-equipped to evaluate theories of phonetics and phonology for intonation languages, starting with the tone interval theory proposed in Chapters 2 and 3. First, we will review the general conception of the phonetics-phonology relation under this theory. Next, we will consider specifically how this theory accounts for the presence of particular types of F0 turning points, as well as the alignment of those points. Finally, we will consider how the theory fares on the whole with respect to the evaluation criteria defined at the beginning of the chapter.

We will first review how tonal information is phonologically represented within tone interval theory, and then we will discuss how this phonological representation gives rise to phonetic properties. As described in Chapters 2 and 3, we assume that the phonological representation is based on tone intervals, which are quasi-mathematical entities which define a relationship between a tone and a referent. Unlike other theories which assume that tones have intrinsic tonal properties like [±high] and [±low], tone interval theory assumes that tones are inherently "toneless". Rather, tones gain "tonal" specification through their participations in tone intervals. Each tone interval specifies one of three primitive relative height relations between a tone and a referent: *higher*, *lower*, or *same*. Rather than intrinsically bearing tonal properties, the primary function of tones in this theory is to associate with particular positions in the metrical structure of utterances. Each position in the metrical structure which is thus "marked" with a tone in turn is associated with some tone-bearing unit, such as a segment and/or a syllable. In this way, tones "point to" particular segments or syllables which ultimately realize the relative height relations specified in tone intervals. The representation is typically sparse, particularly in intonation languages, so that some metrical positions lack tones. A sequence of tones which are thus associated with positions in the metrical grid then join into syntagmatic tone intervals; pairs of adjacent tones on the lowest grid level which form syntagmatic tone intervals then ultimately give rise to F0 turning points. Thus, the phonological representation consistently specifies the underlying relative heights of tones, together with their patterns of association with segments and/or syllables.

At the phonetic implementation stage, each tone in the underlying phonological representation is realized by a target *pitch*. Each target pitch coincides temporally with the segment or syllable which is "marked" by its corresponding underlying tone. Each pitch is chosen so as to realize the relative height relations specified by the underlying sequence of syntagmatic tone intervals. For example, a tone which is *higher* than some referent tone will be produced with a higher pitch than that of the referent tone. Each successive pair of pitches in the sequence is then connected up via monotonic pitch interpolation functions. Because the vocal folds are enlisted to produce an F0 contour realizing the appropriate pitch pattern acoustically, the F0 trajectory will approximately follow the trajectory of target relative pitches. We assume that the listener, for his or her part, "hears out" precisely those pitches which correspond to underlying tones; the relative heights of these pitches are then compared, and the phonological representation is transparently extracted.

How does tone interval theory explain the presence of F0 turning points at a particular location in the segmental string? We assume that the goal of the speaker is to produce an F0 contour which realizes the appropriate relative pitch characteristics specified by the underlying relative height relations in the tone intervals. Here, we assume that at the phonetic implementation stage each tone in the underlying phonological representation is realized by a *target pitch*. Each target pitch coincides temporally with the segment or syllable which is "marked" by its corresponding underlying tone. Each pitch is chosen so as to realize the relative height relations specified by the underlying sequence of syntagmatic tone intervals. For example, a tone which is *higher* than some referent tone will be produced with a higher pitch than that of the referent tone. Each successive pair of pitches in the sequence is then connected up via monotonic pitch interpolation functions. Because there is a high degree of correlation between F0 and perceived pitch (Moore 1997), the trajectory of the F0 curve will approximately follow the trajectory of the sequence of target relative pitches. The phonological representation predicts the occurrence of a particular F0 turning point within a range of timing variation which instantiates a perceptually significant relative pitch level.[44]

We can now understand how an F0 turning point arises at a particular location in the segmental string under this theory. First, consider how the theory explains the presence of an F0 peak near a particular position in the segmental string, s_2. Suppose that a sequence of tones T_1 T_2 T_3 is joined into the tone intervals $I_{1,2} > 1$ and $I_{2,3} < 1$. Then T_2 will be higher than T_1, while T_3 will be lower than T_2. An equivalent is that T_2 will be higher than either T_1 or T_3 in the phonological representation. Now, suppose that T_1, T_2, and T_3 are associated with metrical positions corresponding to the tone-bearing units s_1, s_2 and s_3, respectively. Then at the phonetic implementation stage, s_1, s_2, and s_3 will each be produced with discrete, target pitches which realize the underlying relative height relations specified by the tone intervals. The sequence of discrete pitches will then be connected up by monotonic pitch interpolation functions. Given that T_2 was underlyingly higher relative to either T_1 or T_3, this means that s_2 will be produced with a locally higher pitch than s_1 or s_3. The F0 curve will be produced by the speaker in a way that generates a pitch for s_2 which is higher than that of any other tone-bearing unit in the vicinity. The net result is that an F0 peak will be produced in the temporal vicinity of s_2. More specifically, the theory predicts that an F0 peak will occur precisely within a range of times that generates a locally high pitch on s_2. In other words, the timing of F0 peaks and valleys is only predicted under this theory to within a range that is perceptually significant for relative pitch. A number of studies have shown that variations in F0 peak times on the order of 40-60 ms may not be perceptible (Bruce 1977, Kohler 1987, House 1990).[45,46]

Differences in the *type* of F0 turning point are explained in terms of distinct underlying relative height relations. Consider once again a sequence of tones T_1 T_2 T_3 which are associated with s_1, s_2, and s_3, respectively. Then if the associated tone intervals are $I_{1,2} > 1$ and $I_{2,3} < 1$, T_2 will be higher in the phonology than either T_1 or T_2. In this case, we predict that an F0 *peak* will occur within a range of times that generate a locally high pitch on s_2. On the other hand, suppose the tone intervals are $I_{1,2} < 1$ and $I_{2,3} > 1$, so that T_2 is lower than either T_1 or T_3 in the phonology. In this case, we predict an F0 *valley* within a range of times that generate a locally low pitch on s_2. Finally, suppose that the tone intervals are $I_{1,2} = 1$ and $I_{2,3} > 1$, so that T_2 is at the same level as T_1, while T_3 is higher than T_2. This predicts an F0 *corner* to occur within a range of times that generate a pitch on s_2 which is the same as that on s_1, as well as a monotonic rise from s_2 to s_3.

Differences in the *alignment* of F0 turning point are explained in terms of distinct patterns of association of a tone with positions in the metrical structure. For example, suppose that shifting the timing of an F0 peak across a two-syllable sequence, $s_1 s_2$, gives rise to two distinct categories for listeners. This can be explained in terms of a tone being aligned with s_1 over some range of F0 values and with s_2 over the remaining range of F0 values, together with the assumption that the tone is locally higher than its neighbors in the phonology.

Now that we have described the general relationship between phonology and phonetics under tone interval theory, we can evaluate its input-output relations with respect to a simple rising-falling-rising contour.[47] How clearly does the theory relate this contour with its peaks and valleys to a unique underlying sequence of primitives? First, we will assume that the pattern of ups and downs of this F0 contour produce an alternating sequence of low and high pitches. These pitches can in turn be traced back to a sequence of four tones, $T_1\ T_2\ T_3\ T_4$. The tones are paired into tone intervals which are specified for a particular set of relative height relations. In order to account for the alternating sequence of low and high pitches, we can infer that T_2 is *higher* than T_1, that T_3 is *lower* than T_2, and finally, that T_4 is *higher* than T_3. We can define the underlying phonological requirements formally, as in (4.1). Here, we have indicated tones directly as ratios.

$$\begin{bmatrix} \frac{T_2}{T_1} > 1 & \frac{T_3}{T_2} < 1 & \frac{T_4}{T_3} > 1 \end{bmatrix} \tag{4.1}$$

This phonological representation critically controls the relative heights of every pair of adjacent tones in the sequence. Earlier, we showed that specifying the relative heights of every pair of adjacent tones in the sequence was both a necessary and sufficient condition for defining clear input-output relations that were capable of characterizing F0 data. As a result, the phonetic output of this specification will be a sequence of relative pitch levels which satisfy the underlying relations of relative height.

If we assume that T_1, T_2, T_3, and T_4 are associated with positions in the metrical structure which correspond to tone bearing units s_1, s_2, s_3, and s_4, respectively, then s_2 will have a higher pitch than s_1, s_3 will have a lower pitch than s_2, and finally, s_4 will have a higher pitch than s_3. Monotonic pitch interpolation functions will connect up these target pitches, and the F0 contour will follow suit. Because of the high degree of correlation between pitch and F0, we can expect that there will be an F0 maximum in the vicinity of s_2 in a range that grants s_2 the locally highest pitch. Moreover, we can expect an F0 minimum in the vicinity of s_3 in a range that grants s_3 the locally lowest pitch.

The tone interval theory therefore clearly predicts that T_2 will give rise to an F0 maximum, while T_3 will give rise to a local F0 minimum. These predictions were arrived at precisely by critical specification of relative height relations in the phonology for every pair of adjacent tones. We can conclude that the proposed theory defines a straightforward relation between the input phonology and the output phonetics, and that the theory is capable of accounting for the F0 data described earlier in Section 4.3. Moreover, the theory provides insight into why the representation is the way it is. Evidence from work in auditory perception and music suggests that humans and other animals tend to hear and encode tonal information in terms of the relative heights of the tones – whether a tone is higher than, lower than, or the same level as another tone. Our claim is that this innate tendency has found its way into the phonological system for tone, so that the phonology encodes the relative heights of tones, one to the next. The tone interval theory thus not only meets a critical test for accounting for F0 data, but it also provides insight into why the phonological system should take the form that it is claimed to. We can therefore conclude that the theory fares well with respect to the first of the two criteria for evaluating theories of phonetics and phonology.

How does the tone interval theory fare with respect to the second of our criteria, simplicity? Recall that in order for a theory to be simple, it must not posit any more parameters or mechanisms than necessary to account for the data. Moreover, there ideally should not be any parameters which can't be recovered directly from the speech signal. The proposed system does not assume more parameters than are necessary; the system proposes only three relative height restrictions, which are precisely the number needed to account for critical F0 data. In addition, no parameters are posited which are predicted to not show up in the signal: all the underlying relative levels will have direct phonetic correlates in terms of relative pitch levels and associated F0 values. We can therefore conclude that tone interval theory fares well with respect to the second criterion of simplicity.

To summarize, the proposed theory fares well with regard to both evaluation criteria identified in this chapter for theories of phonology and phonetics. First, the theory defines a clear input/output relationship between the phonology and the phonetics. In so doing, it achieves descriptive adequacy in accounting for the linguistically significant F0 data reviewed in Section 4.3. Second, the theory is simple; it proposes no more parameters than are necessary, and it defines no parameters which are not expected to be heard. Having thus considered the proposed theory, we turn to an evaluation of some other theories of phonology and phonetics.

4.6 An evaluation of some paradigmatic theories of phonology and phonetics

In Section 4.4, we identified some challenges for any theory of phonology and phonetics in accounting for pitch contours in intonation languages. The fact that an F0 contour with a given shape can be scaled in very different parts of a speaker's pitch range means that certain restrictions must be in place on the input-output relation, in order for the theory to account for F0 data without introducing overgeneration and indeterminacy into the description. In particular, a theory based on tones as the phonological primitives must restrict the relative heights of every pair of adjacent tones in a sequence with respect to one another. For example, a theory which posited H and L tones as primitives should not allow L to go higher than adjacent H, or H to go lower than adjacent L. If a theory permitted such variations, it could be logically possible to describe an F0 peak either as arising either from a H tone or from a L tone. This indeterminacy would lead to failure on the part of the theory to account for the crucial F0 data described in Section 4.3.

In the following, we will evaluate some theories of the relation between phonetics and phonology which have been proposed for intonation languages and compare their performance with the tone interval theory. The first theory to be considered will be that of Pierrehumbert (1980). Next, we will evaluate the theory of Pierrehumbert and Beckman (1988). Finally, we will consider some other possible variations on these theories that have not yet been proposed, but which might be.

4.5.1 Pierrehumbert (1980)

We turn first to the theory of Pierrehumbert (1980), which is hereafter referred to as P80. We note that the framework has been used to describe not just English, but a wide variety of languages. How well does this theory fare with respect to the criteria defined earlier as desirable in a theory of phonology and phonetics?

Some background on the theory will be helpful before considering these criteria. The theory described English intonation in terms of variants of two tonal primitives, H and L. English was thus claimed at one level to involve a simpler set of distinctions than previous theories. By contrast, four contrastive tonal levels had been posited by both Pike (1945) and Liberman (1975) for English. In P80, H and L elements could occur as one of two basic types of entities: prominence-lending tonal elements known as pitch accents, or phrase-related tonal elements, known as phrase accents and boundary tones. We will briefly consider some aspects of the inventory of tonal elements.

P80 proposed seven pitch accent types for English: two single-toned pitch accents, H* and L*, and five bitonal accents, H*+L, H+L*, L+H*, L*+H, and H*+H.[48] For the H*+L accent, it was assumed that the "+L" portion is never directly audible. The asterisk or "star" notation indicated that a given tone was aligned with a stressed syllable, while the unstarred tone of a bitonal pitch accent was then assumed to lead or trail a starred tone by a fixed temporal interval. There were also two phrase accents, H- and L-, and two boundary tones, H% and L%. These elements were assumed to combine together in sequence to generate the set of possible shapes for English intonational patterns.

How did this theory conceive of the input/output relation between phonology and phonetics? Part of the answer is that the tonal primitives were understood to be paradigmatic in nature, so that the choice of H or L dictated something about the position of the tone in the speaker's pitch range. In this regard, Pierrehumbert (1980:68) notes that "...there is a paradigmatic distinction in level: ...L* is lower than H* in the same context." However, the consequences for the phonetics of choosing H vs. L are not immediately obvious. In order to understand the phonetics-phonology relation in P80, it is necessary to examine the phonetic scaling rules in that theory, which is undertaken in this section.

Examination of the phonetics-phonology relation in this theory is made most productive by recalling that crucial restrictions must be in place, in order for a theory to account for F0 turning point data. A given tonal pattern in a language like English can be produced in a very high or very low part of a speaker's pitch range; because of this, restrictions on the levels of each pair of tones in a sequence must be in place, in order for a theory to predict F0 turning point data. Given that F0 peaks and valleys were assumed in descriptions of F0 contours in Pierrehumbert's theory to correspond to H and L tones, respectively, it is reasonable to ask whether the theory predicts that L_1 H_2 L_3 H_4 will necessarily give rise to an F0 contour with a peak for H_2 and a valley for L_3.

What phonetic factors governed the scaling of H's and L's in the theory of P80? The crucial input/output relations are defined in a set of phonetic scaling rules described on pp. 144-159. The definition of input/output relations is somewhat complex, so we will guide the reader through them. The question of significance for the evaluation of the theory is whether there are restrictions on the relative heights of H's and L's in sequence, so as to prevent L from rising above adjacent H, for example.

What restrictions on the relative levels of tones are in place in the theory of P80? Because that theory modeled F0, rather than pitch, the rules which scale H's and L's give rise to particular F0 values for each of the tones. In general, F0 values were assumed to be proportional to the ratios of *prominence values* for individual tones. In spite of the choice of terminology, there is little resemblance between this parameter and phonetic or perceptual prominence (Ladd 1990).

In general, the account of the phonology-phonetics relation in P80 assumed that tones were assigned F0 values through approximately 10 phonetic rules. These rules related the prominence values, and hence the F0 values, of tones to one another in the sequence. Depending on the composition of tones in a sequence, the ratios of prominence values could be modified by three additional parameters, *n, k*, and *p*, whose phonetic interpretation is not clear.

The crucial question for determining the adequacy of the input-output definition in this theory is whether there are appropriate restrictions on the relative heights of every pair of tones in the sequence. To understand whether this is the case, consider how the theory assumed the tones to be scaled in a simple and representative sequence: H_1 L H_2, as in H* L+H*. If the input/output relations are defined in an adequate manner, the L tone in this sequence should consistently give rise to a F0 minimum. This will be true only if the relative height of L is necessarily restricted to be lower than both H_1 and H_2.

It turns out that for this sequence, the relative height of the L is not restricted either with respect to H_1 or H_2. As a result, the L can be higher than either H tone.[49] This is clearly problematic, because the theory technically permits H* L+H* to generate either a falling-rising pitch contour with an F0 valley, which is the expected result, or a rising-falling contour with an F0 peak, which is an unexpected result. Thus, the theory overgenerates possible contours that may arise from a sequence like H* L+H*. In addition, the fact that a rising-falling contour with an F0 peak can technically be described in the theory e.g., by either L*+H L* or by H* L+H* suggests that the framework permits a troubling degree of indeterminacy with respect to the phonological description.

Is this an isolated problematic case, or are the difficulties with the theory more pervasive? Recall that in order for a theory to be able to predict the type and timing characteristics of F0 turning points, it was necessary for the relative heights of every pair of adjacent tones in a sequence to be restricted at some level of the input-output relation. Inspection of the set of phonetic rules provided on pp. 144-159 shows that the relative levels of some pairs of tones, but not others, are restricted with respect to one another in the phonetic rules. In other words, there are a number of tone pairs whose relative heights are permitted to vary freely with respect to one another. For such pairs, the relative heights of the tones is not predictable from the input-output relation, so we cannot say what sort of F0 contour shape could arise as a result. We can conclude that the problems cited for the case above are not restricted to just one pitch accentual sequence. Rather, P80 attests pervasive problems in terms of overgeneration of phonetic contours and indeterminacy of phonological forms.

How does this theory then fare with respect to the criteria for a theory of phonology and phonetics described at the beginning of the chapter? The first criterion for a good theory concerned whether it was capable of providing an account for critical phonetic data. The capacity to provide an explanation rested in large part on a theory's providing a clear definition of input-output relations for the phonetics and phonology. Examining the phonetic rules described in P80, it is clear that this requirement is not met. Thus, the theory does not meet the first criterion: it is incapable as it stands of providing an account of F0 data.

How does P80 fare with respect to the second criterion, namely simplicity? The theory assumes that the input-output relations between tones and F0 values are governed by approximately 10 phonetic rules for scaling of "prominence values." These rules appear to be arbitrary in their natures, and prominence values in the theory bear little resemblance to perceptual or acoustic prominence (Ladd 1990). Nevertheless, prominence values are assumed to correspond to ratios of F0 values for tones. Prominence value ratios can be modified by three parameters, *n*, *k*, and *p*, whose phonetic status is unclear. Several parameters are also assumed to not be recoverable from the speech signal. For example, tones are assumed to be scaled with respect to a baseline, which is expected to generally be inaudible. Moreover, the "L" tones in H*+L pitch accents are assumed to never be heard. On the basis of these considerations, we can conclude that the theory does not fare well with respect to simplicity.

In sum, P80 does not perform well with respect to the evaluation criteria described in this chapter. First, it fails to account for phonetic data concerning F0 turning points. Moreover, it posits a number of arbitrary parameters and rules, and therefore cannot be characterized as simple. In contrast, Section 4.4 showed that the tone interval theory met both evaluation criteria. This suggests that the tone interval theory represents a more desirable theory of the phonology-phonetics relation in intonation languages than P80.

4.5.2 Pierrehumbert and Beckman (1988)

Next, we will consider how the theory of phonetics and phonology posited by Pierrehumbert and Beckman (1988), hereafter PB88, fares with respect to the evaluation criteria. The theory of PB88 is notable in that it redefined the phonetics-phonology relation, compared to P80. Like the earlier theory, PB88 assumed that the heights of H and L tones were determined by prominence values. However, in the new version of the phonology-phonetics relation, H and L tones were assigned prominence values ranging from 0 to above 1.0. The prominence value of H tones defined their position with respect to a high reference line *h*, while the prominence value of L tones defined their position with respect to a low reference line *l*. Moreover, PB88 eliminated phonetic scaling rules for tones, so that prominence values alone determined the heights of H and L tones.

We will consider how the theory evaluates with respect to the simplicity criterion first. The fact that PB88 assumed that tones are scaled only with respect to prominence values, rather than a set of phonetic rules, suggests that the theory is simpler than that of P80. However, certain crucial parameters are assumed to be inaudible, such as the reference lines with respect to which H and L tones are scaled, as well as the "+L" of H*+L accents. This suggests that the theory fares only moderately well with respect to the simplicity criterion.

How does PB88 fare in its ability to account for F0 turning point data? To answer this, we need to consider whether the theory crucially defines restrictions on the relative heights of each adjacent tone pair in the sequence, since this was established as a necessary constraint on transparency in input-output relations. Close inspection of the theory suggests that there are *no* constraints in this theory on the scaling of adjacent tones. In other words, the prominence values of H and L, *p*(H) and *p*(L), are not restricted with respect to one another. Thus, L is permitted to be higher than adjacent H. As a result, the theory encounters many of the same problems as that of P80. The lack of appropriate restrictions on the relative heights of tones means that the theory overgenerates possible phonetic contours on the basis of a phonological sequence, and that a given intonation contour is necessarily indeterminate with respect to its phonological description.

In summary, PB88 fails to meet either of the evaluation criteria described earlier in this chapter. First, the theory is insufficient to account for F0 turning point data. This is due largely to its failure to define a clear input/output relation, which depends on the presence of syntagmatic restrictions on the relative heights of tones in sequence. Because of a lack of sufficient restrictions on the scaling of adjacent tones, a L tone can give rise either to an F0 peak, or to an F0 valley. This suggests that the theory generates too many possible F0 contours for a given input, and that a given contour may be indeterminate with respect to its phonological representation. Second, the theory fares only moderately well with respect to the simplicity criterion. While it eliminated the complex system of phonetic rules in P80, it nevertheless posits a number of inaudible parameters. In the following section we consider how these theories might perform with respect to the evaluation criteria if they were to be modified in ways that addressed some of the problems described above.

4.5.3 Other selected theories

In the last section we saw that two versions of the paradigmatic theory of English intonation, P80 and PB88, did not fare well with respect to the evaluation criteria described at the beginning of the chapter. In particular, it was shown that failing to restrict the relative heights of every pair of tones in the sequence led to overgeneration of possible phonetic contours from the phonology and indeterminacy of the phonological representation for phonetic pitch patterns. Although this constitutes the first formal demonstration of this inadequacy, Ladd (1990, 1993) has previously criticized the models of Pierrehumbert (1980) and Pierrehumbert and Beckman (1988) on the grounds that the pitch range parameters are inappropriately unconstrained. He writes (1990: 37): "Unconstrained gradient variability of prominence and other pitch range parameters is, in my view, the most serious empirical weakness of a great many quantitatively explicit models of F0." However, the significance of problems with the lack of constraints on such parameters has not been discussed in the literature, and no comprehensive proposals for tone scaling have been made for how these models of phonetics and phonology might be amended.

In fact, it is possible that some modifications might be made to these theories, which could potentially allow them to overcome some of the problems with overgeneration and indeterminacy described earlier. In this section, we consider how different modifications to existing paradigmatic theories might improve the ability of such theories to meet evaluation criteria described earlier. Ultimately, we would like to evaluate the performance of modified versions of theories which assume exclusively paradigmatic tonal primitives, against the performance of the tone interval theory, which assumes syntagmatic tonal primitives.

There are a number of theories that have proposed in which tones are scaled paradigmatically with respect to abstract references lines in the speaker's pitch range (e.g., Bruce and Gårding 1978, Cohen, Collier, and 't Hart 1982, Pierrehumbert 1980, Liberman and Pierrehumbert 1984, Pierrehumbert and Beckman 1988, Ladd 1990, 1993, van den Berg *et al.* 1992; see Ladd 1996 and Gussenhoven 2004 for discussion). These theories did not necessarily share the same assumptions, nor did they necessarily seek to account for the same types of data.[50] Therefore, the present discussion will be limited to the subset of theories which share an explicit common assumption that H and L tones are each scaled paradigmatically with respect to abstract high and/or low reference lines (e.g., P80, PB88, Liberman and Pierrehumbert 1984). Here, we will consider how certain modifications to these proposals might affect their performance with respect to the evaluation criteria.

The first modification to be considered involves restricting the relative heights of adjacent tones, so that H does not go below adjacent L, and L does not go above adjacent H. Of the two versions of the paradigmatic theory of English intonation discussed above, these changes could be incorporated in a relatively straightforward fashion into the theory of PB88.[51] Recall that in this theory, H and L are assigned prominence values ranging from 0 to above 1.0. The prominence value of H tones defines their position with respect to a high reference line *h*, while the prominence value of L tones defines their position with respect to a low reference line *l*. It turns out that if prominence values of adjacent H and L are constrained so that $1\text{-}p(H) < p(L)$ for adjacent H and L, then L will not be higher than H.

In what ways would such a modification improve the performance of this theory with respect to the criteria described earlier? Although this change would produce a predictable input/output relation, the modified theory still does not perform very well with respect to simplicity. Even though L cannot rise above adjacent H, and H cannot fall below adjacent L, both types of tones are still scaled with respect to inaudible phrasal reference lines. Moreover, the relationship of the reference lines to the tones is unpredictable: the tones may be above or below the reference lines. As a result, a child would have no direct phonetic evidence of this crucial part of the input/output relation.

We can then consider a second modification to this theory. This modification would entail making the phrasal reference lines audible by requiring that H and L be restricted to be scaled *on* the phrasal reference lines. How would such an alteration improve the performance of the model with respect to the evaluation criteria?

The answer is that although such a modification would lead to an improvement with respect to the simplicity criterion, ultimately a theory based on paradigmatic primitives is more complex than a theory based on syntagmatic primitives. In order to see why this is so, consider the fact that all models which involve scaling paradigmatic tones with respect to phrasal reference lines permit these phrasal reference lines to move in F0 space. Indeed, under the assumption that tones are paradigmatic, it is necessary either to assume that the phrasal reference lines can move, or that tones can move vertically with respect to the phrasal reference lines. If we constrain the tones to be on the reference lines, so that a child can hear them, then the reference lines must be able to move around in the pitch space. But then we are back to where we started out, in that we must crucially restrict the relative heights of adjacent phrasal reference lines, rather than the relative heights of adjacent tones, in order to account for F0 turning point data. For example, in order to obtain a predictable input/output relation so that H is higher than adjacent L, while permitting the phrasal reference lines to be heard, we need to constrain the high reference line to be higher than the low reference line. This accomplishes the same ends as requiring H to be higher than L, except that we have more parameters. Even with only a single reference line, say a low tone line, a theory based on paradigmatic primitives evaluates less favorably with respect to the simplicity criterion than a theory based on syntagmatic primitives.

The conclusion of this thought experiment is that any empirically adequate theory of phonology and phonetics which seeks to explain intonation data in terms of paradigmatic phonological primitives is ultimately more complex than a theory which seeks to explain such data in terms of syntagmatic phonological primitives. No empirically adequate theory based on paradigmatic primitives currently exists; this section showed that undertaking modifications to existing paradigmatic theories yields an account in the limit which is ultimately less straightforward than a theory based on syntagmatic primitives. As a result, we can conclude that the tone interval theory outperforms existing *and possible* theories which are based on paradigmatic primitives, both in terms of descriptive adequacy and in terms of simplicity.

4.7 Summary

This chapter evaluated several theories of phonology and phonetics with respect to two evaluation criteria. The first criterion for such a theory is to account for phonological and phonetic data. This can be accomplished by defining a clear input/output relation in the theory. This permits a set of phonological primitives to give rise to a clear-cut set of phonetic attributes, while observed phonetic data can readily be interpreted in terms of a well-defined set of phonological primitives. The second criterion concerned simplicity; a theory should not posit any more parameters than necessary, nor should it posit any parameters for which there is no direct phonetic evidence.

Next, we reviewed a body of phonetic data which any theory of phonology and phonetics of intonation should crucially be able to account for. This data showed that in a wide variety of languages, the types and temporal alignment characteristics of F0 turning points are significant for the phonological representation. It was shown that theories of phonology and phonetics face special challenges in accounting for data from intonation languages in particular, since in such languages, a pitch contour with a given shape can occur in different parts of a speaker's pitch range. In order for a theory to account for F0 turning point data in an intonation language, we established that it should define a predictable input-output relation. In particular, we showed that it is necessary for a theory to somehow restrict the relative heights of each pair of adjacent tones in the sequence. Failure to do so was shown to lead to overgeneration of possible contours on the basis of the phonological string, as well as underdetermination of the phonological description of certain pitch contours.

Having established the restrictions on input-output relations which were necessary for descriptive adequacy, we turned to evaluating several theories with respect to the criteria discussed in Section 4.2. It was shown that the tone interval theory met both of the evaluation criteria. First, it achieved descriptive adequacy by restricting the relative heights of each pair of adjacent tones in the sequence through the choice of syntagmatic phonological primitives. Moreover, the theory presented a simple explanation of the phenomena at hand, positing no more parameters than necessary, as well as no parameters which could not be heard directly. Finally, the theory also achieved the beginnings of explanatory adequacy by proposing that the consistency of the presence and timing of F0 turning points was due to underlying phonological primitives which represent relative height relations, where such representations are the natural outcome of general auditory processes involving the comparison of relative pitch levels of tones in sequence.

Next, the theories of Pierrehumbert (1980) and Pierrehumbert and Beckman (1988) were evaluated. It was shown that neither theory sufficiently restricts input-output relations, leading in turn to phonological overgeneration and phonetic indeterminacy. Moreover, each theory assumed that input-output relations enlist some number of primitives or parameters which are never expected to be heard. This suggests, in sum, that neither theory is capable of describing nor explaining F0 turning point data.

Finally, we considered how these and other theories based on paradigmatic primitives might be modified in order to address some of the theoretical shortcomings. It was found that in the limit, an explanation based on syntagmatic primitives is simpler than an explanation based on paradigmatic primitives. This is interpreted as support for tone interval theory, which assumes the existence of syntagmatic relations in the phonology.

Chapter 5 – An overview of the phonetics and phonology of English intonation

5.1 Organization of the chapter

The last chapter showed that tone interval theory provides a simple and descriptively adequate account of F0 data. On the other hand, the paradigmatic theory of English intonation led to overgeneration of possible F0 contours from a phonological string, as well as indeterminacy in the phonological representation of phonetic pitch patterns. We showed that the advantage of tone interval theory stemmed from the fact that it constrains the relative heights of every pair of adjacent tones in the sequence; in contrast, the paradigmatic theory crucially did not formalize such a constraint. The present chapter shows how the key descriptive insights of the paradigmatic theory can be preserved and even strengthened by adopting a syntagmatic approach to English intonation, as captured in tone interval theory. Moreover, it will be shown that a number of problems which are outstanding in the literature can be dealt with in the tone interval framework. In the following we briefly introduce each of the topics to be discussed in the major sections of this chapter.

First, we will discuss some preliminaries in Section 5.2. In particular, we will show that with a simple change of notation, descriptions of intonation patterns obtained in the tone interval framework are very similar to those of the paradigmatic theory. This similarity then readily permits a re-expression of previous descriptive analyses within the present framework.

Next, Section 5.3 shows through annotated examples that many of the insights of the paradigmatic framework transfer to the present framework, thereby strengthening the overall descriptive apparatus significantly in several ways. First, it will be shown that tone interval theory supports the collective assumptions of linguists about the way tones should and do behave phonetically. By defining a clear input-output relation between phonology and phonetics, tone interval theory places previous descriptive work carried out within the paradigmatic framework on firmer theoretical footing. Second, tone interval theory affords a significantly greater degree of phonetic transparency between the F0 contour and the phonological representation than the paradigmatic framework. Third, it will be shown that the theory can account for almost all of the intonation patterns of English using only six simple primitives stemming from two tone types (starred versus unstarred) and three primitive relations of relative height (*higher*, *lower*, and *same*). Finally, Section 5.3 will also show that this theory presents a solution to several outstanding problems in the literature stemming from discrepancies between phonetic facts and earlier phonological theories.

Section 5.4 addresses some issues related to the phrase-level organization of English intonation. We will propose that the number of phrase-related tones can vary at the ends of intonational phrases. It will be shown that this provides a more phonetically consistent and simpler account of F0 phenomena in English and several other languages, compared with previous proposals. Finally, we will propose some hypotheses regarding factors which restrict tone interval sequences. Supporting evidence for these hypotheses will be provided from the literatures on auditory perception and music cognition.

Finally, Section 5.5 presents quantitative evidence from an experiment by Pierrehumbert (1980), hereafter P80, and Liberman and Pierrehumbert (1984) showing that speakers carefully controlled the relative heights of nonadjacent tones on accented syllables in a short utterance with surprising consistency. It will be shown that tone interval theory provides an account of these findings using the mechanisms proposed in Chapter 3. A summary of the entire chapter will be provided in Section 5.6. We turn now in Section 5.2 to addressing some preliminary issues regarding the examples to be presented later in Section 5.3.

5.2 Some preliminaries to intonational description

One of the strengths of tone interval theory that was demonstrated in the last chapter is that it achieves descriptive and explanatory adequacy in accounting for F0 data. We showed in an abstract way that the phonological representation readily gives rise to phonetic outputs with predictable forms, and conversely, that a set of phonetic observations can readily be interpreted with respect to a clearly-defined set of phonological primitives. Moreover, the theory was shown to also achieve explanatory adequacy by positing that the consistency of the presence and timing of F0 turning points is due to underlying phonological primitives which represent relative height relations as an outgrowth of general auditory processes involving the comparison of relative pitch levels of tones in sequence.

An overarching goal of this chapter is to demonstrate using concrete examples that the tone interval theory achieves phonological and phonetic transparency *in practice*. This section is concerned with addressing some necessary preliminaries to describing such examples. First, we will briefly review the abstract mechanisms by which the theory accounts for data concerning differences in the type and alignment characteristics of F0 turning points. Then we will develop some notation which will ultimately be less cumbersome than the formalisms of Chapters 2 and 3. We will then move on to considering some examples in Section 5.3.

5.2.1 Phonological representations for F0 turning points

In this section we will briefly review how tone interval theory accounts for the presence of a significant F0 turning point of a particular type – a peak, valley, or corner – at a particular location in the segmental string, as described in Chapter 4.[52] In that chapter, we showed that distinct F0 turning point types are associated with distinct combinations of adjacent tone intervals. For a sequence of tones $[T_1\ T_2\ T_3]$, T_2 will give rise to an F0 peak at phonetic output if it is higher than both T_1 and T_3; this will happen if the associated tone intervals are $I_{1,2} > 1$ and $I_{2,3} < 1$. Similarly, T_2 will give rise to an F0 valley at the phonetic output if it is lower than both T_1 and T_3, which will happen if the associated tone intervals are given as $I_{1,2} < 1$ and $I_{2,3} > 1$. Finally, T_2 will correspond to a corner if exactly one of the "surrounding" tone intervals specifies a relation of equality. For example, the corner formed when a level contour is followed by a rise is described by the tone interval sequence $I_{1,2} = 1$ and $I_{2,3} > 1$.

The distinction between starred and unstarred tones aids in accounting for differences in the alignment of an F0 turning point. For a sequence $[T_1\ \ T_2{*}\ T_3]$, the F0 turning point associated with $T_2{*}$ will occur in the vicinity of a metrically stressed syllable, since its corresponding tone is starred. On the other hand, for a sequence of unstarred tones $[T_1\ T_2\ T_3]$, the F0 turning point associated with T_2 will occur in the vicinity of a metrically weak position, since its corresponding tone is unstarred. In this way, tone interval theory can account for phonologically significant differences in F0 turning point alignment. Having reviewed some of the basic aspects of how the theory accounts for F0 curves, we turn now to the issue of notation.

5.2.2 A simplified notation for tone intervals and relational features

This section develops a simplified notation for tones and tone intervals. There are several motivations for such an endeavor. First, a simplified notation will make clearer some of the connections with earlier descriptive linguistic work; such connections will be apparent in part because of the similarity of the notation to be developed as compared to previous descriptive notation. This in turn will illustrate the fact that much of the work previously done within the framework of the paradigmatic theory readily *transfers* to the proposed theoretical framework. In this way the present approach can more readily be seen to build on previous work, rather than to chart a course which is far-flung. In the same vein, developing a simplified notation should make clearer the phonetic transparency of the present system, as well as how its proposals remedy deficiencies in the input-output relations for phonetics and phonology associated with the paradigmatic theory. Finally, the process of simply describing contours is facilitated by a simplified notation as compared to the cumbersome formalisms described in Chapters 2 and 3.

In the simplified notation to be developed, we will replace generic tone labels like T_1, T_2, T_3, etc. with a letter – *H*, *L*, or *E* – which uniquely specifies which of the three relative height relation(s) that each tone shares with its neighbor(s) through syntagmatic tone intervals. An appropriately chosen notation would obviate the need for separately describing tones and tone intervals, since both aspects of the representation could be combined into one. We will therefore adopt the following set of conventions for naming tones and tone intervals. We will replace a tone *T* with *H* when *T* is *higher* than the tone immediately to its left. Moreover, we will replace *T* with *L* when *T* is *lower* than the preceding tone. Finally, we will replace *T* with *E*, when *T* is at the *same* level as the preceding tone. Starred or unstarred notation will be used as appropriate, so that a starred tone T^* becomes starred H^*, etc. We will refer to this notation, which combines tones and tone intervals, as *L/H/E notation.*[53]

This algorithm permits us to uniquely assign a letter to each tone representing the relative height relation it shares with the preceding tone. However, there is one tone which does not get assigned a letter by this means, namely, the first tone in an intonation phrase. How can we determine a letter equivalent for the very first tone in a phrase? The solution is given by the Reciprocal Property discussed in Chapter 2. This property captures the formal equivalence of the statements "T_2 is higher than T_1" and "T_1 is lower than T_2". Thus, if T_2 is higher than T_1, so that T_2 is described by a letter *H*, then T_1 is lower than T_2, and by this logic T_1 should be described by a letter *L*. However, we do not want to confuse an *L* chosen to reflect a relation with respect to the *following* tone, with an *L* which was named according to a relation with a *preceding* tone. Therefore, we will distinguish this tone from the others by assigning a subscript, L_0, when the tone is initial in an intonational phrase. Indeed, this special notation is warranted by the special status of the initial tone – it alone is not preceded by any other tone in the phrase. In the same way, an initial tone which is followed by a *L* will be written H_0, and an initial tone which is followed by an *E* will be written E_0. Moreover, starred initial tones will be notated with asterisks, as might be expected.

Because we think that English probably represents only syntagmatic relations, it seems possible to describe English intonational patterns using a very limited number of primitives. In particular, we will show that it is possible to describe almost all the intonational patterns in English using only six simple primitives based on two types of tones (starred and unstarred) and three relative height relations (*higher*, *lower*, and *same*, or *H*, *L*, and *E*).[54] Moreover, we will show that a description in such terms is more transparent phonetically than previous theories have afforded.

Throughout the chapter, tone interval theoretic notation will be italicized to distinguish it from the notation of the paradigmatic theory, which will be given in a non-italicized font. However, the examples to be presented in Section 5.3 will make clear that there is a close connection between tone intervals in *L/H/E* notation and the way in which L's and H's were used within the paradigmatic framework to describe F0 curves. The last point to be addressed in these preliminary remarks concerns how the examples to be discussed in Section 5.3 were selected and generated.

5.2.3 Genesis of the illustrative examples

We have laid the groundwork for the examples to be discussed in this chapter in two respects. First, we reviewed the relationship assumed under this theory between phonetics and phonology, as exemplified in the relation between F0 turning points versus tones and tone intervals. Second, we developed a simplified notation which permits a more direct comparison with previous descriptive work while also putting that work on firmer theoretical ground. Before we turn to these examples, it is useful to say a few words about why they were chosen and how they were created.

The intonation contours which are presented in this chapter are intended to serve several purposes. First, they illustrate that tone interval theory is characterized by phonetically transparency. As we will show, F0 turning points are consistently described by tones, and tones consistently give rise to F0 turning points.[55] Such a transparent input-output relation is noted by Ladd (2000) to be a desirable characteristic in a system of intonational description, and this transparency was also shown formally in Chapter 4 to be lacking in the paradigmatic theory. Second, these examples show that tone interval theory makes significant connections with previous work by describing key attributes of English intonation patterns in a very similar way to earlier intonational descriptions; this similarity is made plainer by the use of *L/H/E* notation in a way that will become clear through the examples. In this way the present theory can be seen to build on *and strengthen* earlier descriptive work by codifying in a solid theoretic manner the implicit, common-sense assumptions about the relations between phonology and phonetics that guided previous linguistic descriptions under the paradigmatic framework. A third purpose of the examples to be discussed in this chapter is to justify some of the claims and theoretical devices which were introduced earlier in Chapters 2 and 3. At that time we presented descriptive formalisms with little remark, with the intent of providing later justification for those formalisms in Chapter 4 and the present chapter. It will be shown that those claims and devices permit an explanation of a number of outstanding phonetic facts about English intonation which previously have had no account.

Finally, we can say a few words about how the examples were produced. The intonation contours to be presented were spoken by the author and recorded using Praat software (Boersma and Weenink 2002). Praat was also used to generate the F0 contour displays used in these examples. Each utterance was produced with a specific pitch pattern in order to generate an F0 contour which could be described by a particular sequence of tones and tone intervals. In some cases a wider than normal pitch range has been used in order to illustrate the significant turning points of the F0 contour more clearly. We note that all F0 values in the examples are shown on a logarithmic scale, since the measurement scale for pitch perception is logarithmic (Moore 1997). These examples have been annotated using the *L/H/E* notation discussed in Section 5.2.2. We turn now to a discussion of these illustrative examples and the phonological representations which we claim to underlie them.

5.3 A practical guide to tone intervals and tones

5.3.1 F0 peak and valley alignment

Figures 5.1 and 5.2 illustrate how contours with F0 peaks and valleys are described in tone interval theory, and how different patterns of alignment for these points are captured. These and other examples in the chapter display a spectrogram with a superimposed F0 contour, where significant segmental landmarks are indicated below the spectrogram (Stevens 2002). In the utterance *It's Mother Maria* in Figure 5.1, the accents are marked with F0 peaks, which are described by *H** tones. Recall that *H** indicates a starred tone which is higher than the tone just to the left. In general, any tone which is a *H** or *H* will correspond to an F0 maximum in the context of a rightward *L*. The unaccented low points in the utterance in Figure 5.1 are labeled with *L* tones. Recall that *L* indicates an unstarred tone which is lower than the tone to its left. In general, any tone which is a *L* in the context of a rightward *H* will correspond to an F0 minimum. The medial *L* tone is given a "+" symbol to indicate that it is a "rightward-aligning" unstarred tone; in other words, the unstarred *L* tone associates phonologically with a metrically weak position preceding a metrically strong position that carries a starred tone, as discussed in Chapter 3. The entire contour is described as L_0 H^* $L+$ H^* $L\%$.[56] Note that the "%" symbol is indicated after the final tone in an intonational phrase to indicate that the tone occurs at the edge of an intonational phrase boundary, as discussed in Chapter 3. Note that this does not indicate a distinct category of tonal element; rather, we will claim in Section 5.4.2 that tones which happen to be at the edges of intonational phrase constituents are realized with distinctive phonetic properties. Finally, note that there is a smooth drop in F0 from the last peak, which is marked with H*, to the last low, which is marked L%; there is no other phonetic evidence of any additional tones at the end of the phrase. Phrase-final phenomena in general are discussed in Section 5.4.

The equivalent phonological description for the sequence L_0 H^* $L+$ H^* $L\%$ in *L/H/E* notation is written as in (5.1) in standard tone interval notation. Clearly, the notation in terms of *L* and *H* is less cumbersome, and we will prefer it for the remainder of this work.

$$\left[\frac{T_2\,*}{T_1}>1 \quad \frac{T_3\,+}{T_2\,*}<1 \quad \frac{T_4\,*}{T_3\,+}>1 \quad \frac{T_5}{T_4\,*}<1\right] \tag{5.1}$$

A number of examples in this chapter will be presented in pairs, where each member of the pair is the "mirror image" of the other. By this we mean that flipping one contour upside down would approximately yield the other contour. It will be shown that the tone interval framework generates a parallel description for each such pair, suggesting a degree of symmetry in the intonation system of English which had been obscured under previous descriptions. For example, the mirror image of the contour in Figure 5.1 is shown in Figure 5.2; the former intonation pattern is typical of a statement, while the latter is typically of a question. Here, the accents are now marked with F0 valleys, which are described by L^* tones. Conversely, the unaccented high points in the utterance are labeled with *H* tones. This is significant, because the high points in Figure 5.1 contrastingly corresponded to metrically prominent syllables. By comparing these two figures, we can see that a H^* tone followed by $L^{(*)}$ describes an F0 peak on a metrically prominent syllable, while a *H* tone followed by $L^{(*)}$ describes an F0 peak on a metrically weak syllable. Finally, note that there is a smooth rise from the last F0 valley to the high point at the end of the utterance; there is no evidence of any additional tones other than the ones already marked. The entire contour is described as H_0 L^* $H+$ L^* $H\%$, which is indeed the mirror image of the description of 5.1 (L_0 H^* $L+$ H^* $L\%$).[57]

Taken together, Figures 5.1 and 5.2 illustrate several things. First, these examples show that F0 peaks and valleys are associated with unique and distinctive underlying phonological representations in the tone interval framework. Moreover, they illustrate that phonologically distinctive patterns of F0 alignment of peaks and valleys with the syllables of an utterance are captured by whether a tone is starred or unstarred. Starred tones are associated with F0 turning points which are in the vicinity of metrically prominent syllables, while unstarred tones are associated with F0 turning points which are in the vicinity of metrically nonprominent syllables, as discussed in Chapter 4. In addition, these examples help to illustrate that there is a predictable relation between the phonology and phonetics in this system; for example, a *H* tone in the context of a rightward *L* will give rise to an F0 maximum. Finally, the pair of examples in 5.1 and 5.2 are shown to represent a mirror image pair, a fact which is reflected in the coordinate analysis which is afforded to each in the tone interval framework. This symmetry is made apparent by the choice of a simpler set of phonological primitives than previously assumed. In the following section, we show how this treatment of peaks and valleys addresses a problem noted with the paradigmatic theory which has been described by Ladd and Schepman (2003).

5.3.2 A consistent phonological treatment for F0 "dips"

In this theory, F0 valleys are consistently treated as arising from *L* tones. This consistent treatment of such "dips" resolves a problem with previous descriptions of English intonation that has been noted by Ladd (2000) and Ladd and Schepman (2003). In the paradigmatic description of English intonation, F0 "dips" between high accents were claimed to arise through one of two distinct types of mechanisms: a non-monotonic interpolation function which generated a low "sagging transition", or the presence of a *L* tone. The former low points were predicted to occur about halfway between H* accents, while the latter were expected to align more consistently in the vicinity of the second H* accent. Ladd and Schepman showed in a production study that the F0 valley between high accents consistently exhibited a single type of behavior: it consistently aligned with a position just before the upcoming high accent. This behavior suggested that the claim of two separate mechanisms for producing F0 valleys between high accents was incorrect. Ladd and Schepman proposed that the distinction between the two types of contours should be eliminated from the description of English.[58] However, no comprehensive proposals had been made regarding how this should be accomplished, until now.

The present theory presents a formal account of why an F0 valley on an unaccented syllable in a contour like that of Figure 5.1 arises from a *L* tone. In this theory, such an event consistently arises through the presence of an unstarred tone, which is required to be in a timing slot adjacent to a metrically strong position occupied by a starred tone. This tone is assumed to be joined into two adjacent syntagmatic tone intervals which specify that it is lower than either of its neighbor tones. Each tone is associated with a target pitch, and the pitch associated with the *L* tone is connected up to the target pitches of other tones in the sequence by monotonic pitch interpolation functions. This ensures that all significant F0 turning points in the contour arise from underlying tones. The next section demonstrates another way in which the present proposal permits increased transparency between the phonological representation and the phonetic output.

5.3.3 Bitonal accents and phonetic alignment

One aim of this chapter is to illustrate that facts discussed in Chapter 4 concerning F0 turning point alignment in English can be captured using a simpler inventory than was assumed in the paradigmatic theory. A specific goal for this section is to show that the intonational facts which led P80 to propose complex primitives consisting of two tones are more transparently captured under the assumption that those two tones are independent but align in a coordinated fashion. In this regard, it will be demonstrated that the assumption that the two tones of bitonal accents assumed under the paradigmatic theory are independent tonal entities not only presents a simpler description of the intonational inventory, it also permits an account of facts about alignment in English which have now been demonstrated in several studies but until now have had no account.

It is noteworthy that facts concerning the coordinated alignment of F0 peaks and valleys in adjacent positions in English intonation curves originally led P80 to propose the concept of the "bitonal pitch accent". Such accents corresponded to five of the seven pitch accent types originally proposed in the paradigmatic system. Those bitonal accents were: L+H*, L*+H, H+L*, H*+L, and H*+H. A special mechanism of autosegmental association was claimed to underlie the coordination of tones in bitonal accents, as discussed in Chapter 3. In that type of association, the unstarred tones of bitonal pitch accents did not associate directly with tone-bearing positions in the phonological structure, as other tones did; rather, they were assumed to associate with the starred tone of the bitonal pitch accent. This special type of autosegmental association was justified by the claim that the two tones in these accents showed special phonetic properties; in particular, they were claimed to align with respect to one another by a fixed temporal interval.

The claim that the two tones in the paradigmatic framework's bitonal pitch accents align a fixed temporal distance apart with respect to one another has now been shown to be false in several phonetic studies in English, Greek, and Dutch (Arvaniti, Ladd, and Mennen 1998; Ladd, Faulkner, Faulkner, and Schepman 1999; Ladd, Mennen, and Schepman 2000; Dilley, Ladd, and Schepman, forthcoming). These studies have shown that the F0 turning points assumed to correspond to the tones in bitonal accents instead are consistently aligned with respect to the segmental string. This is true even when the speech rate varies (Ladd *et al.* 1999). This suggests that the F0 turning points described by tones can be separated by variable temporal distances, yet they remain anchored to specific segmental positions. To illustrate this point more carefully, Dilley *et al.* examined the alignment of the two tones in contours described as L+H* pitch accents. They found that the F0 valleys and peaks described by the "L+" and "H*" tones of L+H* pitch accents generally behaved as if they were in fixed positions with respect to the segmental string, rather than being separated by a fixed temporal interval with respect to each other.

The proposals in Chapter 3 permit tone interval theory to reconcile these phonetic facts with the theoretical description of English intonation. In that chapter, we asserted that unstarred tones must align with metrically weak positions which are adjacent to metrically strong positions occupied by a starred tone. In other words, unstarred and starred tones are treated as formally distinct entities which nevertheless have coordinated association phonologically and coordinated F0 alignment characteristics phonetically. For example, contours described as L+H* in the paradigmatic framework are described here as consisting of a sequence of a *L+* tone and a *H** tone, where the two tones are required to associate with adjacent metrical positions in the underlying phonological structure. We will show through examples that each of the alignment facts that the paradigmatic theory sought to capture through the use of bitonal accents are accounted for in a simpler and more phonetically transparent manner when we assume that the two tones in such contours are independent, yet coordinated, tonal entities. We will refer to such contours as *bitonal accentual sequences*.

Another remark is in order concerning the proposed phonological description afforded here. The descriptive apparatus of the bitonal accent has been employed widely in linguistics since the early 1990's, and a number of insightful observations about the intonation systems of different languages have been made based on this descriptive approach. It is important to point out that there is a *transparent relationship* between the intended descriptive use of P80's bitonal accents and the descriptions proposed here, so that most all of this earlier descriptive work readily transfers to the proposed descriptive approach. In this way, the present proposals can been seen to build on *and extend* earlier descriptive work by defining a clear relationship between phonetics and phonology. In so doing, the tone interval framework serves to strengthen earlier descriptive work by eliminating the problems of phonetic indeterminacy and phonological underdetermination associated with the paradigmatic theory. The tone interval theory does this by explicitly codifying the implicit assumptions shared by linguists about the behavior of L's and H's in the paradigmatic framework's bitonal accents.

In the following, we consider some examples illustrating the behavior of starred and unstarred tones in English. It will be shown that unstarred tones behave like independent tonal entities in their alignment behavior. These examples therefore will provide anecdotal evidence in support of the claim that the two tones in the paradigmatic theory's bitonal accents are actually independent tones with coordinated association.

5.3.4 The coordinated association of starred and unstarred tones

We presented two arguments in the last section for our claim that the two tones in bitonal accents in the paradigmatic tonal inventory instead involve the coordinated phonological association of independent starred and unstarred tones. First, a number of phonetic studies have now provided evidence against one of the key claims of the paradigmatic theory which justified treating such contours as bitonal primitives. These studies have shown that the two tones in the paradigmatic framework's bitonal pitch accents do not, in fact, align by a fixed temporal distance with respect to one another. Rather, they show characteristics of independent tones by aligning individually with the segmental string in adjacent metrical positions. This suggests that the special theoretical status afforded to such accents in autosegmental theory was unwarranted, and that a simpler assumption, namely that the unstarred tone associated directly with some underlying tone-bearing unit, is closer to correct. Second, we argued that an inventory based entirely on single-toned primitives is simpler than an inventory based on complex bitonal primitives.

In this section we present a third line of argument in support of re-expressing the two tones in the paradigmatic theory's bitonal accents as independent starred and unstarred tones. We will show through examples that doing so yields not only a simpler description of the phenomena which the paradigmatic theory attempted to describe, but also shows that the phonology of English intonation exhibits more symmetry and systematicity than previously revealed under the bitonal accent analysis.

The relative simplicity of the tone interval description as compared with that of the paradigmatic theory is readily shown by considering the example in Figure 5.2. In this contour, the sequence of the low F0 valley, high F0 peak, and low F0 valley is described in terms of three independent yet coordinated tones: *L* H+ L**. Here, the fact that the F0 peak occurs on the unstressed syllable prior to the *L** is explained by the assumption that the unstarred *H* tone is rightward-aligning, *H+*, and that it therefore associates phonologically with a position just before a stressed syllable.

In contrast, the paradigmatic theory is indeterminate with respect to whether the sequence of the F0 valley, peak, and valley should be described as L*+H L* or as L* H+L*. Because of the close temporal proximity of the H and the second L*, the latter description would probably be favored by practitioners of the theory. However, the fact that the size of the supposedly constant temporal interval between the tones in a L*+H accent (or any of the accents for that matter) was never specified, we cannot say with certainty that this contour is not L*+H L*. Moreover, the description afforded within the ToBI framework would probably also permit this sequence of points to be described as L* L*.

If unstarred tones really are independent entities in English, then we might expect that they could freely align either rightward or leftward with respect to a starred tone. In Figure 5.2, we saw an example in which a medial unstarred high tone has aligned *rightward* with respect to a starred low tone in the sequence *L* H+ L**. Figure 5.3 shows a "minimal pair" illustrating that it is also possible for an unstarred tone to align *leftward* with respect to a starred tone, giving the sequence *L* +H L**.

The paradigmatic framework obscures the symmetric nature of these examples by describing the difference between Figures 5.2 and 5.3 as a difference in complex bitonal pitch accent types. On the one hand, the valley-peak-valley sequence in Figure 5.2 would be described as L* H+L*, while on the other hand, the sequence in Figure 5.3 would likely be described as L*+H L*.[59] At least two drawbacks to this description can be cited. First, the description is clearly more complex than the tone interval account, due to the fact that we must assume that this was not due simply to the rightward vs. leftward movement of an unstarred tone. Rather, we must assume that the identity of the tones in each of the *starred* positions was different in 5.2 and 5.9 as well. This brings us to the second drawback, which has to do with the fact that the description obscures the fact that this appears to be a minimal pair. Under the paradigmatic description, this is not a minimal pair, because it involves *two* contrasts in *two* positions, rather than just one contrast.

We can now consider the mirror image case to determine if an unstarred *L* tone shows the same alternation between leftward and rightward alignment with respect to surrounding starred tones. Consideration of Figure 5.1 alongside Figure 5.4 suggests that *L* tones also show independence in being able to align either leftward or rightward. In Figure 5.1, the medial unstarred low tone has aligned rightward with respect to a starred tone; the peak-valley-peak sequence is thus described as *H* L+ H**. The near-minimal pair presented by Figure 5.4 reveals that an unstarred tone may also align leftward with respect to a starred tone. In this case, the peak-valley-peak sequence is described as *H* +L H**. These examples therefore suggest that unstarred *L* tones, like unstarred *H* tones, can associate either leftward or rightward with respect to a starred tone.

In contrast, it is not clear how we can account for the difference between Figures 5.1 and 5.4 in the paradigmatic framework. One possibility which might come to mind for readers who are familiar with this theory is to represent Figure 5.1 as H* L+H* and Figure 5.4 as H*+L H*. However, this is not a possible description of the difference between these contours. This is because the L tone in H*+L was assumed to never be realized phonetically in English in all versions of the paradigmatic theory; thus, the F0 valley cannot be attributed to the L tone of a H*+L accent. Another possibility which might be considered would be to describe both contours in terms of an accentual sequence H* L+H* and to assume that the L has different alignment characteristics in Figures 5.1 and 5.4. However, this is also not a viable way of describing the difference, since it was claimed that the two tones in bitonal accents always align with respect to one another by a fixed temporal distance. Thus, a L+H* would be expected to behave in the same way in both cases. Furthermore, to propose that the "L+" has the ability to align independently with segments weakens further the claim that there are bitonal accents in English. This is because the "bitonality" of the accentual sequence would no longer serve any function, if the L were to be permitted to associate independently with different positions in the phonological structure. However, this is just what tone interval theory proposes to do. Finally, we note that the difference cannot be described in terms of different distributions of intermediate intonational phrase tones, since none of the tones in question coincides with a word boundary. At any rate, it is clear that the descriptive devices of the paradigmatic theory tend to obscure what appears to be a fundamental symmetry in the behavior of unstarred *H* and *L* tones.

It is worth noting that the kinds of paired distinctions which were given in P80 in support of bitonal primitives for English can be readily described in terms of the simpler inventory proposed in tone interval theory. For example, consider the contours in Figures 5.5 and 5.6 for the utterance *Only a millionaire.* These examples show two distinct patterns of F0 peak and valley alignment. Here, the difference is represented in terms of distinct sequences of independent, coordinated starred and unstarred tones. In Figure 5.5, the sequence of the F0 valley on an unstressed syllable and F0 peak on the stressed syllable of millionaire is described as a sequence of a rightward-aligning unstarred tone plus a starred tone, *L+ H**. In contrast, Figure 5.6 shows a sequence of an F0 valley during the initial stressed syllable of *millionaire,* followed by an F0 peak on the poststress syllable. This portion of the contour is described as a sequence of a starred tone plus a leftward-aligning unstarred tone, *L* +H*. In this way the present system is able to capture a distinction which was treated in the paradigmatic theory as a difference in bitonal accent type: L+H* vs. L*+H.

Moreover, the relative freedom of tones to associate with different metrical positions leads to configurations like those in Figures 5.7 and 5.8. In these two examples, a starred tone is surrounded by an unstarred tone on either side. In Figure 5.7, a *H** accent is flanked by a leftward *L+* and a rightward *+L*, giving rise to the appearance of a "tritonal" accent. The mirror image of this contour is shown in Figure 5.8. Here, a *L** accent is flanked by a leftward *H+* and a rightward *+H.*

The central observation here is that by assuming that tones are independent entities with coordinated alignment, we can explain facts about English intonation using only two types of tones (starred and unstarred) and three relative height relations (*higher*, *lower*, and *same*). This framework thus presents a simpler description of English than that proposed by Grice (1995), who also observed that contours such as those in Figures 5.7 and 5.8 can occur in English. However, her description involved the assumption of tritonal pitch accentual primitive, and she proposed a total of eight pitch accents for English. It should be clear that the proposed theory presents a simpler view of English intonational phenomena.

In sum, we have demonstrated that assuming that the two tones in contours described as bitonal pitch accents in the paradigmatic theory are actually independent, coordinated tonal entities yields greater transparency and simplicity in the phonological description. Moreover, certain symmetries in the intonation of English which were previously obscured are revealed when we assume that the inventory consists of simpler primitives. We showed that contours which appeared to attest "tritonal accents" can be accounted for by assuming that the tonal entities are independent yet coordinated.

We have focused to this point on tones like *H* and *L*. In the next section we will examine some contours which exhibit a tone *E* or *E**. It will be shown that admitting the tonal relation *same* into the descriptive framework for English permits a more transparent treatment of intonational phenomena that was possible under the paradigmatic description.

5.3.5 Consistent treatment of F0 corners

In the examples in Figures 5.1-5.4 we illustrated how an unstarred tone could align either with respect to a leftward starred tone or a rightward starred tone. There are also contours in English which simultaneously attest both leftward- and rightward-aligning unstarred tones. For example, consider the utterance *Anna LeMay* in Figure 5.9. In this contour, there is a pitch drop on the second syllable of *Anna*, which is described by a leftward-aligning unstarred tone, *+L*. There is a low level plateau through the first syllable of *LeMay*, and then the F0 suddenly rises. This "corner" is described by an unstarred rightward-aligning tone, *E+*. Thus, this contour attests a sequence of leftward- and rightward aligning unstarred tones in the context $H_0* +L\ E+\ H*$. Figure 5.10 shows the same intonation pattern in the utterance *Natalie May*, which has the same metrical structure as *Anna LeMay* but a different distribution of word boundaries. Here again, we see a sequence of two unstarred tones in the sequence $H_0* +L\ E+\ H*$.

The paradigmatic framework is indeterminate with respect to how these two contours would be described, due to the fact that the associated phrases entail a different distributions of word boundaries. F0 obtrusions at word boundaries are candidates for unstarred "phrase accent" tones marking intermediate intonational phrase boundaries, which are discussed in Section 5.4.1. In particular, the drop in pitch on the last syllable of *Anna* in Figure 5.9 could be described either as a L- phrase accent or as the unstarred L tone of a L+H* pitch accent.[60] Because the corresponding low point in *Natalie* in Figure 5.10 does not coincide with a word boundary, this low region would probably be attributed to the unstarred L tone of a L+H* pitch accent. Again, we see that the earlier descriptive framework tends to obscure similarities and highlight differences while introducing phonological indeterminacy into the description.

The mirror images of these utterances are shown in Figures 5.11 and 5.12. In Figure 5.11, there is a pitch rise on the second syllable of *Anna*, which is described by a leftward-aligning unstarred tone, *+H*. There is a high level plateau through the first syllable of *LeMay*, and then the F0 suddenly falls. This "corner" is described by an unstarred rightward-aligning tone, *E+*. In this way, the low-high-plateau-low sequence is described by the sequence $L_0* +H\ E+\ L*$. Figure 5.12 shows the same pattern on the phrase *Natalie May*. The difference in the positions of the word boundaries does not affect the properties of the intonational contour, nor does it affect the description of that contour, which identical to that of in Figure 5.11.

Permitting level contours to be described as arising from a tone at the right edge of the level region increases the phonetic transparency of descriptions of English intonation. To see this, we will consider a series of examples which were originally described in P80, which have been re-recorded and reanalyzed by the author. One example highlighting the increased phonetic transparency afforded by a tone interval description concerns the utterance *Legumes are a good source of vitamins* in Figure 5.13. This contour exhibits an initial drop from a high to a low pitch; this is followed by a level contour up through the first syllable of *vitamins*, then a rise to a high pitch. The drop to a low pitch is described as a leftward-aligning *+L* target, while the "corner" that arises at the right edge of the level region is then described by a starred $E*$ tone. Significantly, the final corner is treated as an explicit target. In contrast, the paradigmatic description did not treat this F0 turning point as a tone. The decision to not to do so was based on *a priori* assumptions about the nature of the grammar in English intonation. The decision to weigh theory-internal considerations more heavily than phonetic observations resulted in a loss of phonetic transparency, on two counts. On the one hand, clearly identifiable F0 turning points, such as the corner which is marked in Figure 5.13 with an $E*$, were assumed to not be tones. On the other hand, putting theory-internal concerns above phonetic explicitness also led to the assumption that some phrase accents had no identifiable F0 correlates. In Section 5.4.2 we will propose instead that there are a variable number of phrase-related tones between the last, strongest accent and the end of the phrase. In that section we will also consider the status of starred tones in what might be termed "postnuclear" positions, namely positions after the last, strongest accent of the phrase.

There are two lines of evidence supporting the view that F0 corners such as the one in Figure 5.13 are indeed tones, counter to the claims of the paradigmatic theory. First, evidence from Greek, German, Romanian, and several other languages shows that F0 "corners" behave like tones. For example, such points reportedly show consistent alignment with respect to segments, and differences in their alignment can correspond to distinct meanings (Grice, Ladd, and Arvaniti, 2000; Lickley, Schepman, and Ladd, forthcoming). Second, contours with a corner were shown to be heard by Italian listeners as categorically distinct from contours with a rise and no corner (Grice and Savino 1995). These arguments suggest that the proposal to treat F0 corners such as the one in Figure 5.13 as arising from a tone is well-advised.

We will briefly digress here to explain why treating F0 corners such as the one marked $E*$ in Figure 5.13 as arising from a tone addresses another issue in the literature. In order to account for their data from Greek, German, Romanian and other languages, Grice *et al.* followed the proposal of Gussenhoven (1993) by treating F0 corners in terms of a mechanism whereby a tone is copied from a phrasal node (following the proposals of PB88) to the boundary of the final syllable and the post-nuclear stressed syllable simultaneously.

Grice *et al.*'s proposal accounts for their data; however, there are several points that can be made about it. First, it presents a picture of a complex phonological mechanism with multiple steps. A simpler proposal is to assume that the phonology permits the number of post-nuclear tones to vary; such a proposal obviates the need for copying processes. Moreover, Grice *et al.*'s proposal technically creates a structure that appears to be ungrammatical with respect to the proposals of Beckman and Pierrehumbert (1986), hereafter BP86, which assumes that there is only a single phrase accent at the end of an intonational phrase. Finally, it is unclear when the copying mechanism should be applied, and when it should not. This suggests that Grice *et al.*'s account requires additional rules to cover the cases in which copying does not apply, ultimately increasing the complexity of the account. Each of these problems can be avoided by simply assuming that the phrasal grammar permits variable numbers of phrase-final tones.

