Language Learning & Language Teaching

Discourse Intonation in L2

From Theory and Research to Practice

Dorothy M. Chun

Discourse Intonation in L2

Language Learning and Language Teaching

The *LL&LT* monograph series publishes monographs as well as edited volumes on applied and methodological issues in the field of language pedagogy. The focus of the series is on subjects such as classroom discourse and interaction; language diversity in educational settings; bilingual education; language testing and language assessment; teaching methods and teaching performance; learning trajectories in second language acquisition; and written language learning in educational settings.

Volume 1

Discourse Intonation in L2: From theory and research to practice
by Dorothy M. Chun

Discourse Intonation in L2

From theory and research to practice

Dorothy M. Chun
University of California, Santa Barbara

John Benjamins Publishing Company
Amsterdam/Philadelphia

Library of Congress Cataloging-in-Publication Data

Chun, Dorothy M.

Discourse Intonation in L2 : From theory and research to practice / Dorothy M. Chun.
p. cm. (Language Learning & Language Teaching, ISSN 1569-9471; v. 1)
Includes bibliographical references and indexes.
1. Language and languages--Study and teaching. 2. Intonation (Phonetics) 3. Discourse analysis. I. Title. II. Series.

P53.C514 2002
418.0071--dc21 20020118794
ISBN 90 272 1691 6 (Eur.) / 1 58811 168 7 (US) (Hb; alk. paper)
ISBN 90 272 1692 4 (Eur.) / 1 58811 169 5 (US) (Pb; alk. paper)

To Joe, Steven, and Daniel

Acknowledgments

My interest in the melody of language began when I was introduced to the tones of Chinese as a child. Once I began studying other languages and eventually started teaching German as a foreign language, I realized that intonation and rhythm are critical features of spoken language, yet are rarely addressed in the teaching and learning of languages. I hope that this book will demonstrate the importance of intonation as a critical element of linguistic systems and the need for intonation and rhythm to be included in language instruction.

I would like to thank the following colleagues for their very helpful input and discussions: Wallace Chafe, Elizabeth Couper-Kuhlen, H.S. Gopal, Eric Keller, Martha Pennington, Stephen Schuetze-Coburn, and Ann Wennerstrom. The students in a graduate seminar "The Acquisition and Teaching of L2 Phonetics, Phonology and Intonation" that I taught in Winter Quarter 2001 at UCSB provided me with timely and useful feedback, for which I am grateful.

I owe a tremendous debt of gratitude to Stephen Lasswell, who had the unenviable task of clarifying my prose and who did an outstanding job of editing two versions of the manuscript. His expertise both in linguistics and in editing and the painstaking care with which he read every sentence and suggested modifications are reflected in the superb readability of the final product. All errors and infelicities in content are of course my own.

My colleagues in the Office of Instructional Development at the University of California, Santa Barbara who helped create the CD-ROM are the most talented group I have had the pleasure to work with, and they have my special thanks: George Michaels, Steven Brown, and Teja Ream.

Kees Vaes at John Benjamins provided me with constant encouragement and assistance, and I have very much appreciated working with him and his colleagues and with the editors of this series, Birgit Harley and Jan Hulstijn. The reviewers of the original manuscript provided constructive suggestions which improved the text immensely.

I would like to thank my parents for their unconditional love and support,

for instilling in me the value of hard work and perseverance, and for teaching me that "it is not what you say but how you say it."

Finally, I wish to express my heartfelt thanks to my wonderful family, Joe, Steven, and Daniel, to whom this book is dedicated, for their unfailing patience, support, and encouragement.

Dorothy M. Chun
Santa Barbara, California
February 2002

Contents

Introduction

> *"It is very noticeable that, although some research has been devoted to (sentence) intonation, most work has hitherto concentrated on what traditionally have been considered lower level speech phenomena — the phoneme, the distinctive feature, the syllable, and so on. For a fuller understanding of L2 speech acquisition, this concentration will need in the future to be balanced by closer attention to higher level patterning in the discourse domain."*
>
> (Leather & James, 1991, p. 332)

Intonation, stress, and general "melody" of speech are said to be among the first aspects of speech that human infants attend to, react to, and produce themselves (cf. Lieberman, 1986, p. 239). It has also been shown that child learners of a second language have little difficulty in acquiring native-like pronunciation and intonation in the L2 (Felix, 1978). Intonation is thus an important aspect of language that seems to be easily, if not automatically, acquired by children in both L1 and L2. Moreover, as observation and experience amply demonstrate, it is easy for adults to maintain and retain in the L1, yet difficult, if not impossible, for adults to learn in an L2. Due to its inherent complexity and to the corresponding difficulty in