Consideration of other examples from P80 provides additional evidence that the tone interval description increases phonetic transparency in the description of English intonation. For example, consider the contour in Figure 5.14, which shows a level contour in phrase-initial position. The extended level region is treated as arising from a single tone interval marked at its edges by two tones, E_0* and *E*+. (Recall that the initial tone is named in a manner that reflects its relative height specification with respect to the following tone.) Thus, the entire level region is treated as arising from two tones at its edges, just as we had done in the case of the level region in Figure 5.13.

In contrast, the description of the contour in Figure 5.14 which is proposed in the paradigmatic theory shows a lack of phonetic transparency. First, it assumed that there are *no tones at all* in the contour until the stressed syllable of *vitamins*, which is described as a H*. In contrast, the level region in Figure 5.13 was treated as having a single tone at the beginning of the level contour. We note that the assumption of no tones being present until the word *vitamins* in Figure 5.14 constitutes a special kind of problem with the paradigmatic account of the phonology-phonetics relation. This was because no mechanism was defined by which the initial portion of this contour could have obtained its F0 specification. It could not have arisen from direct tonal specification, since as we said, there are no tones until the word *vitamins*. Moreover, it could not have arisen by phonetic interpolation, because once again, there are no tones in the initial portion of the contour between which interpolation could have taken place. Thus, the issue of how the initial portion of the contour arose thus constitutes a gap in the theory. We note that tone interval theory addresses this technical problem by assuming that every tonal contour begins and ends with a tone, as discussed in Chapter 3.

Another example of how the tone interval description shows greater phonetic transparency in the description of level contours comes from Figure 5.15. Here, there is a level contour which extends from the beginning of the utterance up through the stressed initial syllable of the word *vitamins*, which realizes a corner as there is a transition into a rise to a final high pitch. The level contour is again treated with consistency in the tone interval description; it is assumed to arise from two tones at either end, E_0 *E**.

In contrast, there are two ways in which the paradigmatic description for the example in Figure 5.15 fails to reflect phonetic transparency. First, the level portion of the contour is described as having a L* tone at the right edge and no tone at the left edge. This highlights the generally inconsistent treatment of level contours across the examples we have seen. For level contours, sometimes a tone is assumed to be at the right edge, but not the left edge. At other times, it is assumed that there is a tone at its left edge of a level contour, but not the right edge. At still other times, it is assumed that there is a tone neither at the right, nor at the left edge of a level contour. In contrast, the tone interval theory always assumes that there is a tone at both the left and right edges of the level region. Second, the paradigmatic theory describes the rising contour at the end of the utterance in Figure 5.15 in terms of two tones, H- and H%, where the H- tone does not give rise to any identifiable features of the F0 contour. This can be contrasted with the tone interval description, which describes the F0 rise as involving one tone at the corner and one tone at the end of the rise.

In sum, the tone interval description significantly increases the phonetic transparency of the description of level contours as compared to the paradigmatic theory. While level contours are always described as arising from one tone at each edge of the level contour in the tone interval framework, we see general inconsistency in the treatment of level contours in the paradigmatic theory. Specifically, these contours were sometimes described as having one tone at a left edge, sometimes as having one tone at a right edge, and sometimes as having no tones at either edge. Consideration of other examples in P80 suggests that level contours were sometimes treated as arising from a tone at both left and right edges. Even when we ignore the difficulties posed by the input-output relations discussed in Chapter 4, the paradigmatic theory clearly assumed a many-to-many mapping between tones and F0 turning points. That is, tones sometimes gave rise to turning points and sometimes not; conversely, turning points sometimes arose from tones, but other times they did not. In contrast, tone interval theory presents a straightforward and phonetically transparent description of F0 characteristics, in which turning points consistently arise from tones. Moreover, the theory assumes that tones consistently give rise to turning points. In this latter regard there is one particular case which deserves special mention which follows readily from our previous assumptions about the phonology-phonetics relation. This the case of an F0 slope change, and we turn to it next.

5.3.6 Other relations between phonology and phonetics

So far we have presented examples which illustrate that in this framework, every significant F0 turning point – whether a peak, valley, or corner – arises from a tone. In this way, tone interval theory greatly increases the phonetic transparency of the description of English intonation as compared to the paradigmatic framework. That theory, by contrast, treated significant F0 turning points with inconsistency, even when we assumed that its input-output relations operated as they have been presumed to in descriptive linguistics. In contrast, tone interval theory formalizes a relationship between phonology and phonetics such that every significant F0 turning point is assumed to arise from a tone.

This section discusses a prediction of tone interval theory concerning the converse relationship – namely, how tones are manifested in terms of F0 characteristics. All of the examples considered thus far have involved the phonetic manifestations of adjacent tone intervals whose values were *different*. We saw that such cases give rise to F0 peaks, valleys, and corners. However, we have not yet discussed the case in which adjacent tone intervals have values which are *identical*. Indeed, it is a natural prediction of this theory that adjacent tone intervals could indeed specify that two adjacent tone pairs have the same relative height relation. We will show that in this case alone, the relationship between phonology and phonetics is not obvious. This is because, as we will see, adjacent tone intervals with the same values do not always give rise to significant and identifiable obtrusions in the F0 curve, though they often do. Coming to understand how such cases arise has implications for our conception of the relationship between phonetics and phonology. We will show that the predictions of tone interval theory are supported by observational evidence from English, as well as Greek.

In the following, we will first briefly discuss by way of introduction an outstanding problem in the literature concerning the case of Greek prenuclear accents as reported by Arvaniti, Ladd, and Mennen (2000). Next, we will review some of the assumptions of tone interval theory, focusing on its prediction that adjacent tone intervals may specify the same relative height relation. We will then consider the implications of this state of affairs for phonetics. This will lead us to examine some F0 contours in English, which will be shown to support the predicted phonology-phonetics relation. Supported by evidence from English, we will then propose a solution to the question of how to describe Greek prenuclear accents. Finally, we will propose some restrictions on tone interval sequences, which we believe to be universal. These proposals will be supported with evidence from the auditory perception literature. We turn now to a discussion of these issues.

The first issue to be discussed concerns observations about a puzzling set of facts for phonology and phonetics from Greek. In particular, Greek prenuclear accents exhibit F0 alignment characteristics which are schematically represented in Figure 5.16. The striking characteristic about these accents is that they consistently show an F0 valley on a weak prestress syllable and an F0 peak on a weak poststress syllable (Arvaniti and Ladd 1995, Arvaniti *et al.* 1998). Arvaniti *et al.* (2000) have discussed the fact that such accents are difficult to account for under standard phonological theories. For example, under standard assumptions of the paradigmatic theory, L and H tones correspond, respectively, to F0 valleys and peaks. There are two bitonal accents which consist of a sequence of a L and a H tone: L+H* or L*+H. Arvaniti *et al.* show that neither of these possibilities is appropriate for describing Greek accents, since neither the peak nor the valley is aligned with the stressed syllable, and they consider five different ways of modifying the representation so as to accommodate the case of Greek accents. They finally conclude (p. 130) that:

> …some of the five possibilities seem more promising than others, but the fact remains that the data are compatible with all five representations, and therefore the choice among them is not based on solid principles. This conclusion cannot but expose a weakness in the phonological basis of starredness and the definition of its phonetic exponents. The fact that our data do not allow us to reach a firm conclusion suggests that not only can we not infer phonetic alignment from phonological association, but – more importantly – we cannot use phonetic alignment with the stressed syllable as the defining characteristic of starred tones, i.e., of their phonological association.

In other words, the case of Greek prenuclear accents cuts to the very heart of what it means to be a “starred tone.” We will show that Greek prenuclear accents can be accounted for in a straightforward way under tone interval theory. In order to show how this is accomplished, we will first review how tone interval theory accounts for the occurrence of an F0 peak, valley, or corner. Recall that a sequence of tones $[T_1\ T_2\ T_3]$, T_2 will give rise to an F0 peak if $I_{1,2} > 1$ and $I_{2,3} < 1$, while T_2 will give rise to an F0 valley if $I_{1,2} < 1$ and $I_{2,3} > 1$, and so forth. In general, a significant F0 peak, valley, or corner will occur in the phonetics if the values of adjacent tone intervals are *different*. What is the expected phonetic outcome when the values of two adjacent tone intervals are *identical*? For example, suppose that the associated tone intervals are $I_{1,2} > 1$ and $I_{2,3} > 1$, or $I_{1,2} < 1$ and $I_{2,3} < 1$, etc. Then what sorts of F0 contours might we expect?

Before we answer this, we will briefly describe three phonological restrictions which we think are in place that limit possible sequences of tone intervals; our proposals will be justified at the end of the section. These restrictions have the effect of dramatically reducing the number of contexts in which identical-valued tone intervals are expected to arise. The first restriction that we believe to be in place is that the phonology permits no more than two consecutive tone intervals to specify identical relative height relations.[61] The second restriction is that tone intervals specifying identical relative height relations may occur only in a very limited set of metrical contexts. In particular, we propose that identical relative height relations are permitted on two consecutive tone intervals only when the middle tone of the associated three-tone sequence is starred. Third and finally, we propose that the phonology permits a sequence of two adjacent tone intervals to be specified as ***higher higher*** or ***lower lower***, but not ***same same***.

We can summarize the effects of these restrictions quite simply. First, identical tone intervals may occur in exactly two kinds of sequences: ***higher higher*** or ***lower lower***. Second, these tone intervals must be associated with a sequence of three tones, the middle member of which must be starred. In sum, we claim that the phonology only permits the sequences like $H^*\ H^{(*)}$ and $L^*\ L^{(*)}$. In contrast, the sequences $E^*\ E^{(*)}$, $L\ L^{(*)}$, $H\ H^{(*)}$, and $E\ E^{(*)}$ are not permitted, nor are longer sequences of identical tone intervals permitted.

Given an understanding of these restrictions, we can now consider what sorts of phonetic outcomes we might expect from identical tone interval sequences, as well as in which contexts they will arise. Recall that phonetically, tone intervals are implemented by assigning each tone a target pitch which realizes the relative height relation that is specified in the phonology. Moreover, each tone has a phonological association with a position in the metrical grid, where that grid position is also associated with some segment or syllable. Then we can expect two kinds of dimensions to vary phonetically: the "vertical" pitch distance between the target pitches, and the "horizontal" temporal distance between the segmental positions which realize them. This variability is illustrated in Figure 5.17, which schematically depicts the levels of target pitches [P_1 P_2* P_3], corresponding to underlying tones [T_1 T_2* T_3]. Moreover, these pitches are connected up by monotonic pitch interpolation functions, as indicated by the lines connecting them. The acoustic result of this phonetic-perceptual representation is an F0 contour which approximates the shape of the curves.

There are three important observations that can be made about the sequences in Figure 5.17. First, the pitch in the middle of the sequence arose from a starred tone, even though this point does not correspond to an F0 peak, F0 valley, or corner. This shows that the theory predicts that *for a highly restricted set of phonological sequences, starred tones may be realized as locations of a change in the slope of the pitch curve,* and hence the F0 curve. Starred tones will correspond to locations of changes in slope in the majority of cases in Figure 5.17 – specifically in cases (a), (b), (d), and (e). The second point is that we have now defined a theoretical basis for identifying changes in the slope of an F0 curve as significant for the phonological representation. As we will see, this permits an account not only for the phonology-phonetics relation of attested contours in English, but also of Greek prenuclear accents. A third point, however, is that for a minority of cases, a sequence of identical tone intervals may not correspond to a location of a change in F0 slope, as indicated by Figures 5.17(c) and (f).

While this introduces a small degree of indeterminacy into the theory, we note that there is still a quite transparent relationship overall between tones and identifiable F0 obtrusions in the theory. Earlier, we focused on the relationship between the phonology and F0 turning points, which we defined as F0 peaks, valleys, and corners. Such points are consistently treated in this theory as arising from a sequence of adjacent tone intervals with distinct values. Thus, it is still correct to say that all F0 turning points in this theory correspond to underlying tones. However, we have added a small qualification to the *converse* of this statement, in that *most* (but not all) underlying tones correspond to F0 turning points. We now see that for a restricted set of cases, starred tones, but not unstarred tones, may correspond to F0 slope changes phonetically. Moreover, for an even smaller number of cases, phonetic variation may cause some starred tones to show up as smooth rising or falling curves across an accented syllable.[62] We will argue later on the basis of auditory perceptual principles that contours which show smooth rises or falls across accented syllables, as in 5.17(c) and (f), will probably tend to be avoided, since they provide unreliable phonetic cues to the presence of a starred tone. The level of uncertainty in this theory is thus rather small, and it is quite insignificant when compared to the level of indeterminacy that existed in the paradigmatic theory. In that theory, no clear relationship between phonology and phonetics was defined, as shown in Chapter 4; thus, an F0 peak could technically be described as either a H tone or a L tone. Even when it is assumed that the tones correspond to specific shapes, as is assumed in descriptive linguistic work and in this chapter, it is clear that F0 obtrusions, such as peaks, valleys, and corners, do not reliably correspond to underlying tones, while tones do not reliably correspond to F0 obtrusions. The tone interval theory thus presents a strikingly more transparent alternative to the paradigmatic treatment of F0 data.

If there truly is a phonological basis for treating F0 slope changes as arising from underlying starred tones, as predicted by tone interval theory, then we should expect not infrequently to observe cases in which metrically prominent syllables realize F0 curves with a "hook" as in 5.17(a), (b), (d) and (e). We do not need to look far. Indeed, English intonation appears to attest such examples. We will briefly consider two such cases.

First, consider the example in Figure 5.18. This contour shows a rise to a high F0 peak on an unstressed syllable which is labeled *H+*. This is followed by a shallow fall across the following stressed syllable of *waiting*, which is labeled *L**. Next, there is a steep fall to a *L%*. This example thus realizes in English the prediction of tone interval theory that we will find cases of [*lower lower*]; this is observed in the present example where the *L** marks a change in slope.

This portion of the contour would have been described as a sequence H+L* L-L% in the paradigmatic theory.[63] Tone interval theory has the advantage over this earlier approach in an important way. Specifically, the older theory provided no theoretical basis in the phonetics-phonology relation for predicting that the location of the slope change corresponded to an underlying tone, even though descriptive work assumed that it did. (See e.g., examples and discussion in Beckman and Pierrehumbert 1986.) Moreover, the lack of phonetic transparency associated with the previous theory can be observed by the fact that the final, steep fall is described in terms of three tones, rather than just two.

Another example of a phonologically significant change in slope is shown in Figure 5.19. Again, this illustrates a contour with an accent which would have been described as H+L* in the paradigmatic framework. Here, the word _Lemming_ starts with a high F0 peak which falls to an accented syllable. Significantly, this accented syllable changes slope and proceeds to fall more steeply to the end of the utterance. This portion of the contour is therefore described in this framework as _H+ L* L%_, so that it realizes the consecutive tone intervals [_lower, lower_].

Looking more closely, we see that there is another slope change point in this example which occurs early in the utterance. In particular, the contour begins with a shallow rise; on the accented syllable _Mama_ the rise becomes steeper, so that the accented syllable of _Mama_ realizes a slope change point. The sequence of tones is thus described as L_0 _H* H+_, so that it realizes the consecutive tone intervals [_higher, higher_]. In this way, the entire example can be described L_0 _H* H+ L* L%_. We note that there are no more than two consecutive identical relations in the sequence.

It is worth noting that the contour in Figure 5.19 is indeterminate under the paradigmatic description of English, even when we make the standard assumptions of descriptive linguistics about the phonetics-phonology mapping. In particular, two descriptions of this contour are possible (Shattuck-Hufnagel, Dilley, Veilleux, Brugos, and Speer 2004). On the one hand, the F0 peak could be attributed to the H tone of a H+L* accent. On the other hand, the F0 peak could be attributed to a H* accent on the first stressed syllable in the utterance. In the latter case, the fact that the F0 peak came after the stressed syllable would be explained in terms of within-category phonetic variability for a H* accent. In this regard tone interval theory makes the clear prediction that the late peak corresponds to a tonal target which is separate from the one on the stressed syllable; supporting evidence for this prediction is shown through formal experiments in Chapter 6.

Moreover, the initial portion of the contour in Figure 5.19 suggests a strong similarity to the schematic diagram of Greek prenuclear accents as shown in Figure 5.16. Indeed, we think that the two contours can be described in the same way phonologically. The solution proposed here is to describe this contour in terms of a sequence of tone intervals specifying identical relative height relations of _higher_ across the accented prenuclear syllable. Thus, we can describe Greek prenuclear accents as crucially involving a sequence $L_{(0)}$(+) _H* +H_ [$L^{(*)}$]... This accounts for the shape of Greek prenuclear accents by assuming that the F0 valley on the unaccented prestress syllable realizes a rightward-aligning unstarred _L+_ tone, which is lower than an immediately preceding tone. Next, the starred tone is specified to be _higher_ than the preceding tone, yielding _H*_. Moreover, the unaccented poststress syllable realizes a leftward-aligning unstarred _+H_ tone, which is higher than the preceding starred tone. Finally, if the _+H_ is realized phonetically as an F0 peak, there must be a following tone, starred or unstarred, which is _lower_ than the _+H_ tone.

At this point, we have demonstrated several significant results. First and foremost, we have demonstrated that tone interval theory provides a clean and phonetically well-founded definition of starredness in terms of the phonological association of a tone with a metrically prominent position. We have shown that by defining a very abstract notion of starredness, we can account for a wider range of tonal patterns in a more theoretically well-founded way than has been done previously. Our abstract definition of starredness permitted us to predict that when tones join into syntagmatic tone intervals, tone intervals can sometimes have the same value, giving the sequences _higher higher_ or _lower lower_. When this happens, the corresponding starred tone is predicted to be manifested phonetically not as an F0 peak or valley, but as a rise or a fall, usually with an accompanying "hook" or change of slope. Such is the case with certain intonational patterns in English, including the one defined in the paradigmatic theory by the sequence H+L* L-L%; here; the "L*" is realized by a fall. We saw that English also attests contours in which stressed syllables are realized by rises to a late peak in the vicinity of a following weak syllable. This led us to a proposal for the representation of Greek prenuclear accents, which are described here by a sequence $L_{(0)}$(+) _H* +H_. Finally, we proposed some restrictions on tone interval sequencing which we believe lead to adjacent tone intervals having the same value only in a very limited set of contexts. In this last regard, we will briefly review some results from the auditory perception literature which seem to justify our proposals.

We think that these restrictions stem from two kinds of general auditory perception constraints. First, a general rule of thumb in auditory perception is that events which are *different* from their contexts draw attention, while events which are *the same or similar* fail to draw attention (Handel 1989, Bregman 1990). The more different an event is from its context, the more likely that it will stand out perceptually; conversely, the more similar an event is to its context, the less likely it is to stand out perceptually. Second, positions associated with metrical prominence tend to draw attention, both because they have greater perceptual salience, as well as because humans have the ability to learn temporal patterns and to anticipate when salient events will occur (Jones and Boltz 1989, McAuley 1996, McAuley and Jones 2003).

These principles help to explain why sequences of identical tone intervals are in the limited distributions claimed earlier. Our first claim was that only starred tones, and not unstarred tones, may be realized phonetically as slope changes. Simply put, we do not think that the perceptual system can reliably detect slope changes in metrically nonprominent positions, so that the phonological system probably avoids the configurations that give rise to slope change points. Our second claim was that only *higher higher* or *lower lower* is permitted, but not *same same*. This can be explained by the fact that the slopes of contours realizing *same same* would always identically be zero. Thus, there will be no changes in direction to alert the perceptual system to the presence of a tone.[64] It therefore stands to reason that, for an intonation language at least, there can be no tones which participate in adjacent tone intervals which both specify *same*, since such a distinction could not be reliably differentiated from a case in which there was no tone and only phonetic interpolation. Our third claim, then, was that there can only be two identical relative height relations in sequence. Auditory perceptual principles provide insight on such restrictions, due to the fact that "similarity" and "dissimilarity" tend to be cumulative. Two events which are similar to one another, such as two rises, are heard as more similar with repetition. Because phonetic cues to the presence of a tone probably involve detecting local differences at the locus of the slope change, repetition of rises would increasingly undermine "difference" in favor of "similarity". Thus, we posit that the phonological system avoids configurations which would tend to override the very differences which are necessary for perceptual recovery of the phonological representation.

In sum, we have defined an abstract notion of starredness which permits us to provide a more robust account of the relation between phonology and phonetics. It also permits an account of a central outstanding problem in the intonation literature, concerning how to account for Greek prenuclear accents. Having discussed issues of accentuation at length, we now turn our attention to the topic of phrase-final phenomena in English.

5.4 Phrase-level phenomena in English

This section deals with a number of outstanding issues concerning the description of English intonation which relate in one way or another to intonational phrase-sized constituents. The first topic to be addressed concerns why the present theory does not assume the existence of any intonational constituents smaller than an intonational phrase; this is addressed in Section 5.4.1. Next, there are a number of phenomena which occur at the edges of intonational phrases in English, which we deal with largely in Section 5.4.2. Finally, Section 5.4.3 discusses the likelihood that universal principles of perceptual organization restrict sequences of tones, thereby preventing overgeneration in the theory. A two-part hypothesis is presented concerning the nature of tone sequencing in English and other languages. We turn now to these issues.

5.4.1 Phrasal constituents in tone interval theory

A number of examples also suggest that the phrasal structure of English intonation is simpler than relatively recent work has suggested. Prior to 1986, there was good agreement among descriptions of English intonation about the basic phrasal unit type. Different authors have used different terminology to name this unit; it has alternately been termed the intonational phrase (Pierrehumbert 1980, Liberman 1975), the tone unit (Crystal 1969), the sense group (Armstrong and Ward 1926, Vanderslice and Ladefoged 1972), the tone group (Ashby 1978, Halliday 1967), and the breath group (Lieberman 1967).

In contrast to this body of work, a smaller unit of intonational description has more recently been proposed, termed the *intermediate intonational phrase*, hereafter referred to as an intermediate IP (Beckman and Pierrehumbert 1986). Since its proposal, the intermediate IP has been adopted by many researchers seeking to describe intonation in English and other languages. The hallmark of an intermediate IP is the presence of a phrase accent at the right edge of a word that is final in the intermediate IP domain. Its phonetic characteristics are assumed to vary depending on the type of phrase accent and various phonetic factors (Pierrehumbert and Beckman 1988).

We could identify a number of examples which suggest that more generalizations are possible about tonal patterns through tone interval theory, which posits only one kind of unstarred tone. In contrast, permitting two different kinds of unstarred tones compromises our ability to generalize across contours, and F0 patterns become indeterminate with respect to their phonological descriptions. To see why, consider that the intonation pattern in Figure 5.2 which was shown earlier appears to be the same as the patterns in Figures 5.20 and 5.21. Each contour realizes two accents with low valleys; moreover, each contour shows a late-aligned F0 peak in between the two accents. The valley-peak-valley sequence is described unambiguously in each case as *L* H+ L**. Crucially, the phrases on which these contours are realized have the same (or similar) metrical structure. However, there is different placement of word boundaries. Our intuition is that the difference in word boundary placement does not affect the tonal pattern, which seems to be the same in each case.

In contrast, BP86 would suggest that the contours in Figures 5.2, 5.20, and 5.21 should be described differently. In particular, the high F0 peak on the last syllable of *womanly* in Figure 5.19 coincides with the word boundary, which makes it a candidate for the phrase accent of an intermediate IP. However, other descriptions are also possible. For example, this H tone might be described as the "+H" portion of a L*+H pitch accent. In contrast, the possibility of the F0 peak in Figures 5.2 and 5.21 being described as a phrase accent does not arise, because of the different distribution of the word boundaries. The main point is that the descriptive framework of BP86 tends to obscure similarities in contours and to introduce indeterminacy in the description of intonational patterns.

We could find other examples of the increased generalization that comes about by giving up the assumption that there is an intermediate IP constituent in English. For instance, consider the two contours shown earlier in Figures 5.9 and 5.10. We saw already that these two contours have a very similar intonational pattern and metrical structure. The similarity of their tonal patterns is readily explained in a tone interval description, which treats all of these contours as examples of the sequence H_0* +*L E*+ *H** L%. The contour in Figure 5.22 also appears to exhibit the same tonal pattern as the contours in Figures 5.9 and 5.10. Again, this similarity is readily captured in a tone interval description. Next, consider the corresponding mirror image contours. Two of these contours were shown earlier in Figures 5.11 and 5.12. The same basic intonational pattern is shown in Figure 5.23. Again, the tone interval framework captures the fundamental similarity across all these examples, describing each as a sequence L_0* +*H E*+ *L* H%.*

On the other hand, admitting unstarred phrase accent tones into the inventory would force us to treat the contour in Figure 5.22 differently from the contours in Figures 5.9 and 5.10. This is because the fall to a low pitch on *'I'* coincides with a word boundary in Figure 5.22 but not Figures 5.9 and 5.10 and thus it can be described by a L- phrase accent. Moreover, we would have to treat the contour in Figure 5.23 differently from the contours in Figures 5.41 and 5.40, since the rise to a high pitch on *'I'* coincides with the word boundary.

It has been claimed that auditory impressions can aid in judging whether an intermediate IP is present (Pierrehumbert and Beckman 1988). However, listeners trained in the ToBI system (Silverman *et al.* 1992) often have a very hard time determining whether such a boundary is being heard.[65] Moreover, a number of cases of apparent durational lengthening at the site of purported intermediate IP's seem to be explainable by appealing to the rhythmic properties of metrical grids. This makes sense, because factors involving lengthening must necessarily derive from the component of the linguistic grammar which deals with time. In practice, decisions about whether a phrase accent occurs at a particular word boundary often hinge on vague factors, such as whether the candidate boundary coincides with a syntactic phrase boundary of some sort. In the examples above, we saw that permitting the intermediate IP as a descriptive phrasal unit in English tends to obscure the striking similarity of these three examples. Moreover, it appears that allowing this unit in descriptions of English results in a significant level of indeterminacy in the phonological description. In contrast, the tone interval framework reveals the similarity across these three examples while eliminating the indeterminacy associated with the intermediate IP analysis. For these reasons, we conclude that the intermediate IP is not a necessary component of the description of English intonation.

5.4.2 Some remarks on phrase-final phenomena in English

There are a certain number of phenomena which occur at the right edges of intonational phrases which any description of English intonation should be able to account for. Of particular importance is the question of how to characterize the types and sequencing restrictions of phrase-final tones. In previous paradigmatic descriptions of English, every intonational phrase was assumed to end with a sequence of a single phrase accent for the intermediate IP embedded in the intonational phrase constituent, and a single boundary tone. However, we have already seen at least two difficulties with this assumption. On the one hand, a number of cases were cited in which this claim led to a lack of phonetic transparency. On the other hand, if we take evidence of F0 turning points following a nuclear accent as *prima facie* evidence of a tone, then there is good evidence that more than two tones can follow the nuclear accent (Gussenhoven 1999; Grice, Ladd, and Arvaniti 2000).

We propose that between one and three "phrase-final" tones can follow the last (nuclear) accent in English intonational phrases. For example, consider the utterance *Anna!?* in Figure 5.24, which attests three tones; this is notated as *L* +H L H%*. The contour in Figure 5.25 shows that all of these tones can even be fit onto a single syllable, as in *Anne!?* Two phrase-final tones are seen in examples like Figure 5.26. Finally, one phrase-final tone occurs in the majority of the examples in this chapter.

These tones do not have any special status in the theory; they are simply located at a special position in the intonational phrase. Nevertheless, there are a number of additional phenomena associated with the ends of intonational phrases which we have yet to explain. The insight here is that we do not need to posit special tonal categories like "phrase accent" and "boundary tone" in order to account for such phenomena. Rather, we will see that phenomena related to the edges of phrases can be accounted for in terms of the special properties of tone intervals. We will return to these topics in Section 5.4.3.

One important contour which has not yet been discussed is the calling contour (Liberman 1975, Ladd 1980, Gussenhoven 1984). This contour is interesting, because it represents a case in English in which two meanings seem to be differentiated on the basis of the size of the interval. Indeed, the calling contour seems to involve a small interval of around 3 s.t. This suggests that it can be accounted for in terms of a syntagmatic tone interval with a size which is restricted to about 3 s.t. This can be represented as a special tone interval category, $1 > I > \delta_{-}$, where δ_{-} represents a cutoff value that is determined empirically. We will use the symbol $^{!}T$ to represent a tone which steps down by a restricted amount. The contour is shown in Figure 5.27, together with the description *H* +E* $^{!}T$ *E%*.

In support of the claim that there is a restriction on the interval size for the calling contour, consider the fact that a meaning difference appears to result when the excursion size is much larger than 3 s.t. For example, the contour in Figure 5.28 has the same shape as the one in Figure 5.27: both contours show a level region, followed by a step down. However, the size of the step down is much larger in Figure 5.28 than in Figure 5.27. The contour in Figure 5.28 seems to convey either dissatisfaction with a person being addressed, or possibly a mild reprimand. We will refer to the contour in Figure 5.28 as the *scolding contour* to distinguish it from the calling contour.[66] This pair of examples indicates that differences in excursion size previously attributed to distinct paradigmatic primitives can be re-expressed in terms of restrictions on the sizes of syntagmatic tone intervals.

At this point we will move on to consider another topic related to phrasal edges, which will require a digression into the nature of the primitives in this system. In this theory, we distinguish only starred and unstarred tones. These two types of tones have different distributional properties with respect to the metrical structure: starred tones associate with metrically prominent positions, while unstarred tones associate with metrically nonprominent positions.

In recent theories, starred tones have been equated with the notion of pitch accent. However, the present theory conceptually distinguishes between the phonological properties of starred tones, and their phonetic effects. Phonologically, starred tones are restricted to mark prominent positions in the metrical structure, by which we mean grid positions of height 2 or higher. Phonetically, starred tones generally create an impression of increased prominence or accentuation on a metrically prominent syllable.

Distinguishing between the phonological properties of starred tones and their associated phonetic and perceptual effects has two major implications for the theory. First, it increases the phonetic transparency of the theory by allowing us to treat F0 turning points which occur on metrically prominent syllables consistently as starred tones, even when they occur near the ends of phrases. This permits a straightforward account for the fact that F0 turning points at the right edges of phrases are attracted to post-nuclear secondary stressed syllables (Grice, Ladd, and Arvaniti 2000; Lickley, Schepman, and Ladd, forthcoming). A less desirable theoretic alternative is to treat F0 turning points near the ends of phrases as unstarred tones, and then to posit a special rule or category to explain why some unstarred tones align with metrically prominent positions, while others do not. A second significant implication of separating the alignment properties of starred tones from their accentual effects phonetically is that the theory is then free to ask questions regarding why certain tonal movements on stressed syllables generate the perception of pitch accent, while others do not.

To see why it might be useful to separate the notions of "starredness" and "accentuation", consider the contour shown in *Abercrombie* in Figure 5.29. This utterance would have been described as L*+H L-H% in the paradigmatic theory; in this system it is similarly described as $L_0*\ {+H}\ L*\ H\%$. Notably, this example shows that the secondary stress syllable realizes a significant local pitch drop which in many other contexts might be considered to be accent-lending. It turns out that in English, a downward pitch excursion is permitted on the secondary stressed syllable, while an upward pitch excursion in the same position apparently is not. Indeed, if there is a rise to a high pitch on the secondary stress syllable, then the resulting contour sounds to English listeners as if the word is being mispronounced as having main stress in the location of secondary stress. Why is the high pitch peak on the secondary stress syllable illegal, while the low pitch valley in this same position in $L_0*\ {+H}\ L*\ H\%$ is acceptable, particularly when either can cue the perception of accent in other contexts? An interesting clue comes from the observation that while this contour is ill-formed in English, it appears to be perfectly acceptable in Swedish and Italian (Ladd 1996). Separating the notions of accent and starredness thus lead us to the insight that English probably has a language-specific restriction on $H*\ L$ sequences, but not $L*\ H$ sequences, in phrase-final position: a $H*$ tone followed by L in phrase-final position is interpreted as an accent. Viewed in these terms, we can begin to reformulate the observation of Pierrehumbert (1980) and others that there are "no pitch accents after the nuclear accent" in terms of restrictions on sequences of tones and tone intervals occurring after the last starred tone which is intended as an accent.[67] Indeed, the question of sequencing restrictions in general is an important one and brings us to the topic of the next section.

5.4.3 Toward an account of tone interval sequencing restrictions

We have already seen that tone interval theoretic constructs can account for phenomena in English intonation in a more phonetically transparent way than previous theories using a simpler set of primitives. The obvious challenge for such a theory is how to characterize restrictions on tone sequencing in order to prevent the theory from overgenerating possible contours. Rather than proposing a definitive grammar which restricts such sequences, we will instead propose a general hypothesis concerning the nature of tone interval sequencing restrictions. If correct, this hypothesis could potentially aid in explaining a host of other phenomena in intonation, including final lowering (Liberman and Pierrehumbert 1984, Truckenbrodt, forthcoming) and intonational dependency (e.g., Crystal 1969; Ladd 1986, 1996).

Previously, restrictions on tone sequencing were to some degree implicit in the set of complex primitives that were proposed to account for English intonational patterns. For example, a L*+H pitch accent entailed a temporal sequencing restriction such that "L*" was necessarily followed by "+H". However, even if the paradigmatic theory's input-output relations operated as they have been assumed to in descriptive linguistics, it is clear that the restrictions on tone sequencing implicit in bitonal pitch accent relations did not adequately capture the sequencing restrictions which are observed in natural language. Pierrehumbert was aware of the inadequacy of sequencing restrictions in the paradigmatic theory and noted in a (2000) paper that not all the possible sequences in P80 and BP86 are attested. In particular, she noted that based on the grammar proposed in BP86, a phrase with two accents generates 36 accentual combinations, while a phrase with three accents generates 216 combinations. She remarks (2000: 27):

> ...nothing like the full set [of accentual combinations] generated by the grammar has ever been documented. For three accent phrases, the typical pattern is either to use the same accent type in all three positions, or else to use one type of accent in both prenuclear positions and a different type in nuclear position.

Indeed, there have been many authors who have noted that accents tend to be repeated in sequence (e.g., Palmer 1922; Schubiger 1956; Kingdon 1958; Trim 1959; Schachter 1961; O'Connor and Arnold 1973; Crystal and Quirk 1964; Crystal 1969; Bolinger 1965, 1986; Gibbon 1976; Ladd 1986, 1996; Pierrehumbert 2000). Ladd (1996) described this repetition as *intonational dependency*. The fact that pitch patterns tend to repeat in sequence suggests that there may be a set of general restrictions in place which prohibit certain logically possible tonal sequences from being realized. In spite of the fact that such repetition has been described by many authors for English intonation over the years, the issue of why tones tend to repeat in sequence has rarely been the object of empirical study, with a notable exception being Crystal (1969).

Why might tonal sequences exhibit a repeating pattern? We do not need to look very far outside of linguistics for a possible explanation for the existence of tone sequencing restrictions in language. A large body of work in auditory perception has shown that humans and other animals are extremely well attuned to patterns in the environment (Jones 1976, Handel 1989, Bregman 1990). Auditory patterns have been shown to have two distinct kinds of effects for listeners perceptually. First, when tone sequences repeat, listeners infer metrical patterns. That is, tones which form repeating sequences are heard as being alternately strong or weak, and they are heard as being grouped in particular ways (Bolton 1894; Woodrow 1909, 1911, 1951; Handel 1989). Second, patterns of regularity cause listeners to develop expectations that the sequence will continue. If the pattern changes, a listener's attention is drawn to the temporal location of the change (Jones 1976, 1986; Jones and Boltz 1989).

The implication of this work is that humans have an innate ability to produce and perceive patterns, which we believe surfaces in the kinds of tonal sequences that are observed in natural language. Such patterns probably have two primary types of effects in language. First, listeners will naturally tend to interpret repeating tone sequences as having particular metrical patterns. Second, a listener's attention will be drawn to locations in utterances where there is a change in a tonal pattern.

These observations, taken together, lead us to a two-part hypothesis about the nature of universal tone sequencing restrictions across languages. First, we propose that tones tend to be ordered into regular, repeating patterns, which have been shown to evoke particular metrical structures for listeners; we then predict that these evoked metrical structures will be congruent with the metrical structures of the words and phrases being produced. Second, we propose that speakers naturally produce sequences of tones which form repeated patterns, such as *HLHLHL*.... This has the effect of entraining a listener's attention, so that the expected pattern can be violated precisely at locations of salient information in the speech stream, such as a new word in the discourse, a major syntactic boundary, an intonational phrase edge, etc.

There are a number of intonational phenomena that have been reported which are consistent with this two-part hypothesis. We will briefly review three such phenomena below. We think that the fact that these phenomena are generally consistent with work in auditory perception and music cognition supports our proposal that perceptual proclivities effectively restrict the kinds of tonal sequences which are produced. Moreover, the consistency of these intonational phenomena with known effects reported in the auditory literature suggests the fruitfulness of focusing on sequencing restrictions as a serious topic for further research. We turn now to a brief description of these three intonational phenomena.

Accentual sequencing. As noted earlier, many linguists have reported that intonational patterns tend to be repeated in sequence. This is exemplified in the fact that the Greek prenuclear accents appear often to take the same form, as distinct from nuclear accents in this language (Arvaniti and Ladd 1995; Arvaniti *et al.* 1998, 2000). The auditory perception literature suggests that repeating tonal patterns evoke particular metrical structures involving grouping and prominence relations; it remains to be seen whether the kinds of tonal patterns reported for speech evoke the perception of similar kinds of metrical structures as have been reported for tone sequences. Moreover, there have been no studies examining whether those natural metrical properties associated with tonal sequences will show congruence to the metrical properties of texts. In general, we would predict that the natural metrical properties associated with tone sequences should match the metrical properties of texts, rather than conflict with them.

Pitch excursions at phrasal edges. It has been noted that the edges of intonational phrases are often associated with very high or very low pitches (Ladd 1980, Pierrehumbert 1980, Bolinger 1986). This has sometimes been accounted for by positing paradigmatic primitives which were scaled very high or very low in the pitch range, e.g., the H% and L% boundary tones in the paradigmatic theory. However, it is well-known that listeners are attuned to frequency range characteristics, so that they develop implicit expectations about the frequency range of upcoming events (Handel 1989, Bregman 1990). When incoming events have a much higher or lower pitch than expected, listeners' attention is drawn to that location in a sequence. It makes sense that speakers would make use of the natural pairing of frequency range and attention, in order to draw listeners' attention to significant parts of utterances, such as phrasal boundaries.

Final lowering. Liberman and Pierrehumbert (1984) reported that for simple utterances produced with a "downstepping," list-style intonation, the amount of lowering on each non-final accent was suggestive of an exponential function. Based on an extrapolation of this function, they observed that the last accent in the sequence had a lower pitch than expected; this was termed *final lowering*. A number of reports of this phenomenon can be found in the literature (Prieto, Shih, and Nibert 1996; Laniran 1992; Truckenbrodt, forthcoming). We would explain this phenomenon based on auditory perceptual principles. As mentioned earlier, it is well-known that listeners are extremely good at detecting patterning in tonal sequences. In particular, a repeating pattern causes listeners' attention to be entrained; when the pattern changes, a listener's attention is drawn to the location of the change. These observations can help us to understand the nature of final lowering. The effect of producing a repeating pattern of steps down by about the same amount followed by a larger step down than expected is to draw attention to the accent which violated the pattern. Because the accent violating the pattern is generally last, the listener is thereby alerted to the impending finality of the phrase. Our hypothesis also helps to explain why the phenomenon here is *final* lowering, rather than *initial* lowering. The change must come at the end, because the repeating pattern of steps down has to be created in the first place in order for a listener to generate an expectation about how the trend will continue.

These observations concerning apparent connections between phenomena in intonation and phenomena in general audition have led us to the hypothesis that the principles which restrict tonal sequences in language are both language-universal and language-specific in nature. The language-universal restrictions in our view stem from facts about the innate perceptual organization of tone sequences by humans and other animals. We were able to connect these principles of perceptual organization in a preliminary way to three well-known phenomena in intonation: accentual sequencing, pitch excursions at phrasal edges, and final lowering. The probable connection between these phenomena and deeper principles of auditory processing presents a compelling argument for not putting forth a complete "grammar of intonation" at this early stage in our theorizing. Crystal (1969) also noted the probable significance of tone sequencing for understanding language-universal principles. In an extended passage, he expresses a very similar view to the one taken here, namely, that proposing a tonal grammar would not only be premature, it would also mark a missed opportunity for the linguistic community to understand similarities and differences among languages. He writes (1969:240):

> What are the recurrent patterns of tone-unit sequence in connected speech? Until one can give some answer to this question phonologically, one lacks the ability to assess the degree of prosodic 'uniqueness' of any given grammatical structure, and thus loses a great deal of the predictive power of the descriptive statements about co-occurrence which might be made. Phonological sequences of a fairly restricted type do exist independently of grammar, both within major grammatical structures and, less frequently, between structures, and one ought to be aware of the more important tendencies at work here before embarking upon any process of grammatical integration.

Following Crystal's suggestion, we will therefore leave the grammar of English intonation as a matter to be dealt with at some point in the future, once more is known about sequencing restrictions more generally across languages.

5.5 Long-distance syntagmatic relations in English

In this section we discuss some findings originally reported by P80 and Liberman and Pierrehumbert (1984), hereafter LP84. These results show that speakers scale the relative heights of accented tones across intonational phrase boundaries in English. Long-distance scaling effects of relative height are precisely the kinds of phenomena that are predicted by the mechanisms of the tone interval grid-matrix complex derived in Chapter 3. Moreover, the quantitative form of this data supports the claim that tones are scaled according to ratios of F0 values. The experiment and results are reviewed in Section 5.5.1. Moreover, we will briefly discuss P80's and LP84's models and some other models which have been proposed. Finally, in Section 5.5.2 we present a phonological account using the mechanisms of the tone interval grid-matrix complex of Chapter 3.

5.5.1 Quantitative evidence for tone intervals

Evidence for tone interval representations comes from a phonetic study reported by P80, which was subsequently expanded and reported in full in LP84. This experiment investigated the relative levels of accents in a short utterance which speakers reproduced in different parts of their pitch ranges. The results for four speakers conclusively showed that speakers varied the levels of accents relative to one another approximately according to a specific ratio, precisely as predicted by tone interval theory. In the following we review some of the details of the study.

In these experiments, speakers produced the utterance *Anna came with Manny* with one of two intonation patterns, which are shown in Figures 5.30 and 5.31. P80 and LP84 report that the intonation pattern described in Figure 5.30 constitutes an appropriate answer to the question *What about Manny? Who came with him?*. This pattern will be referred to as the "AB" contour. Moreover, the contour in Figure 5.31 reportedly corresponds to an appropriate answer to the question *Who did Anna come with?* This pattern will be referred to as the "BA" contour. These two intonation patterns are assumed in the literature to equate with narrow focus on either *Anna* or on *Manny*, respectively.[68] Of particular interest with respect to these two contours is the relative height relationship among the accents. In particular, the contour in Figure 5.30 shows that the accent on *Manny* is lower in height than the accent on *Anna*, while the contour in Figure 5.31 shows that the accent on *Manny* is at about the same height as the accent on *Anna*.

The two speakers who were not the authors were instructed via imitation how to produce the desired intonation pattern. Each speaker then produced the pattern with different levels of emphasis ranging on a scale from 1-10. The results showed that as speakers varied their overall F0 levels, they maintained an extremely consistent relation of relative height between the two accents, as shown in Figure 5.32. In this figure, the F0 level of the first accent is shown on the x-axis, while the F0 level of the second accent is shown on the y-axis. Imitations produced using the "AB" intonation pattern are indicated by open triangles, while imitations produced using the "BA" intonation pattern are shown with open circles.

A significant observation concerns the precise manner by which subjects scaled the accents relative to one another. In the case of both intonation patterns, the relation between the two accents for the most part appears to be described not just by a straight line, but by a *consistent ratio of F0 values*. In other words, subjects' imitated versions generally fit a special form of a linear equation in which the y-intercept is zero. In such cases, the slope of the line is given directly by the ratio of the x and y values. This is precisely the form of a tone interval relation, thereby suggesting not only qualitative, but quantitative support for these phonological constructs.[69]

Little remark was made in discussions of these results in P80 or LP84 regarding the consistency of tone scaling demonstrated by the speakers. Due to predetermined theoretical ideas about how the two accents were related, both P80 and LP84 chose to model these data in a rather complex way. We note that the data shown in Figure 5.32 can be described in terms of a simple and straightforward linear relation – indeed, in terms of a simple ratio, y/x. However, P80 described the data in terms of four equations and at least six parameters, while LP84 modeled these data in terms of a nonlinear, exponential relation with 11 parameters. Not only did these analyses obscure the simplicity and systematicity of the relation between the accents, but the large number of parameters and relative freedom of their values eliminated any predictive power which accounts based on a more restricted set of relations would possess. In contrast, the tone interval theory not only predicts this consistent relation, but also explains it in terms of consistent control of frequency ratios arising from general auditory perceptual principles.

In general, a significant drawback of any account of data like that in Figure 5.32 in terms of phonetic scaling rules is that it tends to "explain away" the striking and consistent effects observed in the experiment. This point has previously been made by Ladd (1990, 1993, 1996), who has argued that systematicity of the kind seen in P80's and LP84's experiments requires a phonological account. Consistent control of relative height relations has also been demonstrated in English by Ladd (1988), who showed systematic long-distance effects in the scaling of relative heights of accented syllables according to the syntactic structure of utterances. Long-distance effects in the scaling of relative heights of nonadjacent accents have also been shown in other languages, including Swedish and Japanese (Bruce 1982, Pierrehumbert and Beckman 1988).

A number of phonological models have been proposed for data of this kind in which the relative heights of tones are scaled with respect to phrasal reference lines (van den Berg, Gussenhoven and Rietveld 1992; Ladd 1988; Truckenbrodt 2002). There are three drawbacks of such models. First, these models are based at least in part on the assumptions of earlier theories of the relation between phonology and phonetics put forward by P80 and PB88. It was demonstrated in Chapter 4 that the lack of relative height restrictions on adjacent tone pairs in sequence in P80 and PB88 causes these models to be technically unable to explain even simple contours. As a consequence, models which are based even partly on the assumptions of P80 and PB88 inadvertently suffer from some of the same drawbacks. Second, models based on the scaling of phrasal reference lines assume that the scaling effects will be distributed throughout the entire phrase. However, Ladd (1993) acknowledges that scaling effects seem to be largely limited to individual accents, rather than entire phrases. Third, and finally, it was shown in Chapter 4 that in the limit, any model based on paradigmatic tonal primitives plus phonetic scaling rules provides a more complex account of tonal phenomena, including the scaling of tones across long distances, than an account based on syntagmatic primitives.

Tone interval theory both predicts and explains the results obtained by P80 and LP84. The consistent maintenance of relative height relations demonstrated in these experiments is attributed to the fact that the heights of accents are controlled according to tone intervals, which are abstractions of frequency ratios. A large body of work on general auditory perception has shown that listeners encode the relative heights of tones, even when those tones are not temporally adjacent (Handel 1989, Bregman 1990). The fact that data in P80 and LP84 were scaled along straight lines with slope zero effectively proves that the two tones in this experiment were scaled according to a ratio, just as predicted by tone interval theory. The phonological mechanisms by which this long-distance scaling took place were sketched in Chapter 3. In the following section, we provide a phonological account of this data within a tone interval framework.

5.5.2 Re-analysis of Pierrehumbert (1980) and Liberman and Pierrehumbert (1984)

In this section we will show how the descriptive mechanisms developed in Chapter 3 concerning the interactions between tones and the metrical grid permits us to account for P80's and LP84's findings in a simpler way than they proposed. In the following we will first describe the tonal contours used in P80's and LP84's experiment in terms of tone interval theoretic constructs. It will then be shown that this description alone accounts for the data that were produced.

Inspection of the contours in Figures 5.30 and 5.31 which subjects in P80's and LP84's experiment reportedly produced reveal a number of differences in the shapes of the two contours. We note that because the target intonation patterns cannot be listened to, it is impossible to provide a fully reliable description in the tone interval framework, which ultimately assigns a phonological representation on the basis of perception of the contour. However, due to the close correspondence between F0 and perceived pitch (Moore 1997), we can infer that the description of the "AB" contour in Figure 5.30 is H_0* *L% H_0* L+ H* +L H%* in *L/H/E* notation, while the description of the "BA" contour in Figure 5.31 is H_0* *+L H% E_0* E+ H* L%*.

The *L/H/E* notation only captures part of a tone interval description, however. Recall from Chapter 3 that tones associate with columns of X's in metrical grids; relative height relations between tones which come into adjacent positions on higher grid levels are then represented in a coordinate structure called the tone interval matrix. *L/H/E* notation presents a way of capturing relative height relations on the *lowest* level of the tone interval matrix. However, the full tone interval matrix, together with the metrical grid, presents a way of representing hierarchical tonal relations among tones occupying metrically prominent timing positions as well.

In order to capture the relative height relations shared by the accents in the "AB" and "BA" contours in Figures 5.30 and 5.31 through a tone interval matrix, we must use standard tone interval notation as introduced in Chapter 3. The first step is to convert the *L/H/E* representations for each contour given above into standard tone interval notation, which will yield a sequence of tones and tone intervals corresponding to the lowest level of the tone interval matrix. This is shown for the "AB" and "BA" contours in Figures 5.33 and 5.34. The sequence of tone intervals shown captures the relative height relations among the tones giving rise to these contours. For example, $I_{1,2} < 1$ indicates that T_2 is lower than T_1, while $I_{2,3} > 1$ indicates that T_3 is higher than T_2, etc.

Crucially, the full tone interval matrix permits us to represent the relative heights of nonadjacent metrically prominent positions in the phonology. It is this aspect of the representation which permits a capture of P80's and LP84's data. The Syntagmatic Adjacency Restriction of Chapter 3 described the fact that tones which are adjacent on any row of the metrical grid form syntagmatic tone intervals. In the metrical grids shown in Figures 5.33 and 5.34, there are three metrically prominent syllables: *An-* in *Anna*, *came*, and *Man-* in *Manny*. The starred tones on these syllables form tone intervals on higher rows of the tone interval matrix. Significantly, the tones are associated with the stressed syllables of *Anna* and *Manny* are adjacent on row three of the metrical grid.

Entries on the third row of the tone interval matrix in particular capture the data in P80's and LP84's experiments. We will consider how the tone interval representation describes the "AB" data first. In Figure 5.35 the entry on the third row of the tone interval matrix indicates that T_1* and T_5* are joined into a tone interval given by the relation $1 < I_{1,5} < \delta_{-}$. This indicates that T_5* is lower than T_1* by no more than a small interval, δ_{-}. Comparing this result with the production data elicited in the experiment for this condition in Figure 5.32, which is given by the open triangles, we see that the points lie just below the line $y = x$, indicating that T_5* was consistently lower than T_1* by a small amount. The phonological representation therefore predicts the form of the phonetic data that the speakers produced rather well.

Next we will consider how the tone interval representation describes the "BA" data. In Figure 5.36 the entry on the third row of the tone interval matrix indicates that T_1* and T_6* are joined into a relation $I_{1,6} = 1$. This indicates that T_5* is phonologically at the same level as T_1*. Comparing this result with the production data from the experiment for the "BA" condition in Figure 5.32, as given by the open circles, we see that the points lie approximately on the line $y = x$, indicating that T_6* was consistently at about the same level as T_1*. Again, we see that the phonological representation predicts the form of the phonetic data that the speakers produced.

In this way, tone interval theory is able to provide a phonological account for the striking consistency in production data seen in the results of P80 and LP84. Our account offers two primary advantages over approaches based on phrasal reference lines plus phonetic scaling rules (van den Berg, Gussenhoven and Rietveld 1992; Ladd 1988; Truckenbrodt 2002). The first concerns the fact that such an account better fits with extant phonetic data on how tones are scaled. For one thing, tone intervals represent ratios, and in this sense they quantitatively capture the form of the relation between the accents on *Anna* and *Manny* seen in Figure 5.32. In contrast, no other theories predict this level of phonetic explicitness. For another thing, the tone interval account predicts that scaling effects will be largely restricted to the accented syllables themselves, rather than an entire phrase. Phonetic data in support of this prediction was shown by Ladd (1988), who carried out a production experiment in which speakers read sentences with syntactic forms like *A and B but C* vs. *A but B and C*. Ladd's data confirmed in English that accents are scaled with respect to each other, and that their scaling is influenced by the syntactic structure. However, this data also showed that the locus of tone scaling was restricted almost entirely to the accents themselves, and not, as accounts based on phrasal reference lines would predict, based on the scaling of entire phrases. The second major advantage of the tone interval account over other accounts is that it is simpler than other explanations, as we showed in Chapter 4.

In sum, tone interval theory provides a phonetically explicit, phonological account of some striking production data reported by P80 and LP84. It does so in a way which is both simpler and better in line with the balance of phonetic data than other proposals. In the following we summarize the major points of this chapter.

5.6 Summary

This chapter demonstrated that tone interval theory builds on previous linguistic work by providing a solid theoretical underpinning for descriptive work carried out within the paradigmatic framework, which was previously lacking. It was demonstrated that tone interval theory increases the phonetic transparency of the description of English intonation compared to the paradigmatic description, even when problems discussed in Chapter 4 with the phonology-phonetics relation of that theory were neglected. Moreover, we showed through examples that F0 turning points (peaks, valleys, and corners) are consistently treated as tones in this theory. In contrast, such points are sometimes treated as tones in the paradigmatic theory and sometimes not.

We also described how tone interval theory presents an account for a number of outstanding problems in the literature. First, tone interval theory accounts for data from Ladd and Schepman (2003) by describing all F0 "dips" as consistently arising from *L* tones. Second, the theory presents an account for findings that the two tones in contours described as bitonal accents by P80 do not align with respect to one another by a fixed temporal interval (Arvaniti *et al.* 1998; Ladd *et al.* 1999, 2000; Dilley *et al.* forthcoming). Our account proposes that the two tones described as bitonal accents in the paradigmatic theory of English instead are independent tonal entities which align in adjacent metrical positions. The independence of these tones was supported by examples showing that a given unstarred tone can align either to the left or to the right of a starred tone. Third, we proposed that the number of intonation phrase-final tones varies from one to three. This permits a simpler and more phonetically transparent account of contours in English, as well as the data of Grice *et al.* (2000). We argued that the smallest intonational constituent was the intonational phrase, and that the evidence presented by BP86 to justify smaller constituents can probably be explained in terms of two other devices which are universally part of intonational phonology: the metrical grid, and tone interval sequencing restrictions. Fourth, we presented an abstract definition of starredness which accounts for the failure of accents in English and Greek to show temporal alignment with F0 peaks and valleys. This treatment clarifies the phonology-phonetics relation further by defining a theoretical basis by which changes in the slope of an F0 contour may be interpreted as evidence of an underlying starred tone.

To account for phrase-final phenomena in this theory, we proposed a two-part hypothesis about the organization of tone sequences universally. First, we predicted that tones will tend to be ordered into regular patterns of alternation, which have been shown to evoke particular metrical structures for listeners; we predict that these evoked metrical structures will be congruent with the metrical structures of the words being produced. Second, we proposed that speakers naturally produce sequences of tones which form repeated patterns, such as *HLHLHL*, where this regularity has the effect of entraining a listener's attention. We also suggested that speakers will change a pattern at locations of salient information in the speech stream, such as at a new word in the discourse, at an intonational phrase edge, etc., and that this has the effect of drawing a listener's attention to informationally significant positions in a phrase.

Finally, we presented quantitative evidence from an experiment by P80 and LP84 for the existence of tone intervals in English. This experiment demonstrated that speakers scale the relative heights of nonadjacent tones on accented syllables according to a ratio. This provides not only qualitative, but also quantitative support for tone interval theory. Finally, we showed that tone interval theory readily provides an account of these findings; the account makes use of the metrical grid-matrix complex developed in Chapter 3. We noted that this proposal is better in line with phonetic data and simpler than earlier proposals. In the remaining chapters, we present the results of several experiments on English intonation testing the claims and predictions of tone interval theory against those of the paradigmatic theory and other theoretical views.

Chapter 6 – Phonetic correlates of phonological categories I: F0 extremum alignment

Chapter 4 reviewed evidence from a number of phonetic studies which have demonstrated the relevance of F0 alignment to the phonological representation of intonation cross-linguistically. In the present chapter, four experiments are described in which the temporal alignment of F0 extrema (maxima and minima) is systematically manipulated. The experiments test the claims of the theory of Pierrehumbert and colleagues (Pierrehumbert 1980, Beckman and Pierrehumbert 1986, Pierrehumbert and Beckman 1988) regarding the phonological categories of English. This theory will be referred to in this chapter and Chapter 7 as the *paradigmatic theory*. The results will be interpreted with respect to the paradigmatic theory and the syntagmatic theory of English intonation proposed here.

There were several reasons for focusing on F0 alignment in our experiments. First, the temporal alignment of an F0 maximum or minimum has been shown to be a relevant phonetic dimension for making phonological distinctions in a number of languages, including Swedish, Japanese, Italian, and German (Bruce 1977; Kohler 1987; House 1990; d'Imperio, 2000). However, very little work has been done to investigate the role of F0 alignment for phonological distinctions in English (but see Pierrehumbert and Steele 1989).[70] Second, the paradigmatic theory makes a number of claims about the mapping between F0 contour shapes and phonological categories which have largely not been tested. Focusing on F0 alignment provides a way of testing some of these claims. Finally, previous experimental studies have focused on F0 maximum alignment, while little work has been done to investigate F0 minimum alignment, in spite of the fact that consistency in F0 minima has been observed in production data, suggesting that this dimension, too, may give rise to categorical effects. The present study considers the role of alignment of both types of extrema in eliciting phonological contrast.

We will begin by considering some methods in Section 6.1 for testing phonological categories in intonation. In this section, we will consider the issue of what constitutes evidence for a phonological category. In Section 6.2 we review some predictions of the paradigmatic theory and the syntagmatic view. We will then turn in Sections 6.3 and 6.4 to the experiments.

6.1 Methods for investigating intonational categories

What methods are available to the investigator for studying the categories underlying intonation? In general, studying the representation of suprasegmental aspects of language is challenging, partly because of the difficulty of defining intonational meaning. (See Liberman 1975, Ladd 1980, Bolinger 1989 for discussions.) As a result, using meaning to study intonational categories usually meets with mixed success. Nevertheless, some investigations of intonational categories based on semantic inference have been attempted (e.g., Bolinger 1961, Nash and Mulac 1980).

The task of studying intonational categories has been made somewhat easier by the discovery that certain techniques for studying segmental categories can apparently be applied to studying intonational categories (e.g., Kohler 1987, House 1990, Grice and Savino 1995, D'Imperio 2000). In the following, we review some findings relevant to the study of segmental categories. We then explain how these findings give rise to techniques which can be applied to the study of intonational categories.

The study of segmental categories first generated insight into a phenomenon known as *categorical perception* (Liberman et al., 1957; see Repp 1984 for a review). This phenomenon is now well-known, and it has been demonstrated for both auditory and visual modalities using a wide range of stimulus types. The standard method for testing categorical perception in auditory experiments involves presenting subjects with stimuli which lie along an acoustic continuum. Even though all stimuli along the continuum are different from one another by objectively the same amount, listeners often do not perceive all steps along the continuum as differing by the same amount. Rather, subjects' perceptions reflect their learned categories and linguistic experience. In one standard task, called the "AX paradigm," subjects listen to pairs of stimuli which are objectively the same or different. They respond "different" when they hear two stimuli as being different and "same" when they hear two stimuli as being the same. (Pairs of stimuli which are the same simply serve as foils, keeping subjects attention.) When categorical perception obtains, listeners are not equally good at discriminating among all pairs which are objectively different by the same amount. Instead, they will hear stimulus pairs which straddle a linguistic category boundary as being *more different* than stimulus pairs which fall within a linguistic category boundary. As a result, their discrimination performance will be better for stimulus pairs which straddle a linguistic category than for stimulus pairs which fall within a linguistic category. This differential performance in the AX task is then an index of underlying linguistic category boundaries.

A number of experimenters have carried out discrimination tasks involving stimuli in which an F0 target, such as an F0 maximum, is shifted along a continuum (Kohler 1987; House 1990; Grice and Savino 1995; D'Imperio and House 1997; Ladd and Morton 1997; Remijsen and van Heuven 1999). Many of these experiments have shown that there is differential discrimination along intonational continua. However, the discrimination peaks elicited in these studies are broad, similar to those observed for vowel stimuli.[71] Recent work has shown that whether a broad discrimination peak is observed or not may depend on linguistic experience (Hallé, Chang, and Best, 2004). These experiments show that discrimination tasks can be useful for studying the categories which underlie intonation.

Moreover, categorical effects have also been observed in production tasks using intonational continua (Pierrehumbert and Steele 1989; Gussenhoven and Rietveld, 2000). In Pierrehumbert and Steele's study, listeners imitated stimuli in which the timing of an F0 maximum was shifted in equal steps. The resulting productions showed a discrete distribution for F0 maximum times, rather than a continuous distribution, reflecting the influence of categories. It is interesting to note that in this regard, a further analogy is possible between suprasegmental and segmental domains. That is, categorical effects in production have also been observed when listeners are instructed to imitate vowel continua (Viechnicki 2002).

The finding that categorical effects in perception and production obtain for intonational continua is important for several reasons. First, it suggests that intonation is comprised of categories which are analogous at some level to segmental categories. Second, it suggests that certain techniques developed for studying segmental categories can be applied to the study of intonational categories. Third, findings showing categorical effects in perception and production set a precedent for defining a working, operational definition which can be used in intonation experiments to define what is meant by a "phonological category". A phonological category is therefore defined here in terms of a distinction *for which there is evidence from both perception as well as production data*. This definition is implicit in the work of Pierrehumbert and Steele (1989) who used a production paradigm to provide support for the distinction between L+H* and L*+H pitch accents.[72] This dual criterion is adopted here; that is, we approached the issue of investigating intonational categories in this chapter by searching for categorical effects in *both* perception and production data in response to a stimulus continuum. This permitted us to test the predictions of the paradigmatic theory regarding its claims about phonological categories in English.

6.2 Theoretical predictions

Having set forth an operational definition of a "phonological category", we can turn to the issue of predictions made by different theories. First, we will consider the predictions of the paradigmatic theory regarding the relation between phonology and phonetics, as well as the phonological categories of English intonation. We will then turn to the predictions of the syntagmatic view on these points.

An important issue is that the paradigmatic theory does not define a clear relation between phonology and phonetics, as discussed in Chapter 4. In order to test the theory, certain assumptions had to be made about the phonetic shapes which could arise from a particular phonological sequence. In this regard, the following assumptions were made in Experiments 1-4: (1) the level of a L tone has a lower F0 value than an adjacent H tone, (2) a low F0 valley arises from a L tone, and (3) a high F0 peak arises from a H tone. These assumptions are standard and implicit in the literature on the topic. Moreover, we were conservative in that the contrasts which were tested were restricted to cases in which we could find written descriptions of the phonetic criteria by which particular distinctions are assumed to be made.[73]

Given these assumptions, two points are noteworthy concerning descriptions of the mapping of contours with F0 peaks to pitch accent categories. The first point is that the theory assumes that varying F0 peak alignment about a syllable boundary will give rise to a phonological distinction inconsistently, that is, in some cases but not in others. Consider Figure 6.1, which shows three different alignment patterns of an F0 maximum with respect to a stressed syllable. Based on descriptions in various sources (Pierrehumbert 1980, Beckman and Pierrehumbert 1986, Beckman and Ayers-Elam 1997), the paradigmatic theory assumes that contour (1) in Figure 6.1 should phonologically contrast with contour (2); the former is an example of a H+L* accent, while the latter is a H* accent.[74,75] On the other hand, contour (2) is not predicted to contrast with contour (3); both are examples of a H* accent.[76] In other words, varying an F0 peak about a WS syllable boundary should give rise to a phonological contrast, while varying an F0 peak about a SW syllable boundary should not.

The second issue of note concerned a detail of written descriptions of how contours with an F0 peak correspond to pitch accent categories. In particular, for a SW(W)S syllable sequence, a F0 contour with a peak on a W syllable is ambiguous under written descriptions: it could be a H* or it could be a H+L*.[77] In the former case, the starred tone is on the first strong syllable, while in the latter case, the starred tone is on the second strong syllable. It appears that this is a matter of redundancy or indeterminacy in the system. However, the important conclusion to draw is that maximally two categories are predicted by the paradigmatic theory when a F0 peak is shifted across a SWWS syllable sequence: H* and H+L*. This point is relevant for the interpretation of data in Experiments 3 and 4.

We could find few explicit references to the timing of an F0 minimum for distinguishing between categories, or indeed as a phonetic criterion for any particular accentual category.[78] However, a standard assumption consistent with descriptions in the ToBI guidelines is that F0 minima which are low in the speaker's pitch range arise from underlying low tones. Figure 6.2 shows three different alignment patterns of an F0 minimum with respect to a stressed syllable. Given the assumption that an F0 minimum corresponds to a low tone, then contour (1) in Figure 6.2 is predicted to phonologically contrast with contour (2); the former corresponds to L+H*, while the latter corresponds to L*.[79] Moreover, contour (2) is predicted to phonologically contrast with contour (3); the former would be a L* accent, while the latter would be a L+H* accent.[80]

These predictions stand in contrast to those of the syntagmatic view, which defines a clear relation between phonology and phonetics. Tone interval theory claims that the dominant phonetic correlate of relational tone interval primitives is the *perceived pitch level* of tonally-marked timing positions.[81] This leads to predictions regarding the effects of manipulating F0 extremum alignment across syllable sequences. In particular, this manipulation is predicted to consistently give rise to *perceptual category boundaries*. This is because the alignment of an F0 extremum is predicted to affect the perceived relative pitch levels of syllables. For example, if an F0 maximum is aligned with the midpoint of syllable A, then syllable A will probably have a higher pitch than syllable B. Conversely, if an F0 maximum is aligned with the midpoint of syllable B, then syllable B will probably have a higher pitch than syllable A. By extension, these two cases, which are differentiated at one level on the basis of F0 alignment, are predicted to have distinct phonological representations; hence, F0 alignment differences should give rise to categorical effects. Thus, F0 alignment differences are predicted to give rise to categorical effects *to the extent that the alignment differences have perceptually salient consequences for relative pitch level.*

In any given language, it is assumed that some of the perceptual category boundaries will map to phonological categories. However, not all perceptual categories will be attested as phonological categories in the language. Thus, the syntagmatic view does not make predictions about how the contours in the present experiment will correspond to phonological categories. Rather, it takes a "wait and see" approach to phonological description. That is, sequences of phonological categories which are shown to be attested or not attested can be readily accommodated by the descriptive approach afforded under the theory, since the primitives are simply defined in terms of the relations *higher*, *lower*, and *same*.

To summarize, the paradigmatic theory does not define a clear relation between phonology and phonetics. In order to test the theory, it was necessary to make some assumptions, which were described. Moreover, the paradigmatic theory predicts an inconsistent mapping of F0 alignment with particular syllables to phonological categories, and it makes no express provision for the role of perception in the phonology-phonetics relation. In contrast, the present proposal defines a clear relation between phonology and phonetics. The primary phonetic correlate of phonological contrasts is assumed to be based on perceived relative pitch level. In this way, the theory defines a clear role for perception, and it predicts a consistent mapping cross-linguistically between F0 alignment with syllables and phonological categories. However, this theory does not make predictions about the precise number or types of categories that will be attested in a specific language.

6.3 Experiments 1 and 2

Complementary perception and production experiments (Experiments 1 and 2, respectively) were designed in order to test the predictions of the syntagmatic view and the theory of Pierrehumbert and colleagues. The results for the two experiments will be discussed in tandem, since the same set of stimuli were used in both experiments, and the same set of intonational categories are investigated in both.

6.3.1 Stimuli

Four stimulus series were constructed for use in both Experiments 1 and 2. Stimuli consisted of short phrases which contained a target two-syllable sequence with a specific stress pattern. Two of the four series involved time-shifting an F0 maximum; one series was based on the phrase *To Monrovia*, which contains the WS sequence [mən.ˈro], while the other series was based on the nonsense phrase *Too minglingly* (from the SW verb *to mingle*), which contains the SW sequence [ˈmiŋ.gləŋ]. These two series will be referred to as the "*Monrovia*-max" and the "*minglingly*-max" series, respectively. The overall intonation pattern for the *Monrovia*-max and *minglingly*-max series was rising-falling, which is a common pattern for English statements. The remaining two series involved time-shifting an F0 minimum; one series was based on the phrase *To Monrovia?*, which again contains the WS sequence [mən.ˈro], while the other was based on the phrase *They're nonlinguistic?*, which contains the SW sequence [ˈnaŋ.ləŋ]. These two series will be referred to as the "*Monrovia*-min" series and the "*nonlinguistic*-min" series, respectively. The global intonation pattern for these two stimulus series was falling-rising, which is a common pattern for English questions. Target WS or SW syllable sequences were made up of exclusively voiced segments which were mostly sonorant; a nasal-consonant sequence facilitated spectral identification of the syllable boundary. Vowels in the target syllable sequences were matched in vowel height, so that both were high or both were non-high; this minimized possible effects of intrinsic F0 on peak or valley location. (See Ohala 1978 for a discussion of intrinsic F0 effects.)

Each phrase was low-pass filtered and recorded with a sampling rate of 22 kHz using a digital audio tape recorder in a sound-attenuated room with a high-quality microphone. F0 continua were created from a single example of each phrase using the Praat speech analysis and synthesis software package (Boersma and Weenink 2002). The F0 contours were first stylized as a sequence of straight line segments, which provides a reasonable approximation to actual F0 curves (de Pijper 1983, d'Alessandro and Mertens 1995). The F0 maximum or minimum was then shifted in 25 ms increments through the target sequence. A total of 16, 21, 15, and 16 stimuli were created for the *Monrovia*-max, *minglingly*-max, *Monrovia*-min, and *nonlinguistic*-min continua, respectively (Figure 6.3).

6.3.2 Methods – Experiment 1

Participants: The participants were 20 students and staff (6 males, 14 females) at MIT and nearby colleges who were native English speakers and reported normal hearing. Some subjects had participated in other experiments.

Task: The task utilized an AX (same-different) paradigm. Participants were told that should use a computer mouse to click the appropriate button on a computer screen in front of them, and that when they did, they would hear two sound files; the stimuli were presented over headphones via a computer which was running MATLAB software. Participants were asked to decide whether the two sound files were the same or different. They responded by using the mouse to mark a check box on the computer screen corresponding to their selected answer ("Same" or "Different"). Participants then selected a box to go on to the next trial. Each stimulus pair was presented only once. The entire experiment lasted about 20 minutes.

Design: Stimuli from each of the four stimulus series were presented as a block. The order of presentation of blocks was randomized across participants. Ten practice trials preceded each block of experimental trials. Casual listening suggested that stimuli in the *Monrovia*-min and *nonlinguistic*-min series were harder to discriminate than stimuli from the other series, and that a minimum of three steps' separation was necessary to discriminate pairs drawn from these two series. To generate a range of overall levels of discriminability, stimuli from both series which were 3, 5, or 7 steps apart along the continuum were paired. For the *Monrovia-* and *nonlinguistic-* min series, stimuli which were eight steps apart along the continuum were also paired. Approximately 80% of stimulus pairs presented were "different" pairs, while the rest were "same" pairs.[82] Both kinds of stimulus pairs were presented in a fixed, random order within each block.

Analysis: The percentage of correct responses was calculated according to the method described in Ch. 6 of MacMillan and Creelman (1991).[83] Because the four stimulus series apparently differed in the level of difficulty they presented to participants, each stimulus series was analyzed separately. Participants whose false alarm rates equaled or exceeded their hit rates were excluded from further analyses.[84] As a result, data from 20, 19, 17, and 15 participants was included in the *Monrovia*-max, *minglingly*-max, *Monrovia*-min, and *nonlinguistic*-min series, respectively.

6.3.3 Methods – Experiment 2

Participants: Participants were 21 MIT students and staff (8 males, 13 females), who were self-reported native English speakers with normal hearing. All were paid a nominal sum for participation.

Task and Design: The stimuli were the same as in Experiment 1. Stimuli were recorded onto CD's for presentation to subjects. Each stimulus series was blocked, and the order of stimuli within each block was randomized. Moreover, each of the four stimulus series was presented in three separate blocks on the CD, for a total of twelve blocks. The order of presentation of blocks was fixed for all participants. Each block of stimuli within a series was preceded by a set of practice trials consisting of a subset of stimuli drawn from the upcoming block. An imitation paradigm was employed following the method of Pierrehumbert and Steele (1989), with minor modifications. Participants were told that they would hear a phrase over headphones, which they should then imitate as closely as possible in a pitch range that was comfortable. Stimuli were then presented to participants over headphones at a comfortable volume. The text of each phrase was simultaneously displayed on a computer screen. The resulting utterances were digitized at 16 kHz using MARSHA software by Mark Tiede. The total duration of the experiment was approximately 50 minutes.

Analysis: The boundaries of the two target syllables in subjects' productions were determined and marked in Praat by identifying associated discontinuities in amplitude in the spectrogram and waveform and/or by identifying the location of a rise in frequency of F2 or higher formants. The temporal locations of F0 peaks and valleys were determined automatically using a Praat script, and all were subsequently inspected visually for accuracy. If the F0 peak or valley did not occur within the time spanned by the target SW or WS syllable sequence, the token was discarded. In the event that a maximum or minimum appeared to be segment-related, i.e. due to transient pressure buildup at a nasal-liquid boundary, then the next highest maximum or minimum across the target SW or WS syllable sequence, if available, was taken as the location of the peak or valley, respectively. Non-modal voicing was present in some speakers' imitations. In the event of diplophonia without evidence of other voicing irregularities, the F0 maximum or minimum within the diplophonic region was determined. If other voicing irregularities were present, the token was discarded; see Redi and Shattuck-Hufnagel (2001) for a discussion of voicing irregularities.

The normalized timing of the peak or valley was then determined with respect to the onset of the first syllable and the offset of second syllable of the target SW or WS syllable sequence using the formula given in Equation (6.1). In this formula, t is the time of the peak or valley, t_0 is the start of the first syllable in the target SW or WS syllable sequence, and d_1 and d_2 are the durations of the first and second syllables, respectively, in the target syllable sequence. Thus, the normalized turning point location (T_N) ranged from 0 to 1. Various methods of normalization were attempted, but the patterning of the data was unaffected.

$$T_N = \frac{t - t_0}{d_1 + d_2} \qquad (6.1)$$

Two factors appeared to affect the consistency of responses across participants and stimulus sets. The first was the identity of the individual speaker; participants showed differential ability to produce consistent responses to the stimuli. The second was whether the stimuli involved manipulation of an F0 maximum or minimum; it appeared that the stimuli from the *Monrovia*- and *nonlinguistic*-min series were considerably harder for participants to imitate than the stimuli from the *Monrovia*- and *minglingly*-max series. To quantify the consistency of the pattern of productions across participants and stimulus series, bivariate correlations were calculated for all pairs of subjects for each of the four series. A large value for the correlation coefficient r meant that the two subjects produced a similar pattern of responses for that stimulus series. Responses from subjects who were not correlated with 50% or more of the other subjects were judged to be unreliable imitators and not included in the final analyses reported below.[85] Based on this analysis, the final numbers of participants for the *minglingly*-max, *Monrovia*-max, *nonlinguistic*-min and *Monrovia*-min series were 21, 20, 7, and 10, respectively. This preliminary analysis confirmed our initial impression that the *Monrovia*- and *nonlinguistic*-min series were more difficult for subjects to imitate.

6.3.4 Results

The results from Experiments 1 and 2 are reported together, in order to facilitate comparison between the two complementary studies.

minglingly-max series

Experiment 1: Perception task. Figure 6.4 shows the results from the discrimination task for the *minglingly*-max series. The y-axis shows the percentage of correct responses, while the x-axis shows the average of the numbers of two presented stimuli in a given stimulus series. For example, in a 3-step series, the percentage of correct responses to pairings of stimuli 1 and 5 is plotted at an x-axis value of 2.5. In results will be discussed by noting the average of a stimulus pair, followed by a listing of the pair in parentheses, e.g. 2.5 (1, 5). For most discrimination results, the 3- and 5-step series turned out to be the most useful for determining the locations along stimulus series of probable category boundaries and exemplars. This was because the differences in 7-step series were so much above threshold that discrimination was generally quite high across the entire series. As a result, we will focus in the discussion of the results on discrimination for the 3- and 5-step series.

In general, discrimination maxima are consistent with locations of category boundaries, while discrimination minima are consistent with locations of category exemplars. For the 3-step series there was a maximum at 12.5 (11, 14), and there were minima at 7.5 (6, 9) and 19.5 (18, 21). For the 5-step series, there were maxima at 5.5 (3, 8) and 12.5 (10, 15) and minima at 3.5 (1, 6), 6.5 (4, 9), and 18.5 (16, 21). Finally, the highest level of discrimination achieved for the 3-, 5-, and 7-step series was 76%, 82% and 87%, respectively.

Figure 6.5 summarizes the locations of discrimination maxima and minima in this stimulus series (indicated by an "x" or an "o", respectively) with respect to segments for stimulus pairs which were 3 and 5 steps apart. The positions noted correspond to the *average* of the stimuli in the series. For example, to indicate the discrimination maximum at 12.5, there is a "x" halfway between stimuli 12 and 13.

Experiment 2: Production task. Figure 6.6 shows a scatterplot of normalized turning point time, T_N, in imitation of individual stimuli in the *minglingly*-max series. Median values of T_N are shown by open boxes, and whiskers indicate one semi-interquartile range (SIQR) above and below the median.[86] Stimuli 1-11 maintain a rather constant median T_N of approximately 0.4. A large shift in T_N values occurs around stimuli 12-15, then the median T_N for stimuli 16-21 again achieves approximate constancy at around 0.8.

Discussion. A phonological category boundary was defined earlier in terms of categorical effects in perception and in production. Based on this definition, alignment of a F0 maximum across the SW syllable sequence [ˈmiŋ.gləŋ] gives rise to distinct phonological categories. This is shown by the presence of a discrimination maximum in perception data, together with a break point in production data at approximately the same position in the stimulus series. This outcome is not predicted by the paradigmatic theory, which would claim that the stimuli in this series arise from a single category, H*. However, the results do not conclusively rule out the descriptive analysis offered by the paradigmatic theory, a point to which we return in the overall discussion of Experiments 1 and 2.

Monrovia-max series

Experiment 1: Perception task. Figure 6.7 shows the results from the discrimination task for the WS Max series. For the 3-step series, there is a maximum in discrimination at 9.5 (8, 11), and there are minima at 3.5 (2, 5) and at 14.5 (13, 16). For the 5-step series, there is a discrimination maximum at 9.5 (7, 12), and there are discrimination minima at 3.5 (1, 6) and 13.5 (11, 16). The locations of discrimination maxima and minima along the stimulus series for the pairings of stimuli which were 3 steps and 5 steps apart are summarized in Figure 6.8. Just as for Figure 6.5, the positions of discrimination maxima and minima along the stimulus series (indicated by an "x" or an "o", respectively) correspond to the average of the paired stimuli. Finally, the highest level of discrimination achieved for the 3-, 5-, and 7-step series is 74%, 84% and 88%, respectively.

Experiment 2: Production task. Figure 6.9 shows a scatterplot of normalized turning point times, T_N, for imitated versions of individual stimuli for the *Monrovia*-max series. Stimuli 1-9 maintain a relatively constant T_N of approximately 0.3. This is followed by a region of transition for stimuli 10-11. Finally, stimuli 12-16 give a much higher and relatively more stable T_N.

Discussion. Data from perception and production experiments provided complementary evidence consistent with a category boundary for the *Monrovia*-max stimulus series at approximately stimulus 10-11. Under standard assumptions, the paradigmatic theory predicts this pattern of results. In particular, evidence of a category boundary is consistent with the distinction between H+L* and H* pitch accents.

<u>*nonlinguistic*-min series</u>

Experiment 1: Perception task. Figure 6.10 shows the results from the discrimination task for the *nonlinguistic*-min series. For stimuli which were 3 steps apart, there are several local maxima in discrimination; the most significant seem to be located at 4.5 (3, 6), 7.5 (6, 9), and 14.5 (13, 16). Moreover, there are a number of local minima in discrimination; the most significant appear to be located at 3.5 (2, 5), 5.5 (4, 7), and 11.5 (10, 13). The broad shape of the discrimination curve for the 5-step series shows many of the same features as for the 3-step series, but there are more discrimination maxima and minima than in the 3-step series. In particular, there are discrimination maxima at 4.5 (2, 7), 7.5 (5, 10), and 10.5 (8, 13); moreover, there are minima at 3.5 (1, 6), 6.5 (4, 9), 9.5 (7, 12), and 11.5 (9, 14). The locations of discrimination maxima and minima along the stimulus series for the pairings of stimuli which were 3 steps and 5 steps apart are summarized in Figure 6.11. Finally, the average level of discrimination achieved for the 3-, 5- and 7-step series is 65%, 75% and 80%, respectively.

Experiment 2: Production task. In general, the production data were quite unruly, in that subjects as a group showed very different patterns of responses.[87] The results for the 7 subjects who showed the greatest inter-subject consistency are shown in Figure 6.12. There is some consistency in normalized turning point time, T_N, for stimuli 1-8, then there is a general trend toward an increase in T_N values. However, there is no evidence of a break point.

Discussion: Production data show no evidence of a break point. Based on the definition of a phonological category boundary in Section 6.1, these results do not provide support for a phonological category distinction. Thus, the present results do not provide support for the paradigmatic theory's distinction between L+H* and L*. However, the acoustic differences in these stimuli appeared to be less salient perceptually than for either of the series in which F0 maxima were manipulated. We return to these points in the overall discussion of Experiments 1 and 2.

<u>*Monrovia*-min series</u>

Experiment 1: Perception task. Figure 6.13 shows the average percent correct for discrimination of stimulus pairs in the *Monrovia*-min series. For the 3-step series, there were maxima in discrimination at 7.5 (6, 9) and at 10.5 (9, 12). Moreover, there was a minimum in discrimination at 3.5 (2, 5) and at 9.5 (8, 11), and possibly at 13.5 (12, 15). For the 5-step series, there is a single clear local maximum in discrimination at 7.5 (5, 10), but no other clear maxima. Moreover, there is a probable minimum at 3.5 (1,

6), but it is hard to determine whether there are additional minima. The results regarding locations of maxima and minima in the stimulus series are summarized in Figure 6.14.

Experiment 2: Production task. As for the nonlinguistic-min series, the production data were quite unruly, in that subjects as a group showed very different patterns of responses. The results for the 10 subjects who showed the greatest inter-subject consistency are shown in Figure 6.15. This figure shows fairly large variability for median T_N values, but trends are observable. Stimuli 1-6 have a fairly constant average median T_N value of 0.46, while stimuli 9-15 have a higher but still fairly constant average median T_N value of 0.69. There appears to be a break point around stimulus 7 or 8.

Discussion: There is some evidence of a break point in production data at approximately stimulus 7 or 8, and there is evidence of higher discrimination in this region in the stimulus series as well. These results therefore provide some support for the claimed distinction between L+H* and L*.

6.3.5 Discussion – Experiments 1 and 2

Across all four stimulus series, the results from the perception and production tasks were generally in good agreement. For locations along a given stimulus series where there was evidence of a break point in production data and there was generally a corresponding peak in discrimination in perception data. These cases are interpreted as evidence of a phonological category distinction. In the following, we focus on the implications of these results for the phonological description of the paradigmatic theory. Additional points will be addressed in the General Discussion in Section 6.5.

First, the *minglingly-max* series showed a break point in production data and evidence of a discrimination maximum at a corresponding location in perception data. This is interpreted as evidence for the existence of distinct phonological categories along this stimulus series. These categorical effects do not at first glance appear to be predicted by the paradigmatic theory, which would generally claim that all stimuli in this series arise from a single category, H*. However, there are two other possible explanations for the categorical effects observed in this experiment which can be considered.

The first possibility, suggested by Bob Ladd (pers. comm.), is that the *minglingly*-max continuum is interpreted by participants as corresponding to the distinction between H* and another pitch accent type, L*+H. According to written descriptions, a defining characteristic of L*+H is that its F0 peak occurs late, typically after the stressed syllable. It is consistent with standard assumptions that this pitch accent type is produced with a local low valley on the stressed syllable and an F0 peak on the following weak syllable. To check whether listeners could have interpreted cases with a late peak as L*+H, we examined productions in which the peak was aligned with the weak syllable. For these cases, we compared the average F0 on the stressed syllable ([ˈmiŋ]) with the average F0 of the vowel of the preceding syllable ([tʰu]). In 92% of cases, the stressed syllable had a higher average F0 than the /u/ of [tʰu]. This suggests that subjects were not producing instances of the posited L*+H accent category in response to stimuli with F0 peaks after the stressed syllable.[88]

A second possible explanation that might be considered for evidence of a category boundary is that listeners interpreted the location of main stress on different syllables for "early peak" and "late peak" cases. If so, then the categorical effects could be explained in terms of perceived "movement" of the H* accent from the first syllable to the second syllable. This explanation seemed somewhat plausible, because a few subjects reported hearing the stress shift from *ming-* to *gling-*.[89] One motivation for Experiments 3 and 4 was to attempt to replicate the finding of categorical effects of the present experiments using speech materials with more carefully controlled stress, in order to rule out the explanation of a stress shift for these effects.

Next, we consider the results for the *Monrovia*-max series. This series showed a break point in production data and a local discrimination maximum at approximately the same position along the stimulus series. This is consistent with a phonological category distinction which is accounted for in the paradigmatic theory by the pitch accent categories H+L* and H*. We note that these results are consistent with the findings of Kohler (1987). Kohler also observed categorical effects in a pair of discrimination and labeling experiments using synthetic speech in which an F0 peak was shifted across the WS syllable sequence gelo- the German phrase *Sie hat ja gelogen.*

For the *nonlinguistic*-min series, Experiment 2 failed to show evidence of a break point in production. Thus, the results from this series cannot be interpreted as support for the paradigmatic theory's claimed distinction between L+H* and L*. We return to this issue in the general discussion.

Finally, for the *Monrovia-min series,* there was some evidence of a break point in production data in Experiment 2, and there was a corresponding peak in discrimination data at approximately the same location. These results therefore provide some support for the claimed distinction between L* and L+H* accents, in contrasts to the results from the *nonlinguistic*-min series.

We will return to additional discussion points, including implications for the syntagmatic view, in the General Discussion in Section 6.5. Next, we describe two follow-up experiments in which several issues are addressed. First, these experiments attempted to replicate findings of Experiments 1 and 2 that varying the timing of an F0 peak about the right boundary of a stressed syllable gives rise to evidence of a phonological category distinction in perception and production. Also, we attempted to determine whether synthetic stimuli in which an F0 minimum was shifted would yield clearer evidence of category boundaries in perception and production when the contour shapes were slightly modified compared to those used in stimuli in Experiments 1 and 2.

6.4 Experiments 3 and 4

6.4.1 Stimuli

Four stimulus series were constructed for use in both Experiments 3 and 4. Short phrases were used which contained a target two- or three-syllable sequence with a specific stress pattern, in order to test the predictions of the implemented paradigmatic theory. Once again, two series involved time-shifting an F0 maximum; these were based on the short phrases *For a millionaire*, which contains the SW sequence [ˈmɪl.jə] (the "*millionaire*-max" series), and *In Lannameraine* ([ˈla.nə.mə.ˌreʲn] or [ˌla.nə.mə.ˈreʲn]) which contains the SWW sequence [ˈla.nə.mə] (the "*lanna*-max" series). The overall intonation pattern for these two stimulus series was rising-falling. The remaining two series involved time-shifting an F0 minimum; these were based on the short phrases *Some lemonade?*, which contains SW sequence [lɛ.mə] (the "*lemonade*-min" series), and *They're nonrenewable?*, which contains the WS sequence [nan.rə] (the "*nonrenewable*-min" series). The overall intonation pattern for these two stimulus series was falling-rising. For the *lemonade*-min and *nonrenewable*-min series, the shape of the curve for the F0 minimum was made somewhat broader as compared to the F0 minima series in Experiments 1 and 2. Moreover, the preceding context was set to a level contour, rather than a fall. We hypothesized that these alterations might make the stimuli sound more realistic and increase the chance of observing categorical effects in production and perception tasks. Finally, target WS, SW, and SWW syllable sequences in all stimulus series were made up of exclusively voiced, sonorant segments.

Each phrase was low-pass filtered and digitized using a 16 kHz sampling rate in a sound-attenuated room with a high-quality microphone using MARSHA software by Mark Tiede. F0 continua were created according to the method described for Experiments 1 and 2. F0 maxima and minima were shifted in 30 ms increments through the target sequence. The stimuli are shown in Figure 6.16. One difference in the stimuli with F0 valley manipulation was that the initial part of the contour in the WS and SW Min stimuli had a flat F0, since it was hypothesized that this might be a more common, and therefore easier, contour for subjects to imitate than the contour in Experiment 2. There were a total of 13, 18, 10, and 13 stimuli in the *millionaire*-max, *lanna*-max, *lemonade*-min, and *nonrenewable*-min series, respectively.

6.4.2 Methods – Experiment 3

Participants: Participants were 20 students and staff at MIT and nearby colleges (18 females, 2 males) with self-reported native English speaking abilities and normal hearing.

Task and design: The experimental design and procedure were identical to Experiment 1, except no 8-step values were used. The total duration of the experiment was about one hour.

Analysis: Data from each stimulus series were analyzed separately according to the methods described for Experiment 1. Subjects with a false alarm rate which was higher than the hit rate were again discarded.[90] This left 20 subjects for the *millionaire-* and *lanna*-max series, and 18 subjects for the *nonrenewable*-min and *lemonade*-min series, respectively.

6.4.3 Methods – Experiment 4

Participants: Participants were 17 students and staff at local colleges (5 males, 12 females), who were self-reported native English speakers with normal hearing. All were paid a nominal sum for participation. Some of the subjects also participated in the perception study.

Task and design: The procedures were identical to those used in Experiment 2, except as noted. Presentation of the *Lanna*-max series was broken up into two overlapping blocks, based on the possibility that this series contained three categories. The first block consisted of stimuli 1-11 (*Lanna*-max – Block I), the other of stimuli 8-18 (*Lanna*-max – Block II). This was done in order to ease the demands on participants, who might have difficulty keeping three categories separate in a single block. The procedure for normalization of F0 turning points was the same as in Experiment 2 for all series except the *Lanna*-max series. For this series, stimuli produced in the first block were normalized with respect to the initial strong and following weak syllable in *Lannameraine*, while stimuli produced in the second block were normalized with respect to the two weak syllables in *Lannameraine*.

Analysis: Data from the *millionaire*-max, *nonrenewable*-min and *lemonade*-min series, as well as Blocks I and II of the *Lanna*-max series were analyzed separately. As was done for Experiment 2, the consistency of the pattern of productions across participants and stimulus series was quantified by calculating bivariate correlations for all pairs of subjects for each of the stimulus series. The responses from subjects who were judged to be unreliable imitators using the method described in Experiment were not included in the final analyses. Based on this analysis, the final number of subjects for the *millionaire*-max, *Lanna*-max (Block I), *Lanna*-max (Block II), *lemonade*-min, and *nonrenewable*-min series were 11, 15, 12, 12 and 10 subjects, respectively. It was clear from this preliminary analysis that responses of subjects to the *lemonade*-min and *nonrenewable*-min series were substantially more variable than for the other series, suggesting that as found for Experiment 2, imitating the placement of an F0 minimum was harder than imitating the placement of a F0 maximum.

<u>*millionaire*-max series</u>

Experiment 3: Perception task. Figure 6.17 shows the results from the same-different task for the millionaire-max series. The 3-step series shows two clear maxima in the discrimination curve: at 2.5 (1, 4) and at 7.5 (6, 9). Moreover, this series shows two probable minima: at 5.5 (4, 7) and at 11.5 (10, 13). For the 5-step series, there is a local maximum at 8.5 (6, 11), and there are apparent local minima at 3.5 (1, 6), and 10.5 (8, 13). The maximum percentage correct for the 3-, 5- and 7-step series are 76%, 83%, and 85%. Figure 6.18 summarizes the locations of discrimination maxima and minima along this stimulus series, for stimulus pairs which were 3 and 5 steps apart.

Experiment 4: Production task. Figure 6.19 shows a scatterplot of normalized turning point time, T_N. The medians of stimuli 1-7 have an approximately constant value of T_N of 0.45. However, there is a jump in T_N values at stimulus 8, where there are many more late placements of F0 peaks.

Discussion: There is a break point along this series at approximately stimulus 8 which is matched by a discrimination peak. The production and perception data therefore suggest the existence of a phonological distinction along this stimulus series. The results cannot readily be explained by the paradigmatic theory, which generally assumes that the variation in this series corresponds to a single category, H*. However, there are additional issues which must be considered, which we address in the discussion of Experiments 3 and 4.

Lanna-max series

Experiment 3: Perception task: Figure 6.20 shows two major discrimination maxima in the 3-step discrimination curve: at 6.5 (5, 8) and at 12.5 (11, 14). Moreover, there are discrimination minima in this curve at 3.5 (2, 5), 9.5 (8, 11), and 15.5 (14, 17). For the 5-step series, there are discrimination maxima at 6.5 (4, 9) and at 13.5 (11, 16), and there are discrimination minima at 3.5 (1, 6), 10.5 (8, 13), and 15.5 (13, 18). Figure 6.21 summarizes the locations of discrimination maxima and minima along this stimulus series. We also note that the maximum percentages correct for the 3-, 5- and 7-step series are 78%, 87%, and 92%.

Experiment 4: Production task: Figure 6.22 shows a scatterplot of normalized turning point time, T_N, in imitation of individual stimuli in the *Lanna*-max series for Block I. The medians of the stimuli fall along an S-shaped curve, with a break point occurring at about stimulus 7. Moreover, Figure 6.23 shows a scatterplot of T_N in imitation of stimuli in the *Lanna*-max series for Block II. The medians of the stimuli also fall along an S-shaped curve, with a break point occurring at about stimulus 13.

Discussion. The existence of two break points along the stimulus series for both Blocks I and II in the production data, together with peaks in discrimination at corresponding locations in the perception data support the existence of three phonological categories for this stimulus series. These results are at odds with the predictions of the paradigmatic theory. We return to this point in the discussion.

lemonade-min series

Experiment 3: Perception task: Figure 6.24 shows two distinct local maxima in the discrimination curve for the 3-step series at 2.5 (1, 4) and at 7.5 (6, 9), as well as minima at 5.5 (4, 7) and 8.5 (7, 10). For the 5-step series, discrimination is generally high across the board, with no distinctive maxima and minima. Because the 3-step series shows greater variation in discrimination across the board, we have taken it as a better indicator of probable perceptual category boundaries. The locations of discrimination maxima and minima for the 3-step series are summarized in Figure 6.25. Finally, we note that the maximum percentage correct for the 3-, 5- and 7-step series are 84%, 80%, and 84%.

Experiment 4: Production task: Figure 6.26 shows a scatterplot of T_N in imitation of individual stimuli in the *lemonade*-min series. The medians of the stimuli fall along an S-shaped curve, with a break point occurring at about stimulus 6 or 7.

Discussion. The break point in production data, coupled with the peak in discrimination, suggest the presence of two phonological categories. This can be accommodated by the paradigmatic theory as a distinction between L* and L+H*.

<u>*nonrenewable*-min series</u>

Experiment 3: Perception task: Figure 6.27 shows two major local maxima, one at 6.5 (5, 8), the other at 10.5 (9, 12). Moreover, there are local minima at 2.5 (1, 4), 9.5 (8, 11), and 11.5 (10, 13). As with the *lemonade*-min series, the pattern of greater and lesser discrimination is not well-replicated in the 5-step series. Because the 3-step series shows more variation in discrimination, the locations of discrimination maxima and minima for this series alone are summarized in Figure 6.28. The maximum percentage correct for the 3-, 5- and 7-step series are 63%, 66%, and 72%.

Experiment 4: Production task. The median values of T_N fall along an S-shaped curve. There is evidence of a discontinuity or break point in the vicinity of stimuli 10 and 11, supporting at least one category distinction.

Discussion. The existence of a breakpoint in production data and a discrimination peak in perception results at approximately the same location suggest that this series encompasses distinct phonological categories. This supports the distinction between L+H* and L* claimed by the paradigmatic theory.

6.4.4 Discussion

The results for individual stimulus series are summarized below. For these series, we consider the implications of findings associated with each stimulus series for the paradigmatic theory's claims about the phonological categories of English.

First, we consider the result from the *millionaire*-max series. Evidence from production and perception tasks suggested that two distinct phonological categories underlie this series, as indicated by a break point in production data and discrimination peak in perception data. What are the implications of this result for the paradigmatic theory of English? An ambiguity in written descriptions of pitch accent categories in the theory permits two interpretations of which categories underlie this stimulus series. One interpretation is that there is a single accent category, H*; this accent type is assumed to be variable in its phonetic realization, such that the F0 peak can either occur on a stressed syllable or after it. The present results and those of Experiments 1 and 2 rule out this interpretation of the categories. That is, the evidence of a category boundary in perception and production is incompatible with this series corresponding to a single category. The other interpretation is that there are two categories, H* and H+L*. This interpretation is possible because an F0 peak on a weak syllable could be associated with a H+L* on the upcoming stressed syllable. Evidence from the Lanna-max series is relevant to this interpretation, which we now turn to.

The *Lanna*-max series involved shifting an F0 peak across a SWW syllable sequence. Evidence from production and perception tasks suggested that *three* distinct phonological categories underlie this series. This was indicated by the appearance of *two* break points in production data and corresponding discrimination peaks in perception data. The paradigmatic theory of English has difficulty explaining this result because variation in the timing of a peak across a SWW(S) syllable sequence is assumed maximally to correspond to two categories, H* and H+L*, as described above. The paradigmatic theory has no recourse to explaining the present finding of three categories. These results can be explained under the syntagmatic view, which we will consider in the General Discussion.

We now turn to the series in which F0 minimum timing was manipulated. For the *lemonade*-min series, dual evidence was obtained from production and perception tasks suggesting the presence of two phonological categories along this series. This can be accommodated by the paradigmatic theory as a distinction between L* and L+H*. Finally, dual evidence from production and perception tasks suggests the presence of two phonological categories along the *nonrenewable*-min series. This supports the distinction between L+H* and L* claimed by the paradigmatic theory.

The results of Experiments 3 and 4 therefore confirm those of Experiments 1 and 2 in several ways. First, we replicated the finding from Experiments 1 and 2 that shifting an F0 peak across a SW syllable sequence gives rise to evidence of a phonological category boundary. Moreover, additional data showed that shifting an F0 peak across a SWW syllable sequence gives rise to evidence of three phonological categories. These results cannot be accommodated by the paradigmatic theory of English intonation. Finally, we obtained additional evidence that shifts in F0 minimum alignment can give rise to phonological categories as well.

6.5 General discussion

Experiments 1-4 tested aspects of the paradigmatic theory of English intonation (Pierrehumbert 1980, Beckman and Pierrehumbert 1986). In order to test this theory, it was necessary make some assumptions about the mapping from phonology to phonetics. The assumptions that were made were conservative: we assumed that the level of a L tone is lower than an adjacent H tone, and that low and high F0 extrema arise from L and H tones, respectively. Moreover, we defined evidence of a phonological category boundary an operational fashion, i.e., as discreteness in production data and a discrimination maximum in perception data.

Given these assumptions, the paradigmatic theory has difficulty explaining several aspects of the experimental results. The first aspect of the theory which is inconsistent with experimental findings is that F0 peak timing is assumed to be a phonetic correlate of phonological distinctiveness when a peak is shifted across a WS sequence, but not a SW sequence. In the former case, the distinction is assumed to involve the contrast between H+L* and H*, while in the latter case all phonetic variation is assumed to correspond to H*. (This statement is consistent with standard assumptions, despite the ambiguity between H* and H+L* discussed in Section 6.2.) In contrast to the claims, Experiments 1-4 showed that varying the timing of a F0 peak across either a WS sequence or a SW sequence gave rise evidence of a contrast. In particular, we found that shifting a F0 peak across a SW syllable sequence yielded a discrete shift in production data, together with a discrimination maximum in perception data. Moreover, this finding was replicated in three separate stimulus series (i.e., the *minglingly*-max series of Experiments 1 and 2, and the *millionaire*-max and *Lannameraine*-max series of Experiments 3 and 4). These results are therefore incompatible with the claim that F0 peak timing is a phonetic correlate of phonological distinctiveness only when a peak is shifted across a WS sequence, but not when a peak is shifted across a SW sequence.

The second aspect of the theory which is inconsistent with experimental findings concerns the assumption that a maximum of two phonological categories will be represented by shifting a F0 peak across the SWW portion of a SWWS syllable sequence. Contrary to this claim, the present experiment revealed evidence of *three* phonological categories. Consistent with this, we found evidence of two category *boundaries* along the Lanna-max continuum, as gauged by two break points in production data and two clear discrimination maxima at corresponding locations in the series. These data therefore suggest that shifting a F0 peak across a WS, SW, or WW syllable sequence uniformly gives rise to a phonological category distinction. This suggests the hypothesis that shifting a F0 peak across *any* syllable sequence will give rise to a phonological category distinction.

The present experiments also revealed a difference in the responses of subjects to shifts in F0 maxima, as compared to shifts in F0 minima. In all four experiments, shifts in F0 maxima yielded much clearer results in perception and production tasks than shifts in F0 minima. Experiments 3 and 4 utilized stimuli with a slightly different contour shape for F0 minima, as compared with Experiments 1 and 2; these stimuli elicited responses consistent with the distinctions of L* and L+H* claimed by the paradigmatic theory. We return to the issue of differences between F0 maxima and minima later in the discussion.

Evidence of phonological categories obtained in Experiments 1-4 can be accommodated under the syntagmatic view. Recall that the syntagmatic view defines simple primitives (*higher, lower, same*); in this way it can account for attested phonological distinctions without the problems encountered by theories which assume more complex primitives. Then the phonological category distinctions for which we found evidence are described in the tone interval framework as in Table 6.1 and Figures 6.30-6.35. In the table and figures we show phonological descriptions using the *L/H/E* notation developed in Chapter 5.

There was an interesting difference between perception and production results. In particular, the perception results revealed that subjects could apparently hear more distinctions than they could produce, as indicated by the fact that there were more peaks in discrimination curves than there were break points in production data. We think there are probably several reasons for the differences. First, we take heed of the suggestion of House (1990) that F0 alignment differences universally give rise to natural perceptual categories. One consequence of this view is that F0 alignment differences are predicted to universally give rise to phonological distinctions. Another consequence is that certain perceptual distinctions heard by listeners when F0 extrema are shifted will not necessarily correspond to phonological categories that are used and attested in a given language. These predictions fit the present findings well. The experimental results reported here suggested that F0 alignment differences consistently gave rise to phonological categories; however, listeners heard more distinctions than they could reproduce.

Moreover, the syntagmatic view offers a phonological explanation for the finding that there are apparently more perceptual categories than phonological categories. In particular, variations in F0 alignment are assumed to correspond at the phonological level to different numbers of tones and/or different relative height relations among the tones. Then the fact that there are apparently more perceptual categories than phonological categories is explained by assuming that more sequences of tones and relative height relations are heard as being different than are used contrastively in the language.

Another explanation for the finding that there were more discrimination peaks than production break points is that the production method is too coarse and/or challenging for subjects to reveal evidence of all phonological categories present. Indeed, the production task was generally somewhat challenging for most subjects. Nevertheless, most all subjects were capable of reproducing distinct, categorical timing patterns in response to at least some stimulus series. However, imitating the sorts of sub-syllable timing distinctions that were indicated by certain discrimination maxima might have been too much for most subjects to accomplish "on demand" within this paradigm. Nevertheless, our sense was that some subjects actually did reproduce certain finer-grained timing distinctions, but these effects were obscured in the general pattern of subject data. Future work should address this possibility.

One finding which is not explained by production task difficulty concerns the asymmetry in the responses to stimuli in which F0 peaks were manipulated, as opposed to F0 valleys. Across all four experiments, stimuli in which F0 peak timing was manipulated appeared to be easier to perceive and reproduce than stimuli in which F0 valley timing was manipulated. We suspect that this may be related to psychoacoustic findings that discrimination thresholds are higher for F0 minima compared to F0 maxima in comparable frequency ranges (Demany and McAnally, 1994, Demany and Clément 1995a,b, 1997, de Cheveigne 2000).

Comparing results from stimulus series in which F0 minima were manipulated across the experiments, it is clear that those from Experiments 3 and 4 were more successful at eliciting categorical effects than those from Experiments 1 and 2. We attribute this to the fact that the low F0 minima in stimuli used in Experiments 3 and 4 had a somewhat more rounded shape than those of Experiments 1 and 2. Consideration of more natural speech lends insight into possible reasons why the results were less clear-cut than we expected. In particular, it is our sense that when speakers intend to produce a "low" target on a particular syllable, they do not produce as sharp an F0 minimum as those in our stimuli. Rather, they seem to remain at a lower level; the effect of this is to produce a much more rounded contour, rather than the sharp contour used in these stimuli. Then given psychoacoustic findings that F0 minima are not as easily perceived, it may be that a syllable must remain "low" for a longer duration in order for it to be heard as having a "low tone". This suggests the hypothesis that in order to hear a syllable as a "low tone" that a syllable must have a lower pitch perceptually than surrounding syllables. If so, then the present experiments failed to elicit stronger categorical effects because the phonetic cues were ambiguous in their perceived level. This hypothesis receives support in Chapter 7.

Finally, an interesting issue is whether specific segmental characteristics give rise to acoustic "landmarks" which interact with the F0 contour in a way that readily gives rise to perceptual categories

(House 1990, Stevens 2002). One pattern that we noticed in surveying the results from the discrimination experiments was that the locations of discrimination maxima often fell at the location of a vowel onset or another position of a rapid increase in sonority. This pattern of results agrees with the findings of House (1990).

Experiments 1 and 2	
minglingly-max	[L_0 H* L%] vs. [L_0 H* +H L%][91]
Monrovia-max	[L_0 H+ L* L%] vs. [L_0 H* L%]

Experiments 3 and 4	
millionaire-max	[L_0 H* L%] vs. [L_0 H* +H L%]
Lannameraine-max	[L_0 H* L%] vs. [L_0 H* +H L%] vs. [L_0 H+ L* L%]
lemonade-min	[H_0 L* H%] vs. [E_0 E* L+ H* H%]
nonrenewable-min	[E_0 E* L+ H* H%] vs. [E_0 E+ L* H%]

Table 6.1. Description of phonological categories afforded under the present proposal.

6.6 Conclusions

These experiments showed that continuously shifting an F0 maximum or minimum across a SW, WS, or WW syllable sequence uniformly gave rise to evidence of a phonological category boundary. This was demonstrated by discreteness in production data and a discrimination peak in perception data. The paradigmatic theory of English intonation was unable to account for the pattern of observed results. In particular, a maximum of two phonological categories were predicted to underlie varying a F0 peak through a SWW sequence in a SWWS context. However, evidence of three phonological categories was observed. This is consistent with tone interval theory, in that F0 alignment differences are assumed to give rise to perceived differences in relative pitch level; such differences are assumed to be the direct phonetic correlates of syntagmatic phonological category distinctions.

Chapter 7 – Phonetic correlates of phonological categories II: F0 level

The previous chapter focused on the role of the temporal alignment of F0 turning points in phonological contrast. This chapter focuses on the role of variations in absolute F0 level in the phonology. Two additional experiments are discussed which tested various predictions of the syntagmatic and paradigmatic views of English intonation and those of other theories.

In Experiment 5, the predictions of the paradigmatic view were pitted against those of the syntagmatic view. The paradigmatic view predicted that varying absolute F0 level through the pitch range would elicit evidence of phonological contrast, even when relative F0 level was controlled. Here, the relative F0 level of two syllables is defined as whether a syllable is higher than, lower than, or at the same level as another syllable. The syntagmatic view predicted that varying F0 level through the pitch range should elicit evidence of a phonological contrast only when these variations affect relative F0 level.[92]

Experiment 6 also tested the predictions of the syntagmatic view by focusing on the interpretation of earlier experiments reported in Chapter 6, which had suggested the importance of F0 turning point alignment for phonological distinctions. One possible interpretation of the earlier findings was that manipulating F0 turning point alignment across a two-syllable sequence affected phonological categorization because a second phonetic correlate, namely the relative F0 levels or relative pitch levels of the syllables, was affected as well. The earlier experiments did not make it possible to determine which of these two phonetic cues was responsible for differential phonological categorization of stimuli. Experiment 6 tested a prediction of the syntagmatic view that relative pitch level should be the dominant phonetic cue to phonological contrast. If so, then eliminating cues to the temporal alignment of F0 turning points should present no harm to the effectiveness of the phonological contrast. This is precisely what was demonstrated in Experiment 6.

7.1 Experiment 5

This experiment pitted the predictions of the paradigmatic theory of English intonation against those of the syntagmatic theory outlined in preceding chapters. The following briefly reviews general differences between these two views for the presumed relation between phonology and phonetics. Next, three specific cases are discussed in which the predictions of the paradigmatic and syntagmatic views could be tested against one another directly.

7.1.1 Background and hypotheses

A central difference between paradigmatic and syntagmatic views of English intonation regards the presumed relation of each between phonology and phonetics. The paradigmatic view assumes that the primary phonetic basis of phonological distinctions rests in how tonal elements are related to a speaker's pitch range. In contrast, the syntagmatic view assumes that the primary phonetic basis of phonological distinctions rests in the relative pitch level of a tone with respect to another tone.

We could identify three claimed distinctions for English intonation, to be described shortly, which seemed like a good testing ground for the paradigmatic and syntagmatic views. If the paradigmatic view is correct, then varying the level of a tone with respect to the speaker's pitch range should give rise to evidence of a phonological category under an appropriate manipulation. On the other hand, if the syntagmatic view is correct, then varying the level of a tone with respect to the speaker's pitch range *in and of itself* should not give rise to evidence of a phonological category. Rather, the syntagmatic view predicts that evidence evidence of a phonological category will not be observed unless the relative height of one syllable changes with respect to the level of an adjacent syllable, e.g. if a syllable switches from being higher than to being lower than another syllable.

Before describing the three test cases, it should be pointed out that certain assumptions had to be made regarding how tones may be scaled with respect to one another in order to test the paradigmatic theory. In particular, it was necessary to assume certain constraints on the parameters of the phonetic models of P80 and PB88, just as for Experiments 1-4. In making these assumptions, we relied on written descriptions of pitch accent categories and how they differed from one another.

Based on written descriptions of tonal categories, we could identify three cases in which the paradigmatic view could be tested against the syntagmatic view. The first case involves the distinction claimed between a H* pitch accent and a bitonal L+H* pitch accent (Pierrehumbert 1980). It has been claimed that these two accents can be distinguished on the basis of the F0 level of one or more unstressed syllables, when these syllables are initial in a phrase (Beckman and Ayers-Elam 1997). In particular, the syllables preceding the stress are assumed to be lower in pitch in the case of the L+H* pitch accent, compared with the H* pitch accent. However, both accents are assumed to exhibit a rise from the unstressed syllable(s) to the first stressed syllable. This pair of accentual configurations therefore provides an ideal case for testing the paradigmatic view against the syntagmatic view. In particular, the paradigmatic view predicts that a categorical shift should occur when the F0 levels of phrase-initial unstressed syllables are varied from a very low pitch to a relatively higher pitch, consistent with a shift from the category of L+H* to that of H*. This is because under the paradigmatic view, the initial unstressed syllable(s) are claimed to carry a L tone when they are low in the speaker's pitch range, but they are claimed to lack a L tone when they are relatively higher in the speaker's range. In contrast, the syntagmatic view suggests that no such categorical shift should occur under such a manipulation, because both contour shapes involve a rise up from the preceding unstressed syllable(s) to the stressed syllable.[93]

A second case for which the paradigmatic view can be tested involves the claimed distinction between L*+H and H*. The L*+H pitch accent is assumed to be realized with a low pitch in the lower part of the speaker's pitch range on the stressed syllable, followed by a rise to an F0 peak which usually occurs after the stressed syllable. By contrast, a H* accent is assumed to be realized with a high pitch in the "upper part of the speaker's pitch range" on the stressed syllable. It is assumed that the F0 peak may occur after the stressed syllable for this pitch accent. In other words, it is assumed that both accent types may be realized with a shape in which there is a relatively lower pitch on the stressed syllable which rises to a peak after the stressed syllable. The difference between the two is then held to involve how low the stressed syllable is in the speaker's pitch range. Then the paradigmatic view predicts that a categorical shift should occur when the F0 level of the stressed syllable (and any preceding unstressed syllables) are varied from a very low pitch to a relatively higher pitch, consistent with a shift from L*+H to H*. In contrast, the syntagmatic view predicts that no such categorical shift should occur under such a manipulation, because both contour shapes involve a rise up from the stressed syllable to a post-stress F0 peak.

A third case for which the paradigmatic view could be tested involves the distinction between an initial high boundary tone, %H, versus a lack of tonal specification. The %H is assumed to correspond to a very high pitch on the initial unstressed syllable(s) of an utterance. For the tonally unspecified case, the initial unstressed syllable(s) are comparatively lower than for %H, but they are still high in pitch. Written materials accompanying the ToBI transcription system suggest that %H contrasts with a contour possessing the default, unmarked pattern, where the latter "tends to start in the middle part of the speaker's pitch range." Then the phonetic difference between %H and the unmarked, default case is assumed to rest in the height(s) of the initial unstressed syllable(s) in the pitch range. However, for both contours it is assumed that the initial unstressed syllable(s) have a higher pitch level than a following stressed syllable. Then the paradigmatic view predicts that varying the height(s) of the initial unstressed syllable(s) from a very high pitch to a comparatively lower pitch should yield a phonological contrast, consistent with a shift from %H to the unmarked case. In contrast, no evidence of a category boundary should be observed if an unstressed syllable is shifted from being higher than a following stressed syllable to being lower than that syllable. On the other hand, the syntagmatic view makes the opposite prediction: evidence of a category boundary should be found only when an unstressed syllable is shifted from being higher than a following stressed syllable to being lower than that syllable.[94]

The predictions offered up by these three cases were tested in an imitation task. The paradigmatic view predicted that the imitations of target phrases would reflect categorization on the basis of pitch range. This hypothesis predicted that the *oregano*, *oranges*, and *linguistics* stimuli would each be grouped into two categories. On the other hand, the syntagmatic view predicted that the imitations of target phrases would reflect categorization on the basis of relative level. This predicted that the *oregano* and *oranges* stimuli would each be grouped into one category, while the *linguistics* stimuli would be grouped into two categories in a manner distinct from that predicted under the paradigmatic view. We turn now to a discussion of the experimental methods.

7.1.2 Methods

Stimuli: Three stimulus series were constructed using short phrases containing a target two-syllable sequence with a specific stress pattern, in order to test the claims of the paradigmatic theory. All three series involved shifting the F0 level of a portion of the F0 contour through the pitch space. The series were based on three short phrases. The first was *Some oregano*, which contains the WS sequence [ə.ˈreʲ] (the *oregano* series), which was paired with the overall rising-falling contour common of English declaratives.[95] The next was *Some oranges*, which contains the SW sequence [ˈo.rən] (the *oranges* series), which also has an overall rising-falling contour. Finally, the last series was based on the phrase *Linguistics?*, which contains the WS sequence [lɪŋ.ˈgʷɪs] (the *linguistics* series), which was paired with an overall falling-rising or rising-rising contour. Target WS or SW syllable sequences were made up of exclusively voiced segments which were mostly sonorant; a nasal-consonant sequence facilitated spectral identification of the syllable boundary.

Each phrase was recorded using the methods described in Experiments 1-4, Chapter 6. A single token of each phrase formed the basis for the stimulus series. Praat was then used to stylize the F0 contours of these phrases as for stimuli in Experiments 1-4.

The *oregano* series was created by shifting the F0 level of the pre-stress syllables *Some or-* in the phrase *Some oregano* in parallel steps of 1.5 semitones through the pitch range, as shown in Figure 7.1(a). This interval size was chosen so that the entire pitch range could be covered in a fairly small number of steps, so that the experiment had a reasonable duration. While the F0 across *Some or-* was varied, the F0 on the accented syllable *-reg-* was held fixed at a high value. The F0 level of *-reg-* was higher than the level of *Some or-* by 1 semitone or greater for all stimuli in the series. The upper and lower endpoints of this stimulus series corresponded to contours described as H* and L+H* under the paradigmatic theory.

The *oranges* series was created by shifting the F0 level of the pre-stress plus main stress syllable sequence *Some or-* in the phrase *Some oranges* through the pitch range in parallel steps of 1.5 semitones, as shown in Figure 7.1(b). The F0 level of the post-stress syllable *-an-* was meanwhile held fixed at a high value; the level of this syllable was higher than the level of the preceding syllable by at least one semitone. In this manner the upper and lower endpoints of the stimulus series corresponded to contours described as H* (with a late peak) and L*+H under the paradigmatic theory.

For the *linguistics* series, the F0 level of the pre-stress syllable *ling-* was similarly shifted through the pitch space in increments of 1.5 semitones, as shown in Figure 7.1(c). The F0 level of the following stressed syllable *guis-* was held fixed at a low value. For this stimulus series, the level of the first syllable *ling-* was varied through a range such that for stimuli 1-11, *ling-* was higher than *guis-*, while for the stimuli 12-15, *ling-* was lower than *guis-*. As discussed above, stimuli with a very high initial F0 value on *ling-* were claimed under the paradigmatic theory to correspond to instances of %H, while the remaining stimuli were claimed to correspond to the default, unmarked case.

Participants: Participants were 17 students and staff at MIT or other colleges (5 males, 12 females), who were self-reported native English speakers with normal hearing. All were paid a nominal sum for participation.

Procedure: Stimuli were recorded onto CD's for auditory presentation to subjects. Each stimulus series was blocked, and the order of stimuli within each block was randomized. Moreover, each of the three stimulus series was presented in three separate blocks, for a total of nine blocks. The order of presentation of blocks was fixed for all participants. Each block of stimuli within a series was preceded by a set of practice trials consisting of a subset of stimuli drawn from the upcoming block. Stimuli were presented to participants over headphones at a comfortable volume and the text of each phrase was simultaneously displayed on a computer screen, as in Experiment 2. Participants were told to imitate each phrase as closely as possible in a comfortable pitch range. The resulting utterances were digitized at 16 kHz using MARSHA software by Mark Tiede. The total duration of the experiment was approximately 45 minutes.

Analysis and predictions: The competing hypotheses under study made fundamentally different assumptions about the importance of two phonetic dimensions: pitch range and relative pitch level. The paradigmatic view predicts that the position in the pitch range constitutes the most important phonetic dimension for the phonological representation. In contrast, the syntagmatic view predicts that the relative pitch levels of syllables constitute the most important phonetic dimension for the phonological representation. We used two types of metrics to test these two hypotheses.

The first metric entailed a direct estimate of tonal scaling according to each of the two theoretically relevant phonetic dimensions of interest, i.e., position in the pitch range and relative syllable level, as gauged from subjects' imitations. Both of these phonetic dimensions had been varied continuously in the stimuli, given the definition of each provided later in the methods section. We suspected that the phonetic dimension that most mattered for the phonological representation would give rise to categorical behavior along that dimension. Then if stimuli were categorized according to their position in the speaker's pitch range, a plot of this variable was predicted to show a discrete distribution, as opposed to a continuous one. Similarly, if stimuli were categorized according to the relative pitch levels of syllables, a plot of this variable was predicted to show a discrete distribution, as opposed to a continuous one.

The second metric entailed a measure of overall inter-subject variability along each of the two phonetic dimensions of interest. It seemed reasonable that if a particular dimension constituted the phonetic basis of tonal scaling in subjects' imitations, then estimates of that dimension would be comparatively more consistent across speakers. In contrast, a dimension which did not constitute the phonetic basis of tonal scaling in subjects' imitations would be comparatively more inconsistent across speakers.

I. Estimates of tonal scaling. The first metric entailed determining an F0 estimate of how tones from each imitated stimulus were scaled, according to two different methods. One method provided an estimate of position of a tone in the speaker's pitch range, while the other method provided an estimate of the relative F0 level of a syllable with respect to another syllable. These two methods are described in turn below.

The first method for estimating tonal scaling involved determining the position in the pitch range for target syllables according to the phonetic model of Pierrehumbert and Beckman (1988). Under this method, the location of a tone T within the pitch range is modeled as a proportion of the linear distance between some maximum F0 value representing the upper end of the pitch range, h, and a minimum F0 value representing the lower end or reference line of the pitch range, r. The normalized F0 level according to this method is given as:

$$f_1(T) = \frac{T_0 - r}{h - r} \tag{7.1}$$

We will refer to this method of estimating F0 as the *pitch range method*, because it calculates the position of a tone in the pitch range as a fraction of the linear distance between a high tone line h and reference line r. Because this method of estimating tone scaling constitutes the most recent phonetic model assumed under the paradigmatic theory, it seems ideally suited to assessing predictions of that theory regarding phonetic categorization.

The second method for estimating tonal scaling involved determining a ratio of F0 values. This method was chosen based on the prediction of the syntagmatic view that the F0 level of a tone is scaled primarily with respect to the level of a preceding tone. This method made intuitive sense to us by analogy with music. For example, singers producing a melody in different musical keys determine the F0 level of each succeeding note by scaling it with respect to the preceding note or notes according to a ratio of F0 values. Tonal scaling was therefore calculated according to the following formula:

$$f_2(T) = \frac{T_2}{T_1} \tag{7.2}$$

It can be seen that this formula resembles the structure of a syntagmatic tone interval. Here, T_2 and T_1 corresponded to the levels of two tones in sequence. Thus, in this method, tonal scaling was determined as a ratio of two tones. We will refer to this method of estimating F0 as the *relative level method*.

F0 estimates of h, r, T_0, T_1 and T_2 were therefore determined for each utterance, as follows:

h: h was taken as the maximum F0 value for each utterance. For the *oregano* series, we were interested in the scaling of the unstressed syllable(s) relative to either (i) the F0 maximum on the stressed syllable *-reg-* plus the F0 minimum for each utterance, for the pitch range method, or (ii) the F0 maximum on the stressed syllable *-reg-* for the relative level method. For the *oranges* series, we were interested in the scaling of the stressed syllable *or-* relative to either (i) the F0 maximum on the unstressed syllable *-an-* plus the F0 minimum for each utterance, for the pitch range method, or (ii) the F0 maximum on the unstressed syllable *-an-* for the relative level method. For the *linguistics* series, we were interested in the scaling of the unstressed syllable *ling-* relative to either (i) the F0 maximum and the F0 minimum for each utterance, for the pitch range method, or (ii) the F0 on the stressed syllable *-guis-* for the relative level method. Specific choices for measurement points are noted below.

r: For the *oranges* and *oregano* series, r was the minimum modal F0 value on the last syllable of the imitation. If there was no modal value available, the value of r for a given imitation was taken to be the average of remaining estimates of r. For the *linguistics* series, r was the global F0 minimum for modal portions of the utterance.

T_0: For the *oregano* series, T_0 was the minimum modal F0 value across *Some or-*. The choice of a minimum value was consistent with the assumption that this value would reflect the level of a L tone of L+H*, if it was present.[96] For the *oranges* series, T_0 was the average F0 across *or-*. This was consistent with the assumption that the L* of L*+H is localized to a stressed syllable. For the *linguistics* series, T_0

was the maximum F0 value on *ling-*. This was consistent with the assumption that the F0 maximum was a good estimate of %H, if present.

T_1: For the *oranges* and *oregano* series, T_1 was the average F0 level on *or-*. Subjects produced a relatively more complex set of curve shapes for the *linguistics* series, so the choice of T_1 depended on the shape of the curve. When the F0 on *-guis-* was concave down, T_1 was the maximum F0 on *-guis-*. When the F0 on *-guis-* was concave up, T_1 was the minimum F0 on *guis-*. This is consistent with findings that F0 minima and maxima, respectively, are quite salient perceptually (e.g., d'Alessandro and Mertens, 1995).

T_2: For the *oranges* and *oregano* series, T_2 corresponded to the level of the F0 maximum (i.e., on *an-* or *e-*, respectively, for the *oranges* and *oregano* series). For the *linguistics* series the choice of T_2 depended on the shape of the curve. When the F0 on *ling-* was concave down, T_2 was the maximum F0 on *ling-*. When the F0 on *ling-* was concave up, T_2 was the minimum F0 on *ling-*.

Estimates of F0 at a maximum or minimum were determined automatically using Praat software and checked by hand for accuracy. If the F0 peak or valley in an imitation did not occur within the stressed syllable of *oregano* (for the *oregano* series), or during the first unstressed syllable of *oranges* (for the oranges series), then the token was discarded. Moreover, estimates of F0 which corresponded to an average level across some region were determined by first determining segmental boundaries between target phonemes in subjects' productions.[97] We made note of voicing irregularities which would disrupt F0 estimates; these were identified and classified using the acoustic and perceptual criteria described in Redi and Shattuck-Hufnagel (2001). If non-diplophonic voicing irregularities interrupted phonemes or phoneme sequences from which F0 estimates were taken, then the longest modal or diplophonic portion of the phoneme or phoneme sequence was used as the basis for the F0 estimate, or else the token utterance was discarded from the analysis. For portions of the speech that were pitch-halved due to diplophonia, F0 estimates were multiplied by a factor of 2.

Subjects showed individual differences in ability to perform the imitation task. In order to select the most consistent set of responses, bivariate correlation analysis was conducted on the average of the normalized F0 levels calculated from each of the three productions given in response to each stimulus. This analysis was carried out separately on data resulting from each normalization method for each stimulus series. We discarded data from subjects who were judged to be unreliable imitators, i.e., those individuals who were uncorrelated with 50% or more of the other subjects at $\alpha = 0.05$. For analyses corresponding to $f_1(T)$, bivariate correlation analysis resulted in two subjects being discarded for the *oregano* series, and one subject each for the *oranges* and *linguistics* series, out of 17. For analyses corresponding to $f_2(T)$, the correlation analysis resulted in a single subject being discarded from the *oranges* series, and three subjects each being discarded from the *oregano* and *linguistics* series, out of 17.

II. Estimates of inter-subject variability. The second metric of interest for determining the relative importance of the two candidate phonetic dimensions was inter-subject variability. We assumed that if a particular dimension constituted the phonetic basis of tonal scaling in all subjects' imitations, then estimates of that dimension should be relatively consistent across speakers. In contrast, a dimension which did not constitute the phonetic basis of tonal scaling in subjects' imitations should be relatively more inconsistent across speakers. In order to determine which tonal scaling estimate more closely reflected the process by which subjects were scaling tones, we calculated an *average coefficient of variation, cv_{avg},* for each pitch range measure, $f_1(T)$ and $f_2(T)$. The coefficient of variation is a standard metric for determining variability in a data set. It is defined as the sample standard deviation divided by the sample mean, and it therefore corresponds to a normalized measure of variability. One value of cv_{avg} was calculated for all responses to each stimulus, collapsing across subjects. All subjects' responses were included in estimates of cv_{avg}.

7.1.3 Results

The results obtained for the imitation task in response to each of the stimulus series are discussed here in turn. For each series, we will first present the estimates of tonal scaling, and then we will discuss the estimates of inter-subject variability.

oregano series: Figures 7.2 and 7.3 show the tonal scaling of the initial unstressed syllable(s) in imitations of the phrase *Some oregano*, calculated according to the pitch range method, $f_1(T)$ (Pierrehumbert and Beckman 1988), and according to the relative level method, $f_2(T)$, respectively. In both figures, open squares are median values, and the whiskers give the semi-interquartile range (SIQR). Data in both figures show a gradual, continuous trend which appears to be quite linear.[98] Neither figure exhibits the sort of discontinuous trend seen in earlier experiments. (See e.g. Figures 6.6, 6.9, etc.) Moreover, there is a fairly large spread in data values across all stimuli for both $f_1(T)$ and $f_2(T)$, with no obvious clustering of points.

Next, Figure 7.4 shows estimates of inter-subject variability measure, cv_{avg}, for each of the two methods of estimating tonal scaling. Error bars correspond to the standard error for cv_{avg}. Across all stimuli, cv_{avg} is four to eight times higher for estimates of tonal scaling based on the pitch range method, $f_1(T)$, than estimates based on the relative level method, $f_2(T)$. In addition, cv_{avg} is approximately constant across all stimuli for the relative level method, while there is a general increase in values for the pitch range method as stimulus number increases.

oranges series: Figures 7.5 and 7.6 show the tonal scaling of the stressed syllable *or-* in imitations of the phrase *Some oranges*, calculated according to the pitch range method and according to the relative level method, respectively. The data in both figures show a gradual trend which is approximately linear. Once again, neither figure exhibits the sort of discontinuous trend seen in earlier experiments. Moreover, there is a fairly large spread in data values across all stimuli for both $f_1(T)$ and $f_2(T)$, with no obvious clustering of points.

Next, Figure 7.7 shows estimates of inter-subject variability measure, cv_{avg}, for the two methods of pitch range scaling. Across all stimuli, cv_{avg} is approximately two to five times higher for the pitch range method, $f_1(T)$, than the relative level ratio method, $f_2(T)$. Moreover, cv_{avg} for the relative level method is consistently low and approximately constant across all stimuli, while there is an increasing trend in cv_{avg} for the pitch range method as stimulus number increases.

linguistics series: Figure 7.8 shows the tonal scaling of the unstressed syllable *ling-* in imitations of the phrase *Linguistics?*, calculated according to the pitch range method. The data exhibit a gradual downtrend across the stimulus series. Median values also reflect a general downtrend, but the trend is not as smooth as for data in Figures 7.2 and 7.5. There is no apparent clustering of data points or any clear evidence of a discontinuity in the range of stimuli 1-9. However, there is some evidence of clustering of data points in the range of stimuli 12-15.

Figure 7.9 shows the tonal scaling of the unstressed syllable *ling-* calculated according to the relative level method. When the imitations are plotted according to this method, there is evidence of a discontinuous shift in pitch range scaling at stimulus 12, corresponding to the position in the stimulus series where the relative pitch levels of the unstressed syllable and following stressed syllable undergo a reversal. Moreover, there is clustering of data points in stimulus ranges 1-11 and 12-15. Inspection of Figure 7.8 again shows vestiges of this discontinuity in the region of stimuli 8-13, consistent with the data in Figure 7.9.

Next, Figure 7.10 shows estimates of the inter-subject variability measure, cv_{avg}, for the two methods of pitch range scaling. The value of cv_{avg} is up to three times higher for the pitch range method, $f_1(T)$, than for the relative level method, $f_2(T)$, across most of the stimulus series. There is some increase in values of cv_{avg} for the relative level method across the series, but these values are still smaller than cv_{avg} for the pitch range method for nearly all stimuli.

7.1.4 Discussion

The present experiment tested competing predictions of two theories regarding the primary phonetic dimension most relevant to phonological category distinctions in English intonation. In the following, we briefly review the predictions and claims of each theory. We then summarize the results obtained from each of the three stimulus series for the two metrics which were deemed to be relevant to evaluating theoretical predictions. For each stimulus series, we subsequently consider the implications of the experimental results for each theory and for the phonological description of English in light of the results obtained.

The paradigmatic theory of English intonation entails the claim that the primary phonetic dimension along which intonational categories are distinguished is the position in the speaker's pitch range. Each of the three stimulus series was predicted by the paradigmatic theory to give rise to two distinct phonological categories, based on written descriptions of the tonal categories. To test these claims, estimates of tonal scaling were calculated from imitation data using the pitch range method, based on the formula proposed by Pierrehumbert and Beckman (1988).

On the other hand, the syntagmatic view claims that the primary phonetic dimension along which intonational categories are distinguished is the relative pitch level of timing positions (here, syllables) with respect to one another. Under this view, the *linguistics* series was predicted to give rise to two categories, and a break point in production data was predicted to occur at stimulus 12. On the other hand, the *oregano* and *oranges* series were predicted to give rise to a single category. To test these predictions, estimates of tonal scaling were calculated from imitation data based on the relative level method, which involved determining a simple ratio of F0 values.

We now turn to a discussion of the results from the *oregano, oranges,* and *linguistics* series. For each series, we will consider first the implications of results of estimates of tonal scaling. Then we will consider the implications of inter-subject variability analyses.

oregano series

The results for the *oregano* series failed to support claims of the paradigmatic theory of English intonation that this series traverses two distinct phonological categories. Production data showed gradient changes across the stimulus continuum when plotted according to two methods of estimating tonal scaling, thus failing to show evidence of a category boundary. By extension, these results also fail to lend empirical support for the claimed distinction between the pitch accent categories L+H* and H*.

The present results support the syntagmatic view by suggesting that the *oregano* stimulus series corresponds to a single phonological category. This view predicts that the relative level of tones is the important phonetic dimension which distinguishes one phonological category from another. Given that the *oregano* stimuli each involve initial unstressed syllables which are lower than a stressed syllable, the syntagmatic view predicts that this series should give rise to only a single category. Consistent with this prediction, subjects appeared to interpret these stimuli as involving a single phonological category plus within-category phonetic variation. The gradient variation produced by subjects for this stimulus series can be likened to small but audible differences in the degree of rounding in productions of English /o/.

Further support for the syntagmatic view, but not the paradigmatic view, comes from the fact that inter-subject variability was significantly less when tonal scaling was plotted relative to another tone, rather than relative to the speaker's pitch range. It was argued earlier that greater consistency should obtain across subjects' productions when the data were plotted using parameters which were based on the phonetic dimension (pitch range or relative level) which most closely reflects the phonetic parameters that speakers actually use to scale tones. Using a standard, normalized statistical measure to estimate inter-subject variability, four to eight times more consistency was observed in subjects' production data when parameters were based on relative level than on pitch range. This suggests that relative level is used by subjects to scale tones, consistent with the syntagmatic view.

These findings support and extend those of Ladd and Schepman (2003), who also failed to find evidence of a phonetic basis for a category distinction between H* and L+H*. In their study, Ladd and

Schepman focused primarily on one dimension which supposedly distinguishes H* and L+H*, namely, the timing of the F0 minimum. According to Pierrehumbert (1980), when an F0 minimum between high accents is high in the pitch range, it reflects a nonmonotonic, phonetic interpolation function or a "sagging transition" between adjacent H* accents. This predicts that the minimum should be positioned about halfway between the accents, regardless of the number of syllables intervening. On the other hand, Pierrehumbert also claimed that when an F0 minimum is low the pitch range, it reflects the L tone in a L+H* accent; this L is expected to be consistently aligned with a position preceding the high accent. Ladd and Schepman examined the alignment of F0 minima between high accents, for minima which were high in the pitch range. Contrary to the predictions of Pierrehumbert (1980), Ladd and Schepman found that the F0 minimum did not occur halfway between the high accents. Rather, it was consistently aligned with a weak syllable just before a high accented syllable, suggesting that it behaved like a L tone. Ladd and Schepman's data thus failed to support the prediction of Pierrehumbert (1980) that the timing of the F0 minimum distinguishes L+H* and H*.

The present findings extend Ladd and Schepman's results by failing to show evidence of a category boundary along a second phonetic dimension that supposedly distinguishes H* and L+H*: the F0 level of unstressed syllables within a speaker's pitch range. In this study, when F0 level was varied continuously from a low pitch to a high pitch, subjects imitated the resulting contours in a manner suggestive of gradient phonetic detail. These results, together with those of Ladd and Schepman, suggest that F0 minima before high, accented syllables behave in all cases like low tones. These results then present an explanation for the longstanding observation that H* and L+H* are notoriously difficult to distinguish auditorily, even by trained listeners. That is, it appears that these two accents are simply phonetic variants of a single category.

Ladd and Schepman suggested merging Pierrehumbert's H* and L+H* accents into a single descriptive category, treating F0 dips before high accents in all cases as low tones. By assuming that all pitch interpolation functions are monotonic, the present proposals add substance to Ladd and Schepman's suggestions. Previously, low-pitched F0 minima could arise either through phonetic interpolation or via a low tone, leading to significant ambiguity. The assumption of the present system that all pitch interpolation functions are monotonic, however, entails that all low-pitched F0 minima correspond to tonal targets, effectively eliminating the earlier ambiguity. In the present system, a low pitched tone preceding a high tone are described by a single tone interval in which the starred tone is specified to be *higher* than the unstarred tone. Consistent with naming conventions for tone intervals given earlier, the contours in Ladd and Schepman's study would be treated here as $H_{(0)}{*}\ L{+}\ H{*}$.

The contours in the *oregano* stimulus series bore both similarities and differences with those of Ladd and Schepman's study. These contours were similar to those of Ladd and Schepman's study in crucial ways that permitted comparisons of their results and ours. In particular, both sets of stimuli exhibited low F0 preceding a high accent. However, the present system draws a distinct yet overlapping set of categories compared with Pierrehumbert's system. In particular, we assume that a sequence of two unstarred tones may begin a phrase, consistent with the slightly rising but still low F0 across *Some or-*. Under this analysis, it can be noted that the second of the two unstarred tones is slightly higher, thus engendering a small rise. The naming conventions described earlier then indicate that the initial rising portion across *Some or-* should itself be described as L_0 H+, so that the entire contour would be described as $L_0\ H{+}\ H{*}\ L\%$ across *Some oregano*. Although the naming conventions for tone intervals would tend to emphasize a difference between our stimuli and those of Ladd and Schepman, the critical similarity to their stimuli was the fact that a low pitched region preceded a high accent. It was precisely this similarity which permitted a comparison of our results.

In summary, the present results failed to support the paradigmatic view that the *oregano* stimulus series traverses two categories, H* and L+H*. In contrast, the gradient responses produced by subjects are consistent with a single phonological category, as predicted by the syntagmatic view. These results support and extend the findings of Ladd and Schepman by demonstrating that a second phonetic dimension along which H* and L+H* have been claimed to be distinguished, namely F0 level, apparently fails to elicit evidence consistent with two distinct categories. Finally, the theoretical framework outlined in Chapters 2 and 3 lays the groundwork necessary for accommodating Ladd and Schepman's suggestion that Pierrehumbert's H* and L+H* descriptive categories be merged into a single category.

oranges series

For the *oranges* series, production data failed to lend support to the paradigmatic theory's claims that the *oranges* stimulus series traverses two phonological categories, L*+H and H*. Rather, plotting production data according to either of two methods of estimating tonal scaling showed gradient change across the stimulus continuum, consistent with a single category. These findings therefore failed to lend support to a key prediction of the paradigmatic theory, namely, that the position of a syllable in the speaker's pitch range for the *oranges* series distinguishes two phonological categories.

While the predictions of the paradigmatic theory were not borne out, the predictions of the syntagmatic theory appeared to be confirmed. The syntagmatic view predicted that the *oranges* series should correspond to a single phonological category, since all stimuli in the series involved the same relative height relation between targets on *Some or-* and on *-an-*; the latter in all cases was higher than the former. Consistent with this prediction, gradient change was observed in tonal scaling across the stimulus series. These results therefore supported the claim of the syntagmatic view that the relative levels of tones, and not their positions in the pitch range, comprise the dominant phonetic basis of phonological distinctions.

Results from estimates of inter-subject variability also lend support to the syntagmatic view, but not the paradigmatic view. When estimates of tonal scaling were based on the relative levels of tones, subjects' imitations appeared to be two to five times more consistent than when estimates were based on the positions of tones in the pitch range. The fact that subjects apparently used relative tonal level to scale their imitations, rather than correlates of the speaker's pitch range, is consistent with the syntagmatic view's claim that relative F0 level is the most significant dimension of F0 variation for the phonology. This provides further support in favor of the syntagmatic view and against the paradigmatic view.

What are the implications of these results for accentual categories in English assumed under the paradigmatic theory? One interpretation of these results is that the mappings from phonetics to phonology claimed under the paradigmatic theory do not fully correspond to what speakers of English have in their heads. The paradigmatic theory claims that the position of a syllable in the speaker's pitch range distinguishes both H* from L*+H, as well as H* from L+H*. However, both the *oranges* and *oregano* series failed to support these claims about the phonology-phonetics relation, since the evidence was consistent with only a single category in each case. These results therefore concur with the earlier results from Experiments 1-4 by suggesting that the manner by which speakers actually divide the phonetic space into phonological categories is not captured in the paradigmatic theory's mapping from phonetics to phonology.

How do the results from the *oranges* series bear in particular on the L*+H descriptive accentual category of the paradigmatic theory? Two sets of phonetic characteristics have been claimed to distinguish L*+H from other categories: the position of the stressed syllable in the speaker's pitch range, and the alignment of a F0 peak. Earlier experimental results by Pierrehumbert and Steele (1989) had lent support for L*+H by showing that contours described in terms of this accent type were apparently treated as distinct from contours described as L+H*. However, Pierrehumbert and Steele exclusively investigated the role of F0 peak alignment in phonological contrast; until now, claims about the phonological role of the position of a syllable in the speaker's pitch range have gone untested. The results of Pierrehumbert and Steele's experiment and the present results can be explained by assuming that F0 peak (and valley) alignment is a much more important factor in phonological contrasts in English intonation than the position of a syllable in the speaker's pitch range. Moreover, they suggest that the important characteristic of contours described as L*+H under the paradigmatic system was the alignment of the F0 peak for the +H with respect to the L*.[99]

The syntagmatic theory predicted the present empirical findings. Recall that this view emphasizes the importance of relative pitch level as the single most important phonetic factor in phonological distinctiveness. The theory explains the significance of F0 turning point alignment for phonological distinctions in terms of underlying differences in the relative levels of tones. It should be noted that in the *oranges* stimuli, F0 peak alignment, and hence relative pitch level, was fixed across the entire stimulus series, such that the level of the unstressed syllable was always higher than an immediately preceding stressed syllable.

The syntagmatic theory laid out in preceding chapters describes contours such as those in the *oranges* series which show a rise from a stressed syllable to a late F0 peak as a sequence T^* $+H$. Given the naming conventions described earlier, $+H$ indicates that the two tones share the relation *higher*; the $+H$ could either occur late in the stressed syllable or possibly on a following weak syllable. If the stressed syllable were initial in its phrase, the relation higher shared by the two tones would then be described L_0^* $+H$; this is the approximate equivalent of the L*+H category in the paradigmatic theory. However, the stressed syllable is preceded in *oranges* stimuli by an initial unstressed syllable, such that the unstressed-plus-stressed syllable sequence incurs a small rise. This indicates that the relation between *Some* and *or-* is also *higher*. As a result, the *oranges* stimuli are described in the present framework as L_0+ H^* $+H$ $L\%$ across the phrase *Some oranges*.

In summary, the present results failed to support the paradigmatic view that the *oranges* stimulus series traverses two categories, H* and L*+H. In contrast, the gradient responses produced by subjects are consistent with a single phonological category, as predicted by the syntagmatic view.

linguistics series

The paradigmatic and syntagmatic theories of English intonation made competing predictions about the location of a category boundary along the *linguistics* series. On the one hand, the paradigmatic theory describes the upper end of the *linguistics* continuum as corresponding to an initial high boundary tone, %H, and the lower end of the continuum as corresponding to a tonally unmarked case. Written descriptions of each tonal configuration have made explicit that in the case of both %H and in the tonally unmarked case, the level of an initial unstressed syllable is higher than a following stressed syllable. The difference between the two has been claimed to rest in the height of the initial unstressed syllables (with %H being higher in the pitch range than the unmarked case). The paradigmatic theory thus predicted that a category boundary should fall somewhere in the range of stimuli 1-11, since for this range of stimuli the initial unstressed syllable is higher than the stressed syllable. A more specific prediction about the location of a category boundary along this series was not possible, however. On the other hand, the syntagmatic view also predicted two categories for the *linguistics* series, where stimulus 12 was specifically predicted to correspond to the location of a category boundary. The reason was that stimulus 12 marked the point along the *linguistics* series that the pitch level of the initial syllable changed from higher than the following syllable to lower than that syllable.

The results from the *linguistics* series provide support for the syntagmatic view but not the paradigmatic view. Contrary to the predictions of the paradigmatic view, there was no evidence of a category boundary in production data in the range of stimuli 1-11. Rather, evidence consistent with a category boundary was observed at stimulus 12, as predicted by the syntagmatic view. This is most apparent when the data are plotted according to the relative level method. However, vestiges of the apparent category boundary can also be seen when the data are plotted according to the pitch range method. These results support the claim of the syntagmatic view that the most important phonetic dimension for category distinctions is the relative level of a tone with respect to that of another tone.

Data regarding inter-subject variability for this series also provide support for the syntagmatic view. As in the *oranges* and *oregano* series, subjects' imitations revealed substantially more consistency when tonal scaling estimates were based on relative F0 level, rather than on the level of syllables in the pitch range. This is expected under the syntagmatic view, which claims that the primary phonetic dimension relevant to phonological distinctions is relative pitch level.

Evidence of two categories for the *linguistics* series can readily be accommodated within the syntagmatic framework described in Chapters 2 and 3. The apparent category boundary occurred at stimulus 12, which corresponded to the location where the level of the initial syllable changed from higher than the following stressed syllable to lower than that syllable. This is captured by different tonal analyses within this framework. When the stressed syllable is lower than the preceding initial unstressed syllable, as in stimuli 1-11, the tone interval shared by the initial unstarred tone and following starred tone is *lower*. Thus, stimuli 1-11 are described as H_0+ L^* $H\%$. Similarly, when the stressed syllable is higher than the initial unstressed syllable, as in stimuli 12-15, the corresponding tone interval is *higher*. Thus, stimuli 12-15 are described as L_0+ H^* $H\%$.

To summarize, the present results supported the predictions of the syntagmatic view of English intonation, but not the paradigmatic view. Consistent with the syntagmatic view, evidence of an apparent category boundary was obtained precisely at the location in the series where the initial unstressed syllable changed from being higher than to being lower than a following stressed syllable. It is not clear how the paradigmatic theory could accommodate these results.

7.1.5 Experiment 5 Conclusions

The present experiment showed three lines of support for the syntagmatic view. First, the syntagmatic view accurately predicted the number of apparent categories for each stimulus series: one each for the *oregano* and *oranges* series and two for the *linguistics* series. Second, the syntagmatic view accurately predicted the precise location in the *linguistics* stimulus series where the apparent category boundary occurred. Finally, imitation data showed substantially more consistency when tonal scaling estimates were based on relative F0 level than on position in the speaker's pitch range, consistent with the claim of the syntagmatic view that relative F0 level is the most important phonetic dimension for category distinctions.

In contrast, the present experiment failed to show support for category distinctions predicted by the paradigmatic view in three separate stimulus series. Each stimulus series was predicted to correspond to a distinction between tonal categories under the paradigmatic theory of English intonation: H* vs. L+H*, H* vs. L*+H, and %H vs. unmarked for the *oregano*, *oranges*, and *linguistics* series, respectively. The present experiment thus demonstrated for three separate stimulus series that the position of a syllable in the speaker's pitch range does not give rise to categorical distinctions in the manner predicted under the paradigmatic theory.

The results from Experiment 5 suggest a generalization about the mapping from phonetics to phonology in English intonation. Whenever the relative height of a tonal element *changed* with respect to another tonal element (so that e.g. a tone went from higher than another element to lower than that element), evidence consistent with a category boundary was observed. Conversely, whenever the relative height of a tonal element *remained fixed* with respect to another tonal element (thereby affecting only the distance between two tones or their position in the speaker's pitch range but not whether they were higher or lower than one another), evidence consistent with a category boundary *failed* to be observed. This suggests that differences in relative pitch level of adjacent tonal elements may be a consistent phonetic cue to distinct phonological categories in English.

It was suggested earlier that differences in F0 alignment, which elicited categorical effects in Experiments 1-4, can also be viewed in terms of differences in the relative pitch levels of tones. In Experiments 1-4, varying the timing of a F0 peak or valley across a SW or SWW syllable sequence consistently gave rise to evidence suggestive of category distinctions.[100] Because shifting a peak or valley across a syllable sequence entails a change in the relative pitch levels of those syllables, differences in peak or valley timing might interpreted by listeners in terms of categorical differences in the relative levels of tones. Experiment 6 tested the hypothesis that the relative pitch levels of syllables and the alignment of F0 peaks and valleys are really two sides of the same coin.

7.2 Experiment 6

Experiment 6 tested the possibility that earlier results from Experiments 1-4 showing the importance of F0 peak and valley alignment for phonological contrast could be reinterpreted in terms of relative pitch level. If so, then this would support a key prediction of the syntagmatic view, namely, that relative pitch level is the dominant phonetic cue to phonological contrast in English.

7.2.1 Background and hypotheses

One of the most significant and robust findings concerning the phonetics of intonation in the past three decades concerns the fact that F0 peaks and valleys are consistently aligned with the segmental string, and that this alignment is categorical (Bruce 1977, Kohler 1987, Pierrehumbert and Steele 1989, Arvaniti *et al.*, 1998, Ladd et al., 1999, 2000, Ladd and Schepman 2003). These findings have provided empirical support for theories of intonation which assume that F0 contours arise from the concatenation of sequences of discrete level pitch targets connected by interpolating functions. In contrast, these findings have provided evidence against a class of theory which assumes that F0 contours arise from the concatenation of sequences of dynamic pitch movements. These issues were reviewed earlier in Chapter 4.

Three theories based on discrete level pitch targets could potentially explain evidence of consistent F0 turning point alignment. However, previous empirical work has not made it possible to distinguish among these views. The goal of Experiment 6 was to determine which of these three theories best explain findings concerning consistent alignment of F0 turning points. The manner in which each of the three views might potentially explain F0 turning point alignment is discussed in turn below.

One possible interpretation of the earlier findings was that manipulating F0 turning point alignment across a two-syllable sequence affected phonological categorization because a second phonetic correlate, namely the relative F0 levels or relative pitch levels of the syllables, was affected as well. The earlier experiments did not make it possible to determine which of these two phonetic cues was responsible for differential phonological categorization of stimuli. Experiment 6 tested a prediction of the syntagmatic view that relative pitch level should be the dominant phonetic cue to phonological contrast. If so, then eliminating cues to the temporal alignment of F0 turning points should present no harm to the effectiveness of the phonological contrast. This is precisely what was demonstrated in Experiment 6.

First, the syntagmatic view explains consistent F0 turning point alignment by proposing that the phonological representation of tone is universally based on relational tonal features: ***higher***, ***lower***, and ***same***. Under this view, evidence of categorical effects in production from Experiments 1-4 is explained in terms of distinct relative pitch levels of segments. As the alignment of F0 peaks and valleys was shifted through the segmental string, participants attended to the relative pitch levels of segments, interpreting these relative levels in terms of underlying relational features between tones. Imitation then involved attempting to make certain segments e.g. sound higher or lower than other segments, with phonetic interpolation occurring across phonologically non-significant segments. According to this explanation, then, an F0 peak or valley arises from an attempt to make a segment or segments sound higher or lower, respectively, than adjacent segments. In other words, differential alignment of F0 peaks and valleys is understood in terms of differences in the relative pitch levels of segments. The syntagmatic view thus presents a straightforward explanation of F0 alignment, as discussed in Chapter 4.

A second theoretical approach which potentially provides an explanation for F0 turning point alignment is the paradigmatic theory. This theory only *potentially* explains consistent F0 turning point alignment, since no version of the theory has been articulated to date for which the parameters are sufficiently constrained to predict the output phonetic turning points from the input phonological string. These issues were discussed in Chapter 4. The paradigmatic theory states that the phonological representation is based on a paradigmatic distinction between H and L, where H is higher than L would be in the same position. This view claims that the position of H and L in the speaker's pitch range is the phonetic parameter of interest. A widespread assumption in the literature is that an F0 peak corresponds to a H tone, while an F0 valley corresponds to a L tone. However, there is no explicit phonetic model which has been worked out which codifies this view of the relation between phonetics and phonology, as discussed in Chapter 4.

A third view which may explain consistent F0 peak and valley alignment is the theory of Xu and colleagues (Xu and Wang, 2001, Xu and Xu, submitted). In this theory, the phonological representation is based on two types of articulatory targets: static targets (high, low) and dynamic targets (rising, falling); for this reason we will refer to this theory as the *articulatory view*. The articulatory view has been applied to both Mandarin and English. This theory assumes that there is a phonological target on every syllable. Moreover, the phonetic correlates to target identity are assumed to be the contour shape of a syllable, and the timing of the maximum or minimum. There is no express provision for the role of pitch perception.

Of these three theories, only the syntagmatic view predicts that relative pitch level can explain consistent F0 peak or valley timing. If relative pitch level is indeed the critical phonetic cue to distinct phonological categories, then it should be possible to elicit the sorts of categorical effects observed in earlier production experiments using stimuli in which relative pitch level has been manipulated while cues to F0 peak and valley alignment have been eliminated. If categorical effects in production can be elicited through manipulation of relative pitch level alone, it would provide support for the syntagmatic view, since this theory alone predicts such a result.

Experiment 6 specifically tested this prediction of the syntagmatic theory. In this experiment we eliminated cues to F0 peak and valley timing from stimuli used in an imitation task, so that there were no overt cues to the timing of peaks and valleys. In addition, we eliminated all local cues to F0 contour shape, replacing key syllables with level F0 contours. In this way, relative pitch level was manipulated while cues to the presence and timing of F0 peaks and valleys were eliminated.

Consider the schematic F0 contours in Figures 7.11(a) and (b), which represent the stimuli used in this experiment. In these figures, the naturally occurring F0 contours across two syllables, labeled S for "strong" and W for "weak," have been replaced with locally flat F0 contours. Replacing F0 information across these two syllables with flat F0 contours eliminated local cues to the alignment of an F0 peak or valley.

Not only were cues to the alignment of an F0 peak or valley alignment eliminated from these stimuli; we also eliminated local cues to the *presence* of an F0 peak or valley. In order to do this, it was necessary to eliminate F0 transition information in key positions in stimulus materials. Consider the fact that while the S syllable labeled "1" in both Figures 7.11(a) and (b) is locally flat, the transitions to and from the S syllable first rise and then fall in 7.11(a). In contrast, the transitions to and from the S syllable first fall and then rise in 7.11(b). Because F0 transition information to and from target syllables could potentially be used to infer that there was originally a F0 peak in stimuli like those in Figure 7.11(a) and a valley in stimuli like those in Figure 7.11(b), it was necessary to eliminate this transition information, which was redundant with the presence of a peak or a valley. This was accomplished by splicing in noise before and after target S and W syllables. The portions of the F0 contour corresponding to transitions which were eliminated in this way is shown schematically in Figures 7.11(a) and (b) by dashed lines and shading.

Carrying out these operations eliminated all local phonetic cues both to the alignment, as well as the presence, of F0 peaks and valleys. If the categorical cues seen in earlier imitation experiments arose not from F0 alignment cues *per se* but from the perceived relative pitch levels of syllables, then the F0 contours of imitations produced in response to stimuli like those in Figures 7.11(a) and (b) should show two specific kinds of F0 characteristics. First, F0 peaks or valleys should be produced on target syllables, even though stimuli lack local cues to peaks and valleys due to the presence of flat F0 on these syllables. In particular, subjects should produce peaks in response to stimuli like those in Figure 7.11(a), but they should produce valleys in response to stimuli like those in Figure 7.11(b). Second, the alignment of F0 peaks and valleys should be predictable from the relative pitch levels of the S and W syllables. Then in response to stimuli like those in Figure 7.11(a) we expect that when the S syllable is higher in pitch than the W syllable, as for contour "1", the speaker should produce an F0 peak on the S syllable. On the other hand, we expect that when the W syllable is higher in pitch than the S syllable, as for contour "2", the speaker should produce an F0 peak on the W syllable.

The syntagmatic view thus makes a clear set of predictions regarding the outcome of this experiment. In contrast, neither the paradigmatic theory nor the articulatory view makes a prediction about the experimental outcome. This is due to the theoretical emphasis of both models on local phonetic characteristics associated with individual syllables. Neither theory thus codifies a role for relations *between* syllables in a way that makes possible a prediction in the case of this experiment. We turn now to a discussion of the experimental methods.

7.2.2 Methods

Participants: Participants were 13 students and staff at MIT or other colleges (2 men, 11 women). All were self-reported native English speakers with normal hearing, and all were paid a nominal sum for participation.

Stimuli: Four stimulus series were created based on the phrase *Some lemonade*, which are shown in Figure 7.12(a)-(d). The stimulus series shown in Figure 7.12(a) and (c) were based on a single utterance with an overall rising-falling intonation pattern typical of a statement, while the stimulus series shown in Figures 7.12(b) and (d) were based on a single utterance with an overall falling-rising intonation pattern typical of a question. In each of the four stimulus series, the F0 contour across each of the syllables in the two-syllable sequence *lemon-* was replaced with a level F0 contour. All F0 contours in synthetic speech materials were stylized using a sequence of straight lines using Praat software.

To create the synthetic speech stimuli shown in Figures 7.12(a) and (b) (the Roving-High series and Roving-Low series, respectively), each of the two syllables across *lemon-* was paired with a straight line with zero slope corresponding to one level in a 10-step series of F0 levels. Each successively higher level in this 10-step series corresponded to an increase of ½ semitone. For each of the stimuli in the Roving-High and Roving-Low series, different F0 levels were selected for the first and second target syllables. For example, the first stimulus in the Roving-High series was created by pairing the first target syllable with the highest of the 10 F0 levels and pairing the second target syllable with the lowest of the 10 F0 levels. The second stimulus in the Roving-High series was then created by pairing the first target syllable with the second highest of the 10 F0 levels and pairing the second target syllable with the second lowest of the 10 F0 levels, etc. The Roving-Low series was created using a similar method of pairing each of the two target syllables with distinct F0 values. In this manner, ten stimuli were created for both the Roving-High series and the Roving-Low series, such that each target syllable was paired with each of the 10 F0 levels exactly once.

For the Roving-High series, the first five stimuli (numbers 1-5) involved a higher F0 level on the first syllable compared with the second syllable, so that the maximum F0 value occurred “early”, i.e., on the first target syllable. Moreover, the next five stimuli (numbers 6-10) involved a lower F0 level on the first syllable compared with the second syllable, so that the maximum F0 value occurred “late”, i.e., on the second target syllable. For the Roving-Low series, by comparison, first five stimuli (numbers 1-5) involved a lower F0 level on the first syllable compared with the second syllable, so that the minimum F0 value occurred “early”, i.e., on the first target syllable. Moreover, the next five stimuli in this series (numbers 6-10) involved a higher F0 level on the first syllable compared with the second syllable, so that the minimum F0 value occurred “late”, i.e., on the second target syllable.

To create the synthetic speech stimuli shown in Figures 7.12(c) and (d) (the Fixed-High and Fixed-Low series, respectively), the F0 level of the first target syllable was held fixed, while the level of the second syllable was varied. In the Fixed-High and Fixed-Low series, the first syllable was assigned a flat F0 contour with a high or low F0 value (262 Hz and 202 Hz, respectively), which was fixed for all stimuli in the series. The level of the second syllable was then varied in six equal ¾ semitone steps to create six stimuli each for the Fixed-High and Fixed-Low series.

For stimuli 1-3 in the Fixed-High series, the second target syllable was lower than the first so that the global maximum F0 level occurred “early”, while for stimuli 4-6 the second syllable was higher and the global maximum F0 level occurred "late". For stimuli 1-3 in the Fixed-Low series, the second target syllable was higher than the first so that the global F0 minimum occurred "early", while for stimuli 4-6 the second syllable was lower than the first and the global F0 minimum occurred "late". Finally, the F0 levels of stimuli 3 and 4 in both "fixed" series were chosen such that the F0 level on the first syllable was equidistant on a logarithmic scale from each of these stimuli.

Once these files were generated, the portions of each of the waveforms corresponding to /l/, /m/ and /n/ were spliced out at zero crossings and replaced with Gaussian noise of identical duration. The perceptual effect was that a consonant sound similar to [s] was inserted.

Design and Task: The experimental design was nearly identical to that of Experiments 2, 4, and 5. Subjects were instructed that they would hear the phrase *Some lemonade*, as well as some noise. They were told to ignore the noise and to try to imitate the phrase that they heard as closely as possible. Due to the auditory similarity of stimuli from the Roving-High and Fixed-High series, these stimuli were grouped together and randomized as a single block. This was also done for the Roving-Low and Fixed-Low series.

Analysis: To determine the timing of F0 peaks and valleys in subjects' imitations, syllable boundaries were determined via visual inspection of spectrogram displays. The boundary between /m/ and /l/ was taken as the location of an increase in amplitude corresponding to the right edge of the nasal. When more than one discontinuity was observable, the discontinuity consistent with the greater amount of relatively low frequency energy was taken as the location of the boundary. The start and end of the /n/ were marked separately.

Responses from blocks corresponding to each stimulus series were analyzed separately. Temporal locations of F0 maxima and minima were determined by visual inspection of the F0 contour and marking using Praat. If no F0 peak or valley was found across *lemon-* for an utterance produced in response to the Roving- and Fixed-High or Roving- and Fixed-Low series, respectively, then the data point was discarded. Moreover, utterances for which the subject produced an intonation pattern with the wrong global contour were discarded, e.g. if the subject produced a rising, "question" intonation pattern in response to a statement. Utterances which were too monotone across the target two-syllable sequence for reliable determination of the temporal location of the F0 peak or valley were discarded from the analysis. Data from subjects for whom 50% or more of the utterances did not meet the above criteria was discarded. This resulted in one subject's data being discarded for the two High series, and three subjects' data being discarded for the two Low series. The temporal position of peaks and valleys was normalized with respect to *lemon-* using the normalization method described for Experiment 2.

As was done for Experiments 2 and 4, bivariate correlations were used to quantify the consistency of the pattern of productions across participants and stimulus series. Responses from subjects who were not correlated with 50% or more of the other subjects were judged to be unreliable imitators and not included in the final analyses. Based on these criteria, two subjects each were excluded from analysis for the Roving-High, Fixed-High, and Roving-Low series, while four subjects were excluded from the Fixed-Low series. This left 10, 10, 8, and 6 subjects each for Roving-High, Fixed-High, Roving-Low, and Fixed-Low series.

7.2.3 Results

Producing peaks versus valleys: Even though these stimuli contained no F0 peaks or valleys, subjects nevertheless produced F0 peaks or valleys in response to stimuli in particular stimulus series. With few exceptions, subjects produced peaks in response to the Roving- and Fixed-High stimulus series, and valleys in response to the Roving- and Fixed-Low series. Figures 7.13(a) and (b) show examples of typical contours produced in response to the first stimulus of the Roving-High Roving-Low series, respectively.

Timing of peaks and valleys: Subjects also systematically varied the timing of peaks and valleys in response to particular stimuli. Figure 7.14 shows normalized turning point location, T_N, for the Roving-High series. Open squares are median values and whiskers give the semi-interquartile range. It can be observed that for stimuli 1-5, the peak is timed relatively early, as shown by the relatively low value of T_N. There is a discontinuity in T_N values between stimuli 5 and 6, and then the peak is timed late for stimuli 6-10, as shown by the relatively high value of T_N.

Figure 7.15 shows normalized turning point location for the Roving-Low series. For stimuli 1-5, the valley is timed early. There is a discontinuity in T_N values between stimuli 5 and 6, and then valley is timed late for stimuli 6-10.

Figure 7.16 shows normalized turning point location for the Fixed-High series. For stimuli 1 and 2, the peak is timed consistently early, while for stimuli 5 and 6, the peak is timed consistently late. The aggregate shift from early-timed peaks to late-timed peaks is more gradual than for the preceding two series, with no evidence of a discontinuity.

Finally, Figure 7.17 shows normalized turning point location for the Fixed-Low series. For stimuli 1-3, the valley is timed early. There is a clear discontinuity in T_N values between stimuli 3 and 4, and then the valley is timed late for stimuli 4-6.

7.2.4 Discussion

The goal of Experiment 6 was to determine which of three theories based on discrete level pitch targets could best explain findings concerning consistent alignment of F0 turning points: the paradigmatic view of Pierrehumbert and colleagues, the syntagmatic view proposed here, or the articulatory view. The syntagmatic view alone predicted that relative pitch level, rather than F0 turning point alignment, is the crucial phonetic cue to phonological category distinctions. This prediction was tested by creating stimuli in which phonetic cues to the presence and alignment of F0 turning points were eliminated, replacing the F0 across key syllables with flat tonal contours. This permitted us to address several key theoretical issues.

The first question addressed by this experiment was whether subjects would use global, rather than strictly local, intonational cues, so that subjects produced F0 peaks and valleys on key syllables, even though those syllables had flat F0 patterns in the stimuli. This was clearly the case. Subjects used information from the global F0 pattern to produce F0 peaks in response to Roving-High and Fixed-High stimuli, and to produce F0 valleys in response to Roving-Low and Fixed-Low stimuli. That is, subjects were able to produce peaks in response to Roving-High and Fixed-High stimuli and valleys in response to Roving-Low and Fixed-Low stimuli, in spite of the fact that there were no local F0 shape cues to indicate whether a stimulus originally had a F0 peak or a F0 valley. This is consistent with the syntagmatic view, which predicted that subjects would utilize global cues to determine the relative pitch levels of syllables in context. However, this result is also consistent with the paradigmatic view, which interprets the fact that *lemon-* is high in the pitch range in the Roving-High and Fixed-High series as consistent with a paradigmatic High tone. Similarly, under the paradigmatic view the fact that *lemon-* is low in the pitch range in the Roving-Low and Fixed-Low series is consistent with a paradigmatic Low tone.

The second issue addressed by this experiment concerned whether categorical effects in F0 peak and valley alignment would be elicited when the relative levels of syllables were manipulated. This is precisely what was observed: subjects produced distinct and systematic patterns of F0 peak and valley alignment consistent with the relative F0 levels in the stimuli. In both High series, peaks were consistently timed early when the first syllable had a higher F0 level than the second syllable, and they were consistently timed late when the second syllable had a higher F0 level than the first syllable. Similarly, in both Low series, valleys were consistently timed early when the first syllable was lower than the second, and late when the second syllable was lower than the first.

The third issue which was addressed here was whether the same F0 level would elicit different responses, depending on the context. Although the acoustic properties of the syllable *le-* were always identical within each of the Fixed-High and Fixed-Low series, the responses of subjects to these syllables were very different, depending on the adjacent context. For example, subjects responding to the Fixed-High series produced a peak on *le-* when the following syllable was lower, but they failed to produce a peak on *le-* when the following syllable was higher. (In the latter case, the peak was produced by subjects on *-mon* rather than on *le-*.) This differential treatment of the same acoustic material cannot be explained by either the paradigmatic theory or the articulatory view. This is because both theories place theoretical emphasis on the phonetic properties associated with individual syllables, and neither theory articulates a role for context or for relative pitch level.

The finding that manipulating relative pitch level elicits categorical effects in F0 peak and valley alignment supports the claim the syntagmatic view that the crucial phonetic correlate in F0 categories is not F0 alignment *per se*, but relative pitch level. The influence of relative pitch level on subjects' responses can be seen in the fact that the discontinuities in production data corresponded exactly to locations in the stimulus series where there was a change in the relative pitch levels of two syllables. In three of four stimulus series, a robust discontinuity in turning point timing occurred at the point in the series where the relative F0 level of the first and second syllables reversed, so that the first syllable went from being higher than to lower than the following syllable, or vice versa.

The fact that breakpoints in production data occurred precisely when there was a change in the relative F0 level means that subjects must have compared the F0 levels perceptually in order to know which of the two syllables to produce the F0 peak or valley on. The paradigmatic view cannot explain these findings, since it proposes no explicit phonetic or phonological mechanism for the comparison of the levels of two adjacent syllables.

Another important contribution of this experiment was its demonstration that speakers are capable of producing F0 valleys with differential, categorical timing characteristics. This is the first experiment that we are aware of to demonstrate such a result. Earlier, Experiment 4 demonstrated categorical effects in F0 valley timing when subjects imitated stimuli in which F0 valleys had been shifted along a temporal continuum through a target syllable sequence. In this earlier experiment, however, F0 valleys appeared more resistant to categorical effects than F0 peaks, and the data for valleys were decidedly less "clean". A striking aspect of the present experimental results is that patterns of differential alignment of F0 valleys produced in response to relative pitch cues in the Roving-Low and Fixed-Low series were every bit as categorical as F0 peaks; indeed, they appeared to be even more categorical than the F0 peaks. Our explanation of this result is that clear and robust F0 valley alignment was elicited due to clear and robust phonetic cues in stimuli to the relevant categorical distinctions. If the crucial phonetic cues to category distinctions are relative pitch levels, as the syntagmatic view holds, then subjects would more readily infer the phonological representation of *lemon-*, since categorization is assumed to depend phonetically on relative pitch level. Because the pitches of *le-* and *-mon* relative to one another were presumably clearer in the present stimuli than when the F0 contour was dynamically falling to a minimum and then rising, subjects more clearly identified the categories, and the categorical effects were therefore more robust.

To summarize, the present experiment confirmed several key predictions of the syntagmatic view. First, subjects utilized global pitch cues in order to produce F0 peaks and valleys on key syllables, even though those syllables had flat F0 characteristics in the stimuli. This influence of global context is predicted by the syntagmatic view, but not by the paradigmatic view or the articulatory view. Second, manipulations of relative pitch level in stimuli elicited categorical effects in F0 peak and valley timing in subjects' productions, and the locations of break points in production data were predictable from the relative pitch levels of the syllables. This result is both predicted and explained by the syntagmatic view, while it is neither predicted nor explained by the other theories considered here. Third, even though the acoustic properties of the initial strong syllable were always identical within the Fixed-High and Fixed-Low series, subjects responded to these syllables in categorically different ways, depending on the context. While these results are explained and predicted by the syntagmatic view, they are neither explained nor predicted by the other theories. Finally, this experiment demonstrated that speakers are capable of producing F0 valleys with differential, categorical timing characteristics. This can be explained if the crucial phonetic cue to categorical distinctions was more robust in the present experiment than in experiments. The syntagmatic view holds that this cue is relative pitch level. These results, therefore, collectively support the syntagmatic view, while neither the paradigmatic view nor the articulatory view can predict nor explain the outcome of this experiment. This experiment therefore supports the claim that English represents syntagmatic tonal features.

7.3 General discussion and conclusions

We will pause to briefly summarize the results of Experiments 5 and 6. In Experiment 5, we tested the paradigmatic view's claims regarding three contrasts, where the claimed categories were H* vs. L+H*, H* vs. L*+H, and %H vs. unmarked. The paradigmatic view claims that there is a paradigmatic distinction in level between the two categories. When a syllable or syllables were shifted through the pitch range, we failed to see any evidence of break points in production data at the predicted locations. However, there was a break point in production data precisely at a point where one syllable shifted from higher than the following syllable, to lower than that syllable. While these results failed to provide support for the paradigmatic view, they were nevertheless consistent with the syntagmatic view.

Experiment 6 tested the hypothesis that evidence of categorical effects in F0 peak and valley alignment can be reinterpreted in terms of relative pitch level. Moreover, the experiment distinguished among three current theories capable of explaining earlier data on such categorical effects by eliminating cues to the presence and alignment of F0 peaks and valleys and instead expressly manipulating relative pitch level. Subjects varied the timing of peaks and valleys in a manner which was systematically correlated with the relative F0 levels of syllables. These results were consistent with the predictions of the syntagmatic view, while the paradigmatic view and the articulatory view were each unable to account for the findings in this experiment. Most importantly, the results of Experiments 5 and 6 therefore each provided support for the syntagmatic view while failing to provide support for the paradigmatic view. In the following chapter, we draw conclusions from the results of Experiments 1-6 and summarize the contributions of this thesis.

Chapter 8 – Contributions and future work

This work proposes a theory of tone and intonation which provides a new explanation for why tonal properties alternately appear to be "segmental" in some languages and localized to individual segments, but "suprasegmental" in others and realized diffusely across multiple segments. It was claimed that this duality arises by the fact that phonological representations for tone are based on a construct known as a *tone interval*. A tone interval is an abstraction of a frequency ratio; it defines a relationship of relative height and/or distance between a tone and a referent. We showed that tone intervals permit two kinds of tonal relationships to be defined, thereby giving rise to the alternately "segmental" and "suprasegmental" character of tone. On the one hand, when the referent is another tone, a tone interval defines a syntagmatic relation between two tones. On the other hand, when the referent is the speaker's tonic, a tone interval defines a paradigmatic relation between a tone and the tonic, which is a characteristic of a speaker's pitch range. In the following, we will briefly review some of the major results of this work.

It was shown that defining syntagmatic features in the phonology provides a way of defining crucial restrictions on the relative heights of tones which are necessary for the prevention of certain theoretical weaknesses, such as overgeneration of phonetic contours and indeterminacy of phonological representations. It was shown formally that theories of the phonetics and phonology of English intonation based on paradigmatic primitives (Pierrehumbert 1980, Pierrehumbert and Beckman 1988) do not provide sufficient restrictions on the relative heights of adjacent tone pairs, leading to phonetic overgeneration and phonological indeterminacy. We showed that by building syntagmatic restrictions directly into the phonology, tone interval theory not only achieved a transparent input-output relationship which accounted for F0 data, but it did so in a simpler way than could be achieved by assuming paradigmatic primitives plus phonetic rules for tone scaling.

The tone interval framework was used to develop an account of English intonation based on syntagmatic primitives which avoided the indeterminacy and overgeneration associated with previous theories. We showed that the tone interval account builds on previous work by providing a solid theoretical basis for the reasonable and intuitive assumptions about the relationship between phonology and phonetics which permitted previous descriptive work to be carried out. Moreover, we showed that through a simple change in notation, previous descriptions in terms of H's and L's could be readily translated into tone interval constructs. The tone interval description of English was also shown to be significantly more phonetically transparent than previous descriptions, in that significant F0 turning points – that is, F0 peaks, valleys, and corners – are treated as consistently arising from tones, while tones are assumed to consistently give rise to F0 turning points.[101]

It was shown that the tone interval approach permitted an account of a number of outstanding facts. First, the proposal that tones universally associate with respect to metrical grids provides for the first time a unified treatment of attested cross-linguistic interactions between metrical and tonal properties. Second, the assumption that tones associate with respect to metrical structures also permitted us to account for findings that the two tones in contours described as bitonal accents in the paradigmatic theory do not align with respect to one another by a fixed temporal interval (Arvaniti *et al.* 1998; Ladd *et al.* 1999, 2000; Dilley *et al.* forthcoming). Instead, these two tones are treated as independent tonal entities which must associate with adjacent metrical positions, one prominent and one nonprominent. Third, tone interval theory assumes that all pitch interpolation functions are monotonic, thereby providing a principled means of treating all F0 "dips" as arising from *L* tones, as suggested by Ladd and Schepman (2003). Fourth, the more abstract and robust definition of starred tones provided by this theory permits an account of intonation patterns in English and Greek in which neither a peak nor a valley is aligned with a stressed syllable. Such contours are accounted for by a sequence of two tone intervals specifying an identical relative height relationship across a starred tone. Fifth, evidence of long-distance interactions between accented syllables reported by Pierrehumbert (1980) and Liberman and Pierrehumbert (1984) was shown to provide not only qualitative, but quantitative support for tone interval theory, since speakers in those studies scaled accented syllables according to a ratio. It was shown that these effects can be accounted for in a straightforward way in terms of the mechanisms of interaction between tones and grids proposed in Chapter 3.

Empirical support for the tone interval account of English intonation was provided through several perception and production experiments. These experiments utilized stimuli in which F0 had been varied continuously along some dimension, so as to look for "categorical effects" in perception and production tasks as a means of diagnosing phonological category boundaries. It was shown that, in general, categorical effects were observed in perception and production at positions in stimulus continua where the relative pitch levels of two syllables changed, so that the pitch of one syllable switched from being higher than an adjacent syllable to lower than that syllable. This can be explained in terms of a change from one syntagmatic category to another, for example, from *higher* to *lower*. In contrast, the experiments provided only limited support for claims of the paradigmatic theory of English. Significantly, these results failed to show evidence of category boundaries when the F0 level of a syllable or syllables was shifted through the pitch range while controlling for the relative heights of adjacent syllables, in contrast to the claims of the paradigmatic theory.

This work suggests a number of avenues for future work. In particular, the empirical work established that perceptual experiments using a discrimination task and production experiments using an imitation task can be utilized to investigate the phonological categories in a language. This suggests that such procedures can be useful in other languages for investigating phonological categories of intonation and tone.

Moreover, an important step will be to extend the descriptive analytic methods presented here to lexical tone languages, which have not been treated at all in this work. Such an extension is quite straightforward, as indicated by additional work which could not be included here. It will be particularly important to work out the principles of interaction between paradigmatic and syntagmatic tone intervals at the lexical and phrasal levels.

Moreover, it will be necessary to evaluate the claim that all languages represent syntagmatic tonal relations, while only a subset of languages represent paradigmatic tonal relations. In this regard, several different avenues for experimental work present themselves. First, the paradigm and methods of Hallé, Chang, and Best (2004) could be extended to determine which languages demonstrate categorical effects under similar manipulations. Second, the psycholinguistic methods of Wong and Diehl (2003) might usefully be extended to determine whether and how context affects lexical tone identification. Third, additional work can be done to test whether the paradigmatic distinctions which have previously been claimed for intonation languages materialize in the form of telltale categorical effects in perception and production.

In general, we could identify several goals for future research in this area. First, it will be necessary to develop a better understanding of how F0 is perceived in speech. Indeed, proper testing of all the proposals made here will involve perceptual experiments of some kind. To this end, it will be necessary to develop more and better methods for studying perception of F0 in speech. While discrimination, identification, and imitation tasks are useful, psychophysical research has developed a far wider armory of experimental methodologies which will likely be useful in investigating the linguistic representation of F0. Moreover, it will be necessary to develop a better understanding of how principles of general auditory perception may or may not be responsible for some of the tonal phenomena which are observed across languages.

A related challenge will be to work out the principles by which tone interval sequences are restricted across languages. We proposed that principles of general auditory perception universally restrict the kinds of sequences which are produced. This leads us to expect a predominance of repeated patterns of tones across languages. Moreover, we expect that the kinds of metrical structures that such tone sequences elicit in isolation should be congruent with the metrical patterns of words and phrases. However, these proposals will need to be worked out in more detail in future work. Our sense is that in general, a fruitful approach for future research will be to seek connections between linguistics and other related fields, such as music and auditory psychology. In this way, it will likely emerge that many phenomena which seem otherwise arbitrary may find a common and systematic explanation in terms of processes of auditory perception.

Appendix A.1 – Restrictions on hierarchies in the tone interval matrix

Chapter 3 described a structure, called the tone interval matrix, which stores information about syntagmatic tone intervals and their values. Syntagmatic tone intervals ultimately specify the relative height relations between pairs of tones at all levels of the grid. To understand what sorts of syntagmatic steps can be specified in a tone interval matrix, it is necessary to understand the inherent restrictions that are placed on steps up, down, and sideways between tones on different grid levels.

This section has two goals. The first goal is to address the question of how the relative heights of adjacent tones at different levels of the grid are related. We will derive a principle which formalizes restrictions on relative height relations among tones on different grid rows. This principle entails elementary mathematical relations which we feel are significant for understanding the phonology of tone. The second goal of this section is to illustrate some of the corollaries of the formal principle to be derived for restricting relative height relations between tones on different grid rows. In Appendix A.2 we will show that this formal principle can ultimately provide an account for so-called *stepping relations*, by which we mean patterns of relative height relations among tones, such as downstep and upstep. By showing that an account of downstep and upstep follows naturally from the principle which relates different levels of the grid, we will highlight a strength of the tone interval approach.

To begin, we will suggest that there *must* be restrictions on the relative heights of tones on different grid rows, and hence on syntagmatic tone interval values. To see why, consider the fact that in Figure A.1, tones T_1 and T_2, and T_2 and T_3 are adjacent on row 1 of the grid, while only T_1 and T_3 are represented on row 2. (Note that for the discussion in this appendix we will leave off the "*" and "+" notation for tones, in order to not detract from investigating matrix-internal restrictions.) Crucially, if we step *down* from T_1 to T_2, and then step *down* again from T_2 to T_3, this is incompatible with stepping *up* from T_1 to T_3. This is because stepping down twice cannot result in being higher at the end compared to the start. How can we derive a formal account of these internal restrictions on tone interval relations?

In order to facilitate exposition of the following formalisms, it will be useful to develop a term describing the relation between tones on "consecutive" grid rows. We will refer to this relation as *subtending*. Then for example, we will say that the sequence [T_1 T_3] on grid row 2 *subtends* the sequence [T_1 T_2 T_3] on grid row 1. We can generalize this definition in the following way so that it applies to any two consecutive grid rows.

Subtending. (*Definition*.) For any tones T_i and T_N which are adjacent on row $j+1$ of the metrical grid, T_i and T_N on row j+1 *subtend* the sequence T_i, T_i+1,...,T_N on row j.

To introduce how the relative height relation specified in a subtending tone interval [T_1 T_3] on grid row 2 is related to a subtended sequence of tones [T_1 T_2 T_3] on grid row 1, recall from Chapter 2 that tone intervals are multiplicative, as given by the Multiplicative Property. Then the expression in (A.1) should relate the tone intervals on grid rows 1 and 2:

$$I_{1,3} = I_{1,2} \cdot I_{2,3} \tag{A.1}$$

This expression describes how the steps between T_1, T_2, and T_3 on grid rows 1 and 2 must be related. We can generalize on this result in a way that permits us to formalize how the steps on *any* two rows of the grid must be related, by making use of the Multiplicative Property. We formalize these restrictions in terms of the *Metrical Principle of Syntagmatic Relations*, which is given in (A.2). This principle is simply a generalization of the insight given above in (A.1); we will see that it has a number of important corollaries which will permit an account for various kinds of stepping relations in a unified and simple way.

Metrical Principle of Syntagmatic Relations (MPSR): Let T_i and T_N be tones on row j of the metrical grid, where T_i and T_N are adjacent on row $j+1$. Then $I_{i,N}$ on row $j+1$ of the tone interval matrix is equal to the product of the subtended syntagmatic tone intervals on row j of the matrix:

$$I_{i,N} = I_{i,i+1} \cdot I_{i+1,i+2} \cdot \ldots \cdot I_{N-1,N} \tag{A.2}$$

The Metrical Principle of Syntagmatic Relations (MPSR) provides a general statement of how tone interval values at different levels of the matrix are related. It is not a rule, but simply a statement about the properties of the tone interval matrix. Then the MPSR says that *the relative height and distance between two adjacent tones on some higher grid row is predictable from the relative height and distance between pairs of subtended adjacent tones on the next lower grid row, and vice versa.* More generally, the MPSR expresses restrictions on phonological categories on different rows of the matrix, where the categories correspond to restrictions on syntagmatic tone interval values.

In the following, we will be interested in understanding what insights the MPSR affords regarding the relative height relations associated with contiguous rows. We will see that in many cases, the relative height relation of a subtending tone interval is predictable entirely from the sequence of relative height relations of the subtended tone interval. We will derive several corollaries of the MPSR describing these special cases for which this kind of predictable relationship obtains. However, there are also a number of possible configurations of relative height relations on consecutive grid rows where the relative height relation of a subtending tone interval is *not* predictably related to the relative heights of the subtended tone intervals. These latter cases will be the topic of Appendix A.2.

We will first consider some of the cases where the relative height relation of the subtending tone interval is determined entirely by the relative height relations of the subtended tone intervals. The first such case is when the subtended steps are in the *same direction*. For example, suppose we step down from T_1 to T_2 and from T_2 to T_3. Common sense tells us that we will be lower at the end of the stepping sequence than when we started out. The MPSR captures this intuition formally. In particular, the MPSR gives us that $I_{1,3} = I_{1,2} * I_{2,3}$. Taking two steps down implies that $I_{1,2} < 1$ and $I_{2,3} < 1$. Then $I_{1,3}$ is the product of two terms which are each less than 1, which necessarily yields a product less than 1. For example, if $I_{1,2} = 0.95$ and $I_{2,3} = 0.95$, then $I_{1,3}$ will be approximately 0.9025. We can infer that $I_{1,3} < 1$, so that the relative height relation entails that T_3 is *lower* than T_1. In other words, the MPSR permits us to formalize why two steps down means that the ending point will necessarily be lower than the starting point.

We can consider the analogous situation for a case involving two steps up from T_1 to T_2 and from T_2 to T_3. Taking two steps up implies that $I_{1,2} > 1$ and $I_{2,3} > 1$. Then the MPSR says that $I_{1,3}$ is the product of two terms which are each greater than 1, which must give a product greater than 1. For example, if $I_{1,2} = 1.05$ and $I_{2,3} = 1.05$, then $I_{1,3}$ is about 1.1. Then $I_{1,3} > 1$, which means that T_3 is *higher* than T_1. Once again, a sequence of identical steps has yielded a predictable relative height relation for T_1 and T_3. It is transparent to show that the same situation holds true then $I_{1,2} = 1$ and $I_{2,3} = 1$, so we will not work through this example.

These cases suggest a generalization regarding the relationship between a subtended sequence consisting of identical steps up, down, or sideways, and the direction of the step associated with the tones on the next higher row. We generalize this observation as a corollary of the MPSR, as follows.

MPSR – Corollary 1: If all I_n in the subtended sequence $I_{i,i+1}, I_{i+1,i+2}, \ldots, I_{N-1,N}$ for $n \in \{i, \ldots, N\}$ have a relative height relation of [αsame, (βhigher)], then the relative height relation of the subtending tone interval $I_{i,N}$ will also be [αsame, (βhigher)].

Corollary 1 captures the fact that the relation of relative height between two subtending tones will be identical to that of each of the subtended steps, when the relative height relations associated with those steps are identical. What are the implications of the MPSR for the more general case in which the terms in a subtended sequence have *different* relative height relations compared to one another? There are two cases which we will focus on. First, consider the case in which one of the relative height relations involves a step sideways. For example, what is relative height relation for $I_{1,3}$ if we take a step sideways, followed by a step down? Common sense tells us that $I_{1,3}$ will itself entail a step down and that the ending point will be lower than the beginning point. Similarly, we can infer that taking a step sideways, followed by a step up means that $I_{1,3}$ entails a step up. Here again, the MPSR offers insight into how to formally characterize the relations among steps at different levels of the grid-matrix complex. For example, consider the case of stepping sideways from T_1 to T_2 and then up from T_2 to T_3. From earlier in this section, we know that the MPSR states that $I_{1,3} = I_{1,2} * I_{2,3}$. Given the step sideways from T_1 to T_2, we know that $I_{1,2} = 1$, while stepping up from T_2 to T_3 means that $I_{2,3} > 1$. Multiplication of any term by 1 yields an identity relation. This implies that $I_{1,3} = I_{2,3}$. In general, we can infer that if the relative height relations associated with any of the tone intervals in a subtended sequence is *same*, then the relative height relation of the subtending tones is determined by the relative height relation(s) of the remaining term(s) in the subtended sequence. We can formalize this as a corollary of the MPSR, as follows.

MPSR – Corollary 2: If any I in a subtended sequence $I_{i,i+1}, I_{i+1,i+2}, \ldots, I_{N-1,N}$ has the relative height relation [+same], then the relative height relation of the subtending tone interval will depend on relative height relation's of the remaining subtended tone intervals, for $N \geq 2$.

Corollary 2 captures the "identity" relation associated with multiplication by 1. It simply says that if any term in a subtended sequence is such that $I = 1$, the relative height relation of the ending point with respect to the beginning point will depend on the remaining terms. This corollary is straightforward, and we will have no further comment on it.

We have just shown two special cases predicted by the MPSR in which the relative height relations of a tone interval on a higher grid row is entirely predictable from the sequence of relative height relations of subtended tone intervals on the next lower grid row. However, there is another class of cases in which a sequence of adjacent steps are in opposite directions (such as *down*, followed by *up*), where the relative height relation of two nonadjacent tones on a higher grid row is not predictable from the direction of the steps between adjacent tones on the next row down. In this case, the relative height relation of two nonadjacent tones is specified directly in the phonology, and the MPSR describes the resulting restrictions on the relative magnitudes of the corresponding subtended steps. This ultimately permits a treatment of stepping relations like downstep and upstep. We show how this is accomplished in Appendix A.2.

Appendix A.2 – An account of downstep and upstep

In Appendix A.1, we showed that the Metrical Principle of Syntagmatic Relations (MPSR) captured certain restrictions on relative height relations on different rows of the grid. It turns out that the MPSR has implications for describing a broad set of cross-linguistic observations which were referred to earlier as stepping relations, such as downstep and upstep. In this appendix we explore further properties of the MPSR to show how it provides an account for these phenomena.

Patterns of stepping up, stepping down, and staying level are observed across languages (Pierrehumbert 1980, Hyman 1993, Snider 1999). In a number of cases, distinct stepping relations are shown to contrast within a language. We discussed in Chapter 1 some cases in which two sequential tones gave rise to distinct, contrastive relative height relations. Cases of distinctions involving tones which are not sequential in time are also attested. For example, in Hausa a HLH sequence corresponds to a declarative or an interrogative depending on whether the High tones step down or stay at the same level (Inkelas *et al.*, 1986). In general, control of the relative height relations between tones is observed across a wide variety of languages, in a wide variety of tonal contexts. (See van der Hulst and Snider (1993) for an overview.) For example, stepping down between tones is observed not just between lexical High tones, but also between Low and Mid tones (Snider and van der Hulst 1993). This suggests that any formulation based on a particular sequence of tones is probably not flexible enough to capture the range of cross-linguistic phenomena observed.[102]

Moreover, we note there is a special class of contours which involve a pattern of *repeated stepping*. These are widely attested and have received much attention in the literature. A well-known example in this category involves a sequence of a fall and a rise, which is shown in Figure A.2(a). Repeated stepping patterns of the sort shown in Figures A.2(b) and A.2(c) are additionally attested in a number of languages, including English. Tonal patterns like those in Figures A.2(a)-(c) are highly suggestive of the application of an iterative rule.[103]

In this appendix, we will provide an account for observed stepping relations between *nonsequential* tones. Some examples of distinct relative height relations between nonsequential tonal elements are shown in Figures A.3(a)-(c). These contours show three possible relations of relative height between the initial high point and the final high point.

To begin, recall that in Appendix A.1 we introduced the MPSR, which related syntagmatic tone intervals at different levels of a grid-matrix complex. To account for the three cases shown in Figure A.3(a)-(c), we first note that we can re-express the MPSR in a different way. Recall from Chapter 2 that syntagmatic tone intervals have the properties of fractions. Then we can rewrite the MPSR in terms of the tones which are referred to in the tone interval expressions. In particular, the MPSR states that the subtending tones T_i and T_N on row j+1 gave rise to a subtending tone interval, $I_{i,N}$, which was equal to the product of the subtended tone intervals on row j.

$$I_{i,N} = I_{i,i+1} \cdot I_{i+1,i+2} \cdot \ldots \cdot I_{N-1,N} \tag{A.3}$$

Then each of the terms on the right-hand side of the expression above corresponds to a sequence of two adjacent tones on row *j*. We can rewrite the MPSR in terms of the corresponding adjacent tone pairs on row *j*, as follows.

$$I_{i,N} = \frac{T_{i+1}}{T_i} \cdot \frac{T_{i+2}}{T_{i+1}} \cdot \frac{T_{i+3}}{T_{i+2}} \cdot \ldots \cdot \frac{T_N}{T_{N-1}} \tag{A.4}$$

The observation here is that tones associated with numerators and denominators of adjacent expressions cancel. For example, T_{i+1} occurs in the numerator of the first term and the denominator of the second term, and therefore cancels. In the same way, T_{i+2} cancels out, and so on.[104] In general, all terms from the numerator and denominator cancel, except for T_i and T_N. We are then left with the following relation:

$$I_{i,N} = \frac{T_N}{T_i} \tag{A.5}$$

In other words, a simple manipulation to the MPSR shows that a sequence of adjacent tone interval relations gives rise to a tone interval expression composed of tones which are *not sequential in time.*[105] This is a syntagmatic tone interval like any other; as such it can be specified for a relative height relation, and its values can further be restricted in a language-specific manner. We can express this observation as an additional corollary to the MPSR.

MPSR – Corollary 3. Let T_i and T_N be tones on row *j* of the metrical grid, where T_i and T_N are adjacent on row *j+1*. Then the syntagmatic tone interval $I_{i,N}$ on row *j+1* is given by $I_{i,N} \equiv T_N/T_i$.

MPSR-C3 simply states that the first and last tones in a sequence of subtended tone intervals give rise to a subtending tone interval corresponding to the next higher row of the grid. These syntagmatic tone intervals have properties of other syntagmatic tone intervals – they can be specified for relative height relations, or they may have further language-specific restrictions on their values. MPSR-C3 essentially shows that a relation between nonadjacent tones may arise through relations among a sequence of adjacent tones. Each row of the grid/matrix is related in hierarchical fashion in this way.

This immediately suggests a way to account for stepping relations between nonadjacent tones. In particular, it suggests that the relative height relation between the nonadjacent tones can be specified directly for a tone interval in which two nonadjacent tones participate. However, one requirement – and prediction – is that in order for a relation between nonadjacent tones to hold, the tones must be represented on row two of the grid or higher. In other words, the nonadjacent tones to be joined in a syntagmatic tone interval must be associated with *metrically prominent timing positions.*

Assuming the nonadjacent tones in Figures A.3(a)-(c) are indeed associated with metrically prominent timing positions, then formulating the distinction among these three cases is quite straightforward. In particular, we can use the observations of MPSR-C3 to generate the following relation between T_1 and T_3.

$$I_{1,3} = \frac{T_3}{T_1} \tag{A.6}$$

Turning to the contours in Figures A.3(a)-(c), we first note that in Figure A.3(a), T_3 is at a lower level than T_1. As a result, we can infer that $I_{1,3} < 1$, capturing the fact that the associated relative height relation is *lower*. Moreover, Figure A.3(b) shows that T_3 is higher than T_1. Then we know that $I_{1,3} > 1$, capturing the fact that the associated relative height relation is *higher*. Finally, Figure A.3(c) shows that T_3 is at the same level as T_1. Then we can infer that $I_{1,3} = 1$, capturing the fact that the associated relative height relation is *same*.

In other words, relative height relations may be represented *directly* between tones which are not sequential in time in terms of syntagmatic tone interval values. Nonsequential tones can participate in tone interval constructs only when they are adjacent on a higher grid row, and such tones are only represented on higher grid rows if they are associated with metrically prominent timing positions. The relations between nonsequential tones followed directly from simple mathematical properties associated with tone intervals.

It seems highly plausible to us that relative height relations between nonadjacent tones would be captured in phonological representations, since relations between nonsequential tones are important for the representation of music and other kinds of auditory sequences. There are many illustrative examples. In general, a melody may be carried by nonadjacent tones, so long as the notes which comprise it are metrically prominent. For example, it is common in Baroque music for high-pitched and low-pitched notes to alternate in sequence. Then a melody is carried either by the high or the low notes, depending on which are associated with metrically prominent positions (Lerdahl and Jackendoff 1983). Furthermore, there is a well-known phenomenon in general auditory perception termed *auditory streaming* (Bregman 1990). When tones form auditory streams, listeners are only capable of attending to tones which are not in sequential order.

Given that relations between nonsequential tonal elements are widely attested in music and other auditory domains, it makes sense to us that the same situation would obtain in language. Indeed, many of the tonal phenomena observed across languages is suggestive of congruence between linguistic tonal phenomena and similar phenomena in music and general auditory perception. These connections suggest that a deeper set of processes relating to auditory perception emerge not only in domains like music, but also in the domain of language.

To summarize, we have claimed that the phonology specifies relations of relative height directly between nonsequential tones. This followed from the MPSR, which showed that the tone interval formed by two adjacent tones on a higher grid row was equal to the product of the subtended sequence of tone intervals on the next lower grid row. This permitted us to describe the differences among the contours in Figure A.3(a)-(c) in terms of distinct relations of relative height between T_1 and T_3. In particular, Figures A.3(a)-(c) are described in terms of distinct values for $I_{1,3}$: Figures A.3(a), A.3(b) and A.3(c) correspond to $I_{1,3} < 1$, $I_{1,3} > 1$, and $I_{1,3} = 1$, respectively.

However, there are some additional details to this analysis which remain to be filled in. For example, at the end of Appendix A.1 we noted that when two or more of the steps in the subtended sequence were in opposite directions, the overall relative height relation of the last tone with respect to the first was related to the relative magnitudes of the steps up and down. Then stepping down by a large amount and up by a small amount yields an overall step down, while stepping down by a small amount and up by a large amount yields an overall step up. We have not yet investigated how the relative height relation of a subtending tone sequence is related to the relative magnitudes of the subtended tones. What we will see is that the relative magnitudes of a sequence of subtended steps is lawfully related to the relative height relation of a pair of nonadjacent subtending tones, in a way that is once again given by the MPSR. It is important to understand this relationship between relative height and relative magnitude, because it suggests that there is an alternative phonological formulation for stepping relations in this framework that must be considered. In particular, it might be that the phonology restricts the relative magnitudes of adjacent tone intervals, thereby deriving an overall relation of relative height between the nonadjacent tones. How can we then defend the claim that the relative height relation between subtending tones is specified directly in the tone interval matrix?

In order to defend this analysis, we must first formalize the relationship between the relative heights of a pair of subtending tones, and the relative magnitudes of the subtended steps. The MPSR shows that these two issues are just two sides of the same coin. To see why, consider the contour in Figure A.3(c), which presents the most straightforward example for conceptualizing the relation between relative height and relative magnitude. Relative height is straightforward to conceptualize in tone interval terms; in Figure A.3(c), it is easy to see that the last high point is at the same level as the first high point. In contrast, relative magnitude is rather tricky to conceptualize in tone interval terms. This is because we cannot directly compare the sizes of falls and rises, since these two events are categorically distinct in tone interval theory. Moreover, a fall corresponds to a number which is less than 1, while a rise corresponds to a number which is greater than 1. In order to see how relative height is related to relative magnitude, recall that the Reciprocal Property of tone intervals described in Chapter 2 permits us to describe a fall as a "reverse time" rise. Suppose we think of the shape in Figure A.3(c) as a "V" whose left edge is formed by a "reverse time" rise from the point of the "V," and whose right edge corresponds to the "forward time" rise from the point of the "V". In order to characterize the necessary conditions on relative magnitude for the contour to step down and back up to the level where it started out, we need to consider how large one rise has to be with respect to the other rise. It should be clear that in order for the last high point to be at the same level as the first high point, the size of the "reverse time" rise must be the same as the size of the "forward time" rise.

We have now arrived at a conceptual understanding of the relation between relative height and relative magnitude. Next, we will use the MPSR to show how this can be expressed in formal terms. Indeed, the MPSR will show that observations about relative height and relative magnitude are two sides of the same coin.

For all of the contours in Figures A.3(a)-(c), we know from the MPSR that $I_{1,3} = I_{1,2} \cdot I_{2,3}$. Suppose that we rewrite this expression as $I_{1,3} = x \cdot y$, where $x = I_{1,2}$ and $y = I_{2,3}$. We know that the relative height relation between T_1 and T_3 in Figure 3.10c can be characterized as $I_{1,3} = 1$. Then what does this imply about the relative magnitudes of x and y?

In a nutshell, x and y must be *reciprocals* of one another. To see why this is the case, an analogy from music may be helpful. In particular, we can think of the step down from T_1 to T_2 and back up to T_3 as being somewhat like playing a sequence of notes on a piano keyboard. Suppose we play a sequence G-C-G, where the C is below G by less than an octave. Then stepping down from G to C corresponds to a frequency ratio of 2/3, while stepping up from C to G corresponds to a frequency ratio of 3/2.

This suggests that in order for the endpoint to be at the same level as the beginning point after a sequence of a step down and a step up, the size of the step up must be the reciprocal of the size of the step down. Given this observation, it is straightforward to characterize the relation between relative height and relative magnitude in tone interval terms. Suppose we let $x = \Delta$; then given that x and y are reciprocals, we know that $y = 1/\Delta$. This brings us back to the MPSR, which we earlier characterized as $I_{1,3} = x{*}y$. Then given $x = \Delta$ and $y = 1/\Delta$, we obtain the desired result: $I_{1,3} = \Delta \cdot (1/\Delta) = 1$. Obtaining the result that $I_{1,3} = 1$ shows that we have successfully characterized the relation between relative height and relative magnitude.

We can formalize these observations as Corollary 4 of the MPSR. Here, we have dealt only with the case of stepping relations among three tones, but the conditions can easily be extended to deal with other situations. In particular, consider the following.

MPSR – Corollary 4. Let $I_{i,i+2}$ be a syntagmatic tone interval on row 2 which subtends $I_{i,i+1}$ and $I_{i+1,i+2}$ on row 1, where $I_{i,i+1}$ is [α *same*, (β *higher*)] and $I_{i+1,i+2}$ is [α *same*, (-β *higher*)]. Then if $I_{i,i+1} = \Delta$, the following relations hold:

(i) If $I_{i,i+2} = 1$, then $I_{i+1,i+2} = 1/\Delta$.
(ii) If $I_{i,i+2} < 1$, then $I_{i+1,i+2} < 1/\Delta$.
(iii) If $I_{i,i+2} > 1$, then $I_{i+1,i+2} > 1/\Delta$.

MPSR-C4 formalizes the relationship between relative magnitude and relative height. More specifically, it captures the relationship between the relative height of a subtending tone interval on grid row 2 and the relative magnitudes of two steps in opposite directions. It says that if two adjacent steps have reciprocal tone features, then their relative magnitudes are predictable from the relative height relation of the subtending tone interval. It is a simple exercise to show that MPSR-C4 describes the relation between relative height and relative magnitude not only for high-low-high sequences, as in Figures A.3(a)-(c), but also for low-high-low sequences.

Now that we have formalized the relationship between relative height and relative magnitude in the tone interval framework, we can consider the issue raised earlier concerning our claim that the phonology specifies relative height relations in the phonology directly. MPSR-C4 suggests that an alternative phonological formulation of stepping relations in this framework is to say that there are restrictions on the relative magnitudes of subtended steps. Restricting the relative magnitudes of a sequence of subtended steps would also achieve particular relative height relations on higher grid rows. If the phonology restricted the relative magnitudes of steps in this way, then it would not be necessary to claim that relations between nonadjacent tones were specified directly in the matrix.

We reject this alternative formulation, on several grounds. First, without recourse to higher-level structures, it is not even meaningful to talk about the relationship between relative magnitude and relative height relations. We were only able to formulate the above relation expressed in MPSR-C4 by appealing to a higher-level construct in the first place. Second, it would be less economical for systems to represent rules in terms of relative magnitudes. A phonological system would have to encode rules in the form "if X, then Y" – for example "if you step down by Δ, then step up by an amount less than 1/Δ" and so on. By representing rules in terms of higher-level relative height relations, the system can represent rules more simply and flexibly, for example, as just "step down". Third, the encoding of nonlocal relations of relative pitch is a well-attested phenomenon in other auditory domains, such as music. Therefore, it is not only reasonable, but highly likely that the same phenomenon obtains in language, which is primarily based on auditory processes of transmission. We can therefore reject this alternative formulation on principled grounds.

We have shown that internal properties of the tone interval matrix readily provide an account of stepping phenomena. In this way, such phenomena can be accounted for without positing additional rules for stepping relations. This is made possible by the fact that relations between nonsequential tones are represented directly in terms of tone intervals on higher levels of the grid. We proposed a principle, the MPSR, which formally described the ways in which relative height relations specified on consecutive grid levels were related. This concludes our discussion of stepping relations and of internal restrictions on specifications in the tone interval matrix.

References

Armstrong, L. E., & Ward, I. C. (1926). *Handbook of English Intonation*. Leipzig and Berlin: B. G. Teubner.

Arvaniti, A., & Ladd, D. R. (1995). Tonal alignment and the representation of accentual targets, *Proceedings of the 13th International Congress of Phonetic Sciences* (Vol. 4, pp. 220-223). Stockholm.

Arvaniti, A., Ladd, D. R., & Mennen, I. (1998). Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics, 26*, 3-25.

Arvaniti, A., R., L. D., & Mennen, I. (2000). What is a starred tone? Evidence from Greek. In M. Broe & J. Pierrehumbert (Eds.), *Papers in Laboratory Phonology V* (pp. 119-130). Cambridge, UK: Cambridge University Press.

Ashby, M. (1978). A study of two English nuclear tones. *Language and Speech, 21*, 326-336.

Atterer, M., & Ladd, D. R. (2004). On the phonetics and phonology of "segmental anchoring" of F0: evidence from German. *Journal of Phonetics, 32*(2), 177-197.

Attneave, F., & Olson, R. K. (1971). Pitch as a medium: A new approach to psychophysical scaling. *American Journal of Psychology, 84*, 147-166.

Beckman, M., & Ayers-Elam, G. (1997). *Guidelines for ToBI labeling, version 3.0*: The Ohio State University.

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

Bickmore, L. S. (1995). Tone and stress in Lamba. *Phonology, 12*, 307-341.

Boersma, P., & Weenink, D. (2002). Praat, a system for doing phonetics by computer (Version 4.0.26): Software and manual available online at http://www.praat.org.

Bolinger, D. (1958). A theory of pitch accent in English. *Word, 14*, 109-149.

Bolinger, D. (1965). Pitch accent and sentence rhythm. In I. Abe & T. Kenekiyo (Eds.), *Forms of English: Accent, Morpheme, Order* (pp. 139-180). Cambridge, MA: Harvard University Press.

Bolinger, D. (1986). *Intonation and Its Parts*. Palo Alto, CA: Stanford University Press.

Bolinger, D. (1989). *Intonation and Its Uses*. Palo Alto, CA: Stanford University Press.

Bolton, T. L. (1894). Rhythm. *American Journal of Psychology, 6*(2), 145-238.

Brady, P. T., House, A. S., & Stevens, K. N. (1961). Perception of sounds characterized by a rapidly changing resonant frequency. *Journal of the Acoustical Society of America, 33*, 1357-1362.

Brazil, D. (1997). *The Communicative Value of Intonation in English*. Cambridge: Cambridge University Press.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*. Cambridge, MA: MIT Press.

Broselow, E. (1995). The skeletal tier and moras. In J. Goldsmith (Ed.), *A Handbook of Phonological Theory* (pp. 175-205). Oxford: Blackwell Publishers.

Broselow, E., Chen, S., & Huffman, M. (1997). Syllable weight: convergence of phonology and phonetics. *Phonology, 14*, 47-82.

Bruce, G. (1977). *Swedish word accents in sentence perspective*. Lund: Gleerups.

Bruce, G. (1982). Developing the Swedish intonation model. *Lund University Dept. of Linguistics Working Papers, 22*, 51-116.

Bruce, G., & Gårding, E. (1978). A prosodic typology for Swedish dialects. In E. Gårding & G. Bruce & R. Bannert (Eds.), *Nordic Prosody* (pp. 219-228): Lund University.

Burns, E. M. (1999). Intervals, scales, and tuning. In D. Deutsch (Ed.), *The Psychology of Music* (2nd ed., pp. 215-264). San Diego: Academic Press.

Burns, E. M., & Ward, W. D. (1978). Categorical perception - phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *Journal of the Acoustical Society of America, 63*(2), 456-468.

Chao, R. C. (1947). *Cantonese Primer*. Cambridge, MA: Harvard University Press.

Chen, M. Y. (2000). *Tone Sandhi: Patterns across Chinese Dialects*. Cambridge: Cambridge University Press.

Ciocca, V., & Bregman, A. S. (1987). Perceived continuity of gliding and steady-state tones through interrupting noise. *Perception & Psychophysics, 42*(5), 476-484.

Clark, M. M. (1978). *A dynamic treatment of tone, with special attention to the tonal system of Igbo.* Unpublished Ph.D. dissertation, University of Massachusetts at Amherst.

Clark, M. M. (1993). Representation of downstep in Dschang Bamileke. In H. van der Hulst & K. Snider (Eds.), *The Phonology of Tone: The Representation of Tonal Register* (pp. 29-74). Berlin, New York: Mouton de Gruyter.

Clements, G. N. and Ford, K. C. (1979). Kikuyu tone shift and its synchronic consequences. *Linguistic Inquiry, 10*, 179-210.

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

Connell, B. (2000). The perception of lexical tone in Mambila. *Language and Speech, 43*(2), 163-182.

Connell, B., & Ladd, D. R. (1990). Aspects of pitch realization in Yoruba. *Phonology, 7*(1), 1-29.

Crystal, D. (1969). *Prosodic Systems and Intonation in English*. Cambridge: Cambridge University Press.

Crystal, D., & Quirk, R. (1964). *Systems of prosodic and paralinguistic features in English*. The Hague: Mouton.

d'Alessandro, C., & Mertens, P. (1995). Automatic pitch contour stylization using a model of tonal perception. *Computer Speech and Language, 9*(3), 257-288.

d'Alessandro, C., Rosset, S., & Rossi, J. P. (1998). The pitch of short-duration fundamental frequency glissandos. *Journal of the Acoustical Society of America, 104*(4), 2339-2348.

Dannenbring, G. L. (1976). Perceived auditory continuity with alternately rising and falling frequency transitions. *Canadian Journal of Psychology, 30*, 99-114.

de Cheveigné, A. (2000). A model of the perceptual asymmetry between peaks and troughs of frequency modulation. *Journal of the Acoustical Society of America, 107*(5), 2645-2656.

de Lacy, P. (2002). The interaction of tone and stress in Optimality Theory. *Phonology, 19*(1), 1-32.

de Pijper, J. R. (1983). *Modelling British English Intonation*. Dordrecht: Foris Publications.

de Saussure, F. (1966). *Course in General Linguistics*. New York: McGraw-Hill Book Company.

Demany, L., & Clément, S. (1995a). The perception of frequency peaks and troughs in wide frequency modulations. II. Effects of frequency register, stimulus uncertainty, and intensity. *Journal of the Acoustical Society of America, 97*(4), 2454-2459.

Demany, L., & Clément, S. (1995b). The perception of frequency peaks and troughs in wide frequency modulations. III. Complex carriers. *Journal of the Acoustical Society of America, 98*(5), 2515-2523.

Demany, L., & Clément, S. (1997). The perception of frequency peaks and troughs in wide frequency modulations. IV. Effects of modulation waveform. *Journal of the Acoustical Society of America, 102*(5), 2935-2944.

Demany, L., & McAnally, K. I. (1993). Hearing melodies in continuous frequency modulations. *Journal of the Acoustical Society of America, 93*(4), 2346.

Demany, L., & McAnally, K. I. (1994). The perception of frequency peaks and troughs in wide frequency modulations. *Journal of the Acoustical Society of America, 96*, 706.

Dilley, L., Ladd, D. R., & Schepman, A. (in press). Alignment of L and H in bitonal pitch accents: Testing two hypotheses. *Journal of Phonetics*.

d'Imperio, M. (2000). *The role of perception in defining tonal targets and their alignment.* Unpublished Ph.D. dissertation, The Ohio State University.

d'Imperio, M., & House, D. (1997). Perception of questions and statements in Neapolitan Italian, *Proceedings of Eurospeech 97* (Vol. 1, pp. 251-254). Rhodes, Greece.

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America, 49*(2), 524-531.

Dowling, W. J., & Harwood, D. L. (1986). *Music Cognition*. Orlando, FL: Academic Press, Harcourt Brace Jovanovich Publishers.

Duanmu, S. (1999). Metrical structure and tone: evidence from Mandarin and Shanghai. *Journal of East Asian Linguistics, 8*, 1-38.

Earle, M. A. (1975). *An acoustic phonetic study of Northern Vietnamese tones* (SCRL Monograph Number 11). Santa Barbara, CA: Speech Communications Research Laboratory, Inc.

Fox, R., & Qi, Y. Y. (1990). Context effects in the perception of lexical tone. *Journal of Chinese Linguistics, 18*, 261-283.

Gandour, J. T. (1976). *Aspects of Thai tone.* Unpublished Ph.D. dissertation, University of California at Los Angeles.

Gandour, J. T. (1978). The Perception of Tone. In V. A. Fromkin (Ed.), *Tone: A Linguistic Survey* (pp. 41-76). New York: Academic Press.

Gårding, E., & Eriksson, L. (1989). Perceptual cues to some Swedish prosodic phrase patterns - a peak shift experiment, *Speech Transmission Laboratory - Quarterly Progress and Status Report* (Vol. 1, pp. 13-16). Stockholm: Royal Institute of Technology.

Gibbon, D. (1976). *Perspectives on Intonational Analysis*. Bern: Herbert Lang.

Goldsmith, J. (1976). *Autosegmental phonology.* Unpublished Ph.D. dissertation, MIT, Cambridge, MA.

Goldsmith, J. (1995). Phonological theory. In J. Goldsmith (Ed.), *The Handbook of Phonological Theory.* Oxford: Blackwell Publishing.

Gósy, M., & Terken, J. (1994). Question marking in Hungarian: timing and height of pitch peaks. *Journal of Phonetics, 22*, 269-281.

Grice, M. (1995). Leading tones and downstep in English. *Phonology, 12*, 183-233.

Grice, M., Ladd, D. R., & Arvaniti, A. (2000). On the place of phrase accents in intonational phonology. *Phonology, 17*, 143-185.

Grice, M., & Savino, M. (1995). Low tone versus 'sag' in Bari Italian intonation: a perceptual experiment, *Proceedings of the 13th International Congress of the Phonetic Sciences* (Vol. 4, pp. 658-661). Stockholm.

Gussenhoven, C. (1984). *On the Grammar and Semantics of Sentence Accents*. Dordrecht: Foris.

Gussenhoven, C. (1993). The Dutch foot and the chanted call. *Linguistics, 29*, 37-63.

Gussenhoven, C. (1999). Tone systems in Dutch Limburgian dialects. In S. Kaji (Ed.), *Proceedings of the Symposium on Cross-Linguistic Studies of Tonal Phenomena: Tonogenesis, Typology and Related Topics* (pp. 127-143). Tokyo: Tokyo University of Foreign Studies.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation.* Cambridge: Cambridge University Press.

Gussenhoven, C., & Rietveld, T. (2000). The behavior of H* and L* undervariations in pitch range in Dutch rising contours. *Language & Speech, 43*(2), 183-203.

Halle, M., & Vergnaud, J.-R. (1987). *An Essay on Stress*. Cambridge, MA: MIT Press.

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

Halliday, M. A. K. (1967). *Intonation and Grammar in British English.* The Hague: Mouton.

Handel, S. (1989). *Listening: An Introduction to the Perception of Auditory Events*. Cambridge, MA: MIT Press.

Hayes, B. (1980). *A metrical theory of stress rules.* Unpublished Ph.D. dissertation, MIT, Cambridge, MA.

Hayes, B. (1995). *Metrical Stress Theory.* Chicago: University of Chicago Press.

Hogg, R., & McCully, C. B. (1987). *Metrical Phonology: A Coursebook.* Cambridge: Cambridge University Press.

Honorof, D. N., & Whalen, D. H. Perception of pitch location within a speaker's F0 range. *Journal of the Acoustical Society of America*, submitted.

House, D. (1990). *Tonal Perception in Speech.* Lund: Lund University Press.

Howard, D., Rosen, S., & Broad, V. (1992). Major/minor triad identification and discrimination by musically trained and untrained listeners. *Music Perception, 10*(2), 205-220.

Hyman, L. M. (1993). Register tones and tonal geometry. In H. van der Hulst & K. Snider (Eds.), *The Phonology of Tone: The Representation of Tonal Register* (pp. 75-108). Berlin, New York: Mouton de Gruyter.

Inkelas, S., & Leben, W. R. (1990). Where phonology and phonetics intersect: the case of Hausa intonation. In J. Kingston & M. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* (pp. 17-34). New York: Cambridge University Press.

Inkelas, S., Leben, W. R., & Cobler, M. (1986). The phonology of intonation in Hausa. In S. Berman & J. Choe & J. McDonough (Eds.), *Proceedings of the 16th Annual Meeting of NELS*. Amherst: GLSA, University of Massachusetts.

Inkelas, S., & Zec, D. (1988). Serbo-Croatian pitch accent: the interaction of tone, stress, and intonation. *Language, 64*, 227-248.

Ishihara, T. (2003). A phonological effect on tonal alignment in Tokyo Japanese. In M. J. Solé & D. Recasens & J. Romero (Eds.), *Proceedings of the XVth International Congress of Phonetic Sciences* (Vol. 1, pp. 615-618). Barcelona.

Jakobson, R., Fant, G., & Halle, M. (1952). *Preliminaries to Speech Analysis*. Cambridge, MA: MIT Press.

Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review, 83*(5), 323-355.

Jones, M. R. (1986). Attentional rhythmicity in human perception. In J. R. Evans & M. Clynes (Eds.), *Rhythm in psychological, linguistic and musical processes*. Springfield, IL: C.C. Thomas.

Jones, M. R. (1993). Understanding melody and music: how melody and rhythm fit together. In T. Tighe & W. J. Dowling (Eds.), *Psychological approaches to understanding music* (pp. 67-92). Hillsdale, NJ: Lawrence Erlbaum.

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review, 96*(3), 459-491.

Jones, M. R., & Ralston, J. (1991). Some influences of accent structure on melody recognition. *Memory & Cognition, 19*, 8-20.

Kenstowicz, M. (1994). *Phonology in Generative Grammar*. Oxford: Blackwell Publications.

Kenstowicz, M., & Sohn, H.-S. (2001). Accentual adaptation in North Kyungsang Korean. In M. Kenstowicz (Ed.), *Ken Hale: A Life in Language* (pp. 239-270). Cambridge, MA: MIT Press.

Keyser, S.J. and K.N. Stevens (2001) Enhancement revisited. In *M. Kenstowicz* (ed.), *Ken Hale: A Life in Language* (pp. 271-291). Cambridge MA: MIT Press.

Kingdon, R. (1958). *The Groundwork of English Intonation*. London: Longman.

Kisseberth, C. (1984). Digo tonology. In G. N. Clements & J. Goldsmith (Eds.), *Autosegmental Studies in Bantu Tone* (pp. 105-182). Dordrecht, Holland: Foris.

Knight, R.-A. (2004). *Peaks and plateaux: The production and perception of high intonational targets in English*. Unpublished Ph.D. dissertation, University of Cambridge.

Kohler, K. J. (1987). Categorical pitch perception. In U. Viks (Ed.), *Proceedings of the 11th International Congress of Phonetic Sciences* (Vol. 5, pp. 331-333). Tallinn.

Ladd, D. R. (1980). *The Structure of Intonational Meaning: Evidence from English*. Bloomington: Indiana University Press.

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

Ladd, D. R. (1988). Declination 'reset' and the hierarchical organization of utterances. *Journal of the Acoustical Society of America, 84*(2), 530-544.

Ladd, D. R. (1990). Metrical representation of pitch register. In J. Kingston & M. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* (pp. 35-57). Cambridge: Cambridge University Press.

Ladd, D. R. (1993). In defense of a metrical theory of intonational downstep. In H. van der Hulst & K. Snider (Eds.), *The Phonology of Tone: The Representation of Tonal Register* (pp. 109-132). Berlin, New York: Mouton de Gruyter.

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

Ladd, D. R., Faulkner, D., Faulkner, H., & Schepman, A. (1999). Constant "segmental anchoring" of F0 movements under changes in speech rate. *Journal of the Acoustical Society of America, 106*(3), 1543-1554.

Ladd, D. R., Mennen, I., & Schepman, A. (2000). Phonological conditioning of peak alignment in rising pitch accents in Dutch. *Journal of the Acoustical Society of America, 107*(5), 2685-2696.

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

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

Laniran, Y. O. (1992). *Intonation in tone languages: the phonetic implementation of tones in Yoruba.* Unpublished Ph.D. dissertation, Cornell University.

Laniran, Y. O., & Clements, G. N. (2003). Downstep and high raising: interacting factors in Yoruba tone production. *Journal of Phonetics, 31*, 203-250.

LaVelle, C. R. (1974). An experimental study of Yoruba tone. *Studies of African Linguistics, supplement 5*, 185-194.

Leather, J. (1983). Speaker normalization in perception of lexical tone. *Journal of Phonetics, 11*, 373-382.

Leben, W. R. (1973). *Suprasegmental phonology*. Unpublished Ph.D. dissertation, MIT, Cambridge, MA.

Lerdahl, F., & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge, MA: MIT Press.

Li, Z. (2003). *The phonetics and phonology of tone mapping in a constraint-based approach.* Unpublished Ph.D. dissertation, MIT, Cambridge, MA.

Liberman, A. M., Harris, K. S., Hoffman, H., & Griffith, B. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*, 358-368.

Liberman, M. (1975). *The intonational system of English.* Unpublished Ph.D. dissertation, MIT, Cambridge, MA.

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

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry, 8*(2), 249-336.

Lickley, R., Schepman, A., & Ladd, D. R. Lab speech is real speech: the case of Dutch falling-rising questions. *Language and Speech*, submitted.

Lieberman, P. (1967). *Intonation, Perception, and Language*. Cambridge, MA: MIT Press.

MacMillan, N. A., & Creelman, C. D. (1991). *Detection Theory: A User's Guide*. New York: Cambridge University Press.

Malm, W. P. (1967). *Music Cultures of the Pacific, the Near East, and Asia*. Englewood Cliffs, NJ: Prentice-Hall.

Manfredi, V. (1993). Spreading and downstep: prosodic government in tone languages. In H. van der Hulst & K. Snider (Eds.), *The Phonology of Tone: The Representation of Tonal Register* (pp. 133-184). Berlin, New York: Mouton de Gruyter.

McAuley, J. D. (1996). Perception of time as phase: Toward an adaptive-oscillator model of rhythmic pattern processing. *Dissertation Abstracts International: Section B: The Sciences & Engineering, 56*(9-B).

McAuley, J. D., & Jones, M. R. (2003). Modeling effects of rhythmic context on perceived duration: A comparison of interval and entrainment approaches to short-interval timing. *Journal of Experimental Psychology: Human Perception & Performance, 29*(6), 1102-1125.

McPhee, C. (1966). *Music in Bali*. New Haven, CT: Yale University Press.

Moore, B. C. J. (1997). *An Introduction to the Psychology of Hearing* (4th ed.). New York: Academic Press.

Moore, C. B., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *Journal of the Acoustical Society of America, 102*, 1864-1877.

Nabelek, I. V., Nabelek, A. K., & Hirsh, I. J. (1970). Pitch of tone bursts of changing frequency. *Journal of the Acoustical Society of America, 48*, 536-553.

Nabelek, I. V., Nabelek, A. K., & Hirsh, I. J. (1973). Pitch of sound bursts with continuous or discontinuous change of frequency. *Journal of the Acoustical Society of America, 53*(5), 1305-1312.

Nagano-Madsen, Y., & Eriksson, L. (1989). The location of the F0 turning point as a cue to mora boundary, *Speech Transmission Laboratory - Quarterly Progress and Status Report* (Vol. 1, pp. 41-45). Stockholm: Royal Institute of Technology.

Narmour, E. (1991). *The Analysis and Cognition of Basic Melodic Structures*. Chicago: University of Chicago Press.

Narmour, E. (1999). Hierarchical expectation and musical style. In D. Deutsch (Ed.), *The Psychology of Music* (2nd ed., pp. 441-472). New York: Academic Press.

Nash, R., & Mulac, A. (1980). The intonation of verifiability. In L. R. Waugh & C. H. van Schooneveld (Eds.), *The Melody of Language* (pp. 219-242). Baltimore: University Park Press.

Niebuhr, O. (2003). Perceptual study of timing variables in F0 peaks. In *Proceedings of the 15th International Congress of Phonetic Sciences,* Barcelona, 1225-1228.

O'Connor, R. J., & Arnold, G. F. (1973). *Intonation of Colloquial English*. Bristol, U.K.: Longman Group, Ltd.

Odden, D. (1995). Tone: African languages. In J. Goldsmith (Ed.), *The Handbook of Phonological Theory* (pp. 444-475): Blackwell.

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

Palmer, H. (1922). *English Intonation, with Systematic Exercises*. Cambridge: Heffer.

Perlman, M., & Krumhansl, C. L. (1996). An Experimental Study of Internal Interval Standards in Javanese and Western Musicians. *Music Perception, 14*(2), 95-116.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. Unpublished Ph.D. dissertation, MIT, Cambridge, MA.

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

Pierrehumbert, J., & Beckman, M. (1988). *Japanese Tone Structure*. Cambridge, MA: MIT Press.

Pierrehumbert, J., & Steele, S. A. (1989). Categories of tonal alignment in English. *Phonetica, 46*, 181-196.

Pike, E. V., & Wistrand, K. (1974). Step-up terrace tone in Acatlan Mixtec (Mexico). In R. M. Brend (Ed.), *Advances in Tagmemics, North Holland Linguistic Series* (Vol. 9, pp. 81-104). Amsterdam: North-Holland.

Pike, K. L. (1945). *The Intonation of American English*. Ann Arbor: University of Michigan Press.

Prieto, P., Shih, C., & Nibert, H. (1996). Pitch downtrend in Spanish. *Journal of Phonetics, 24*(4), 445-473.

Prieto, P., van Santen, J., & Hirschberg, J. (1995). Tonal alignment patterns in Spanish. *Journal of Phonetics, 23*, 429-451.

Pulleyblank, D. (1986). *Tone in Lexical Phonology*. Dordrecht: D. Reidel Publishing Co.

Purcell, E. (1976). Pitch peak location and the perception of Serbo-Croatian word tone. *Journal of Phonetics, 4*, 265-270.

Purnell, T. (1997). *Principles and parameters of phonological rules: Evidence from tone languages.* Unpublished Ph.D. dissertation, University of Delaware.

Rausch, R. A., & Plomp, R. (1999). The perception of musical tones. In D. Deutsch (Ed.), *The Psychology of Music* (2nd ed.). New York: Academic Press.

Redi, L., & Shattuck-Hufnagel, S. (2001). Variation in the realization of glottalization in normal speakers. *Journal of Phonetics, 29*(4), 407-429.

Remijsen, B., & van Heuven, V. J. (1999). Gradient and categorical pitch dimensions in Dutch: diagnostic tests, *Proceedings of the 14th International Congress of Phonetic Sciences* (Vol. 3, pp. 1865-1868). San Francisco.

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. In N. J. Lass (Ed.), *Speech and Language, Advances in Basic Research and Practice* (Vol. 10, pp. 243-335). Orlando: Academic Press, Inc.

Rice, K. D. (1987). Metrical Structure in a Tone Language: The Foot in Slave (Athapaskan), *Proceedings of the 23rd Annual Regional Meeting of Chicago Linguistics Society* (Vol. 23, pp. 239-252).

Rose, P. (1987). Considerations in the normalisation of the fundamental frequency of linguistic tone. *Speech Communication, 6*, 343-351.

Schachter, P. (1961). Phonetic similarity in tonemic analysis. *Language, 37*, 231-238.

Schubiger, M. (1958). *English Intonation, Its Form and Function*. Verlag/Tübingen: Max Niemeyer.

Selkirk, E. O. (1984). *Phonology and Syntax: The Relation Between Sound and Structure*. Cambridge, MA: MIT Press.

Shattuck-Hufnagel, S., Dilley, L., Veilleux, N., Brugos, A., & Speer, R. (2004). F0 peaks and valleys aligned with non-prominent syllables can influence perceived prominence in adjacent syllables. In *Proceedings of Speech Prosody 2004*. Nara, Japan.

Siegel, J. A., & Siegel, W. (1977a). Absolute identification of notes and intervals by musicians. *Perception & Psychophysics, 21*(2), 143-152.

Siegel, J. A., & Siegel, W. (1977b). Categorical perception of tonal intervals: Musicians can't tell *sharp* from *flat*. *Perception & Psychophysics, 21*(5), 399-407.

Silverman, K., & Pierrehumbert, J. (1990). The timing of prenuclear high accents in English. In J. Kingston & M. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* (pp. 71-106). Cambridge: Cambridge University Press.

Silverman, K. E. A., Beckman, M., Pierrehumbert, J., Ostendorf, M., Wightman, C. W. S., Price, P., & Hirschberg, J. (1992). ToBI: A standard scheme for labeling prosody, *Proceedings of the 2nd International Conference on Spoken Language Processing* (pp. 867-879). Banff.

Smiljanic, R. (2004). *Lexical, Pragmatic, and Positional Effects on Prosody in Two Dialects of Croatian and Serbian: An Acoustic Study*. New York: Routledge.

Snider, K. (1999). *The Geometry and Features of Tone*. Dallas: Summer Institute of Linguistics and the University of Texas at Arlington Publications in Linguistics 133.

Snider, K., & van der Hulst, H. (1993). Issues in the representation of tonal register. In H. van der Hulst & K. Snider (Eds.), *The Phonology of Tone: The Representation of Tonal Register* (pp. 1-27). Berlin, New York: Mouton de Gruyter.

Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America, 111*(4), 1872-1891.

Stevens, K.N. and S.J. Keyser (1989) Primary features and their enchancement in consonants. *Language, 65*(1), 81-106.

't Hart, J., Collier, R., & Cohen, A. (1990). *A Perceptual Study of Intonation*. Cambridge: Cambridge University Press.

Tenzer, M. (1991). *Balinese Music*. Singapore: Periplus Editions.

Thorsen, N. (1980). A study of the perception of sentence intonation - evidence from Danish. *Journal of the Acoustical Society of America, 67*, 1014-1030.

Trim, J. L. M. (1959). Major and minor tone-groups in English. *Le Maître Phonetique, 112*, 26-29.

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

Truckenbrodt, H. (in press). Final lowering in non-final position. *Journal of Phonetics*.

Turk, A. & Shattuck-Hufnagel, S., (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics, 28*, 397-440.

van den Berg, R., Gussenhoven, C., & Rietveld, A. C. M. (1992). Downstep in Dutch: Implications for a model. In G. J. Docherty & D. R. Ladd (Eds.), *Papers in Laboratory Phonology II: Gesture, Segment, Prosody* (pp. 335-359). Cambridge: Cambridge University Press.

van der Hulst, H., & Snider, K. (Eds.). (1993). *The Phonology of Tone: The Representation of Tonal Register*. Berlin, New York: Mouton de Gruyter.

van Noorden, L. P. A. S. (1977). Minimum differences of level and frequency for perceptual fission of tone sequences ABAB. *Journal of the Acoustical Society of America, 61*, 1041-1045.

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

Viechnicki, P. (2002). *Composition and granularity of vowel production targets*. Unpublished Ph.D. dissertation, University of Chicago.

Wade, B. (1979). *Music in India: The Classical Traditions*. Englewood Cliffs, NJ: Prentice-Hall.

Wang, W. (1967). Phonological features of tone. *International Journal of American Linguistics, 33*, 93-105.

White, B. W. (1960). Recognition of distorted melodies. *American Journal of Psychology, LXXIII*, 100-107.

Williams, E. S. (1971/6). Underlying tone in Margi and Igbo. *Linguistic Inquiry, 7*, 463-484. [Ms. written in 1971].

Williamson, K. (1972). *Igbo-English Dictionary*. Benin City: Ethiope Publishing Corp.

Wong, P. C. M., & Diehl, R. L. (2003). Perceptual normalization for inter- and intra-talker variation in Cantonese level tones. *Journal of Speech, Language, and Hearing Research, 46*, 413-421.

Woodrow, H. (1909). A quantitative study of rhythm. *Archives of Psychology, 14*, 1-66.

Woodrow, H. (1911). The role of pitch in rhythm. *The Psychological Review, 18*, 54-77.

Woodrow, H. (1951). Time perception, *Handbook of Experimental Psychology* (pp. 1224-1236). New York: Wiley.

Xu, Y. (1994). Production and perception of coarticulated tones. *Journal of the Acoustical Society of America, 95*(4), 2240-2253.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics, 25*, 61-83.

Xu, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica, 55*, 179-203.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics, 27*, 55-105.

Xu, Y. (2001). Fundamental frequency peak delay in Mandarin. *Phonetica, 58*, 26-52.

Xu, Y., & Xu, C. X. Intonation components in short English statements, submitted.

Yip, M. (1980). *The tonal phonology of Chinese*. Unpublished Ph.D. dissertation, MIT, Cambridge, MA.

Yip, M. (1993). Tonal Register in East Asian Languages. In H. van der Hulst & K. Snider (Eds.), *The Phonology of Tone: The Representation of Tonal Register* (pp. 245-268). Berlin, New York: Mouton de Gruyter.

Zatorre, R. J., & Halpern, A. R. (1979). Identification, discrimination, and selective adaptation of simultaneous musical intervals. *Perception & Psychophysics, 26*(5), 384-395.

Zec, D. (1999). Footed tones and tonal feet: rhythmic constituency in a pitch-accent language. *Phonology, 16*(2), 225-271.

Zhang, J. (2001). *The effects of duration and sonority on contour tone distribution – typological survey and formal analysis*. Unpublished Ph.D. dissertation, University of California at Los Angeles.

---

[1] The character of segmental opposition was described in the following way by de Saussure (1966): "Every language forms its words on the basis of a system of sonorous elements, each element being a clearly delimited unit, one of a fixed number of units. Phonemes are characterized not, as one might think, by their own positive quality but simply by the fact that they are distinct. Phonemes are above all else opposing, relative, and negative entities" (p. 119).

[2] Pierrehumbert (1980) defined the opposition between H and L by stating that H is higher than L in a given context, by which was meant the context was a single tonal position. In this way, the definition of the tonal distinction was expressly paradigmatic, since it crucially required no reference to another position, in order for the distinction to be made.

[3] Wong and Diehl (2003) note that Chao (1947) actually described Tone 3 as approximately four semitones lower than Tone 1. However, Chao's musical notation of Cantonese tones indicated that Tone 3 was actually three semitones lower than Tone 1, which agreed with the first author's native intuitions about the pitch difference between these tones.

[4] I wish to enthusiastically thank Victor Manfredi and Keith Snider, respectively, for suggesting these examples.

[5] The term used in music theory for this up-and-down patterning of notes is *contour*. We have avoided the use of this term because of the potential for confusability with F0 contours.

[6] One way in which researchers have attempted to account for data showing distinct relative height relations is through a principle of autosegmental theory known as the Obligatory Contour Principle (OCP). Failure of two identical-valued tones to stay at the same level was interpreted as a failure of the OCP, which supposedly mandated phonetically level contours across sequences of identical-valued tones. Differences in tone height were then viewed as the outcome of a phonetic rule. There are three problems with this view, some of which have been discussed elsewhere. First, the account is not adequate to deal with cross-linguistic data, since the three attested cases (i.e., whether a tone is higher than, lower than, or the same level as another tone) cannot be described in terms of a two-way contrast (i.e., adherence to the OCP vs. failure to do so). Second, any attempt to account for all three attested relations would force the claim that rules operate in a language-specific way, which is unattractive. Third, to grant phonetic rules the ability to generate meaning differences is to endow them with a level of power not previously attested.

[7] It was clear from the start that there were problems with applying autosegmental theory across languages. For example, Goldsmith himself noted in an extended footnote on pp. 260-262 that a supposedly universal principle, the Well-Formedness Condition, didn't seem to match data about English!

[8] Snider (1999) explicitly argues for tone features which have a syntagmatic interpretation. He points out that Inkelas *et al.* (1986) and Inkelas and Leben (1990) do not explicitly argue for syntagmatic tonal features, but this is nevertheless implicit in their approach.

[9] Since I began working on linguistic tone and its relation to music, one of the central questions in my mind was always how best represent syntagmatic and paradigmatic features in a unitary theory. The basic ideas of the present theory are much the same as when I started working them out several years ago, but the representation of those ideas has changed dramatically over time. The most significant insight for me was discovering that the phonology of tone and intonation is essentially mathematical in its nature. Early on in this work, I was using "arrow" diacritics to express the categories *higher*, *lower*, and *same*. This notation and the accompanying ideas were derived independently of the work of Mary Clark for Igbo, about whose work I did not learn until very late in the process of producing this thesis. At some level I knew that these entities had a rather straightforward mathematical interpretation in terms of abstractions on frequency ratios. It took me quite a while to realize the usefulness of the abstract frequency ratio constructs. I was in Edinburgh working with Bob Ladd in May, 2003 when it suddenly dawned on me one afternoon that the mathematical properties of these abstract frequency ratios obviated the need for the otherwise arbitrary phonological rules I had been developing, as systematic and regular as the rules had

---

seemed. The system I had been developing took an entirely new turn, and tone intervals were born.

[10] The note at 400 Hz, of course, marks the octave equivalent of the first note in the scale at 200 Hz.

[11] My sincere thanks go to Ani Patel for bringing the work of Perlman and Krumhansl (1996) to my attention, as well as for many interesting and helpful discussions.

[12] The symbol μ was selected to represent the tonic level because in mathematics this symbol is often used to represent the mean or average. In this regard, this symbol seemed like a good choice to represent the "normalizing level" in the speaker's pitch range.

[13] The fact that the reference level in this theory is realized under predictable circumstances sets the current theory apart from a number of other theories, which have instead assumed that the reference level may or may not be realized phonetically (e.g., Pierrehumbert 1980, Pierrehumbert and Beckman 1988). The advantage of the assumption that the tonic is predictably realized is that the theory becomes much more amenable to empirical tests of its validity. Moreover, it suggests a way that listeners might actively accommodate for such tonal effects as declination during the speech perception process, since they can continually update their phonetic definition of the "reference level" whenever an instance of the tonal category associated with the tonic is heard.

[14] We note that a similar idea to a tonic referent level is entailed in the notion of *key* proposed by Brazil (1997).

[15] We note that syntagmatic features have often been characterized as describing a relation to what came before in the utterance (e.g., Ladd 1996). The present formalism makes clear that we could just as well describe a syntagmatic relation in terms of what comes next in the utterance. The choice is arbitrary, but taking a convention which reflects forward-going time (as opposed to backward-going time) makes more sense, given that natural language proceeds in a forward direction.

[16] Note that any statements about the relative magnitudes of the numerator and denominator of a tone interval is conditional on the assumption of the default "directionality" for syntagmatic and paradigmatic tone intervals, which was discussed in section 2.3. Recall that the default convention for a syntagmatic tone interval is that the earlier tone is in the denominator. Then, for example, the relation ***higher*** when specified for a syntagmatic tone interval implies that the later tone is higher than the earlier tone, rather than the other way around. Similarly, the relation ***higher*** for a paradigmatic tone interval specifies that the referring tone is higher than the tonic, rather than the other way around.

[17] Note that because tone intervals are always positive, the relation lower entails that there are no values of I which are zero or negative. So *lower* is really equivalent to the tone interval range $0 < I < 1$. The expression $I < 1$ will be used here with the understanding that no values of I equal to or less than zero satisfy the expression.

[18] Here, the notation is drawn from standard autosegmental theory. It is necessary to point this out, because symbols like H and L are defined in a different way in Chapters 5-7.

[19] The symbol δ was chosen because in mathematics it is often taken to represent a distance or a change.

[20] Given the level of phonetic specificity that the current framework affords, more research on the phonetics of lexical tone languages will be needed in order to determine how lexical tone categories would be represented in the present theory.

[21] A capital delta is used here in order to distinguish paradigmatic from syntagmatic tone interval distances; however, the distinction is one of notation, rather than one of substance.

[22] We note that a given tonal category can be realized at different locations within a given utterance through processes such as declination. Nevertheless, work on the phonetics of lexical tone languages has shown that even when such effects obtain, the tones remain in rather distinct parts of the speaker's pitch range. For example, Laniran and Clements' (2003) work on Yoruba shows that although H and L tones decline throughout an utterance, they nevertheless are restricted to largely non-overlapping frequency "bands" within the speaker's pitch range. This sort of "frequency banding" is expected to be typical of languages which genuinely represent paradigmatic tonal categories.

[23] The implications of the assumption that one tonal category always corresponds to the tonic in paradigmatic systems are especially important when we consider that under this theory, the tonic serves as a referent or normalizing level, in both a phonological and a phonetic sense. It has often been noted that the level of a given

---

lexical tone category can drift down through an utterance through declination (e.g., Laniran and Clements 2003). The assumption that one of the lexical tone categories is always associated with the tonic referent level then provides an explanation for how this referent level can be updated phonetically so as to take account of factors like declination. Moreover, the assumption that one of the tonal categories always permits the theory to be subjected to empirical tests of its validity.

[24] Systems which posit two traditional binary-valued features do not achieve descriptive adequacy in this sense, since such theories predict only four possible tone levels.

[25] We can understand the effects of reciprocal operations more easily if we conceive of these relations in terms of their associated binary-valued features [± higher] and [± same]. First, consider why taking the reciprocal leaves the value of the feature [± same] unaffected. For a tone interval which is [+same], the relative magnitude of the numerator and denominator are equivalent, so that dividing the former by the latter gives $I = 1$. Taking the reciprocal of $I = 1$ gives $I^R = 1$. Because this is the same value as when we started out, we can conclude that taking the reciprocal of a tone interval which is [+same] yields a reciprocal tone interval which is still [+same]. Similarly, if a tone interval is [-same], we can infer that the relative magnitudes of the numerator and denominator are not the same, so that $I \neq 1$. Taking the reciprocal of such a tone interval still gives $I^R \neq 1$. Thus we can conclude that if a tone interval starts out with the feature [-same], taking the reciprocal of that tone interval yields a tone interval which is still [-same]. In sum, taking the reciprocal of a tone interval does not affect the value of the feature [±same].

Next, we can consider how the reciprocal transform affects the feature [±higher]. Consider first the case where the tone interval is [+higher], so that $I > 1$. Taking the reciprocal means that the entity with the smaller magnitude comes to occupy a position in the numerator, while the entity with the larger magnitude comes to occupy a position in the denominator. Then dividing these terms means that the value for the tone interval will now be less than 1, $I < 1$. But this corresponds to a specification of [-higher]! In other words, if we start out with a tone interval which is [+higher], then taking the reciprocal of this tone interval will yield the featural specification [-higher], and vice versa.

[26] Although we assume that tones are aligned with timing slots, we leave it for future work to define exactly what a timing slot is in either phonological or phonetic terms. For the present purposes it suffices to say that each syllable is generally associated with one or two timing slots on a language-specific basis, depending on factors such as lexical stress. Moreover, phrase-final syllables may also have three or more timing slots, consistent with their longer phonetic length (e.g., Klatt 1976, Turk and Shattuck-Hufnagel, 2000). In some theories, timing slots correspond to a constituent known as a mora; however, we will take no position here regarding whether the mora is the unit with which tones associate.

[27] We note that various authors, including Selkirk (1984), have defined the notion of a *silent beat*, which corresponds to a timing slot which is not associated with segmental or syllabic material. We assume that only timing slots which are associated with segmental or syllabic material may anchor a tone.

[28] In the case of paradigmatic tone intervals, tones mark particular positions to fulfill multiple relative height relations simultaneously. That is, a paradigmatic tone interval restriction will require a particular tone to be a certain relative height with respect to the tonic, while one or more syntagmatic tone interval restrictions will require a particular tone to be a certain relative height with respect to another tone or tones in the same utterance or discourse.

[29] To clarify these issues further, we note that this theory does not permit any mechanisms by which tones can "spread" to all tone-bearing units in a given region, as was earlier assumed under autosegmental theory. Tone spreading was represented by drawing multiple lines of association between a tone and tone-bearing units; the phonetic outcome of such a representation was assumed to be a level contour. Rather, phonetic evidence of level tone contours which previously would have described in terms of tone spreading processes will be described here in terms of syntagmatic tone intervals specifying the relation *same*; some examples to this effect will be presented in Chapter 5.

[30] The theory of Pierrehumbert (1980) also claimed that an F0 contour arose through one of two processes: either it came about through interpolation between tonal targets, or it came about through direct instantiation of tonal targets. However, the theory did not require that a tone be present in initial positions in phrases. This led to a theory-internal ambiguity with respect to how this portion of the F0 contour arises across a sequence of initial unstressed syllables, since logically speaking, they could not obtain their F0 values through either of the processes Pierrehumbert

---

described.

[31] The present theory thus contrasts with that of Pierrehumbert (1980), which assumed the existence of both monotonic and non-monotonic interpolation functions.

[32] The intonation phrase under the present theory is similar to all of the following: the intonational phrase of Liberman (1975), the "tone unit" of Crystal (1969), the "sense group" of Armstrong and Ward (1926) and Vanderslice and Ladefoged (1972), and the "tone group" of Ashby (1978) and Halliday (1967).

[33] Among the phenomena which we believe can be insightfully explained in terms of principles of general auditory perception is *final lowering*, a phenomenon whereby the last accent in a phrase is realized with a lower pitch than would be expected based on an earlier pattern of steps down on accented syllables (Liberman and Pierrehumbert 1984; Truckenbrodt, forthcoming).

[34] A simpler expression of the PTA might have been "Every tone must associate with exactly one timing slot; every timing slot may associate maximally with only one tone." However, such a formulation runs into problems under current linguistic theories of timing (see e.g., Broselow, 1995). The primary reason that (b) appears to be permitted in the phonology has to do with the observation that short vowels can realize a contour tone. (See e.g., Odden, 1995.) If it is correct that short vowels always entail exactly one timing position, as claimed by Broselow (1995), then a theory which did not permit the structure in (b) would be unable to account for this observation. We think that ultimately, vowel quality and vowel duration entail quasi-orthogonal aspects of the representation, such that short vowels can be associated with more than one timing position. This would ultimately predict that such vowels would be lengthened phonetically when they were associated with multiple tones. Zhang (2001) has claimed that there is indeed a relationship between phonetic length and the complexity of tonal contours, where complexity is defined partly in terms of the number of pitch targets.

[35] Previously, some of the strongest evidence supporting the existence of floating tones was shown by Clements and Ford (1979) for Kikuyu. A close examination of this data shows that it can readily be accommodated within the present framework, which does not permit floating tones. In particular, it appears that Clements' and Ford's data can be accounted for by assuming a diachronic process which involved a rightward shift of tones, such that the default syntagmatic feature *lower* arising from paradigmatic Low after High came to be realized as a small step down on the following High.

[36] In fact, a number of difficulties with Selkirk's (1984) account of metrical stress and intonation at the phrasal level might have been avoided by assuming that intonation is assigned *after* metrical structure.

[37] It should be noted that the descriptive domain of metrical stress theory has traditionally been the word; far less work has been done on formalizing metrical structures at the phrasal level. This presents certain drawbacks in developing a general theory of tone based on phrase-level metrical structure. Among other issues, it means that certain aspects of the current proposal are difficult to evaluate, particularly as they relate to higher grid structures. However, metrical grids appear to capture quite well native-speaker intuitions about the relative prominence and timing properties of utterances. Therefore, we feel that the prospects for grounding metrical structures in empirical observation and experimentation through future work is quite good, which somewhat mitigates these concerns.

[38] We consider it an important issue for future work to ground representations of metrical structure in empirical work in production and perception of durational patterns. Some empirical work of this kind has already been carried out (e.g., Broselow *et al.*, 1997), but more work is needed. Indeed, without further research to empirically ground the current proposals, it is not possible to properly evaluate these claims through phonetic studies.

[39] We refer the interested reader to work on unstarred tone alignment in phrase-final position by Gussenhoven (1999) and Grice, Ladd, and Arvaniti (2000).

[40] One position in which unstarred tone association appears to not follow the MPAR is at the ends of phrases. In particular, when two or three unstarred tones are in sequence at the end of a phrase, the "penultimate" unstarred tone appears to exhibit dual behavior. By default, it appears to be attracted rightward (*T+*) to a position just before the phrase-final unstarred tone. On the other hand, when a secondary stress syllable is present, it appears to be attracted to that syllable, thereby behaving like a starred tone (Gussenhoven 1999; Grice *et al.*, 2000; Lickley *et al.*, forthcoming).

[41] We note that in relating phonetic data to a set of well-defined underlying primitives, it may be necessary to take

---

into account language-specific adjustments, such as feature enhancements. (See Stevens and Keyser 1989, Keyser and Stevens 2001).

[42] Note that the specific temporal alignment of F0 points differs from one language to another. Moreover, the temporal characteristics of alignment may even differ from one dialect to another (e.g., Atterer and Ladd 2004).

[43] Theories based on discrete tones or targets include Pike (1945), Trager and Smith (1951), Liberman (1975), and the work of Pierrehumbert and colleagues (Pierrehumbert 1980; Beckman and Pierrehumbert 1986; Pierrehumbert and Beckman 1988). Theories which would be classified as based on rises and falls include Bolinger (1951, 1958), the IPO model ('t Hart, Collier, and Cohen 1990) and the British school of intonation (e.g., Halliday 1967; Crystal 1969; Cruttenden 1986). Another important theoretical view which falls into neither of these categories is the work of Xu and colleagues (Xu 1998, 1999; Xu and Wang 2001). In Xu and colleagues' theory, the primitives are articulatory in nature; according to this theory, the speaker attempts asymptotic approximation to underlying dynamic or static targets. The theory of Xu and colleagues, like the tonal target models, is quite consistent with extant phonetic data concerning F0 alignment.

[44] We assume that the listener, for his or her part, "hears out" precisely those pitches which correspond to underlying tones; the relative heights of these pitches are then compared, and the phonological representation is transparently extracted.

[45] There are some subtleties of this view which should be mentioned. First, the fact that this theory assumes that all pitch interpolation functions are monotonic means that all pitch minima and pitch maxima will be associated with positions of underlying tones. Conversely, no local pitch maximum or minimum can arise on a given tone-bearing unit without an underlying tone being located at that position.

[46] It is well-established that F0 variations in speech are differentially salient ('t Hart, Collier and Cohen 1990, de Pijper 1983, House 1990, d'Alessandro and Mertens 1995, D'Imperio and House 1997, D'Imperio 2000). This suggests that some components of the F0 contour are less relevant to the phonological representation than others. The tone interval theory seeks only to explain those variations in F0 timing which have perceptual significance.

[47] In this chapter, we have chosen to evaluate how well the tone interval theory and various paradigmatic theories predict the presence of F0 peaks and valleys from the phonological representation. We focused on these points, because to date far more phonetic work has been done on such points than on F0 corners, although the timing of corners appears to be significant for meaning contrasts (Grice, Ladd, and Arvaniti 2000). We could just as well have evaluated each theory with respect to how well it predicted the occurrence of F0 corners from the phonology, and in this case we would arrive at the same set of conclusions about the ability of each theory to do so. A number of examples are presented in Chapter 5 of how F0 corners are accounted for in terms of tone intervals.

[48] The H*+H accent was later rescinded from the English inventory by Beckman and Pierrehumbert (1986).

[49] The details are as follows. Starting with Eqn. 4 (p. 145), we can first prove that Pierrehumbert's equations permitted L to be higher than an adjacent tone, H_1. Eqn. 4 is given below:

$$f(\mathrm{L}) = n \cdot f(\mathrm{H}_1)\frac{p(\mathrm{H}_1)}{p(\mathrm{L})} \qquad \text{(I)}$$

Rearranging terms, we have the following:

$$\frac{f(\mathrm{L})}{f(\mathrm{H}_1)} = n \cdot \frac{p(\mathrm{H}_1)}{p(\mathrm{L})} \qquad \text{(II)}$$

In this equation, $0 < n < k$. From Eqn. 3, we know that $0 < k < 1$, so collapsing the two inequalities, we have $0 < n < k < 1$. When $p(\mathrm{H}_1)/p(\mathrm{L}) > 1/n$, the F0 level of L will be higher than the F0 level of adjacent H_1. No restrictions are in place to prevent this situation from occurring.

We can also prove that equations permitted the level of the L to be higher than the level of the following adjacent tone, H_2. Rearranging terms in equations (I) and (II), above, and substituting, we obtain:

$$\frac{f(\mathrm{L})}{f(\mathrm{H}_2)} = \frac{n}{k} \cdot \frac{p^2(\mathrm{H}_1)}{p(\mathrm{L})p(\mathrm{H}_2)} \tag{III}$$

This proves that the F0 level of L is higher than the F0 level of H_2 precisely when $p^2(H_1)/[p(L)p(H_2)] > k/n$. Again, no restrictions are in place to prevent this from occurring.

[50] Not all of these theories lead to the problems with phonological overgeneration and phonetic indeterminacy described above for P80 and PB88. For example, the model of Bruce and Gårding (1978) requires tones to be scaled with respect to phrasal reference lines. However, overcoming problems with overgeneration and indeterminacy in this way seems to come at the expense of some phonetic explicitness in specifying the mechanisms for scaling both H and L tones. As we show later in the chapter, a simpler account of tone scaling is possible by assuming that the tones are scaled with respect to one another directly.

[51] The phonetic model of Liberman and Pierrehumbert (1984) is very complicated, and it was not possible to determine how this theory scales H with respect to adjacent L.

[52] Just as in Chapter 4, we are concerned with predicting the *significant* F0 turning points, namely, those turning points associated with the gross shape of the F0 curve across many observations, rather than the fine details of F0 turning points on a single observation.

[53] This notation clearly emphasizes the relations of tones with respect to other tones to the left, as compared to the relations of tones to with respect to tones to the right. This choice to utilize a left-to-right representation is justified by the fact that language unfolds left-to-right in time. Thus it makes sense to favor a representation which more directly reflects how each tone would be compared to a previous tone, over one which depicts how each tone would be compared to an upcoming tone. It is worth emphasizing that in spite of the unidirectional nature of this representation, tone intervals clearly involve a "bidirectional" relation. This bidirectionality permits a more straightforward account of metathesis operations on tones, as discussed in Section 2.6. Moreover, this property also permits an account of stepping relations, as shown in Appendix A.2.

[54] Implicit in this statement is the assumption that contours like *high rise* vs. *low rise* as discussed in O'Connor and Arnold (1973) are not categorically distinct in nature, but rather two endpoints of a phonetic continuum. We point out that if such a distinction is indeed categorical in English, it is not problematic for the current theory, which can accommodate such a result by assuming that English represents paradigmatic primitives as well.

[55] In Section 5.3.6 we discuss another kind of F0 characteristic, namely a *slope change point*, which can arise when a sequence of two tone intervals have identical values: *higher higher* or *lower lower*. This F0 characteristic technically was not counted in our earlier definition of F0 turning points in Chapter 4.

[56] Note the phrase-initial tone does not have a "+." In general, we will not use a "+" for the first or last unstarred tone in a phrase. This is because we assume that the Minimum Tone requirement of Chapter 3 supersedes the requirement that unstarred tones must align next to starred tones.

[57] We note that the symmetry of "mirror image" tone pairs can be described formally in terms of the Reciprocal Property described in Chapter 2, which explains why the inverse of a fall is a rise, and vice versa.

[58] It is clear that theory-internal considerations were what led Pierrehumbert to propose these two distinct treatments for F0 valleys. In particular, her proposal for a non-monotonic interpolation function treated the transition between H* accents as a special case of more general monotonic transitions. Pierrehumbert was clearly aware of the disadvantages of these theory-internal choices, and she expressed dissatisfaction with the exceptional treatment of tone and transition in her model. She states of the latter case (p. 70): "This complication in the interpolation rules is in some ways unattractive, and we have made a serious attempt to get rid of it by developing an account under which the dip…arises from a L tone. It does not appear possible to do this without considerably changing the form of the theory."

[59] Moreover, we note that because the F0 peak coincides with the right edge of the word *mother*, it is technically a possible location of the phrase accent of an intermediate intonational phrase. As a result, the valley-peak-valley sequence could also be described as L* H- L*.

---

[60] It has been claimed that auditory impressions can aid in judging whether an intermediate IP is present or not (Pierrehumbert and Beckman 1988). However, listeners trained in the ToBI system (Silverman *et al.* 1992) often have a hard time determining whether such a boundary is being heard. It seems that many cases of apparent durational lengthening at the site of phrase accents seem to be explainable by appealing to the rhythmic properties of metrical grids.

[61] Here, we assume that these restrictions on identical-valued tone intervals hold only for sequences of adjacent tones on the lowest row of the grid-matrix complex, which was discussed in Chapter 3. In contrast, sequences of accents realizing tone intervals between nonsequential tones on higher rows of the grid-matrix complex are readily permitted to exhibit sequences of identical-valued tone interval relations, thereby accounting for iterative "downstepping" and other contours, as discussed in Appendix A.2.

[62] However, even in cases where there is no local slope change on an accented syllable, the analysis suggests that we may be able to infer the presence of a starred tone in non-phrase-final position by the presence of a downstream $H^{(*)}$ or $L^{(*)}$ tone, which will show up as a local F0 turning point.

[63] In the framework of P80 or BP86 this portion of the contour would be described as H+L* L-L%, while it would be described as H+!H* L-L% in the ToBI system (Silverman *et al.* 1992).

[64] In contrast, we suspect that the situation is different in lexical tone languages. In such languages, tones can probably be maintained in the surface phonological representation even if they realize a syntagmatic relation of same with two adjacent tone intervals. This is because listeners can rely on other cues than pitch differences to determine whether tones are present, such as lexical cues, knowledge of the interaction between tone and stress, etc.

[65] This insight comes from more than five years' personal experience in using the ToBI system to label speech corpora, as well as teaching it to others. It has also been the general impression of others whom I have spoken with who are familiar with the system.

[66] As we noted, the difference in meaning between the contours in Figures 5.27 and 5.28 seems to be given primarily or exclusively by the size of the excursion between the two level regions in each case. However, for contours like that in Figure 5.28 there sometimes seems to be a voice quality change of some sort, such as creak or increased breathiness, at the location of the step down, although this voice quality difference does not appear to be necessary in order to obtain a meaning difference between the two contours.

[67] We cannot offer a complete formulation of this restriction here. However, the form of the restriction is something like "disallow $L\ H^*\ L$ sequences after the last intended accented syllable in the intonational phrase."

[68] We note that the issue of whether these intonation patterns indeed convey narrow focus has not been independently verified. Indeed, an alternative possibility is that the "AB" pattern corresponds to "broad focus" on the entire utterance.

[69] Close inspection of the data for the four speakers in the experiment reported in LP84 indicates that the scaling of accents in the AB case for each speaker was highly consistent. In particular, speakers DWS, MYL, and JBP produced F0 values on Peak 2 and Peak 1 such that the ratio of (Peak 2/Peak 1) was well-fit by straight lines with y-intercepts of zero and slopes of 3/4, 2/3, and 2/3, respectively. Moreover, speaker KXG's data was reasonably well-fit by a straight line with a slope of 3/5 which passes near the origin. While speakers DWS, MYL and JBP produced peaks according to a constant ratio less than 1 throughout the entire experiment, speaker KXG appeared to produce peaks consistently according to a ratio less than 1 and to adopt slightly different values of this ratio over the experiment's duration. The fact that all speakers controlled the ratio of peak F0 values to within a small range less than 1 is exactly what we would expect if speakers manipulate tone intervals for these accents. Moreover, the form of the data is predicted by tone interval theory for the productions of three of the four speakers. The fact that the tone interval approach predicts not just the general shape, but the precise quantitative form of the data provides further support for the theory. An additional note on the data is that the ratios 3/4, 2/3, and 3/5 correspond to musical intervals: a perfect fourth, a perfect fifth, and a major sixth, respectively. This further suggests that the analogy between melodic intervals in music and tone intervals in language is well-motivated. Finally, the fact that speakers adopted slightly different ratios supports the analysis in Section 5.5.2 that the phonological relation involves a narrow range of possible values, such that $1 < I < \delta_{-}$ describes the relation between accents on *Anna* and *Manny*, with $\delta_{-} \approx 0.56$. As a result, we conclude that the differences across speakers in observed values are phonetic in nature for English.

---

[70] We make note additionally of Silverman and Pierrehumbert's (1990) study of rising contours in English in nuclear and prenuclear positions. However, the methodology of this study did not permit examination of the role of F0 alignment in producing phonological contrast.

[71] Somewhat different discrimination behavior is observed when the endpoints of acoustic continua are stop consonants, as opposed to vowels (Repp 1984). For example, when the endpoints of the acoustic continuum are stop consonants, the shift along the continuum from good to poor discrimination performance is quite abrupt, yielding a distinct discrimination "peak". On the other hand, when the endpoints of the acoustic continuum are vowels, the shift from good to poor discrimination performance is more gradual, yielding a broader discrimination "peak".

[72] We make note of the fact that Pierrehumbert and Steele (1989) investigated only F0 peak timing in that study, not F0 valley timing, in spite of the fact that according to Beckman and Ayers-Elam (1997) "the crucial difference between L+H* and L*+H is the timing of the low pitched portion". This is relevant to some of the findings presented here, which found equivocal evidence for low F0 points as a phonetic basis of category distinctions.

[73] In particular, the sources which we drew on for descriptions of the phonetic criteria by which contrasts can be made included Pierrehumbert (1980), Beckman and Pierrehumbert (1986), and training materials for the ToBI intonation transcription method (Silverman *et al.* 1992, Beckman and Ayers-Elam 1997), which is based on the theory of Pierrehumbert and colleagues.

[74] Discussion and examples in Beckman and Pierrehumbert (1986) clearly indicate an assumption that peak timing about the left edge of a stressed syllable is assumed to be the basis of a phonological contrast. An F0 peak on an unstressed syllable preceding a stressed syllable is assumed to be a hallmark of a H+L* pitch accent, while an F0 peak on the stressed syllable is assumed to indicate a H* pitch accent. The distinct alignment patterns of the F0 maximum are noted to be "very salient perceptually" and they are claimed to correspond to "a clear difference in interpretation" (p. 259).

[75] Note that the H+L* accent proposed in Pierrehumbert (1980) corresponds to the H+!H* in the ToBI transcription system. In written ToBI materials, Beckman and Ayers-Elam (1997) state: "The renaming of this pitch accent type [i.e., H+L*] was intended to make the analysis somewhat more concrete and intuitive for the transcriber... The substitution of the letters `!H' for `L' in the name of the pitch accent reflects the fact that the pitch target on the accented syllable is only somewhat lower than the preceding H tone target..."

[76] Training materials for the ToBI system (Beckman and Ayers-Elam 1997) describe H* "peak accents" as involving "at most a small rise from the middle of the speaker's voice range" across some number of preceding unstressed syllables, where the F0 peak for this accent is "timed to occur on the accented syllable". It is also noted that "the actual timing of the F0 peak that realizes the high tone can vary" depending on phonetic and phonological factors, such that "the peak for the high tone can be quite late, sometimes after the actual acoustic end of the syllable."

[77] To see this ambiguity, the descriptions of H* from the ToBI training materials can be compared with the description of H+L* from Beckman and Pierrehumbert (1986). In the ToBI training materials, H* accents are referred to as "peak accents," where "the actual timing of the F0 peak that realizes the high tone can vary" such that "the peak for the high tone can be quite late, sometimes after the actual acoustic end of the syllable." However, in Beckman and Pierrehumbert (1986) the discussion is quite clear that an F0 peak on an unstressed syllable preceding a stressed syllable is assumed to be a hallmark of a H+L* accent.

[78] Written materials accompanying the ToBI transcription system rarely make reference to an F0 minimum or valley. Instead, vague language is used, such as "low tone target," "fundamental frequency value low in the pitch range," "low pitched region," "low F0 target." The only reference to an F0 valley in the descriptions of any accent is that L+H* accent is said to involve a "relatively sharp rise from a valley in the lowest part of the speaker's pitch range" [emphasis provided].

[79] Contour (2) in Figure 6.2 could also possibly be described as L*+H, consistent with the claim that "the crucial difference between L+H* and L*+H is the timing of the low pitched portion" (Beckman and Ayers-Elam, 1997).

[80] The assumption that the contour (3) in Figure 6.2 is a L+H* pitch accent implies that the "H*" portion of the accent must be on some stressed syllable following the F0 minimum.

[81] To some readers, "perceived pitch" may seem like a redundant phrase, but we have selected it deliberately to emphasize the role that perception presumably plays in the phonetics.

---

[82] In this experiment it was necessary to present all "different" pairs. Had we used the same proportion of "same" and "different" pairs, the duration of the experiment would have been prohibitively long. The unequal proportions of "same" and "different" pairs seems unlikely to have affected the outcome of the experiment.

[83] Calculation of standard error was made difficult by the fact that subjects used different decision criteria for "same" and "different". (See MacMillan and Creelman 1991 for a discussion.) Therefore, it was not possible to determine an appropriate method for calculation of standard error within time constraints, which prohibited us from presenting error data.

[84] Due to a computer error, one participant's responses in the *millionaire*-max series were not recorded.

[85] In order to take into account what appeared to be large overall differences in subjects' ability to imitate stimuli across the four series, a slightly more relaxed criteria was used to assess reliability between subjects for the minimum series than for the maximum series. This equated to using alpha levels of 0.25 and 0.1 to assess the inter-rater-reliability of subjects for the series involving manipulation of F0 minimum and maximum, respectively. Using a more stringent criterion for the minimum series would have eliminated most of the subjects.

[86] The semi-interquartile range (SIQR) is a measure of spread or dispersion. It is computed as one half the difference between the 75th percentile (Q3) and the 25th percentile (Q1), and is therefore given as (Q3-Q1)/2. Because it is not much affected by extreme scores, it is a good measure of spread for skewed distributions. Since half of the scores in a distribution lie between Q3 and Q1, the SIQR is half the distance needed to cover half the scores. Thus for a symmetric distribution, an interval stretching from one SIQR below the median to one SIQR above the median will contain ½ the scores.

[87] In particular, one group of subjects tended to place F0 valleys quite early for nearly all stimuli, while a second group of subjects tended to place valleys quite late for nearly all stimuli. A third group of subjects produced an essentially monotone F0, leading to random placement of small local F0 minima. A fourth small group of subjects appeared to have a bimodal distribution of F0 valley placements.

[88] Bob Ladd has additionally suggested (pers. comm.) that there are qualifications on ruling out the L*+H interpretation. That is, the data presented here would ideally need to be supplemented by additional evidence of a categorical distinction between a contour in which the stressed syllable is lower than the preceding unstressed syllable, and a contour in which the stressed syllable is at the same level or higher than the preceding unstressed syllable. Tone interval theory predicts that these three possibilities would correspond to distinctive, although highly similar, phonological representations. We think of the predicted differences as being analogous to slight variations on the same musical melody – the listener would recognize them as distinct, but would nevertheless hear them as being quite similar to one another. Whether each of these possibilities corresponds to an attested distinction within a single language is an issue left for future research.

[89] We note that perceiving a stress shift is also consistent with a prediction of tone interval theory, since in general a peak is predicted to be a "good" match to a pitch accent. (See Chapter 5 for discussion.) Thus, hearing the second syllable as a pitch accent is more likely when the syllable gets a peak, and when knowledge of lexical stress cannot rule out that position as being unaccented. Moreover, the likelihood of stress shift is exacerbated by the fact that the intended stressed syllable, *ming-*, lacks a significant slope change when the peak is late in this stimulus series.

[90] Due to a technical problem, one participant's data for the *nonrenewable*-min series was unavailable.

[91] Alternatively, a contour with a late peak would be described as *L H* L %* with the "H*" on *-gling-*, if the contour is perceived to have its main stress on that syllable.

[92] This prediction of course refers only to systems like English, which is presumed to represent exclusively or predominantly syntagmatic contrasts. As noted elsewhere, syntagmatic representations are assumed to be common to all languages, while paradigmatic representations are assumed to be common to a subset of tonal systems. For that subset of languages which represent paradigmatic contrasts, the level of a syllable with respect to the pitch range then becomes an important phonetic dimension for phonological contrasts.

[93] In Pierrehumbert (1980) and later work, the theoretical status of phrase-initial unstressed syllable(s) preceding the pitch accents H*, L*, H*+L and L*+H is ambiguous. Subsequent work has simply treated such syllables as tonally unspecified. (See e.g. Beckman and Ayers-Elam 1997.) As discussed in Chapter 3, the theoretical ambiguity stems from the claim of Pierrehumbert (1980) that F0 values arise by one of two mechanisms: either by direct specification

---

according to phonetic parameters associated with phonological tones, or else by phonetic interpolation. Because F0 values associated with phrase-initial unstressed syllables preceding H*, L*, H*+L or L*+H can arise through neither of these mechanisms, it is not clear under paradigmatic theories how these syllables achieve F0 specifications. For the purposes of the experiment, however, the claims of the theory are clear. If the initial unstressed syllables are in the low part of the pitch range, the phonological category should be L+H*, but if the initial unstressed syllables are in the mid- to upper part of the pitch range, the phonological category should be H*.

[94] The syntagmatic view also predicts that another possible category could arise, *same*, corresponding to the case in which two syllables were at the same level. However, the experimental stimuli and method in question were not designed to test for evidence of this category. We leave this issue for future work.

[95] Here, the target word had the American pronunciation: [ə.ˈreʲ.gə.noʷ] or [ə.ˈrɛ.gə.noʷ], depending on the speaker's dialect.

[96] In selecting this measurement, we also made note of the fact that the paradigmatic theory of English assumes that the L could be located on either initial unstressed syllable (i.e., on *Some* or on *or*-). Thus, determining the minimum F0 across both syllables provided reasonable way of estimating the L.

[97] The boundary between /r/ and /ə/ in the word *oranges* or between /r/ and /e/ in the word *oregano* was taken as the location of an increase in F3 frequency, or else the location of an increase in amplitude of F2-F4.

[98] It will be noted that estimates of tonal scaling calculated according to the pitch range method show a gradual decrease as stimulus number increases, while estimates of tonal scaling calculated according to the relative level method show a gradual increase as stimulus number increases. The reason for this difference can be understood by considering the equations for calculating tonal scaling. For the *oranges* and *oregano* series, as stimulus number increases, T_0 approaches r so that T_0-r becomes a smaller quantity in relation to $h-r$, resulting in a decrease in the value of $(T_0-r)/(h-r)$ with increasing stimulus number. In contrast, the relative level method calculates tonal scaling estimates as a ratio of T_2/T_1. Because T_1 gets smaller with increasing stimulus number, T_2/T_1 concurrently gets larger with increasing stimulus number.

[99] One attribute of the L*+H accent which is not well-specified in descriptions is the level of the accented syllable with respect to a previous syllable. As noted in Chapter 5, the present proposal would predict that the level of the stressed syllable with respect to a preceding unstressed syllable is an important phonetic dimension along which category distinctions are made. In the *linguistics* stimulus series, we obtained evidence supporting the claim that the level of a stressed syllable with respect to a preceding unstressed syllable is significant for the phonological representation. In particular, evidence consistent with a category boundary was obtained when the level of an initial unstressed syllable switched from higher than to lower than the level of a following stressed syllable.

[100] The phonetic evidence obtained in Experiments 1-4 actually suggests that here, a more accurate statement would be to say that varying F0 peak timing across the vowel onset of an upcoming syllable consistently yields evidence of a category boundary.

[101] As we discussed in Section 5.3.6, this is true unless the values of adjacent tone intervals are the same, which will happen in a highly restricted set of cases. In this situation, tones will then generally correspond to locations of a change in slope.

[102] A number of explanations in terms of a particular tonal sequence, e.g. HLH, have been offered, but for reasons discussed elsewhere this formulation is ultimately untenable. In general, autosegmental frameworks have had difficulty in accounting for stepping phenomena (Odden 1995).

[103] One proposal along these lines that gained momentum through work in African languages in the mid-1970's was that contours like those in Figures A.2(a)-(c) are the result of a particular tones sequence – HLH – plus a phonetic rule. This rule was said to lower each successive H tone with respect to the previous H tone. However, there are a number of problems with such an account. For one thing, it is not obvious how a pattern such as the one in Figure A.2(c) could be described in terms of a sequence of H's and L's, since there is no overt F0 valley in this contour realizing the L. For such contours, the L was assumed to be "floating" and thus not phonetically realized. It is worth noting that the Africanist work of the 1970's influenced the account of Pierrehumbert (1980) in significant ways. For example, Pierrehumbert adopted the "floating L" tone account in order to describe contours such as that of A.2(c) in terms of a sequence of H*+L accents.

---

[104] Here, we have implied that N is at least 5 in order to illustrate a point about the mathematical properties associated with tone intervals. Thus, we have neglected in this expression the fact that N probably is not ever greater than 4. It should be obvious that the length of the sequence does not affect the fact that cancellation obtains nonetheless.

[105] Technically speaking, there is one exception to this. For example, it could be that $N = 2$ and $j = 1$, in which case the tones would be sequential in time. This would correspond to the case of starred tones which aligned with a sequence of two stressed syllables, where no other stressed syllable intervened.

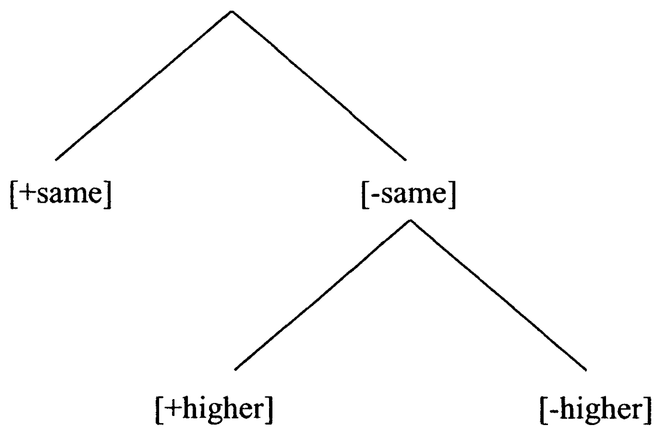
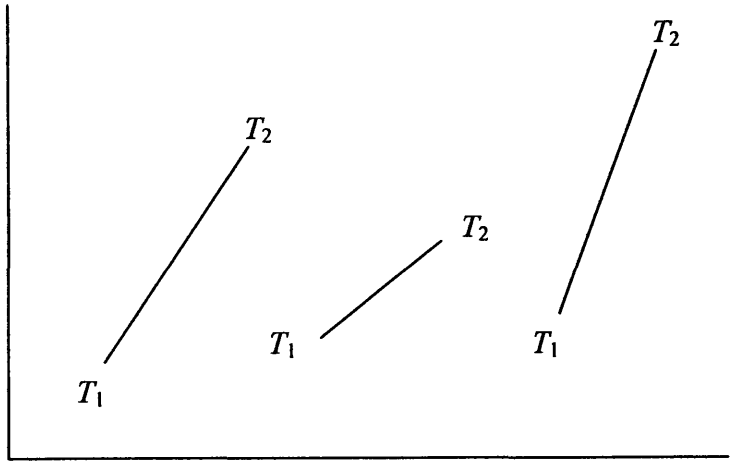


Figure 2.1: Geometry of the phonological feature specifications [±same] and [±higher] for tone intervals.

(a) $I_{1,2} > 1$



(b) $I_{\mu,0} > 1$

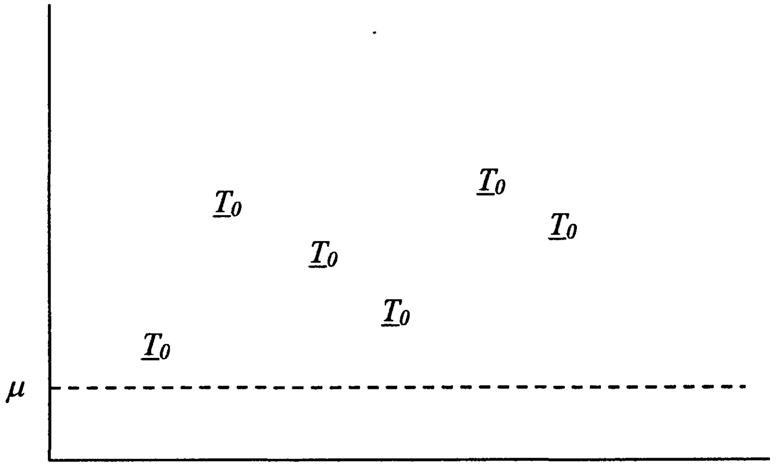
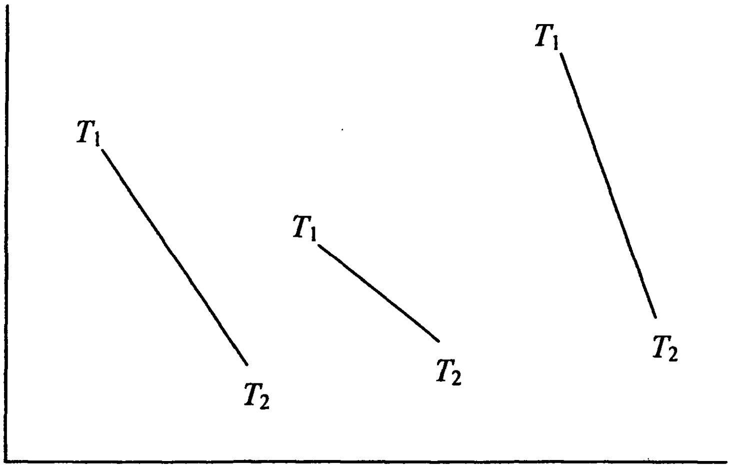
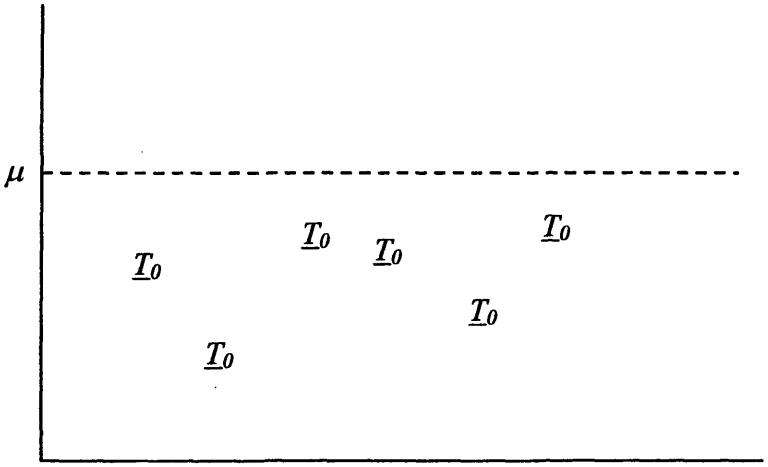


Figure 2.2: Effects of language-specific syntagmatic tone interval restrictions on the scaling of tones with respect to their referents; (a), (b) and (c) indicate three possible restrictions on falling contours, delineated by the syntagmatic intervallic cutoff δ_-, while (d), (e), and (f) indicate three possible restrictions on rising contours, delineated by the syntagmatic intervallic cutoff δ_+. See text.

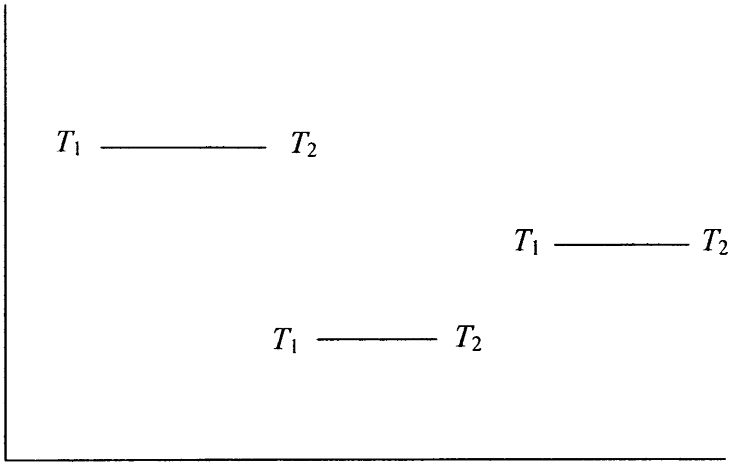
(c) $I_{1,2} < 1$



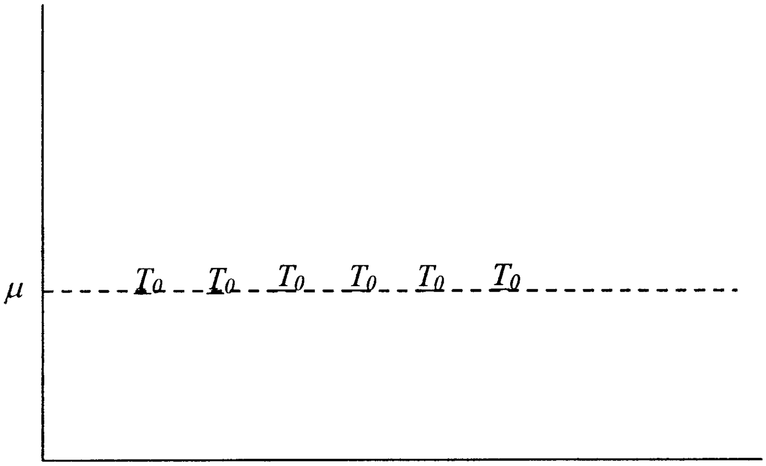
(d) $I_{\mu,0} < 1$



(e) $I_{1,2} = 1$



(f) $I_{\mu,0} = 1$



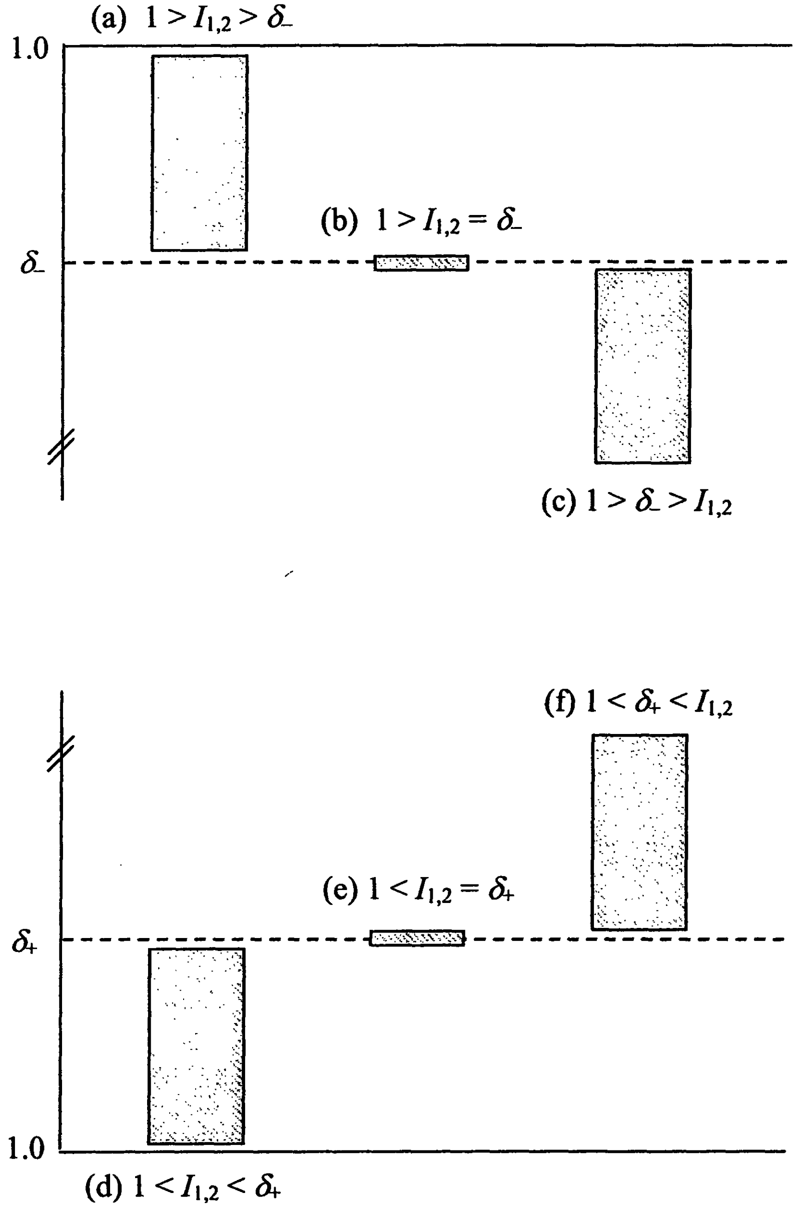


Figure 2.3: Effects of language-specific syntagmatic tone interval restrictions on the scaling of tones with respect to their referents; (a), (b) and (c) indicate three possible restrictions on falling contours, delineated by the syntagmatic intervallic cutoff δ_-, while (d), (e), and (f) indicate three possible restrictions on rising contours, delineated by the syntagmatic intervallic cutoff δ_+.

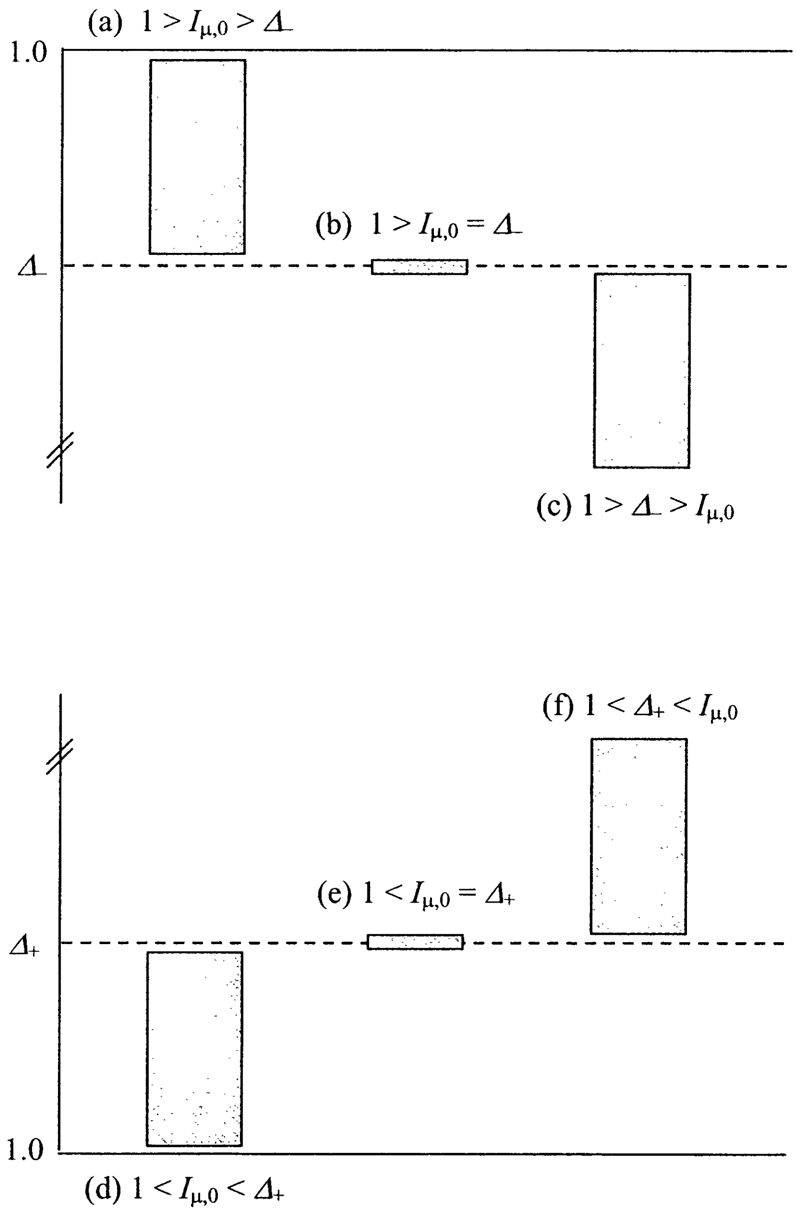


Figure 2.4: Effects of language-specific paradigmatic tone interval restrictions on the scaling of tones with respect to their referents; (a), (b), and (c) indicate three possible restrictions on tone scaling with respect to the tonic, delineated by the paradigmatic intervallic cutoff Δ_-, while (d), (e), and (f) indicate three possible restrictions on tone scaling with respect to the tonic, delineated by the paradigmatic intervallic cutoff Δ_+.

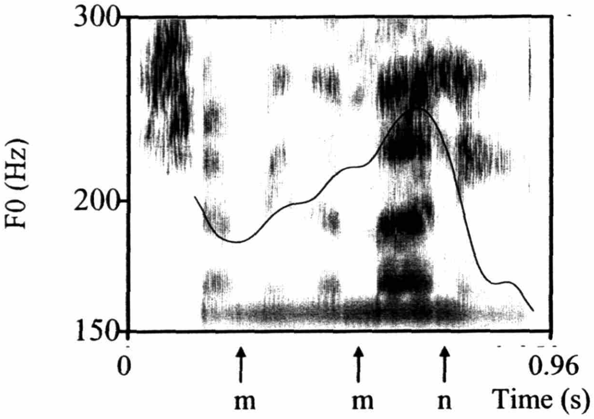


Figure 3.1. A simple declarative sentence: *Show me the money*!

Show me the money!

```
           X
X          X
X   X   X  X X
```

Figure 3.2: A metrical grid structure for the utterance shown in Figure 3.1.

Show me the money!

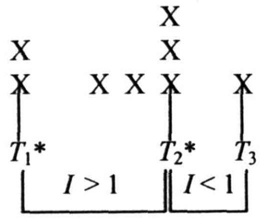


Figure 3.3: The phonological representation of the contour in Figure 3.1, showing the pattern of association between tones and timing positions, with tone intervals indicated.

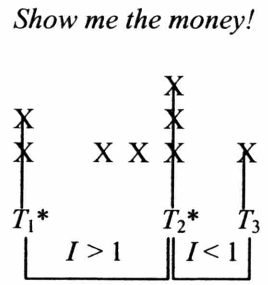


Figure 3.4: The same phonological representation as in Figure 3.3, now showing tones to be associated with all X's in their respective metrical grid columns.

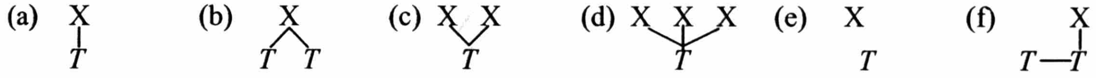


Figure 3.5: Several different patterns of association between tones and timing slots. In this theory, only structures like (a) and (b) are permitted, while structures (c)-(f) are prohibited.

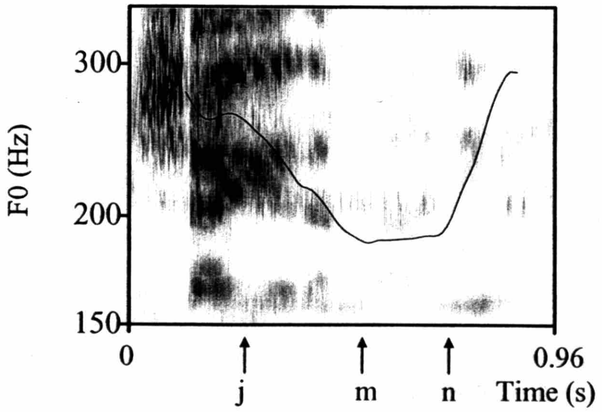
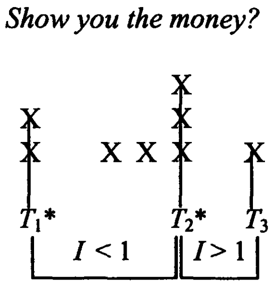


Figure 3.6. A simple interrogative sentence: *Show you the money*?



Show you the money?

Figure 3.7: The phonological representation for the contour in Figure 3.6.

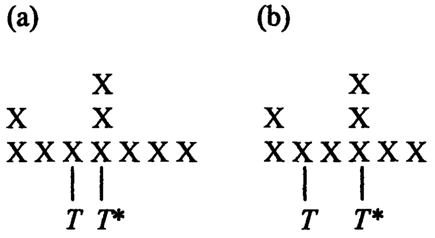


Figure 3.8: Two patterns of association between unstarred tones and grids. The structure in (a) is well formed, while the structure in (b) is not.

Figure 3.9: Patterns of association between unstarred tones and timing slots. (a) is attracted leftward and (b) is attracted rightward.

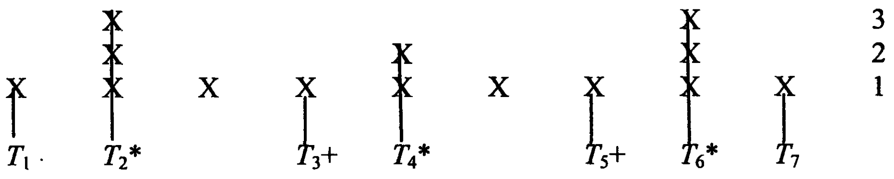


Figure 3.10: A metrical grid with associated tones speared upward to associate with all X's in their respective columns.

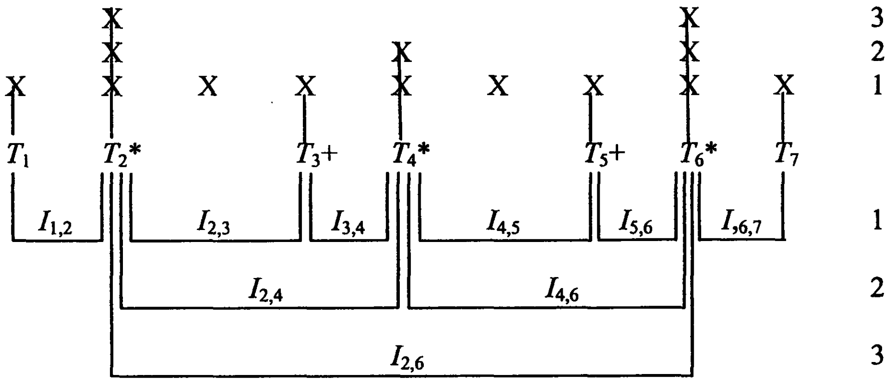


Figure 3.11: The tone interval matrix formed from tones which are adjacent on higher grid rows.

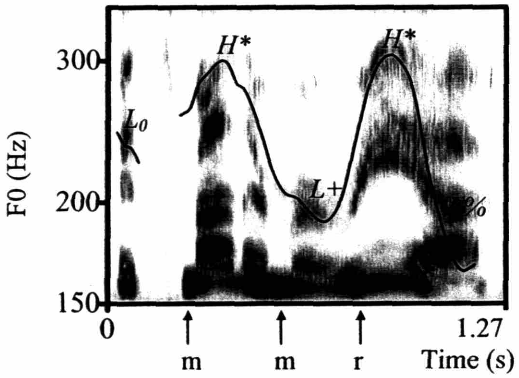


Figure 5.1. *It's Mother Maria!*

| | | | \

[L_0 H* L+H*L%]

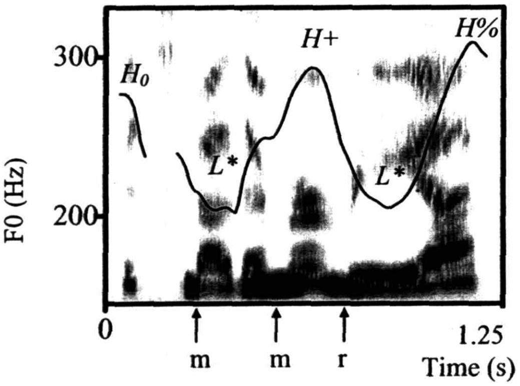


Figure 5.2. *It's Mother Maria?*

| | | | |

[H_0 L* H+L*H%]

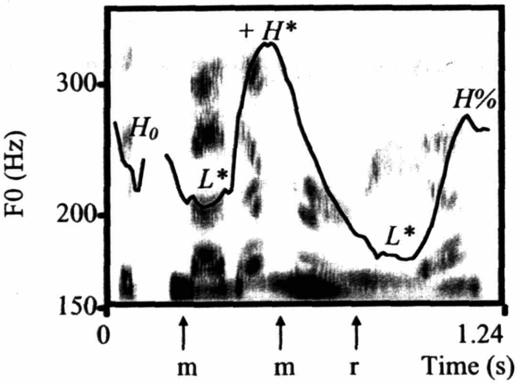


Figure 5.3. *It's Mother Maria?*

| | | | |

[H_0 L* +H L*H%]

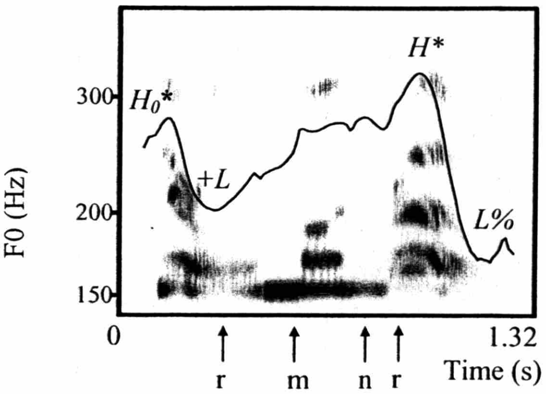


Figure 5.4. *Marilyn Monroe!*

| | |\

[H_0*+L H*L%]

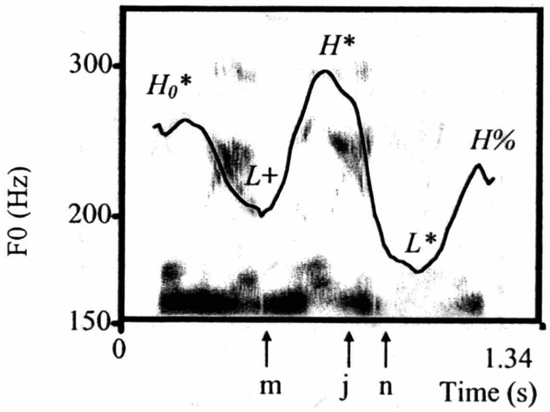


Figure 5.5. *Only a millionaire?*

| | | |\

[H_0* L+ H* L*H%]

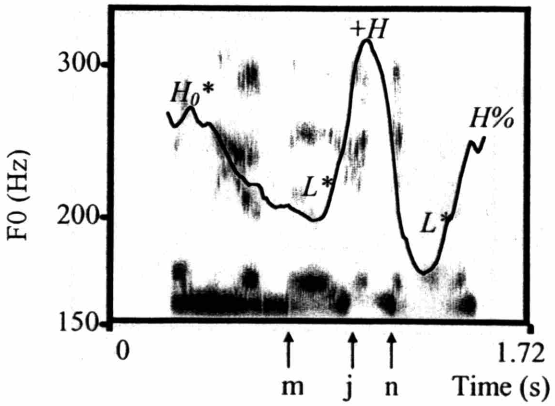


Figure 5.6. *Only a millionaire?*

| | | |\

[H_0* L* +H L*H%]

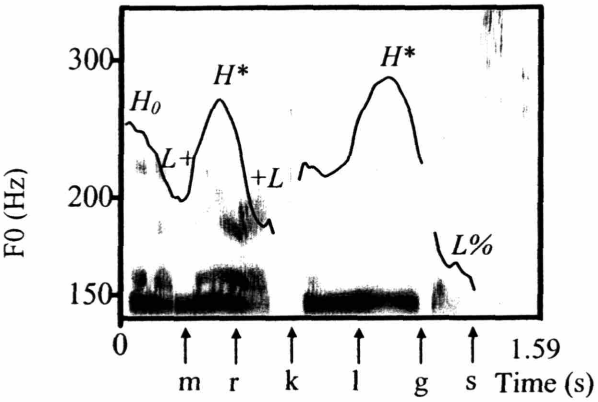


Figure 5.7. *An American linguist.*

| | | \ | |

[H_0 L+H*+L H* L%]

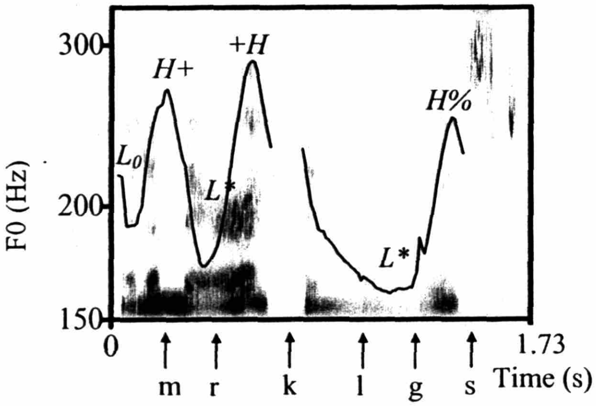


Figure 5.8. *An American linguist?*

| | | \ | |

[L_0 H+L*+H L* H%]

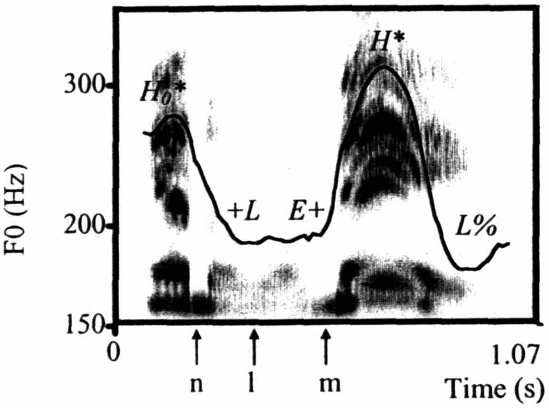


Figure 5.9. *Anna Le May.*

| | | |\

[H_0*+L E+ H* L%]

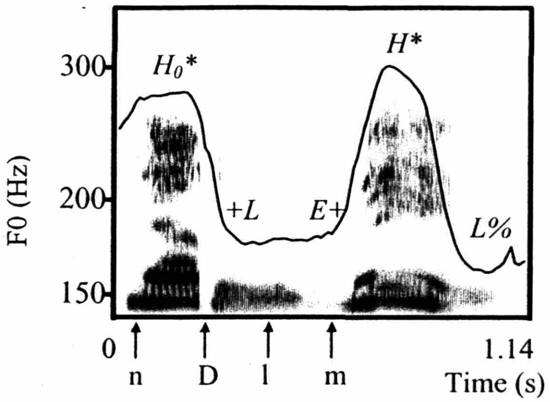


Figure 5.10. *Natalie May.*

| | | |\

[H_0*+L E+H* L%]

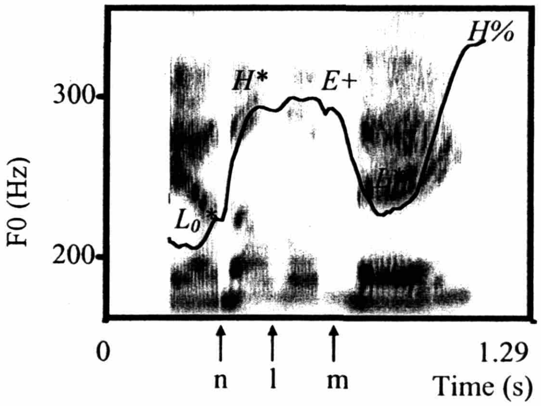


Figure 5.11. *Anna Le May?*

| | | | \

[L_0*+H E+L* H%]

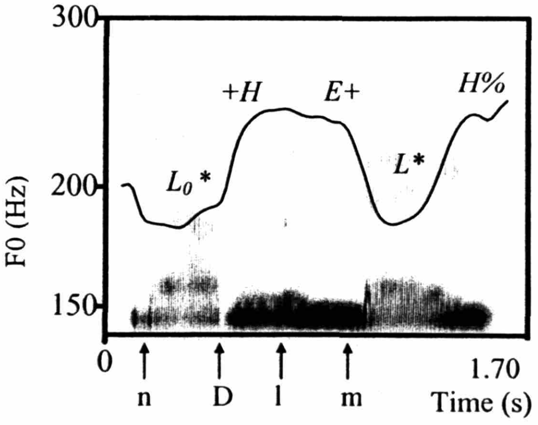


Figure 5.12. *Natalie May?*

| | | |\

[L_0*+H E+L* H%]

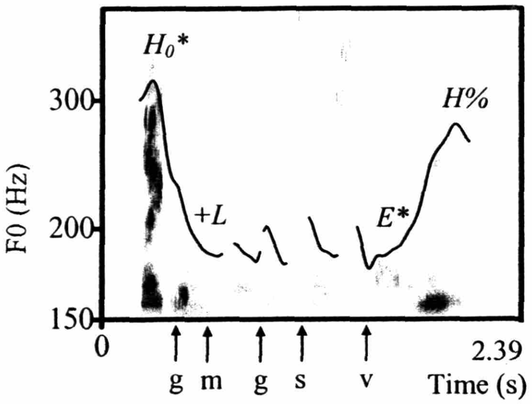


Figure 5.13. *Legumes are a good source of vitamins!*

| | | |

[H_0*+L E* H%]

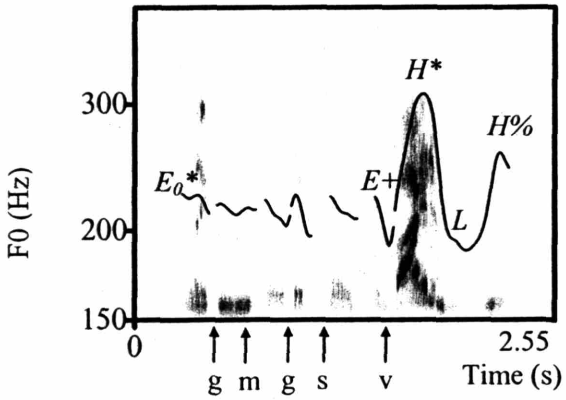


Figure 5.14. *Legumes are a good source of vitamins!*

| | | | |

[E_0* E+H* L H%]

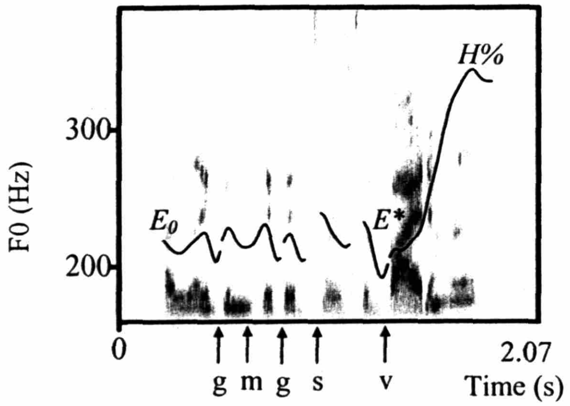


Figure 5.15. *Are legumes a good source of vitamins?*

| | |

[E_0 E* H%]

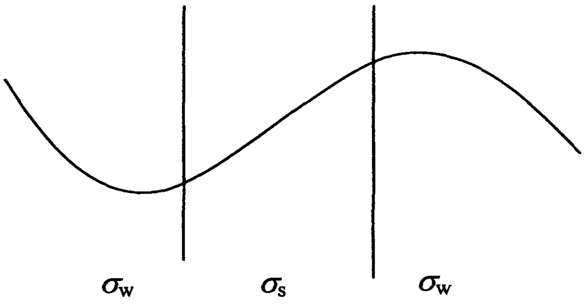


Figure 5.16. Schematic diagram of a Greek prenuclear accent. Vertical lines indicate approximate locations of syllable boundaries. Neither the F0 peak nor the valley is aligned with the stressed syllable.

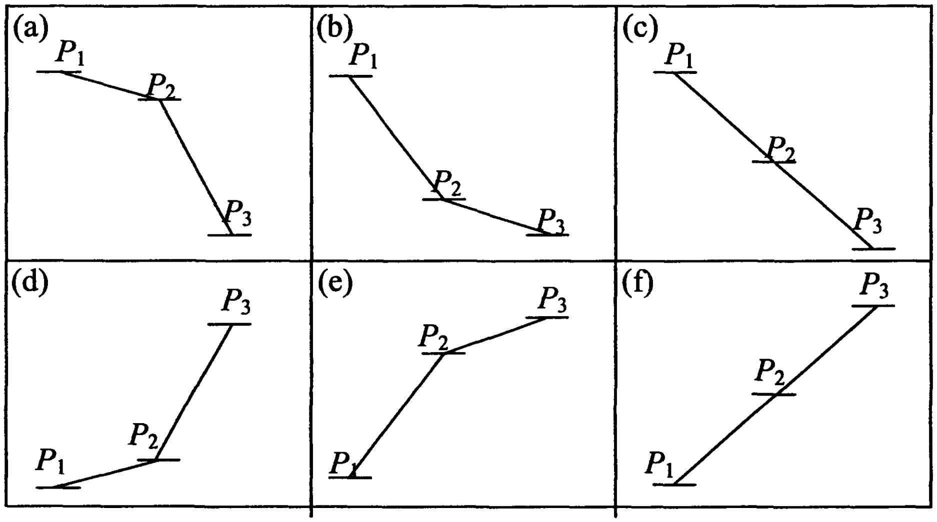


Figure 5.17. Relation between variability in the vertical and horizontal "placement" of a sequence of three target pitches and the shape of the pitch curve connecting them. In (a)-(b) and (d)-(e), the slopes of the pitch interpolation lines connecting pitches P_1 and P_2 and P_2 and P_3 are different, leading to an identifiable location of a change in slope. In (c) and (f), the slopes of the pitch interpolation lines connecting pitches P_1 and P_2 and P_2 and P_3 is the same, so no such change in slope obtains.

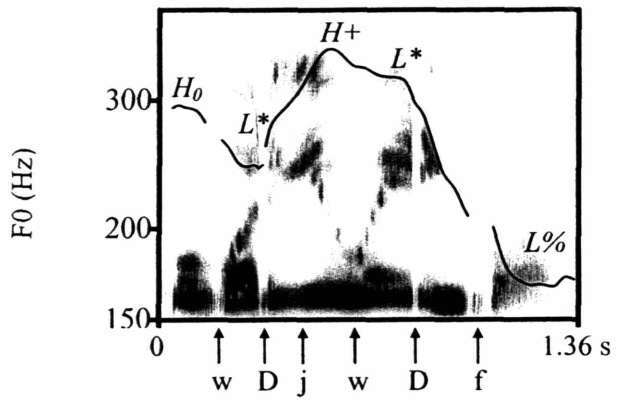


Figure 5.18. *Well, what are you waiting for!?*

| | | | |

[H_0 *L** *H+* *L** *L%*]

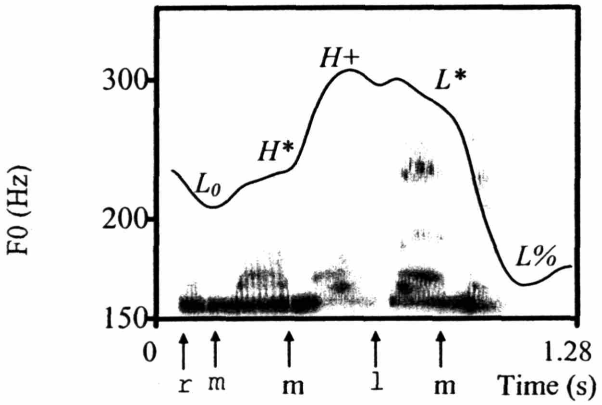


Figure 5.19. *For Mama Lemming.*

| | | | |

[L_0 *H** *H+* *L** *L%*]

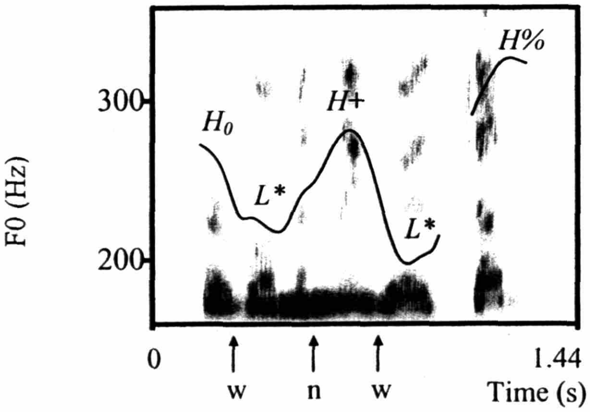


Figure 5.20. *Her womanly wisdom?*

| | | | |

[H_0 *L** *H+* *L** *H%*]

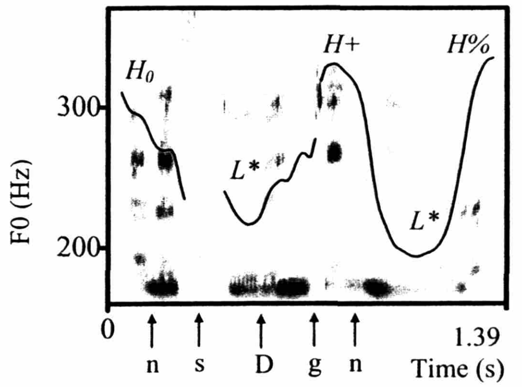


Figure 5.21. *An astute engineer?*
| | | |\
[H_0 L^* $H+L^*H\%$]

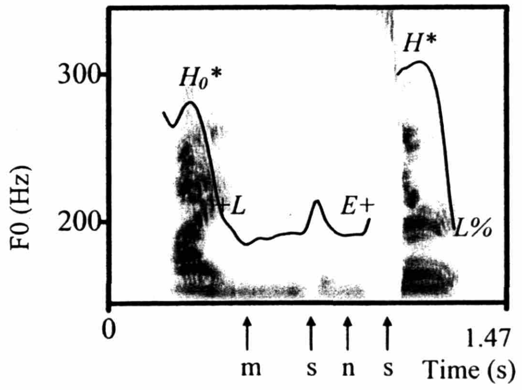


Figure 5.22. *"I" means insert.*
|\ | |\
[H_0^*+L $E+H^*L\%$]

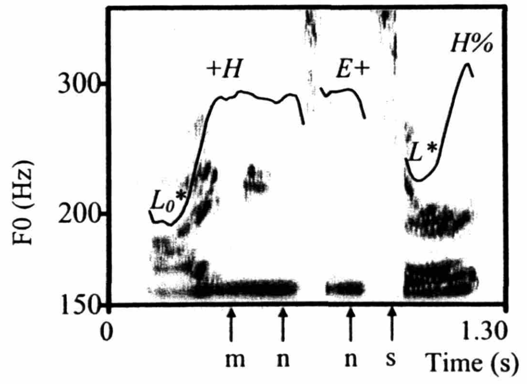


Figure 5.23. *"I" means insert?*
| \ | | \
[L_0^*+H $E+L^*H\%$]

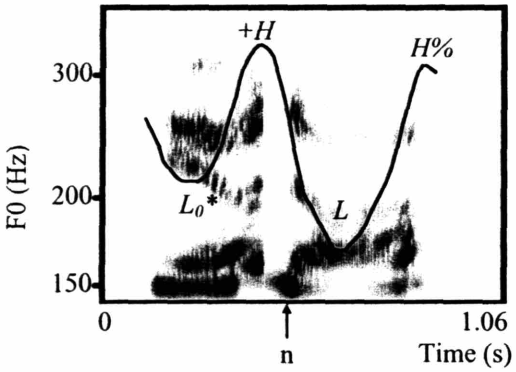


Figure 5.24. *An - na!*
[L_0*+*H* *L* *H%*]

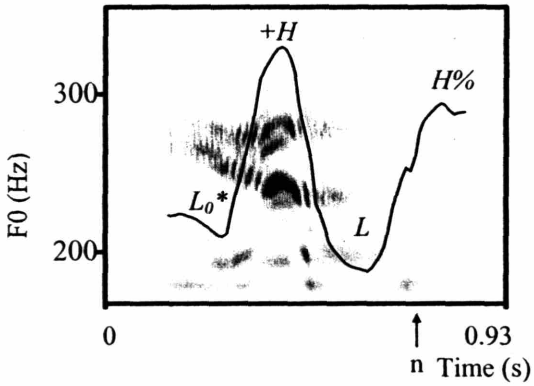


Figure 5.25. *Anne!*
[L_0*+*H* *L* *H%*]

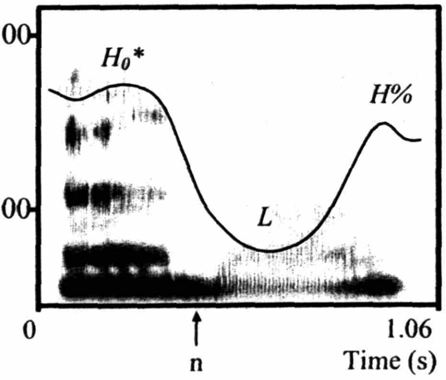


Figure 5.26. *An - na!*
[H_0* *L* *H%*]

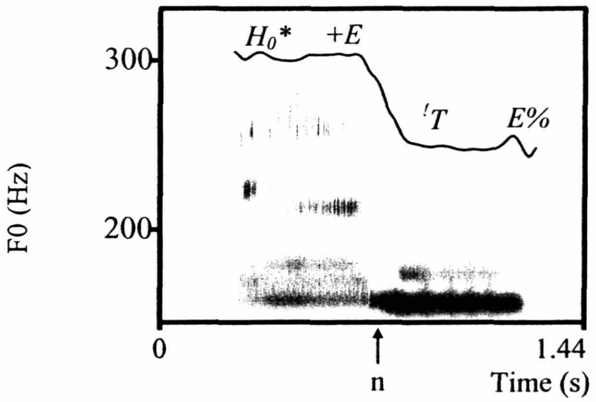


Figure 5.27. *An - na!*

[H_0*+E $^{!}T$ $E\%$]

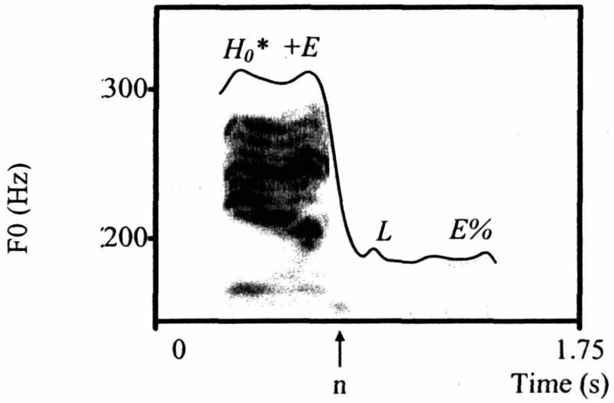


Figure 5.28. *An - na!*

[H_0*+E L $E\%$]

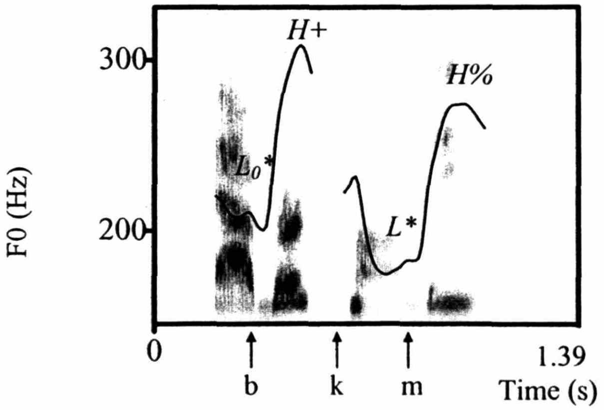


Figure 5.29. *Abercrombie?*

[L_0*+H $L*$ $H\%$]

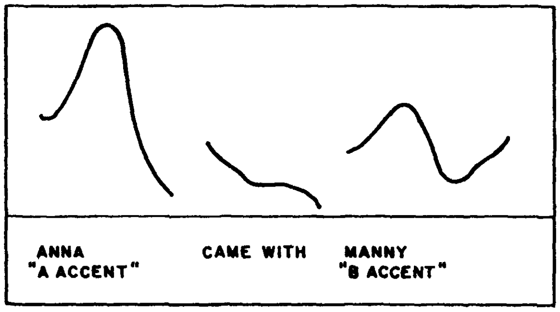


Figure 5.30. F0 contour for the phrase Anna came with Manny in the experiments of P80 and LP84 corresponding to the "AB" condition.

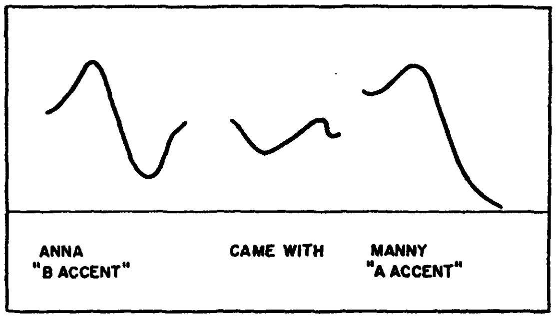


Figure 5.31. F0 contour for the phrase Anna came with Manny in the experiments of P80 and LP84 corresponding to the "BA" condition.

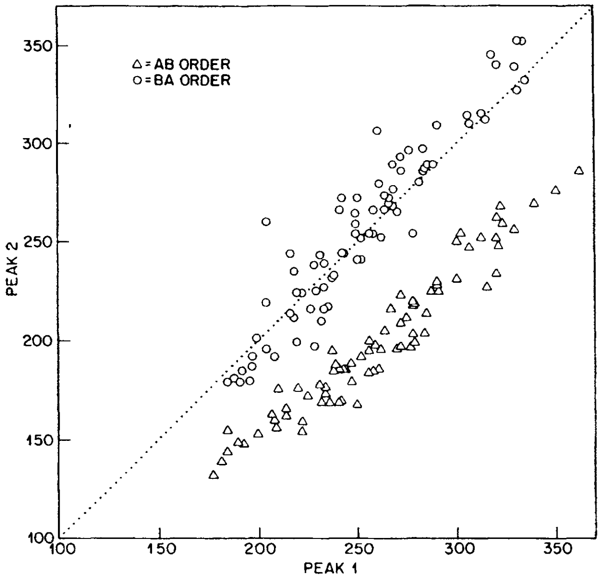


Figure 5.32. Production data for a representative subject from an experiment reported in LP84.

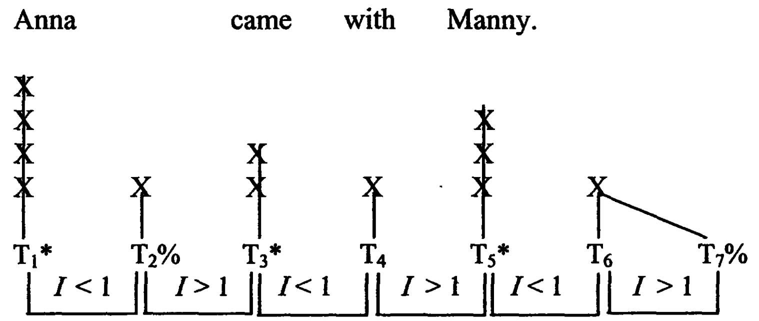


Figure 5.33. Partial tone interval representation of "AB" contours in experiments reported by Pierrehumbert (1980) and Liberman and Pierrehumbert (1984).

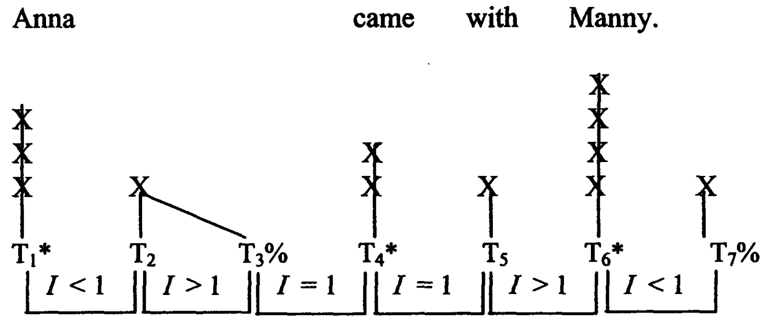


Figure 5.34. Partial tone interval representation of "BA" contours in experiments reported by Pierrehumbert (1980) and Liberman and Pierrehumbert (1984).

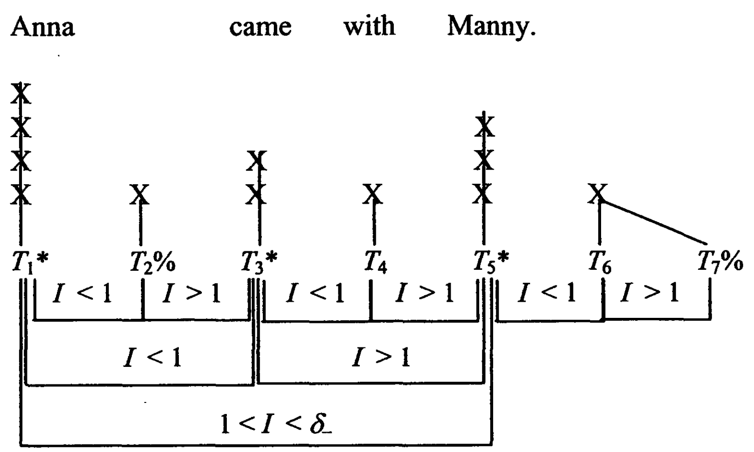


Figure 5.35. Full tone interval representation of "AB" contours in experiments reported by Pierrehumbert (1980) and Liberman and Pierrehumbert (1984).

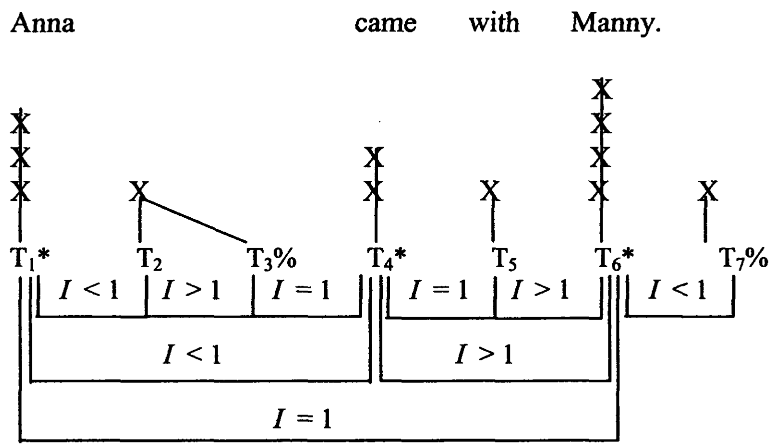


Figure 5.36. Full tone interval representation of "BA" contours in experiments reported by Pierrehumbert (1980) and Liberman and Pierrehumbert (1984).

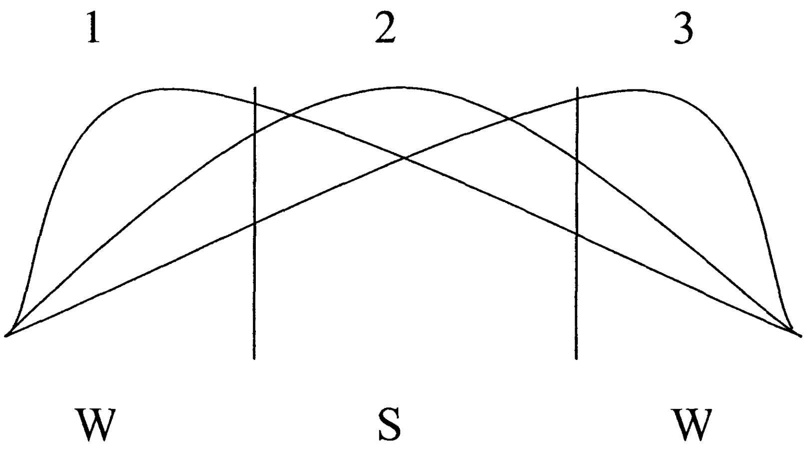


Figure 6.1. Three possible placements of an F0 maximum with respect to a stressed syllable.

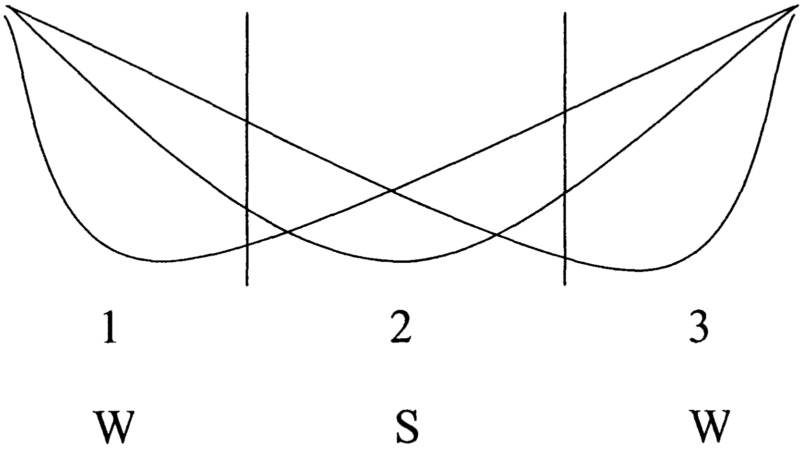


Figure 6.2. Three possible placements of an F0 minimum with respect to a stressed syllable.

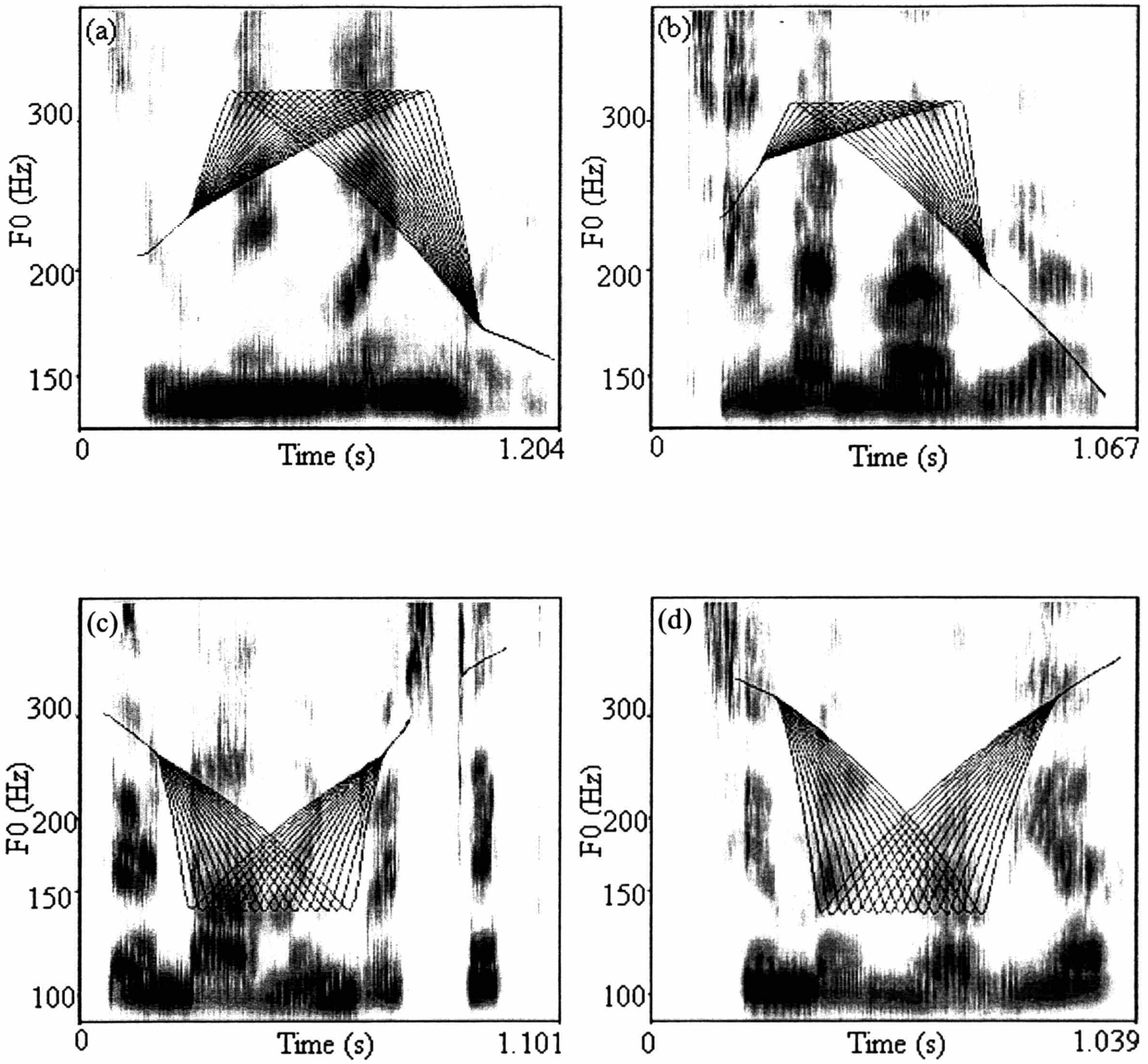


Figure 6.3. Stimuli used in Experiments 1 and 2. (a) *Too minglingly* (*minglingly*-max series), (b) *To Monrovia* (*Monrovia*-max series), (c) *They're nonlinguistic?* (*nonlinguistic*-min series), and (d) *To Monrovia?* (*Monrovia*-min series).

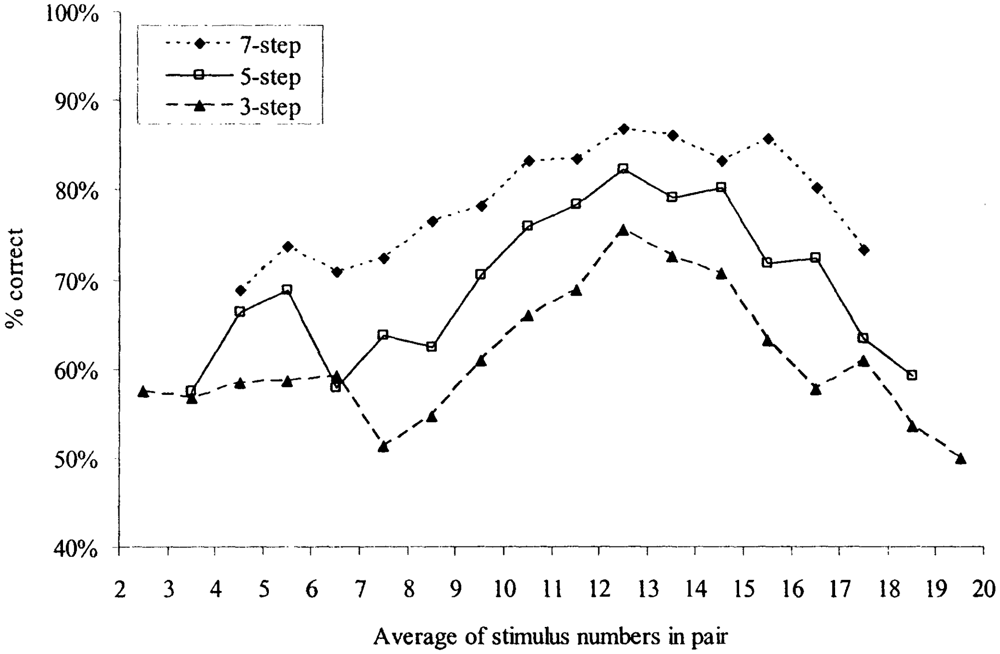


Figure 6.4. Percentage of correct responses for a discrimination task using stimuli drawn from the *minglingly*-max series.

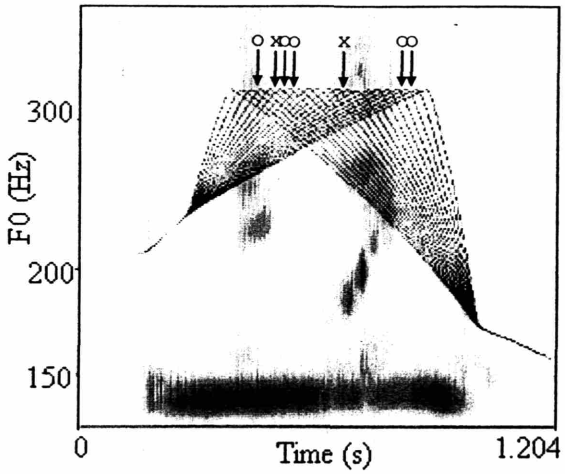


Figure 6.5. Summary of locations of discrimination maxima ("x") and minima ("o") along the *minglingly*-max stimulus series, for stimulus pairs which were 3 and 5 steps apart. See text.

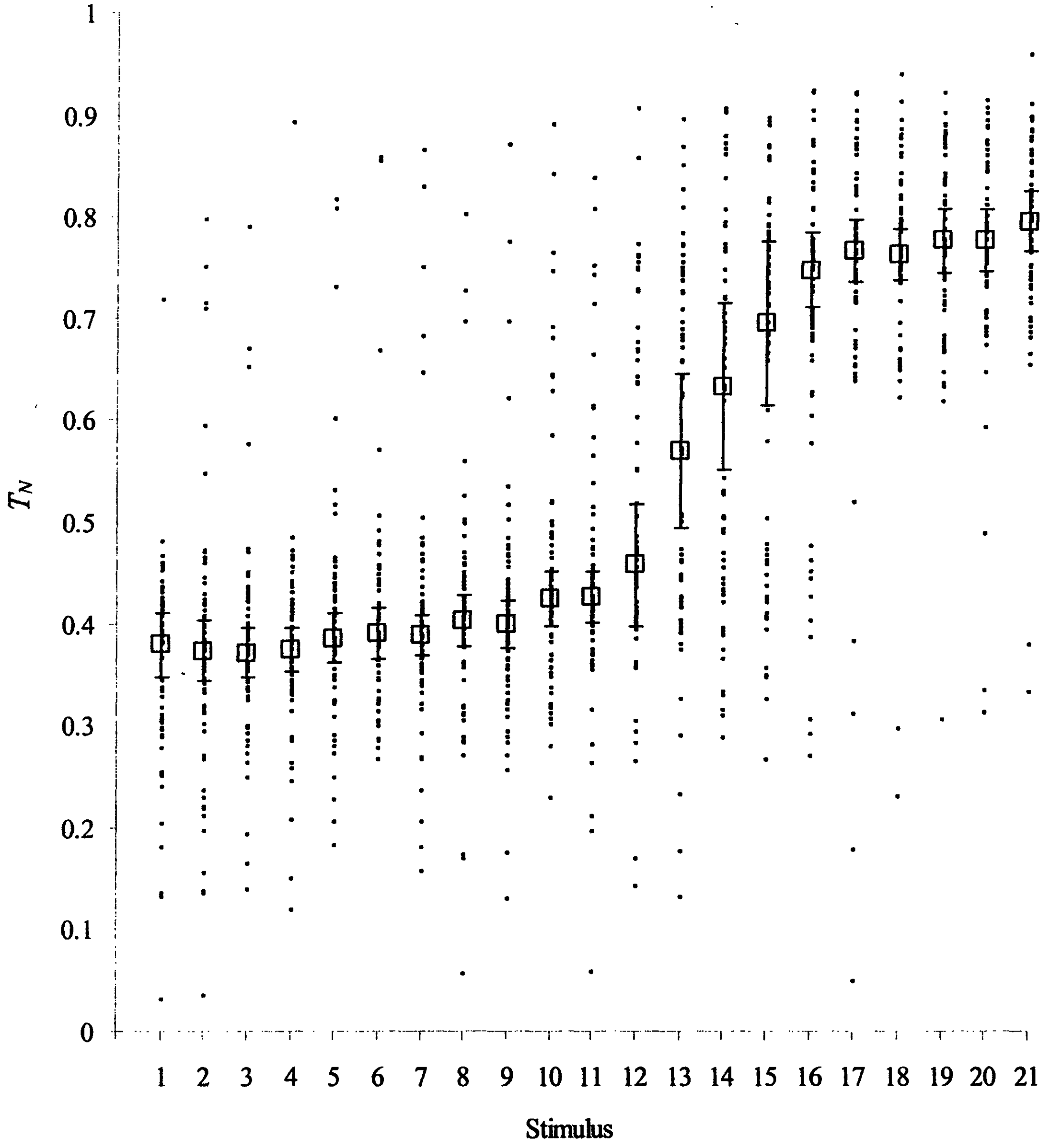


Figure 6.6. Normalized turning point time, T_N, for an imitation task using stimuli drawn from the *minglingly*-max series. Data points represent all observations for subjects in this stimulus series. Open squares show median values, while whiskers indicate one semi-interquartile range.

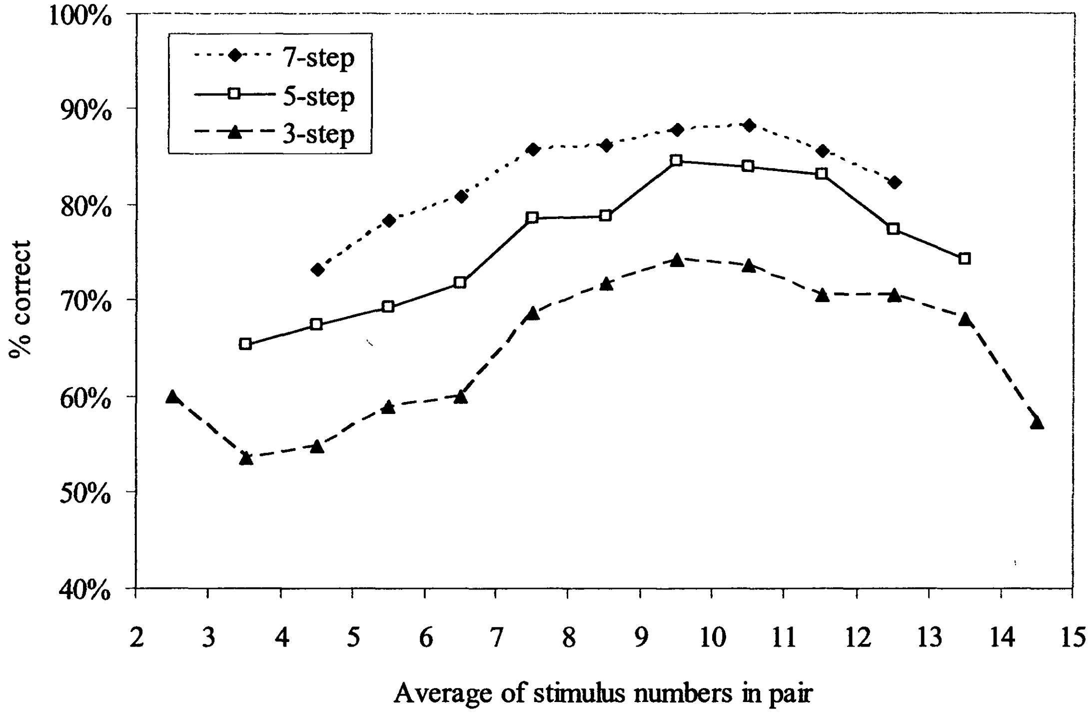


Figure 6.7. Percentage of correct responses for a discrimination task using stimuli drawn from the *Monrovia*-max series.

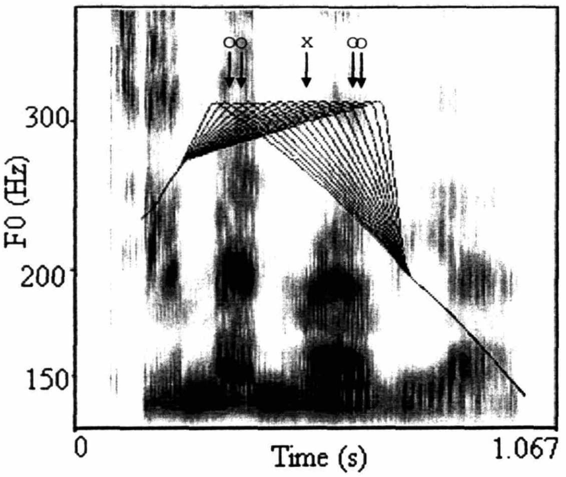


Figure 6.8. Summary of locations of discrimination maxima ("x") and minima ("o") along the *Monrovia*-max stimulus series, for stimulus pairs which were 3 and 5 steps apart. See text.

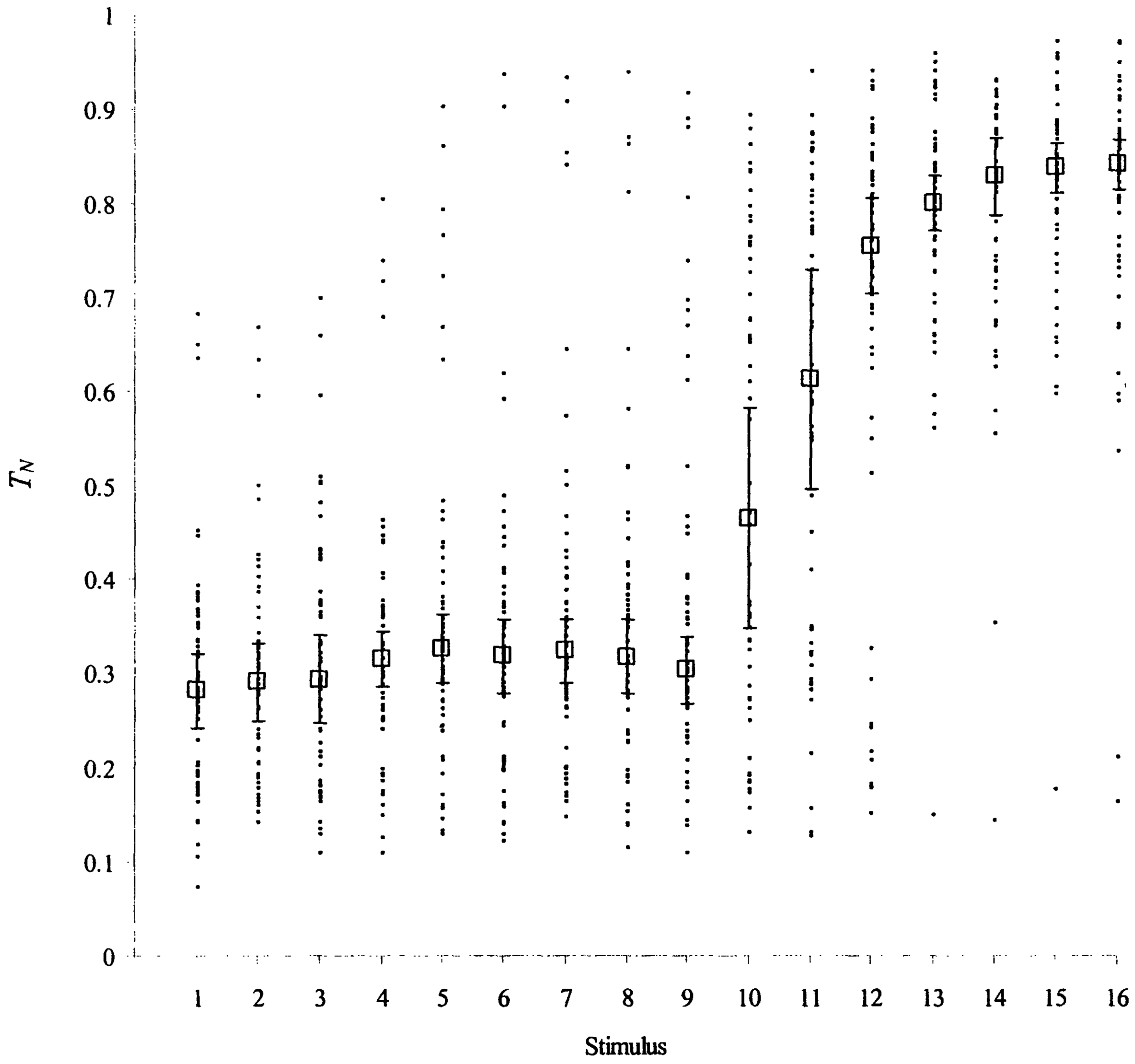


Figure 6.9. Normalized turning point time, T_N, for an imitation task using stimuli drawn from the *Monrovia*-max series. Data points represent all observations for subjects in this stimulus series. Open squares show median values, while whiskers indicate one semi-interquartile range.

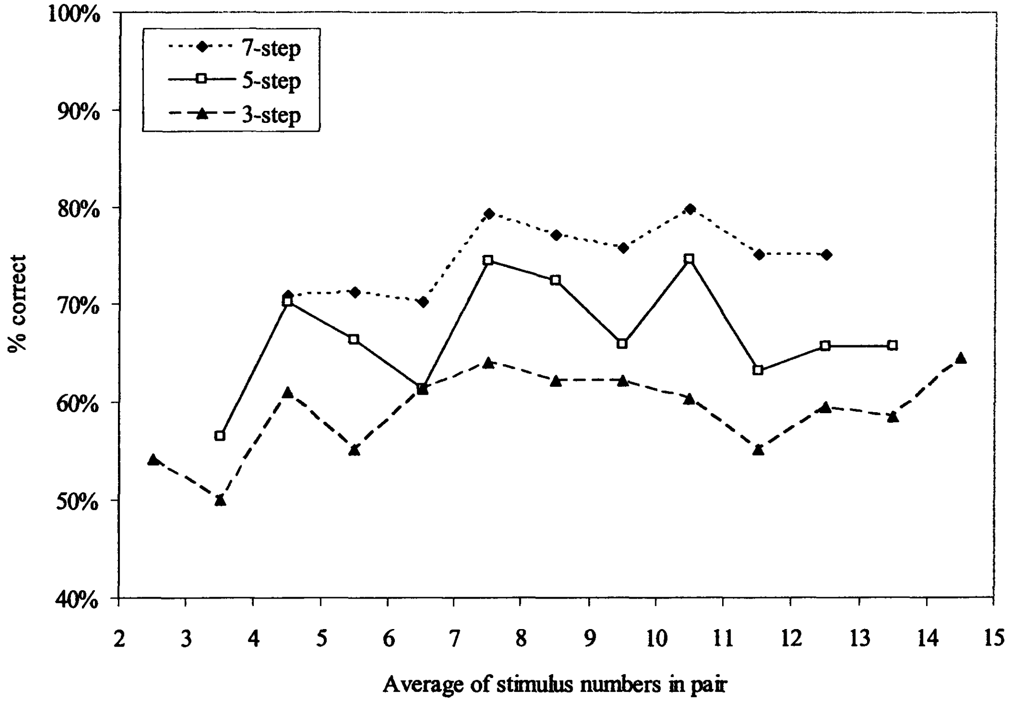


Figure 6.10. Percentage of correct responses for a discrimination task using stimuli drawn from the *nonlinguistic*-min series.

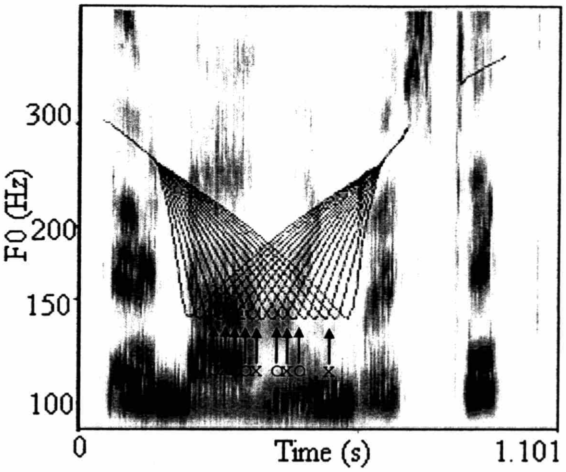


Figure 6.11. Summary of locations of discrimination maxima ("x") and minima ("o") along the *nonlinguistic*-min stimulus series, for stimulus pairs which were 3 and 5 steps apart. See text.

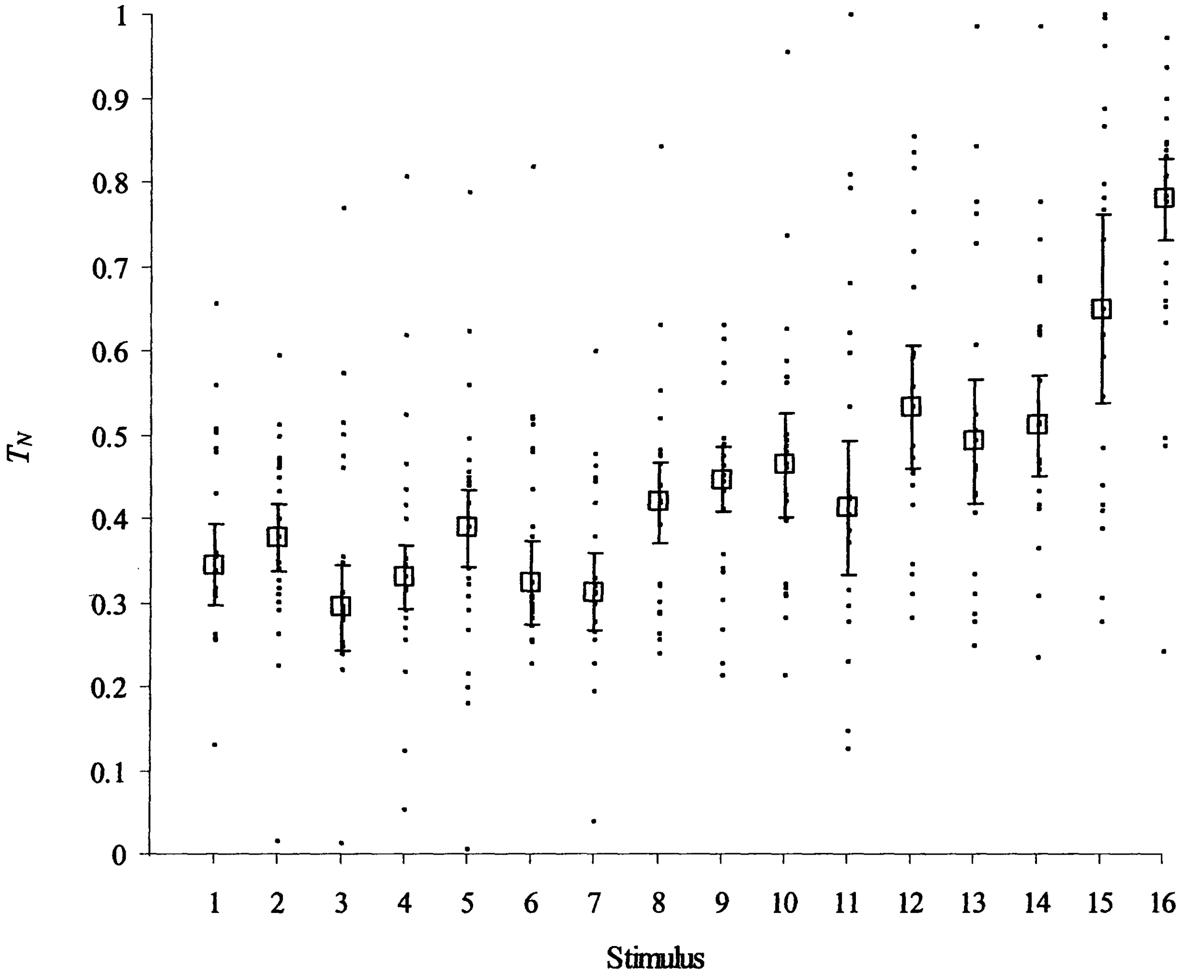


Figure 6.12. Normalized turning point time, T_N, for an imitation task using stimuli drawn from the *nonlinguistic*-min series. Data points represent all observations for subjects in this stimulus series. Open squares show median values, while whiskers indicate one semi-interquartile range.

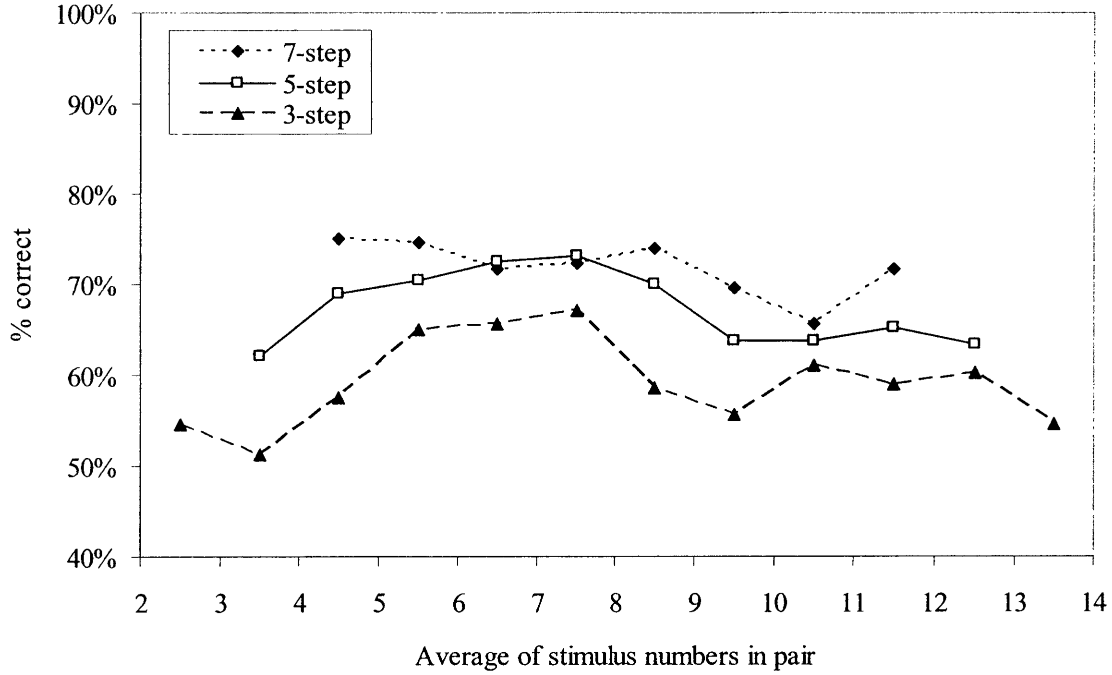


Figure 6.13. Percentage of correct responses for a discrimination task using stimuli drawn from the *Monrovia*-min series.

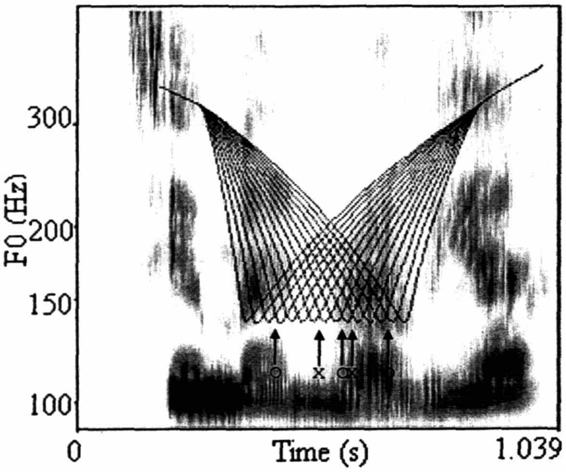


Figure 6.14. Summary of locations of discrimination maxima (“x”) and minima (“o”) along the *Monrovia*-min stimulus series, for stimulus pairs which were 3 and 5 steps apart. See text.

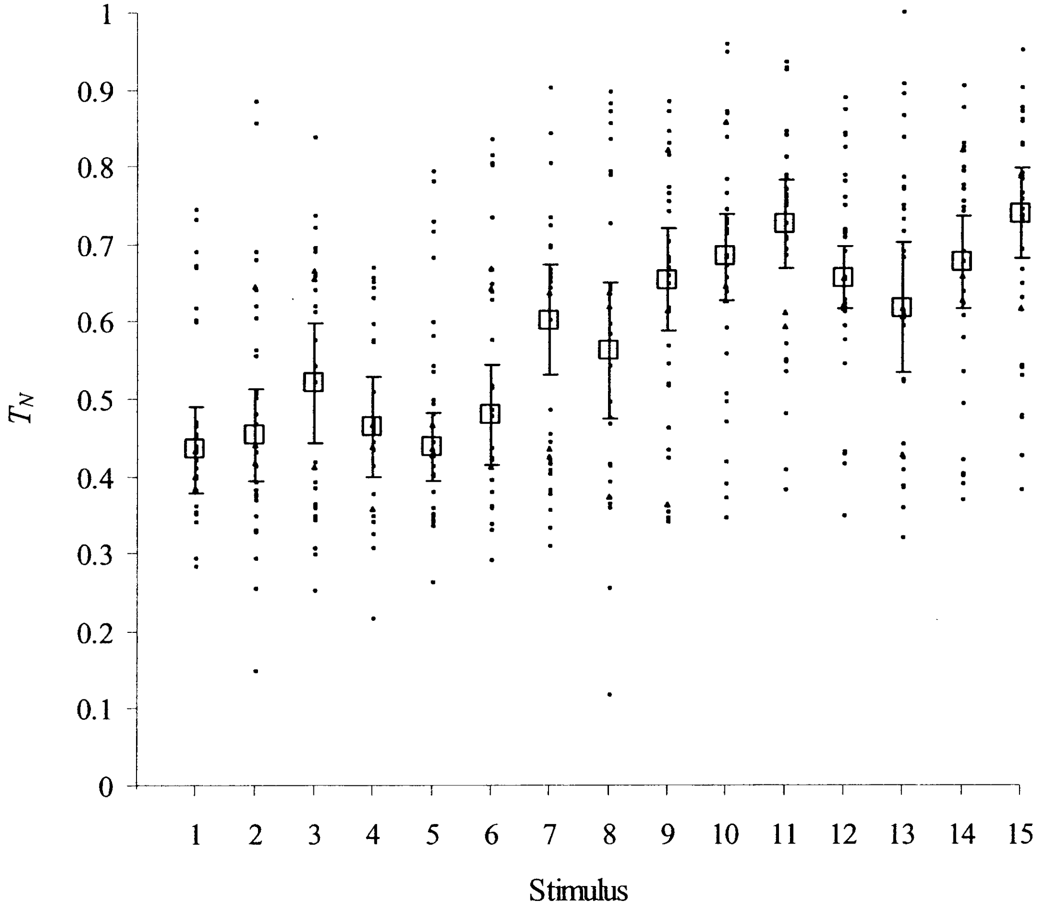


Figure 6.15. Normalized turning point time, T_N, for an imitation task using stimuli drawn from the *Monrovia*-min series. Data points represent all observations for subjects in this stimulus series. Open squares show median values, while whiskers indicate one semi-interquartile range.

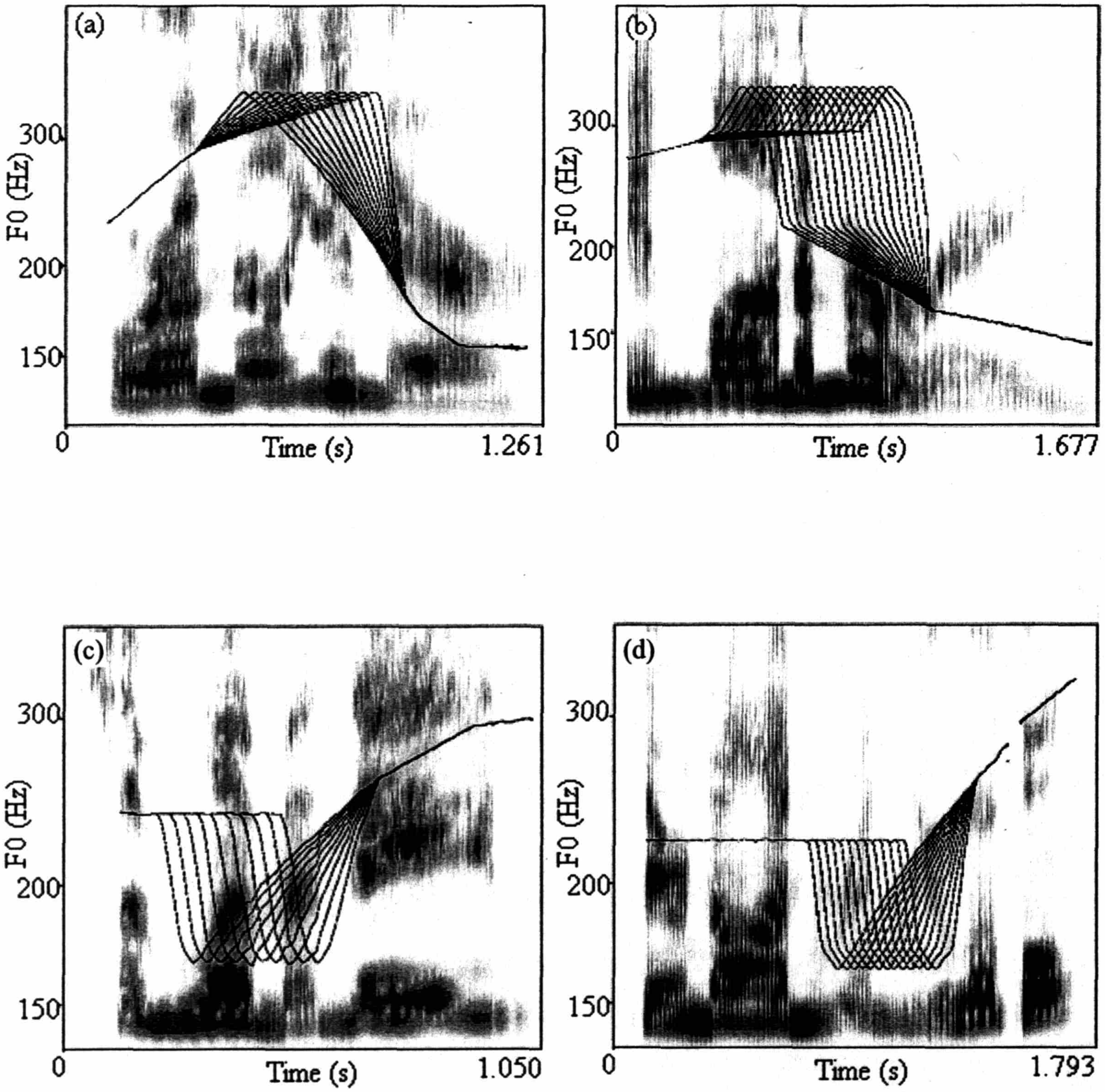


Figure 6.16. Stimuli used in Experiments 3 and 4. (a) *For a millionaire* (*millionaire*-max series), (b) *In Lannameraine* (*Lanna*-max series), (c) *Some lemonade?* (*lemonade*-min series), and (d) *They're nonrenewable?* (*nonrenewable*-min series).

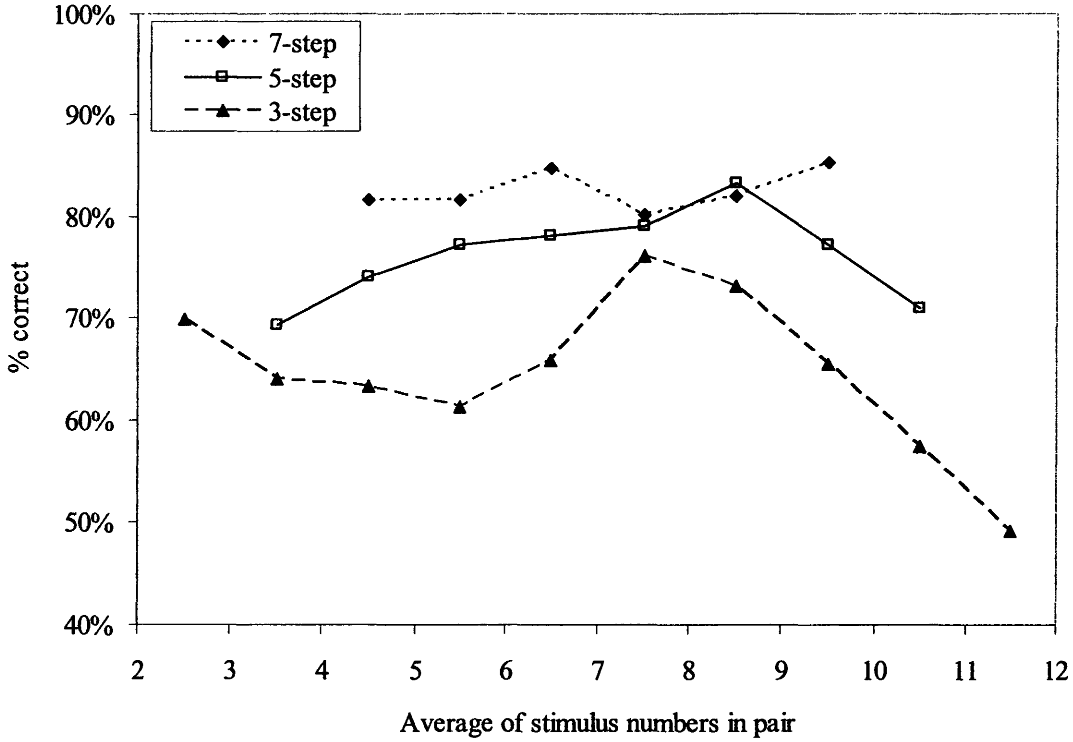


Figure 6.17. Percentage of correct responses for the *millionaire*-max series in Experiment 3.

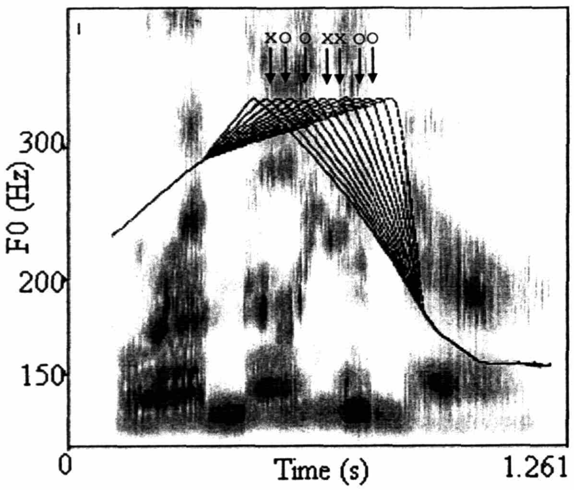


Figure 6.18. Summary of locations of discrimination maxima ("x") and minima ("o") along the *millionaire*-max stimulus series, for stimulus pairs which were 3 and 5 steps apart. See text.

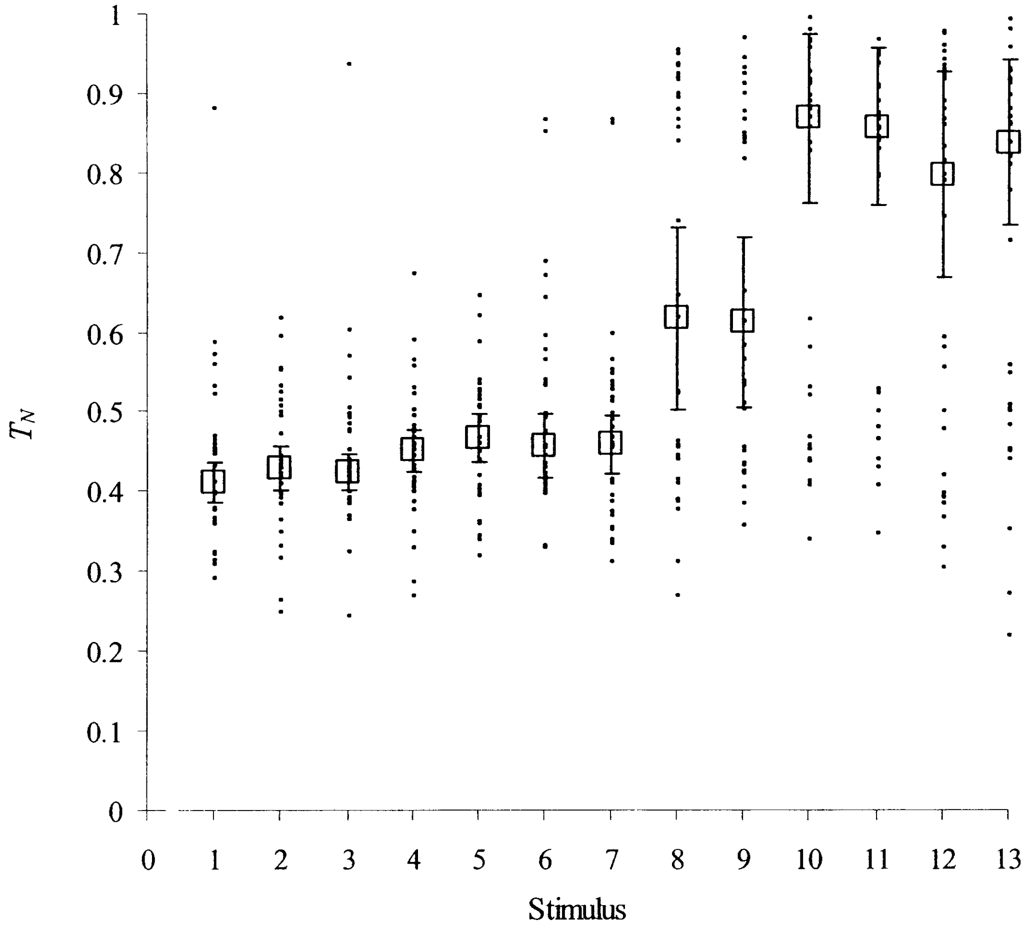


Figure 6.19. Normalized turning point time, T_N, for the *millionaire*-max series in Experiment 4.

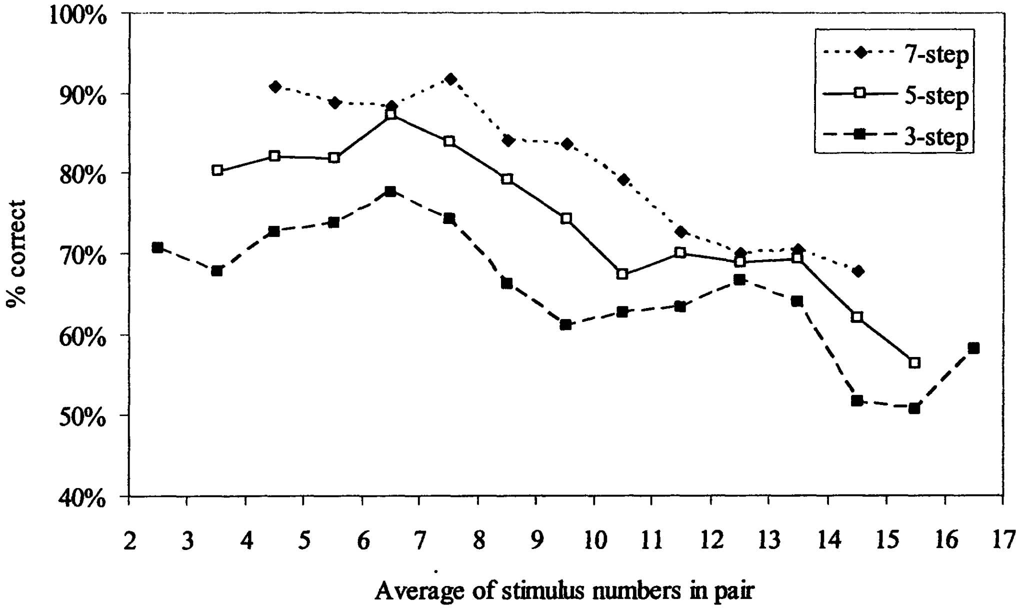


Figure 6.20. Percentage of correct responses for the *Lanna*-max series in Experiment 3.

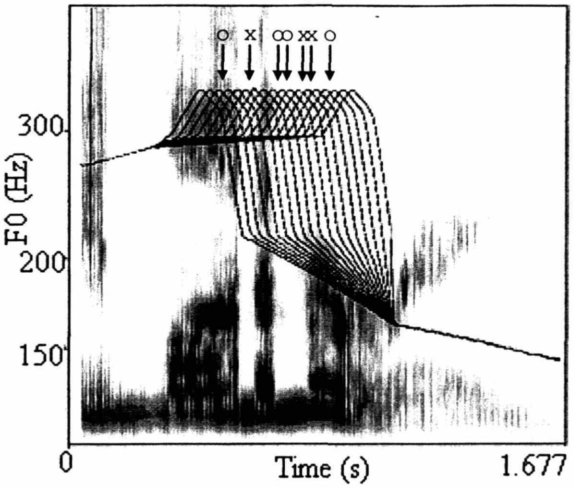


Figure 6.21. Summary of locations of discrimination maxima ("x") and minima ("o") along the *Lanna*-max stimulus series, for stimulus pairs which were 3 and 5 steps apart. See text.

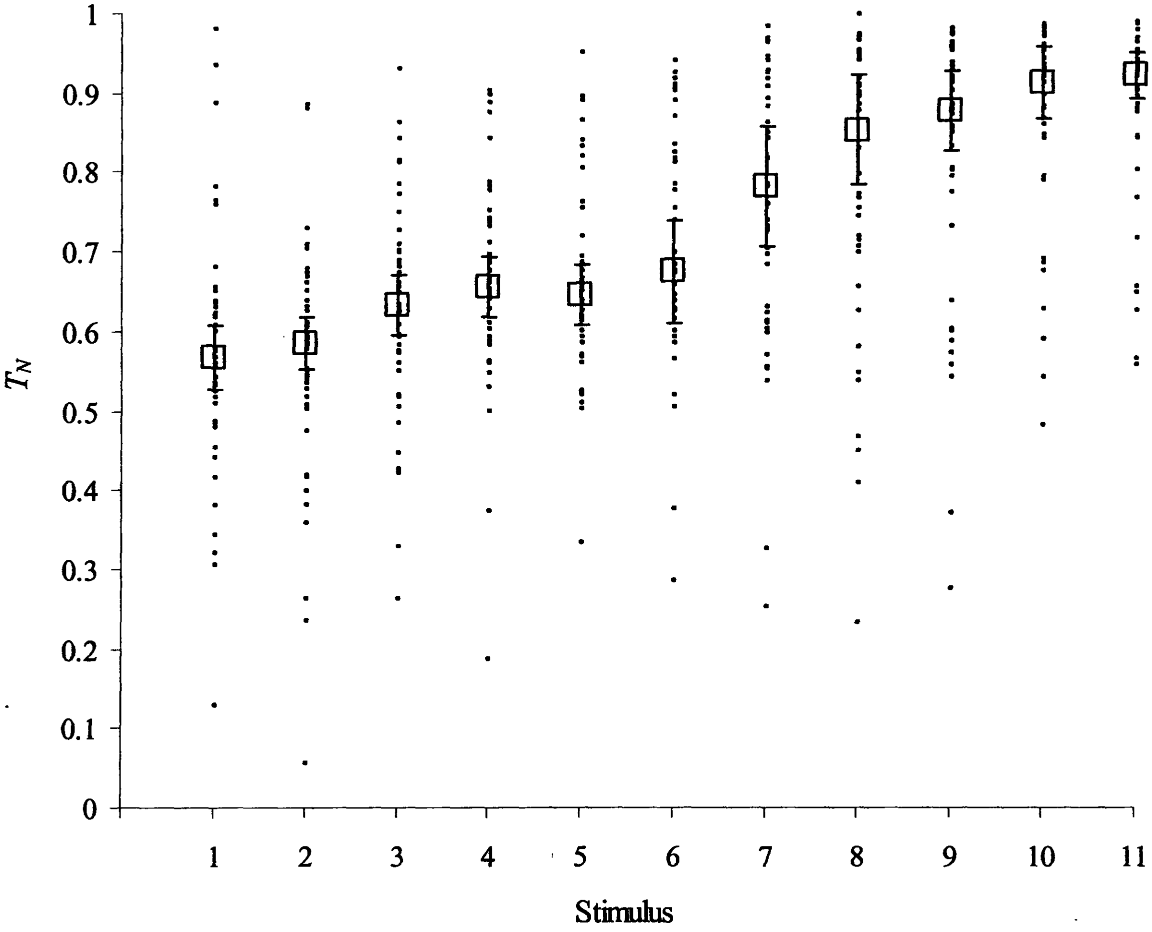


Figure 6.22. Normalized turning point time, T_N, for the *Lanna*-max series (Block I) in Experiment 4.

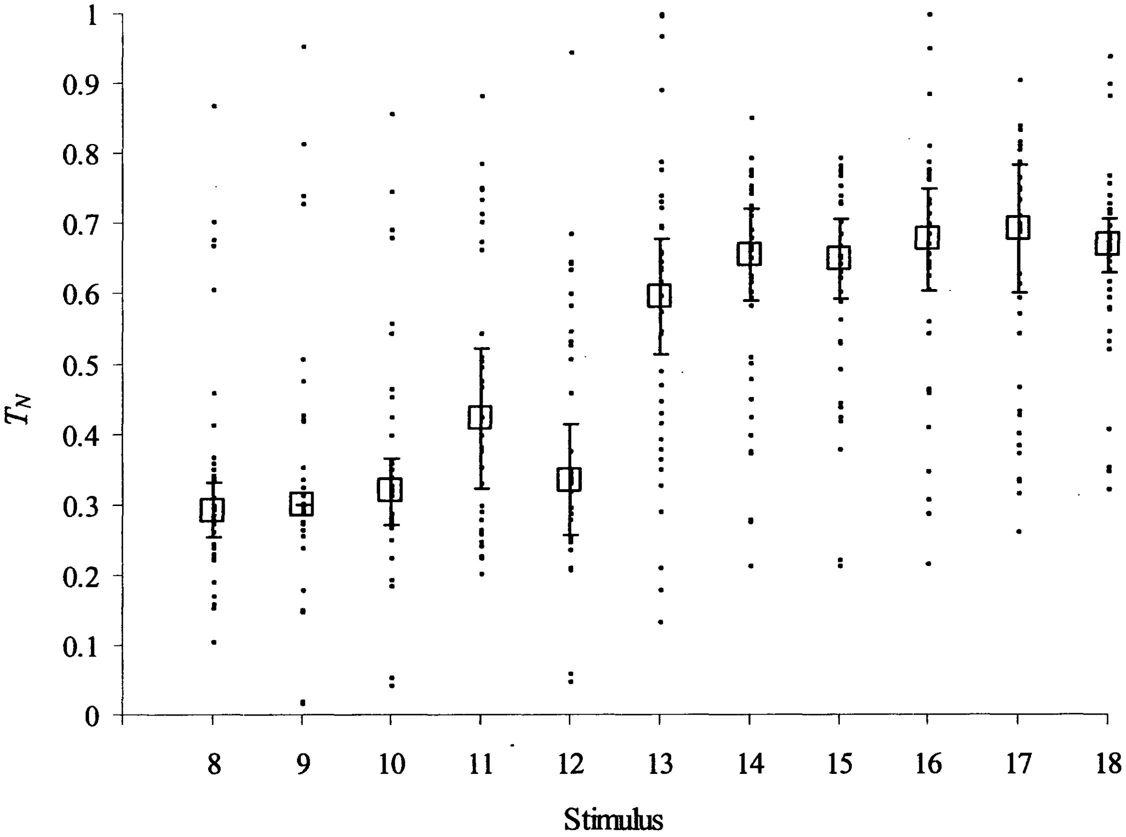


Figure 6.23. Normalized turning point time, T_N, for the *Lanna*-max series (Block II) in Experiment 4.

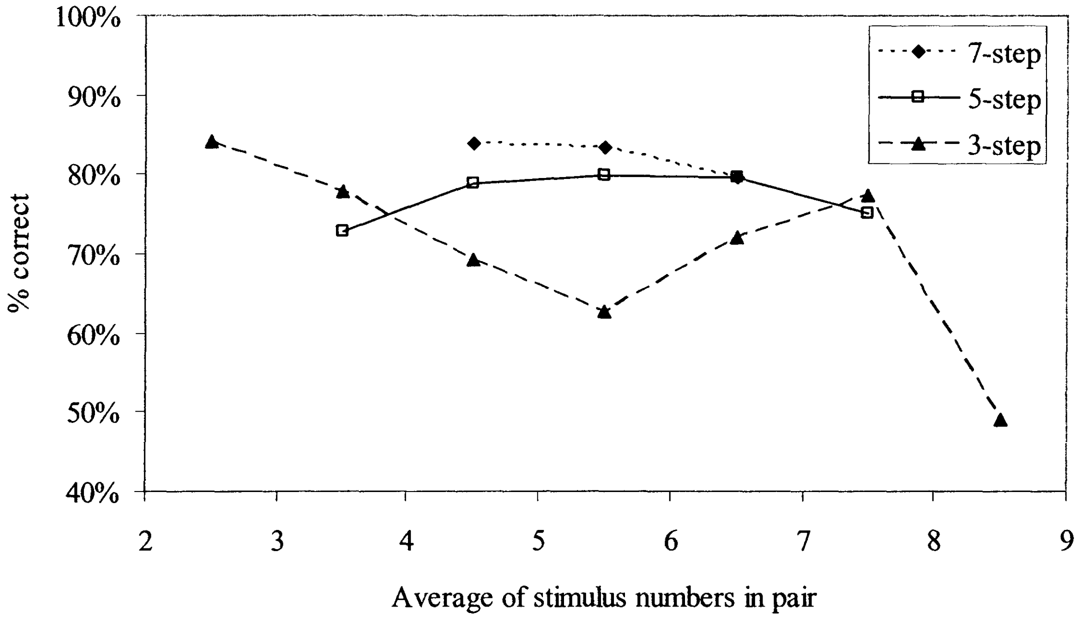


Figure 6.24. Percentage of correct responses for the *lemonade*-min series in Experiment 3.

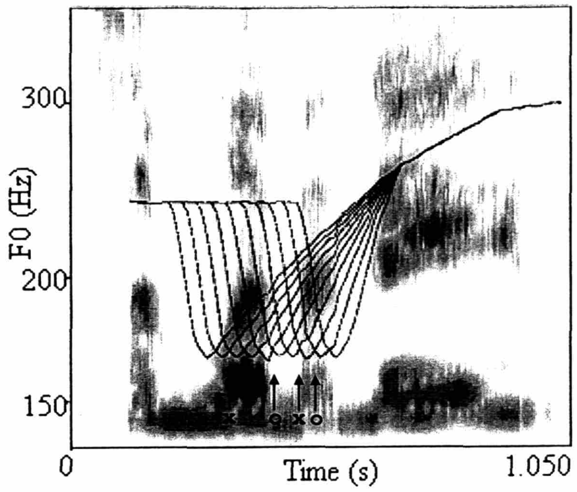


Figure 6.25. Summary of locations of discrimination maxima ("x") and minima ("o") along the *lemonade*-min stimulus series, for stimulus pairs which were 3 steps apart. See text.

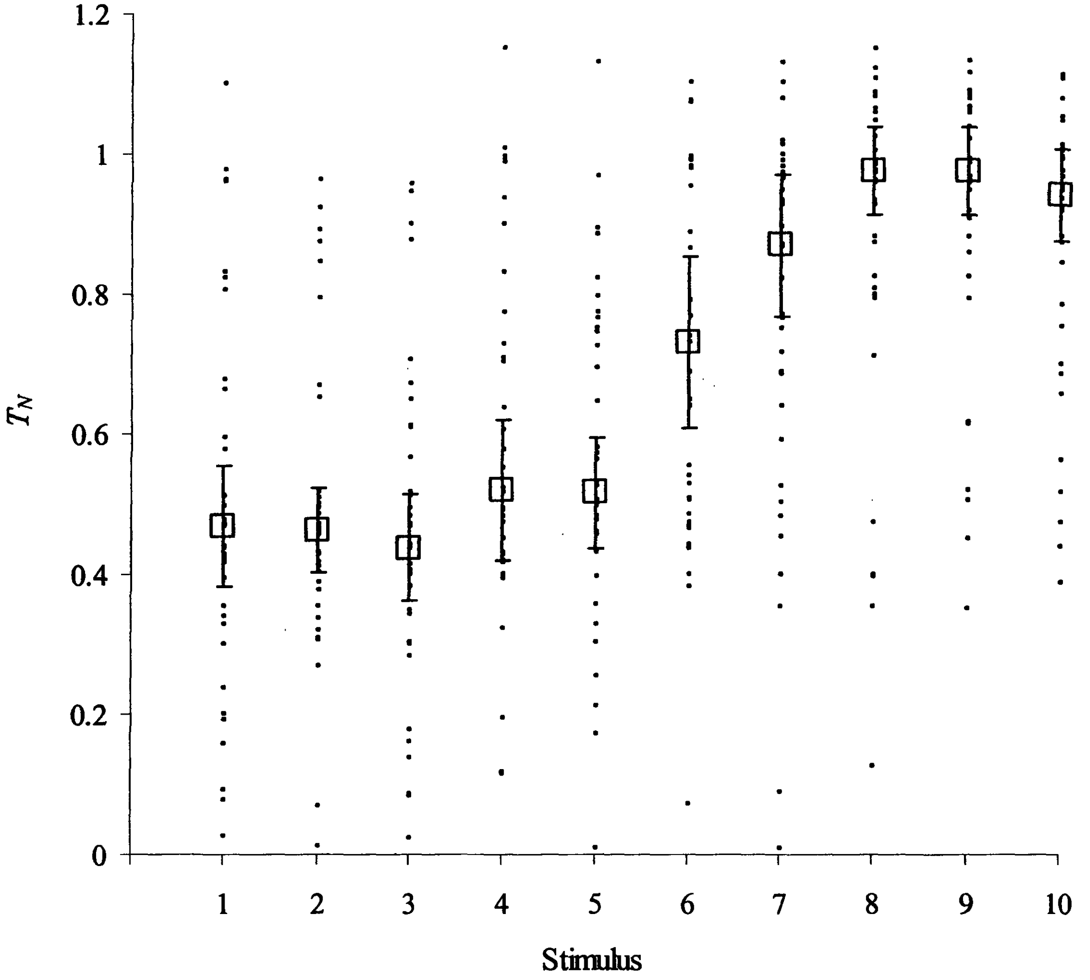


Figure 6.26. Normalized turning point time, T_N, for the *lemonade*-min series in Experiment 4.

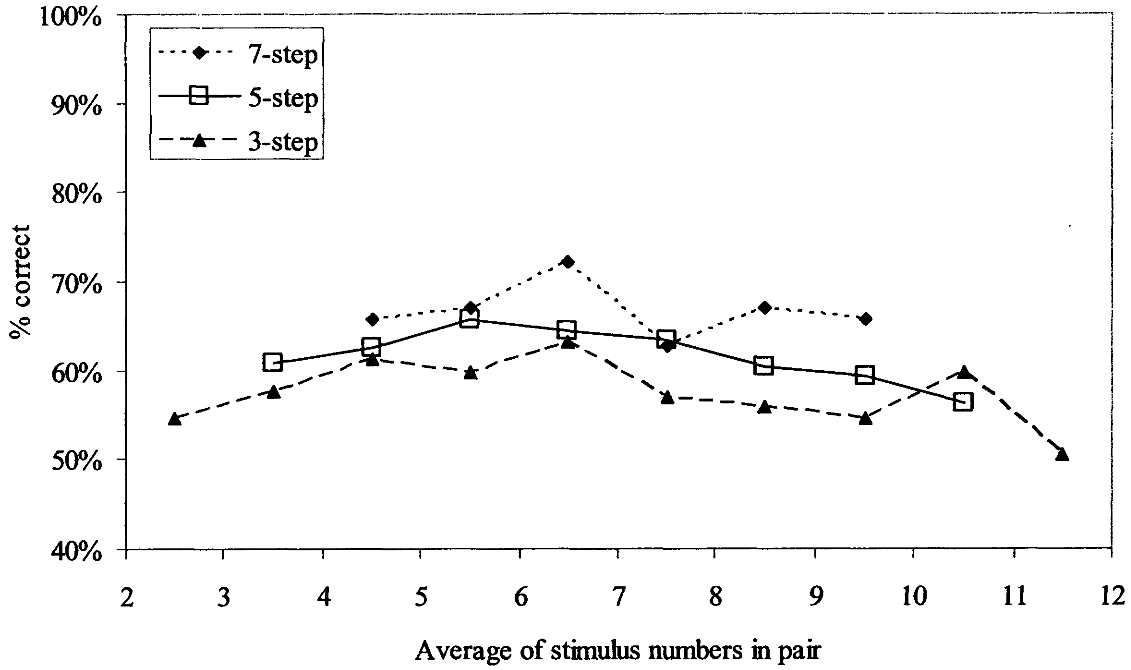


Figure 6.27. Percentage of correct responses for the *nonrenewable*-min series in Experiment 3.

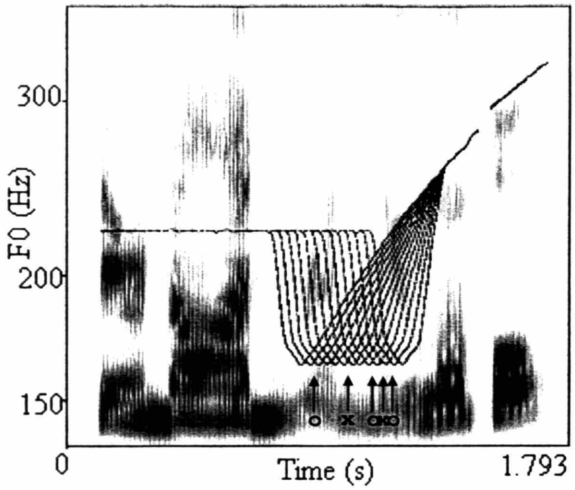


Figure 6.28. Summary of locations of discrimination maxima ("x") and minima ("o") along the *lemonade*-min stimulus series, for stimulus pairs which were 3 and 5 steps apart. See text.

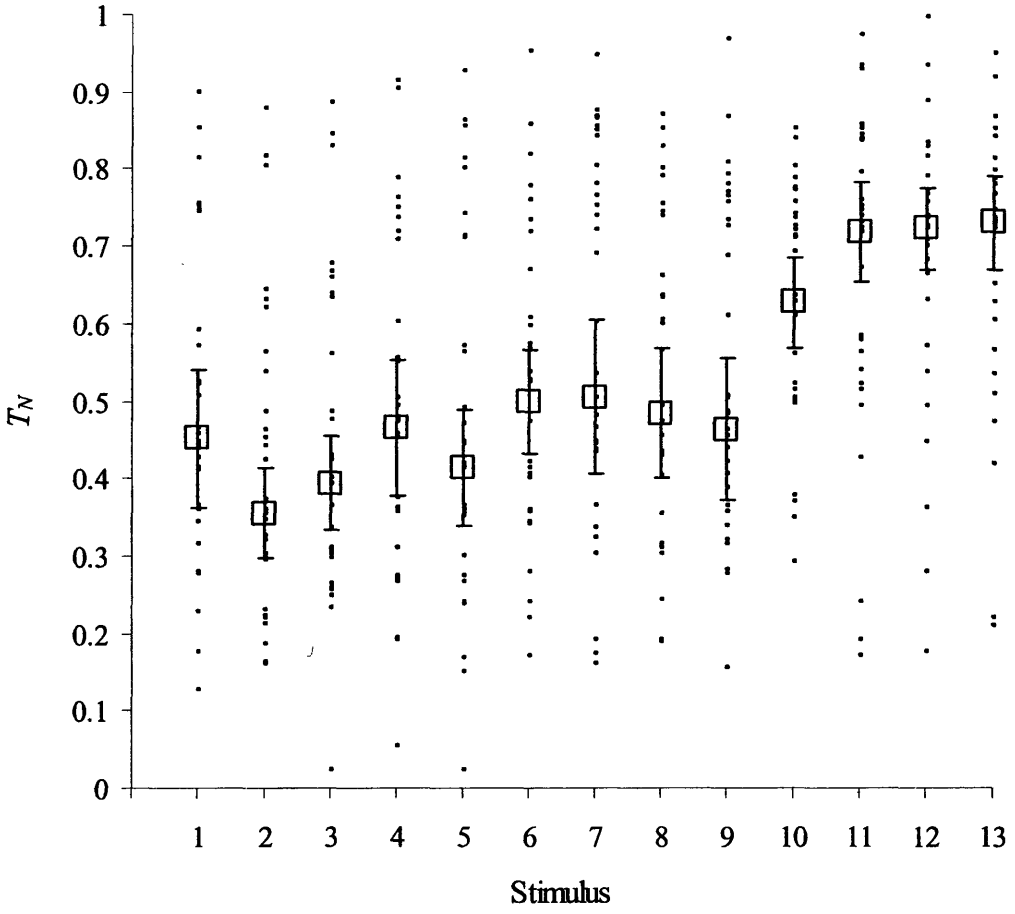


Figure 6.29. Normalized turning point time, T_N, for the *nonrenewable*-min series in Experiment 4.

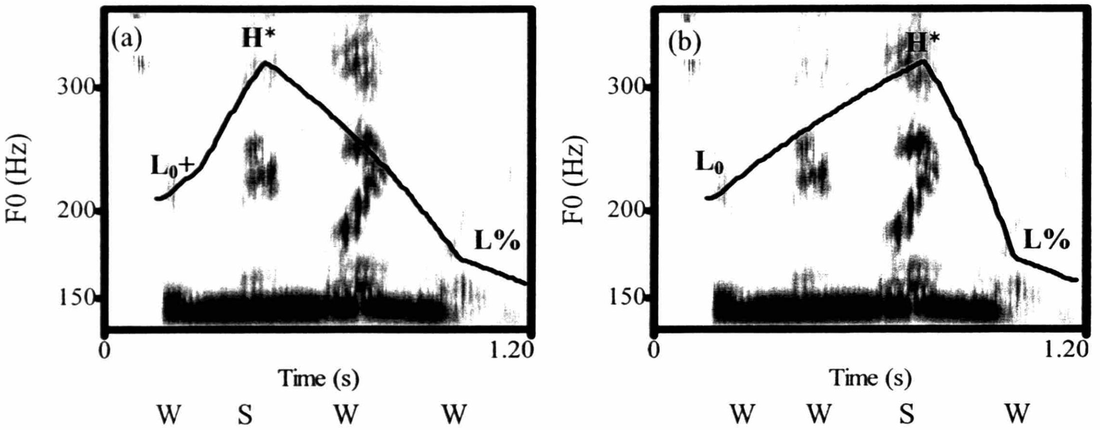


Figure 6.30. Phonological analyses of categories represented in the *minglingly*-max series (Experiments 1 and 2) for a syntagmatic tonal framework. The utterance is *Too minglingly*. 'W' and 'S' indicate the stress associated with each syllable.

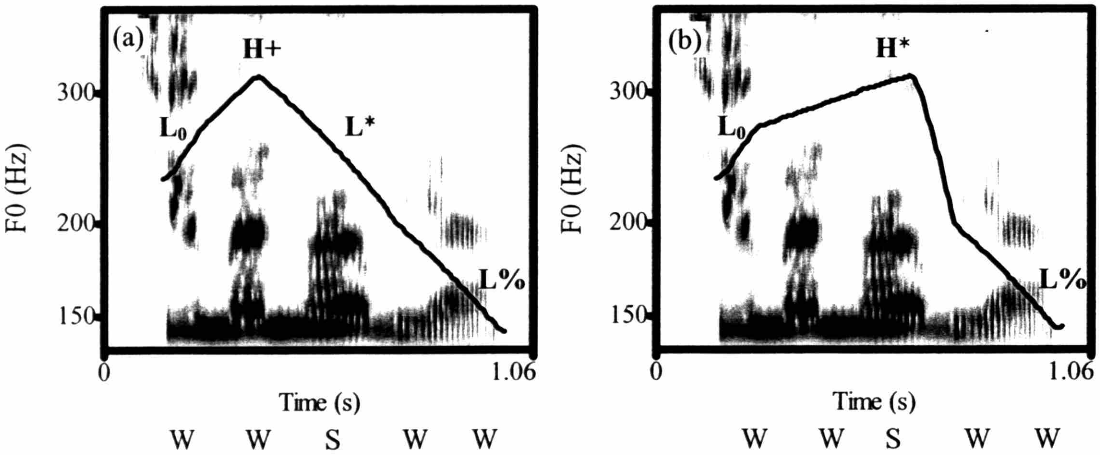


Figure 6.31. Phonological analyses of categories represented in the *Monrovia*-max series (Experiments 1 and 2) for a syntagmatic tonal framework. The utterance is *To Monrovia*. 'W' and 'S' indicate the stress associated with each syllable.

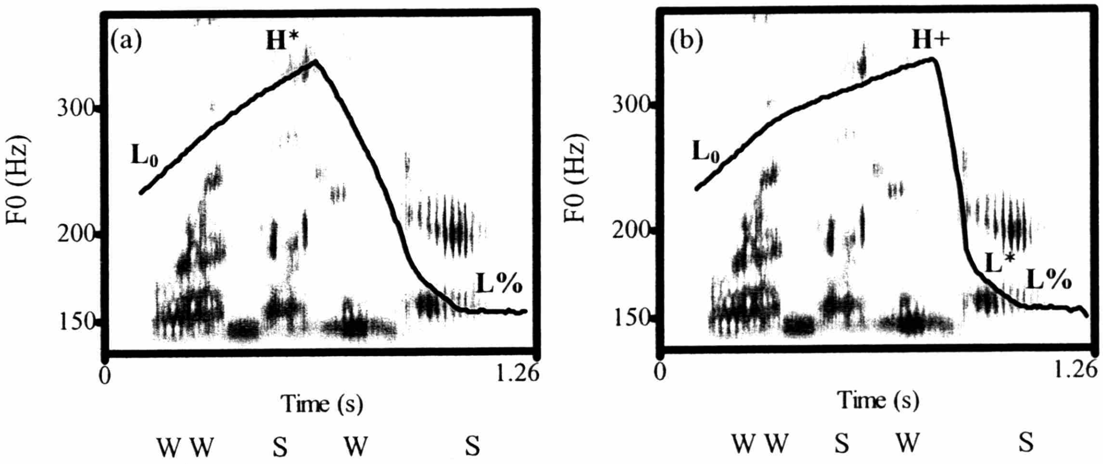


Figure 6.32. Phonological analyses of categories represented in the *millionaire*-max series (Experiments 3 and 4) for a syntagmatic tonal framework. The utterance is *For a millionaire.* 'W' and 'S' indicate the stress associated with each syllable.

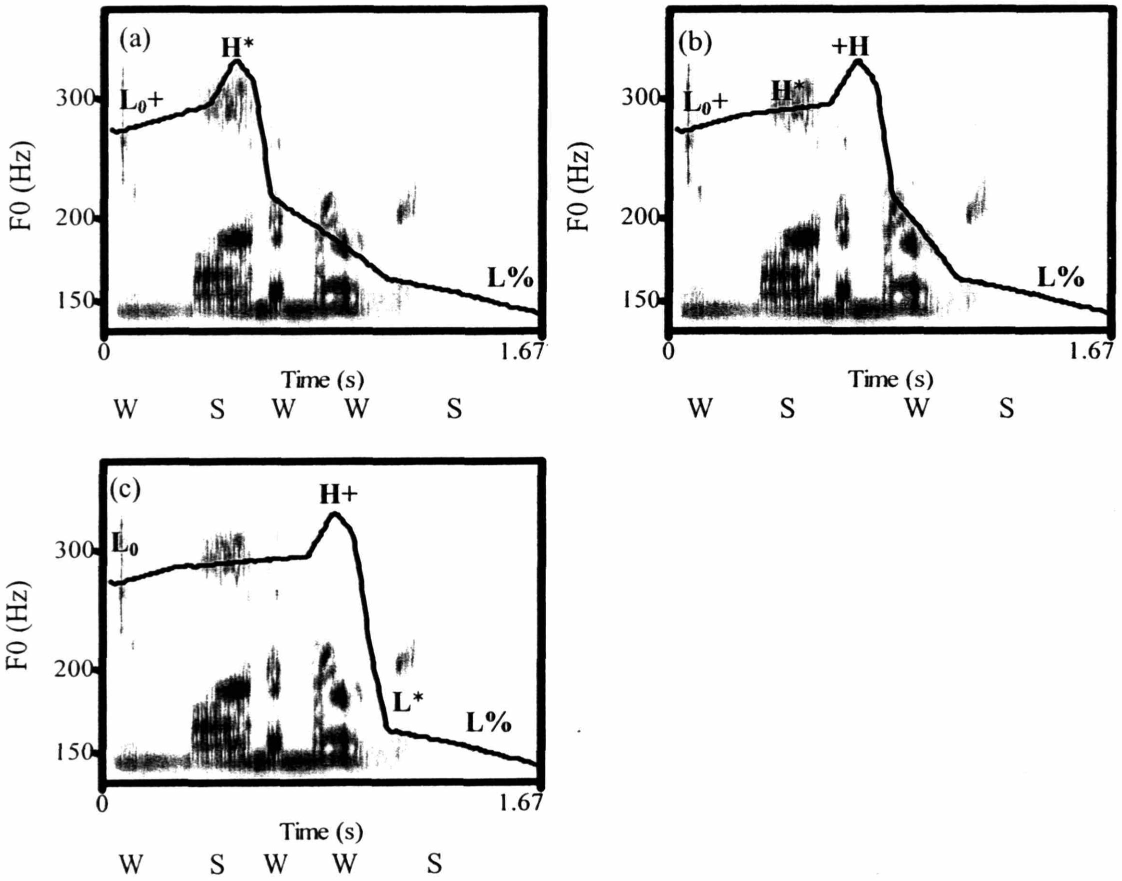


Figure 6.33. Phonological analyses of categories represented in the *Lannameraine*-max series (Experiments 3 and 4) for a syntagmatic tonal framework. The utterance is *In Lannameraine*. 'W' and 'S' indicate the stress associated with each syllable.

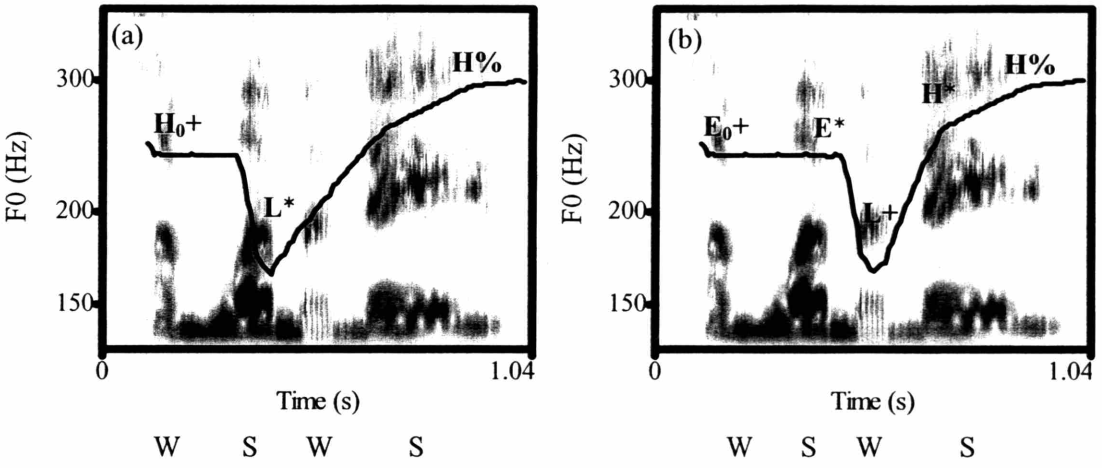


Figure 6.34. Phonological analyses of categories represented in the *lemonade*-min series (Experiments 3 and 4) for a syntagmatic tonal framework. The utterance is *Some lemonade?* 'W' and 'S' indicate the stress associated with each syllable.

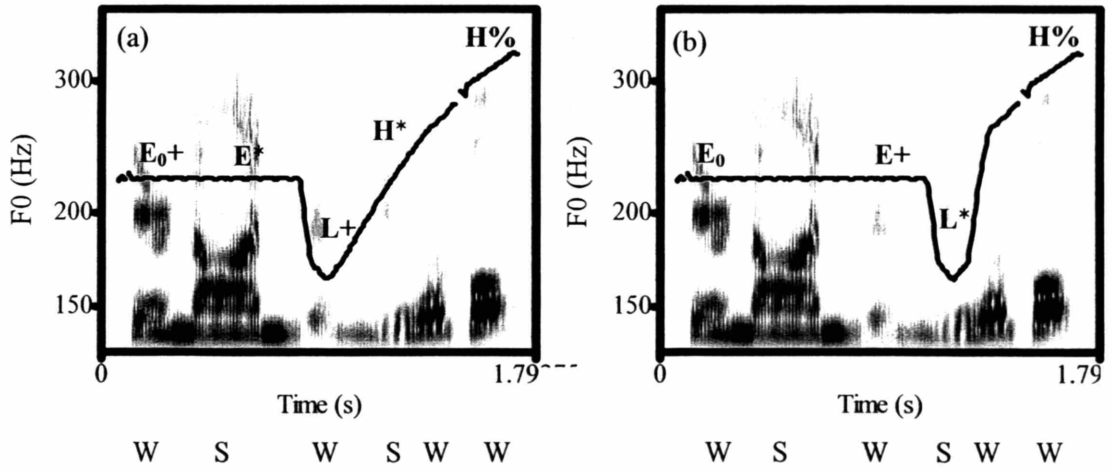


Figure 6.35. Phonological analyses of categories represented in the *nonrenewable*-min series (Experiments 3 and 4) for a syntagmatic tonal framework. The utterance is *They're nonrenewable?* 'W' and 'S' indicate the stress associated with each syllable.

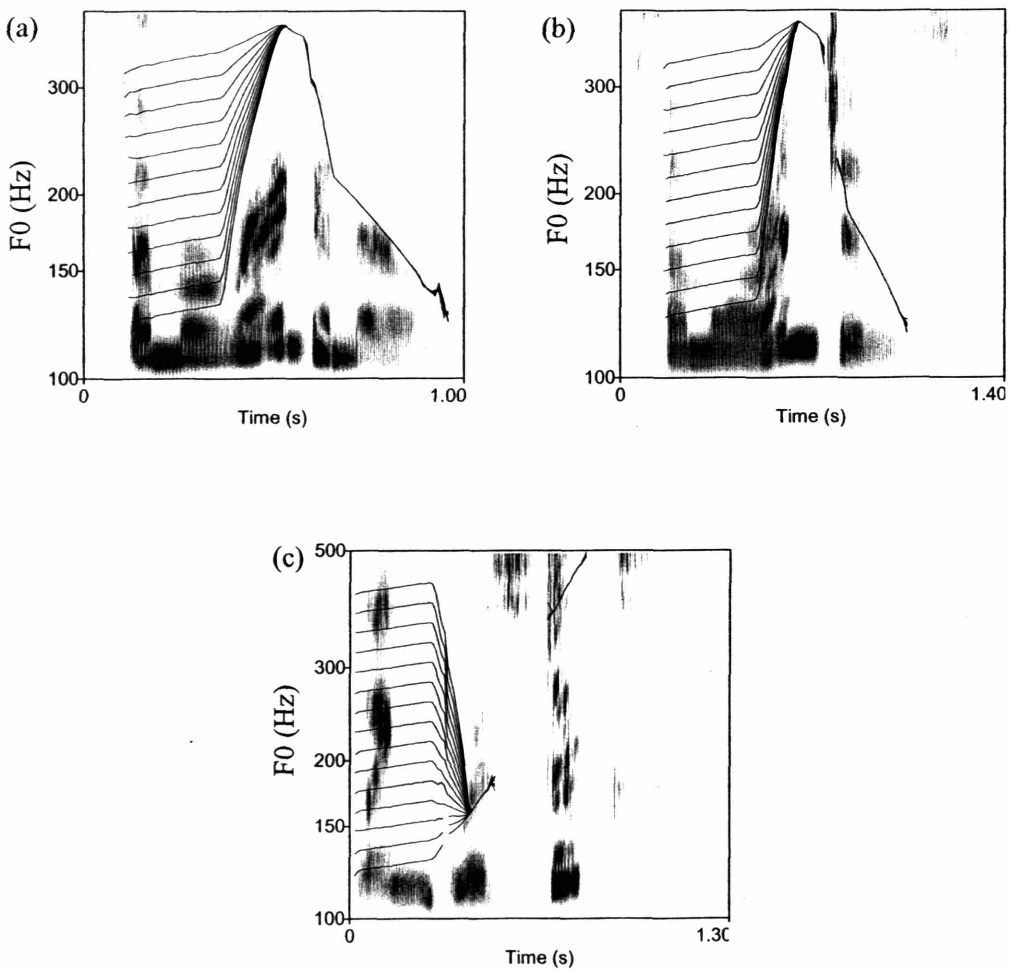


Figure 7.1. Stimuli used in Experiment 5. Clockwise from upper left: *Some oregano*, *Some oranges*, and *Linguistics?*

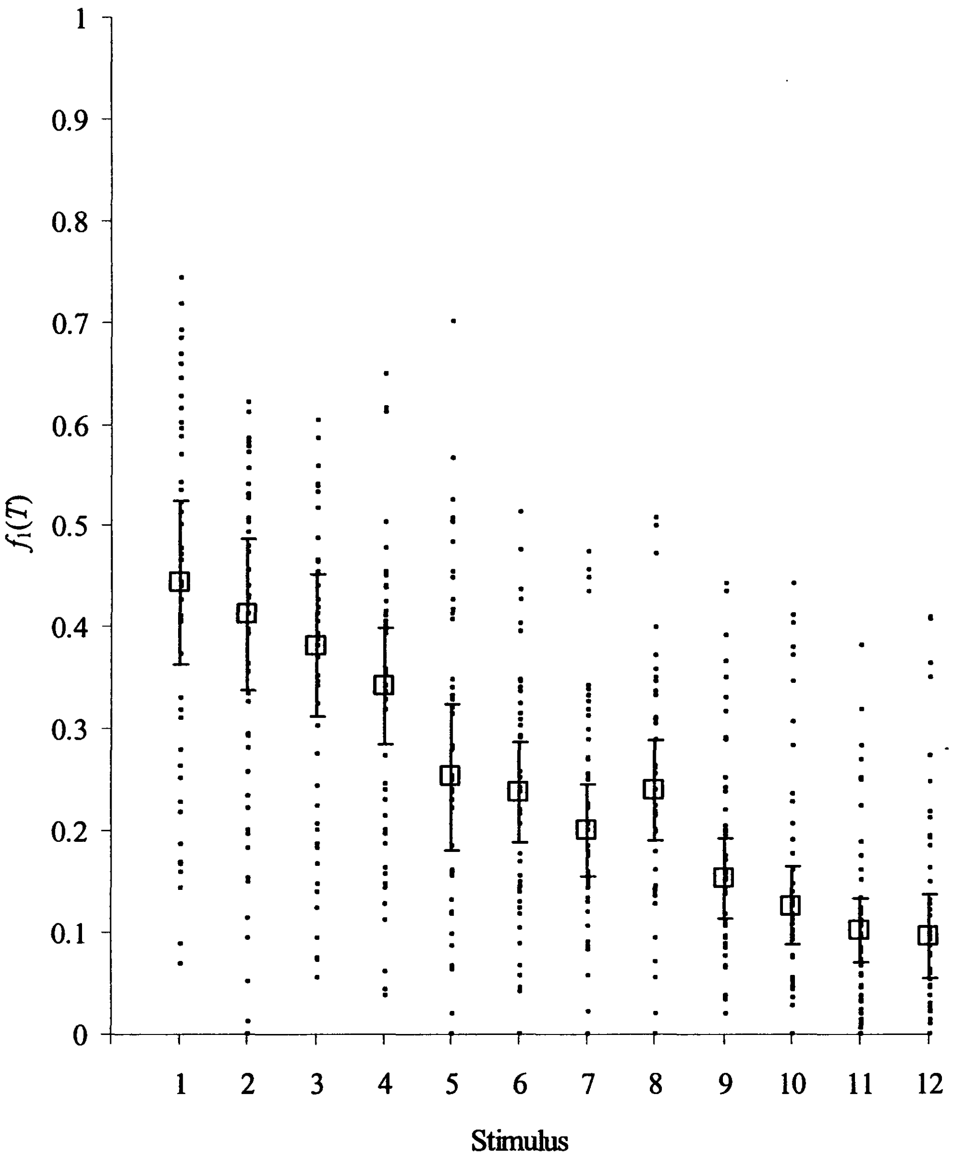


Figure 7.2: Scaling of *Some or-* in an imitation task for the *oregano* series, calculated according to the pitch range method of tonal scaling, $f_l(T)$ (Pierrehumbert and Beckman 1988). Data points (n = 539) represent all observations of $f_l(T)$ meeting criteria for this stimulus series. Open squares show median values, while whiskers show the semi-interquartile range. See text for more information.

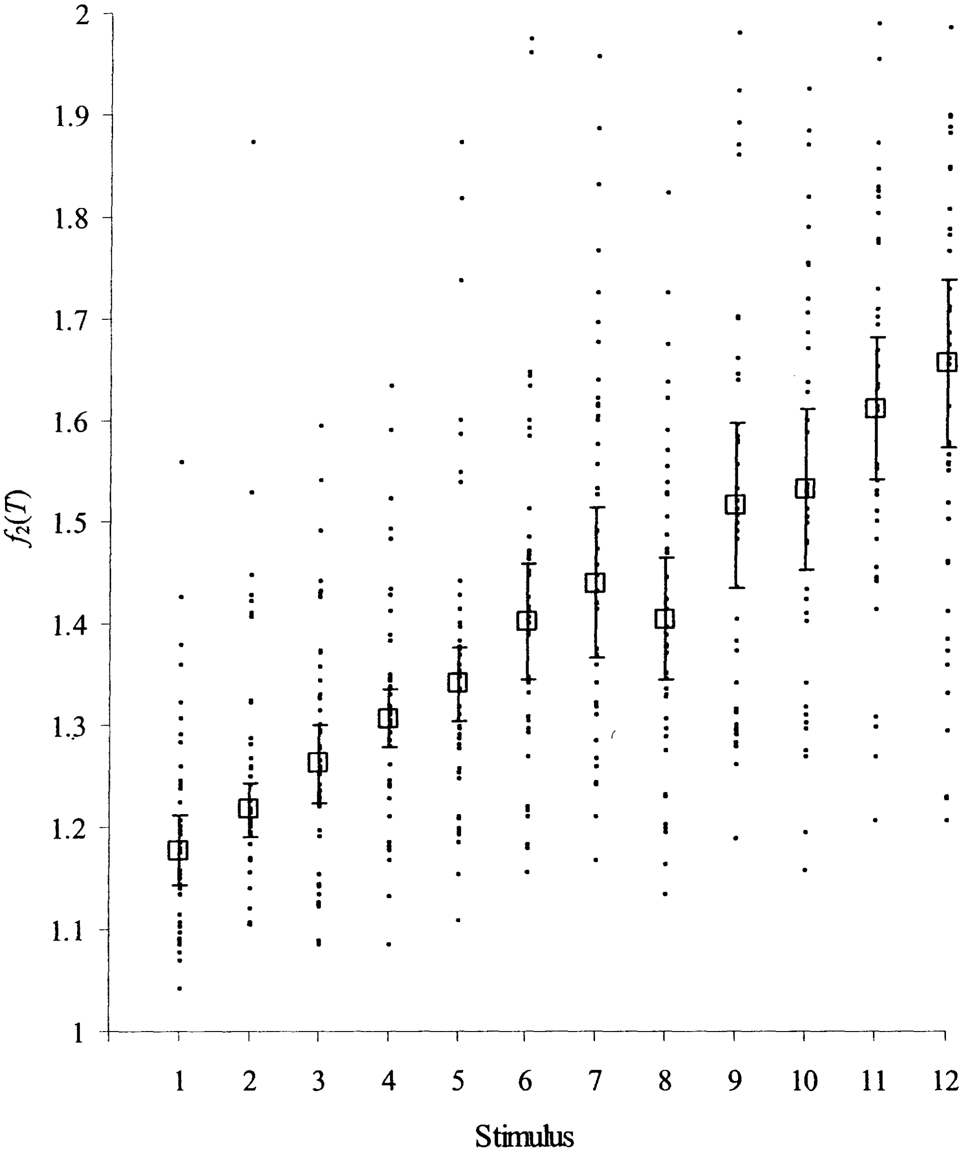


Figure 7.3: Scaling of *Some or-* in an imitation task for the *oregano* series, calculated according to the relative level method of tonal scaling, $f_2(T)$. Data points (n = 509) represent all observations of $f_2(T)$ meeting criteria for this stimulus series. Open squares show median values, while whiskers indicate one semi-interquartile range. See text.

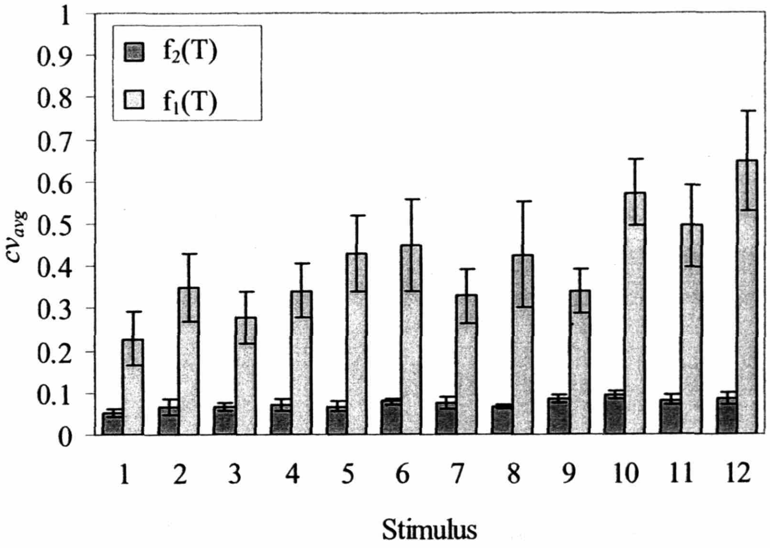


Figure 7.4: Coefficient of variation, cv_{avg}, for scaling of *Some or-* in an imitation task for the *oregano* series, calculated for the two methods of estimating tonal scaling. $f_1(T)$ is the pitch range method and $f_2(T)$ is the relative level method.

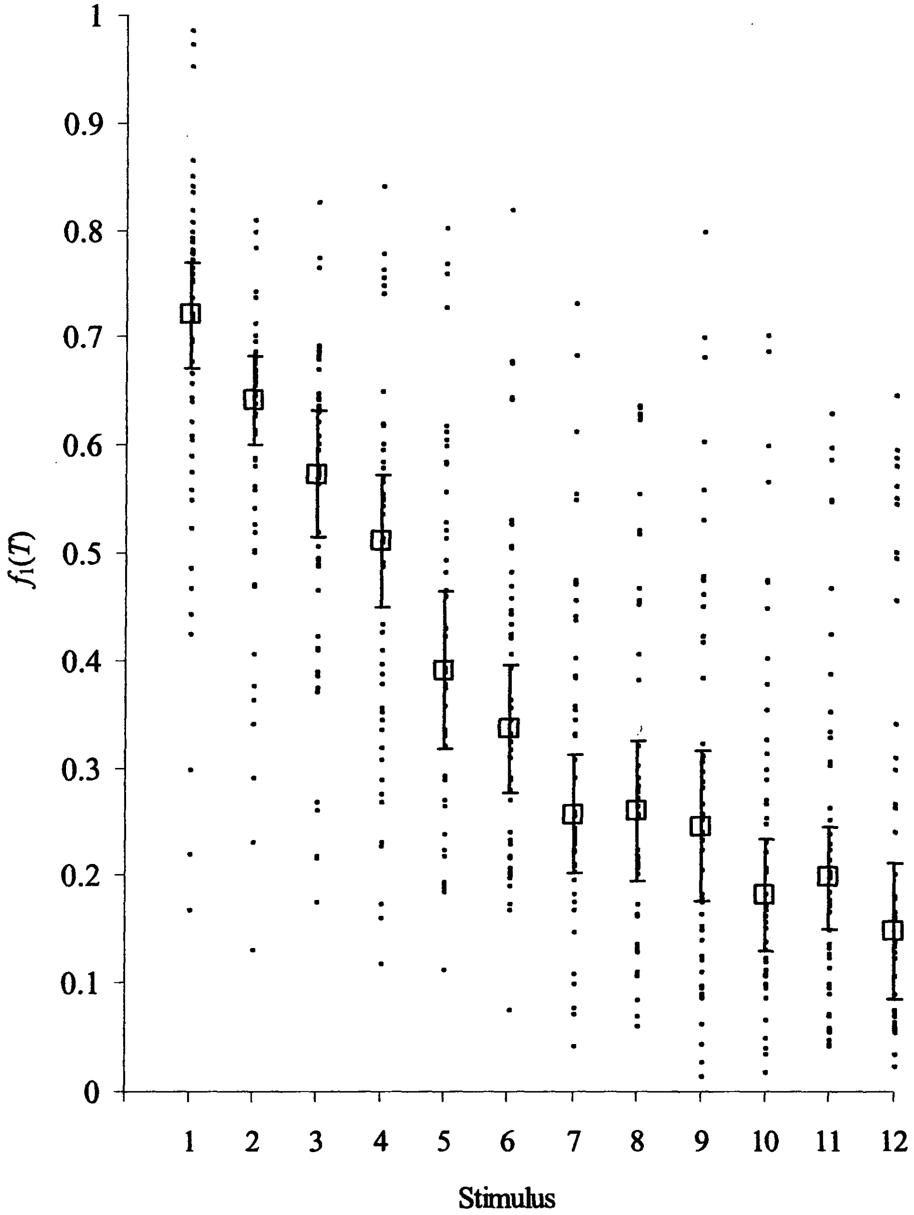


Figure 7.5: Scaling of *Some or-* in an imitation task for the *oranges* series, calculated according to the pitch range method of tonal scaling, $f_1(T)$ (Pierrehumbert and Beckman 1988). Data points (n = 576) represent all observations of $f_1(T)$ meeting criteria for this stimulus series. Open squares show median values, while whiskers indicate one semi-interquartile range. See text.

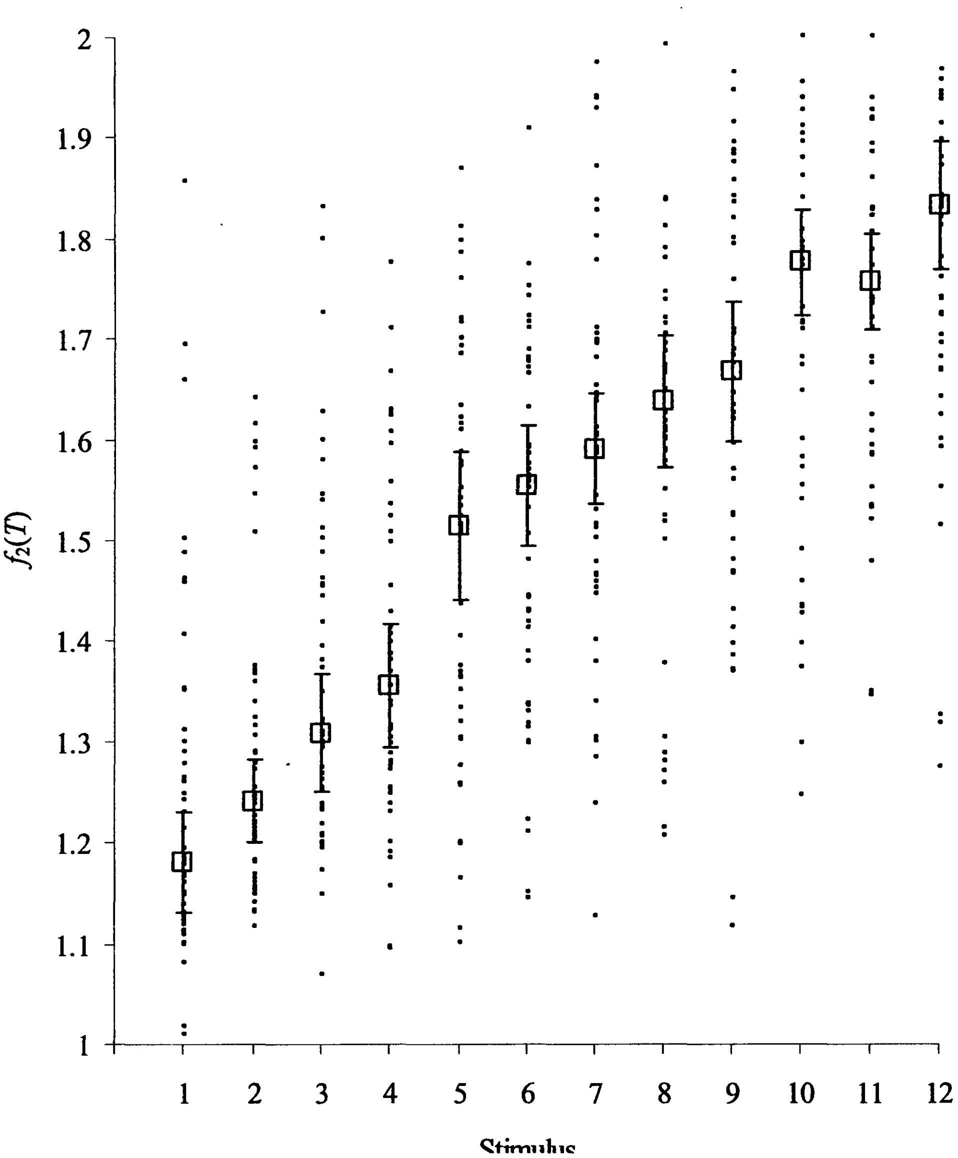


Figure 7.6: Scaling of *Some or-* in an imitation task for the *oranges* series calculated according to the relative level method of tonal scaling, $f_2(T)$. Data points (n = 576) represent all observations of $f_2(T)$ meeting criteria for this stimulus series. Open squares show median values, while whiskers indicate one semi-interquartile range. See text.

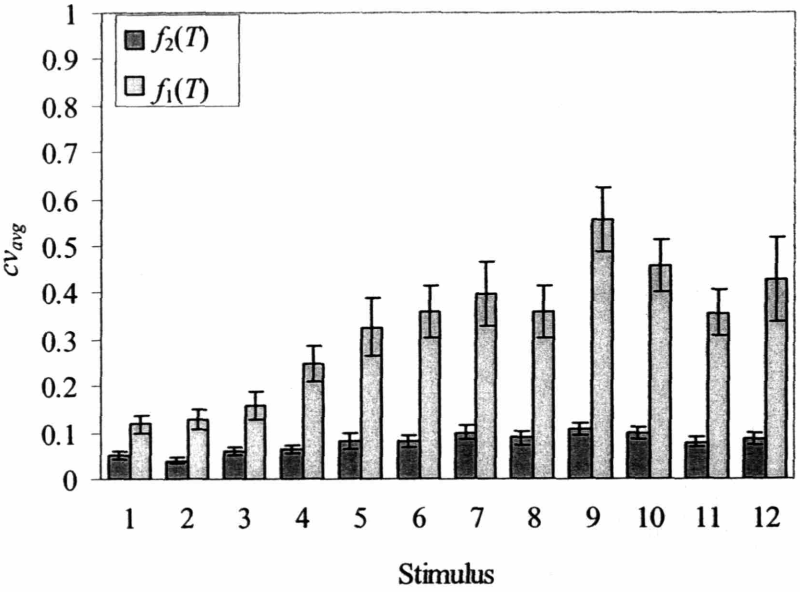


Figure 7.7: Coefficient of variation, cv_{avg}, for scaling of *Some or-* in an imitation task for the *oranges* series, calculated for the two methods of tonal scaling: $f_1(T)$ is the pitch range method and $f_2(T)$ is the relative level method.

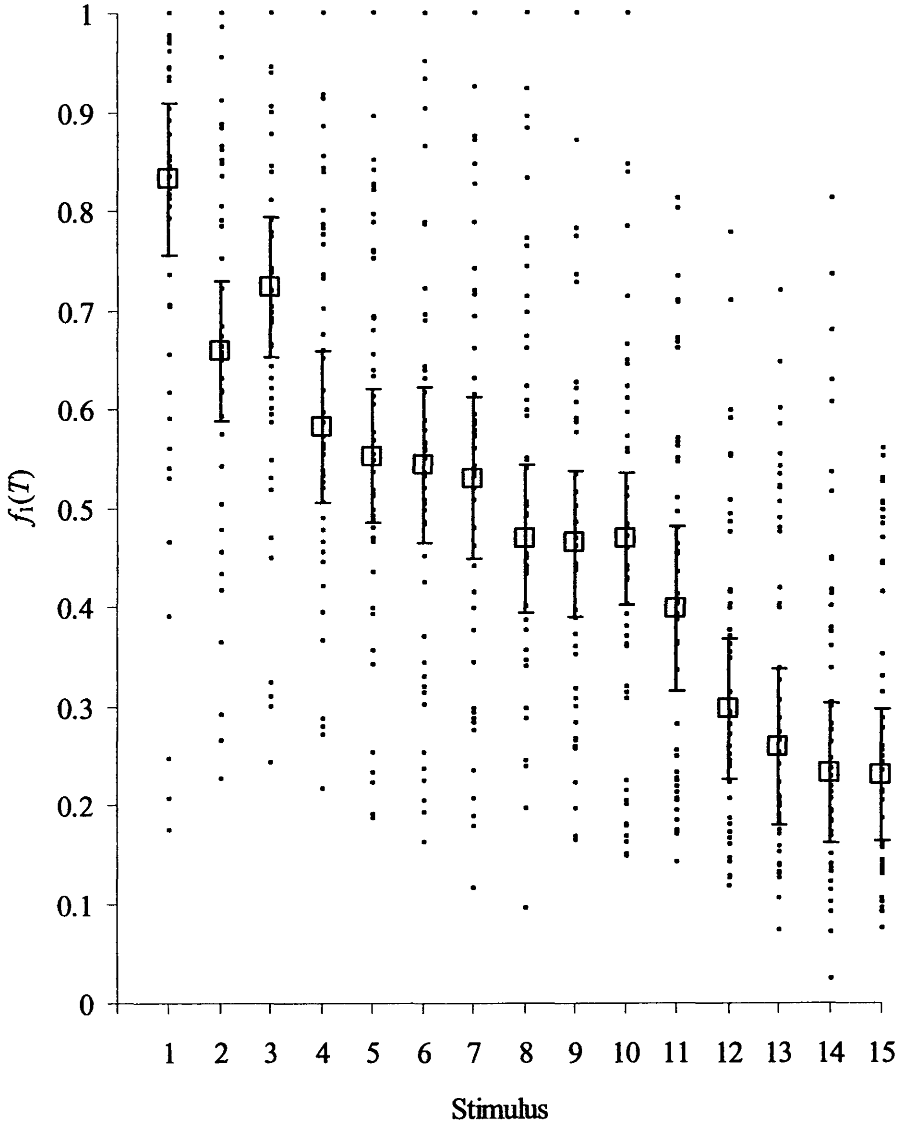


Figure 7.8: Scaling of *Ling-* in an imitation task for the *linguistics* series, calculated according to the pitch range method of tonal scaling, $f_l(T)$ (Pierrehumbert and Beckman 1988). Data points (n = 705) represent all observations of $f_l(T)$ meeting criteria for this stimulus series. Open squares show median values, while whiskers indicate the semi-interquartile range. See text.

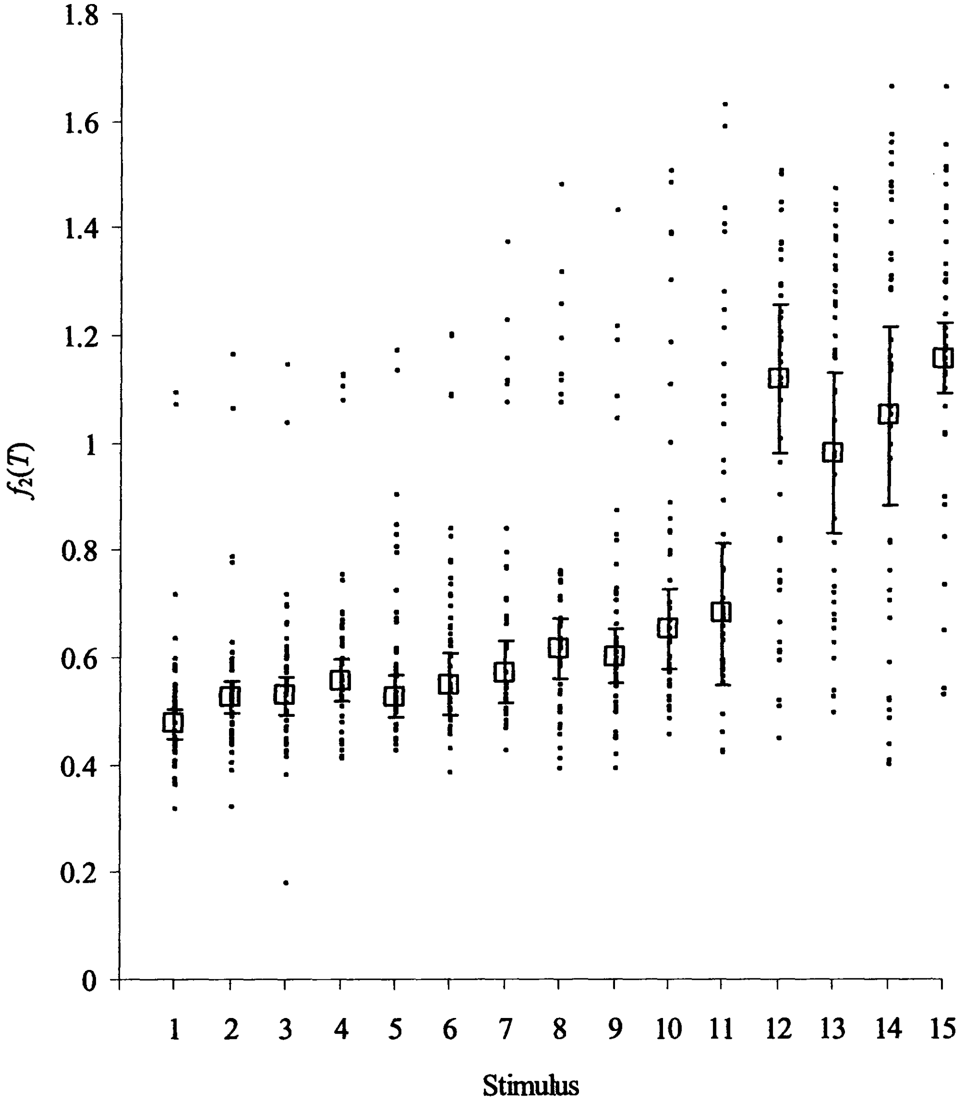


Figure 7.9: Scaling of *Ling-* in an imitation task for the *linguistics* series, calculated according to the relative level method of tonal scaling, $f_2(T)$. Data points ($n = 607$) represent all observations of $f_2(T)$ meeting criteria for this stimulus series. Open squares show median values, while whiskers indicate the semi-interquartile range. See text.

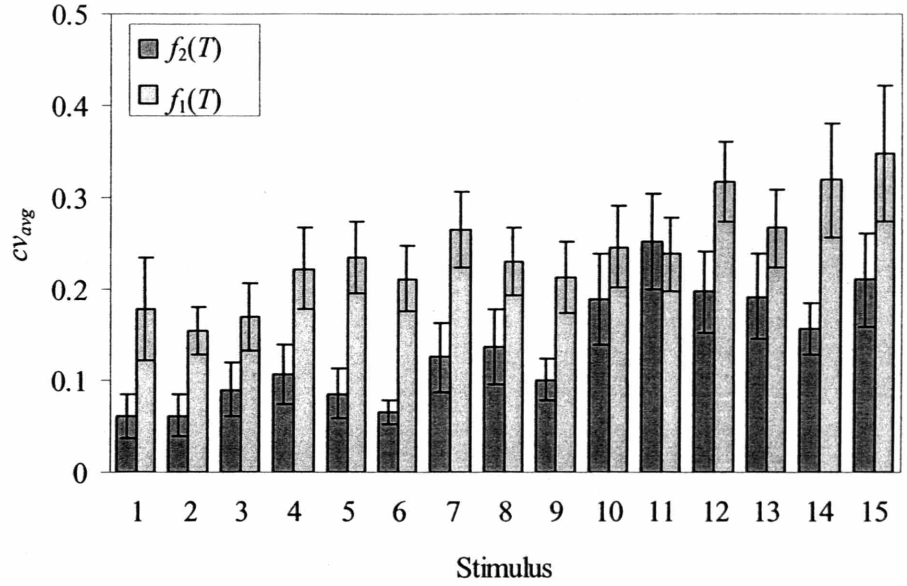


Figure 7.10: Coefficient of variation, cv_{avg}, for scaling of *Ling-* in an imitation task for the *linguistics* series, calculated for the two methods of tonal scaling: $f_1(T)$ is the pitch range method and $f_2(T)$ is the relative level method.

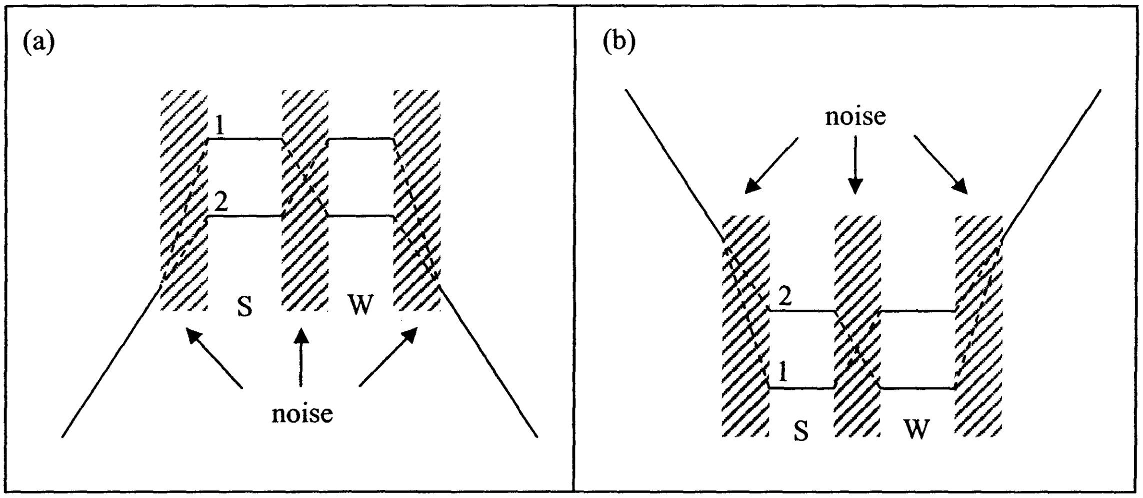


Figure 7.11. Schematic diagram of the stimuli used in Experiment 6.

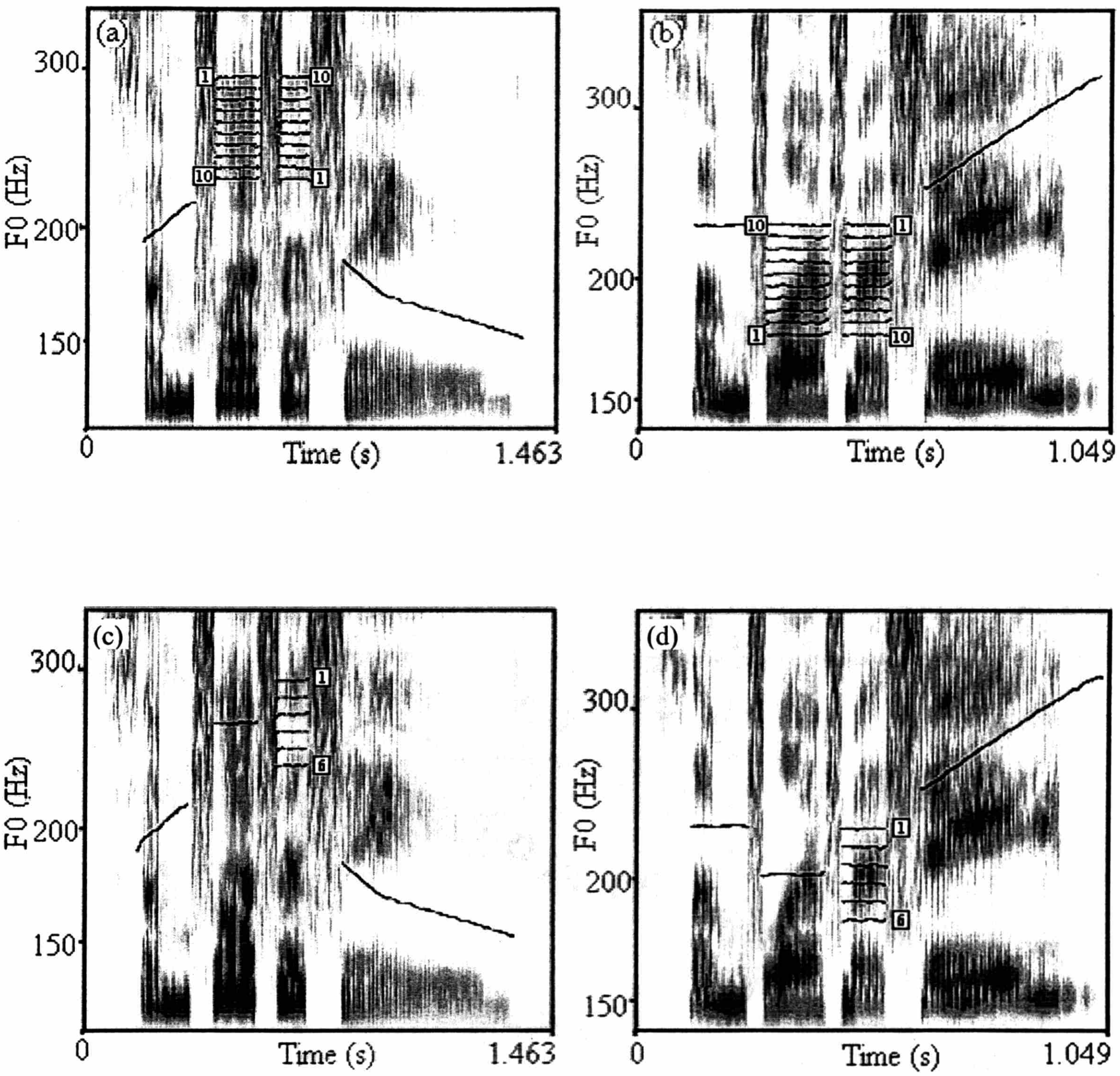


Figure 7.12: Stimuli used in Experiment 6: (a) Roving-High series, (b) Roving-Low series, (c) Fixed-High series, and (d) Fixed-Low series.

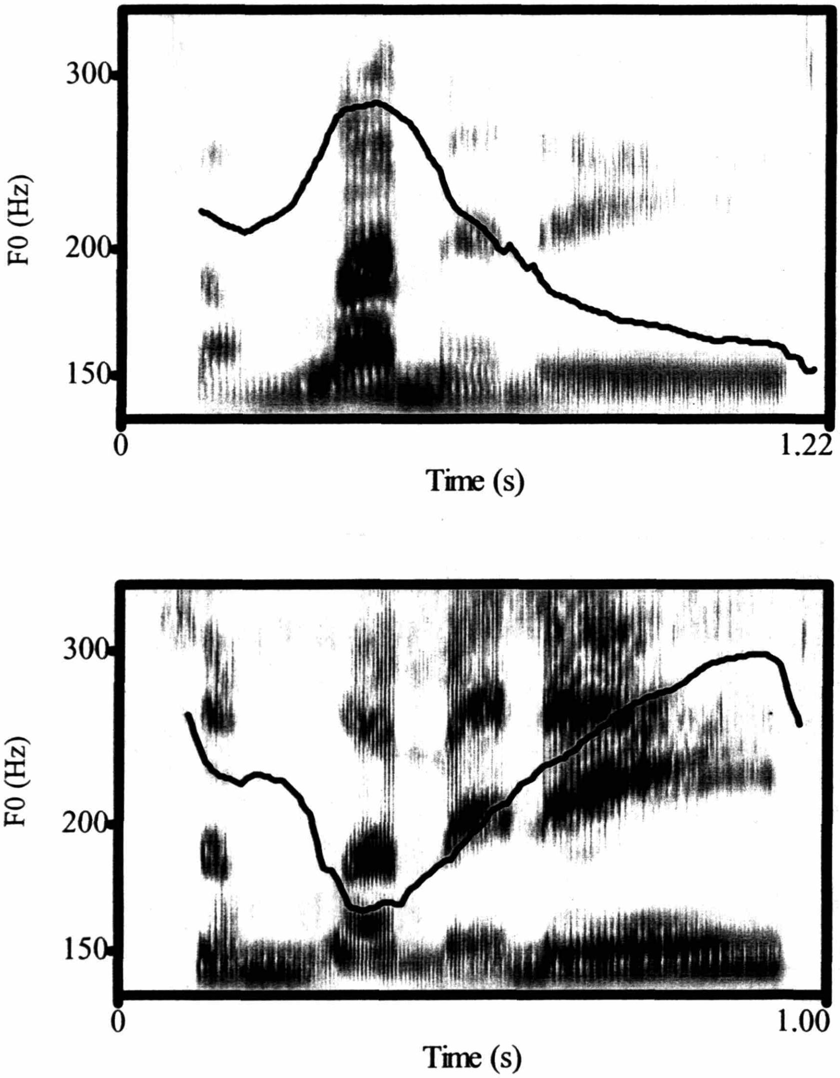


Figure 7.13: Examples of typical productions of the first stimulus in the Roving-High and Roving-Low series.

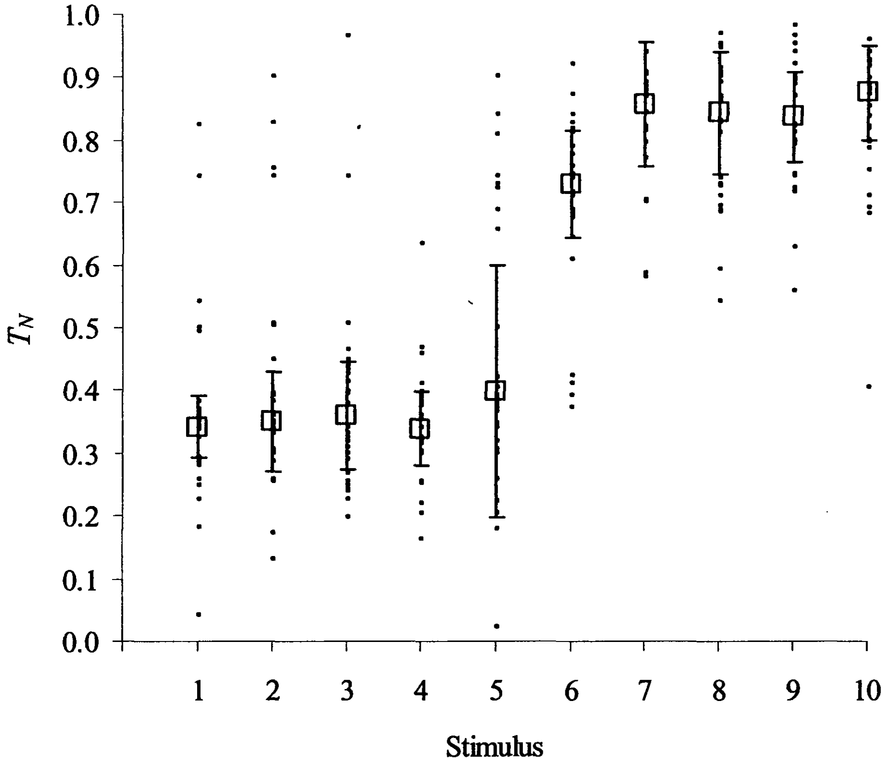


Figure 7.14: Normalized turning point location, T_N, for the Roving-High series. Data points (n = 300) represent all observations of T_N meeting criteria for this stimulus series. Open squares show median values, while whiskers indicate the semi-interquartile range. See text.

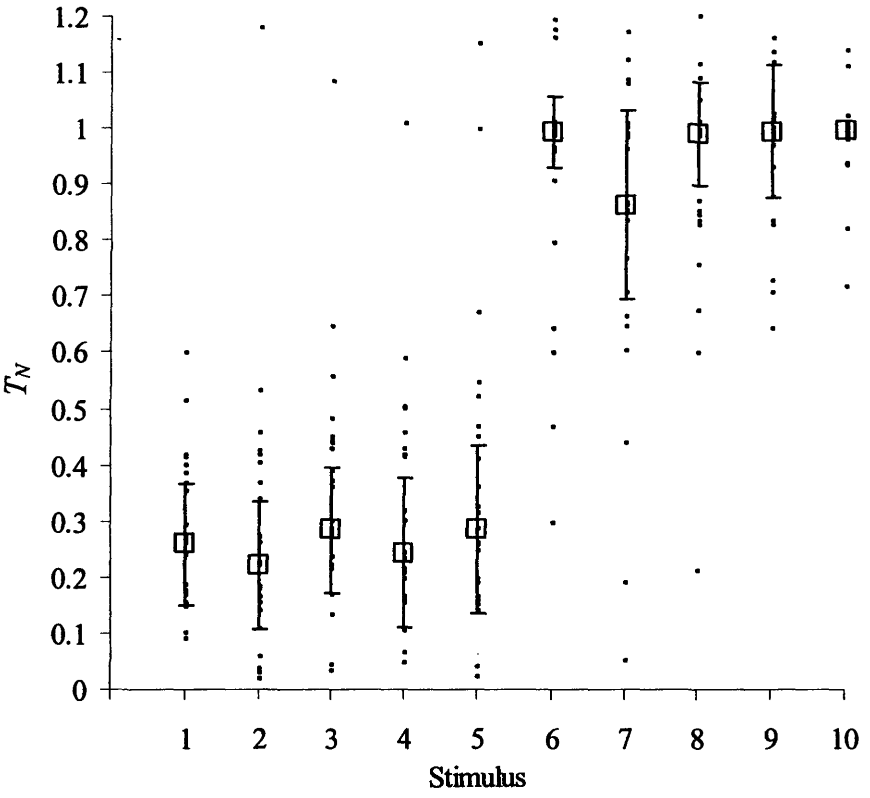


Figure 7.15: Normalized turning point location, T_N, for the Roving-Low series. Data points (n = 237) represent all observations of T_N meeting criteria for this stimulus series. Open squares show median values, while whiskers indicate the semi-interquartile range. See text.

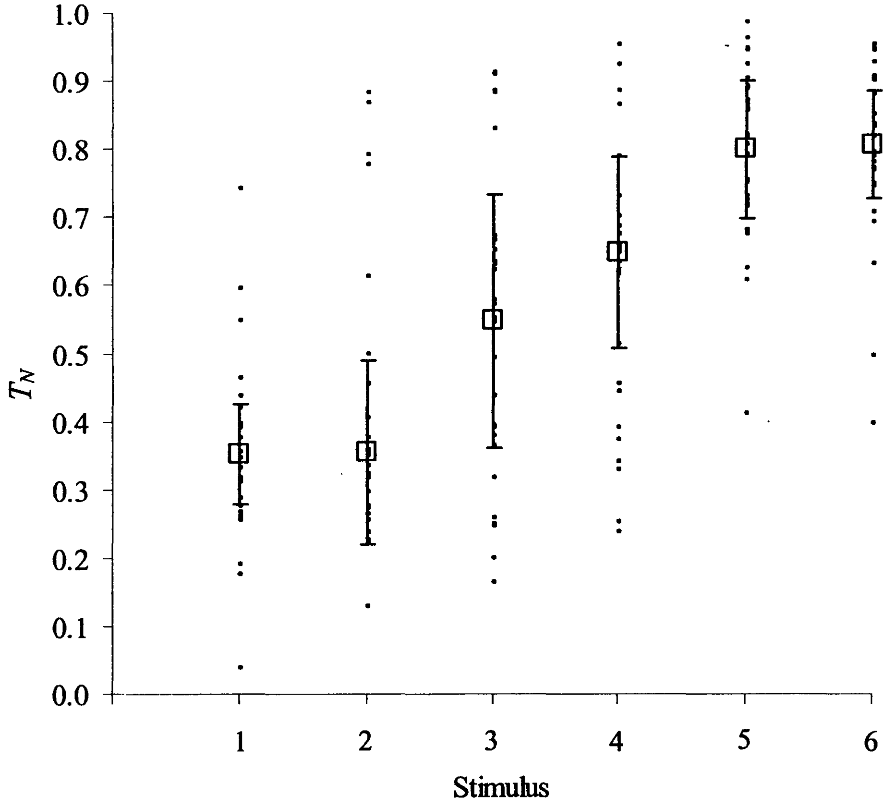


Figure 7.16: Normalized turning point location, T_N, for the Fixed-High series. Data points (n = 180) represent all observations of T_N meeting criteria for this stimulus series. Open squares show median values, while whiskers indicate the semi-interquartile range. See text.

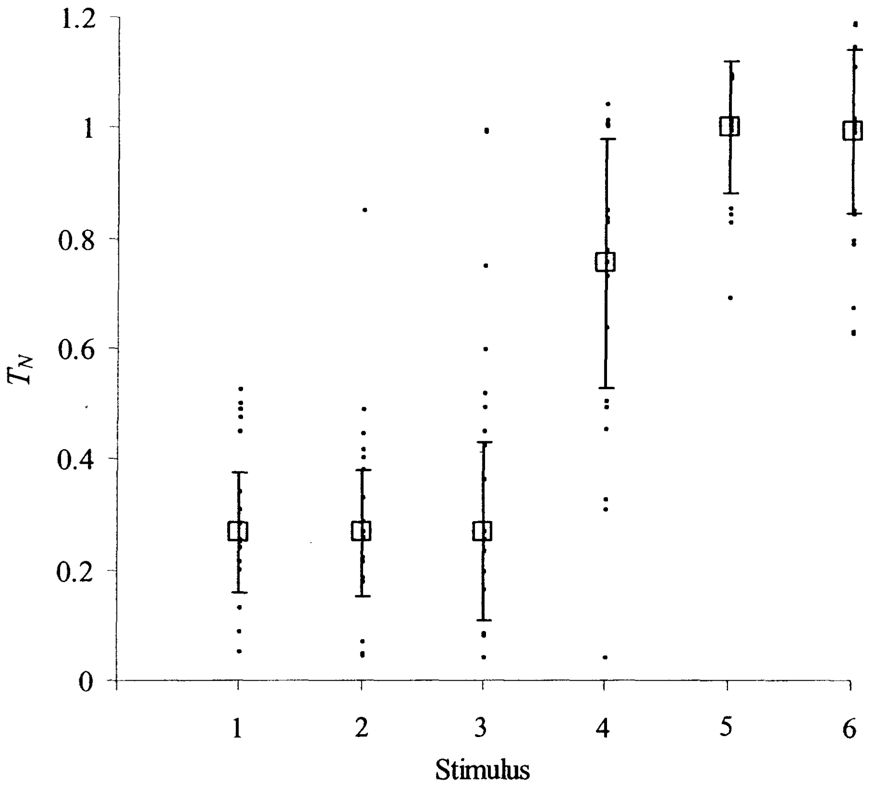


Figure 7.17: Normalized turning point location, T_N, for the Fixed-Low series. Data points (n = 104) represent all observations of T_N meeting criteria for this stimulus series. Open squares show median values, while whiskers indicate the semi-interquartile range. See text.

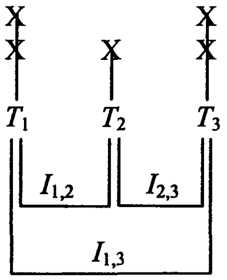


Figure A.1. A minimal subtending unit of the tone interval matrix and grid.

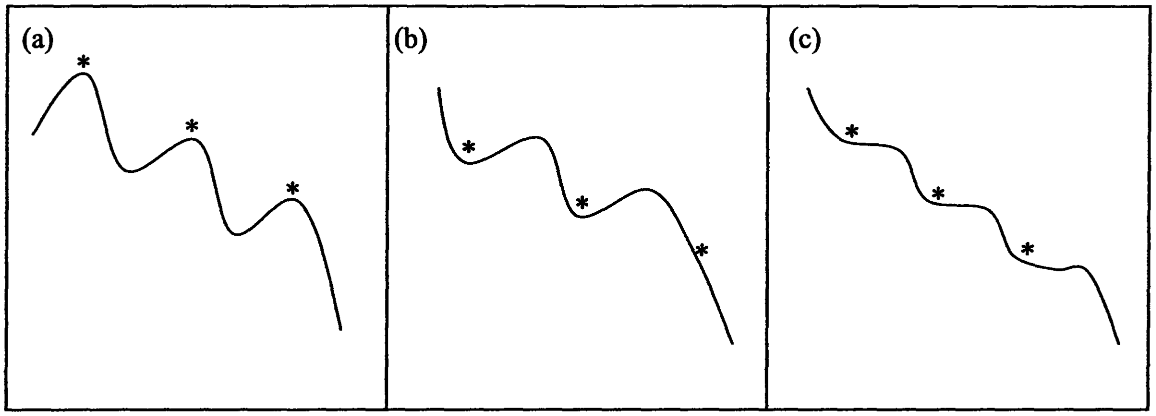


Figure A.2. Three iterative stepping patterns showing a downtrend.

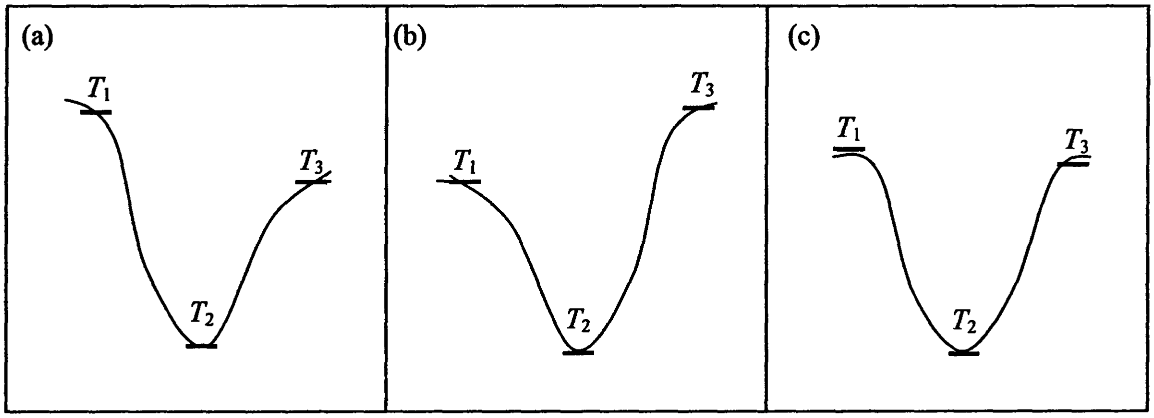


Figure A.3. Three contours which step down and then up. In (a), T_3 is lower than T_1. In (b), T_3 is higher than T_1. Finally, in (c), T_3 is at about the same level as T_1.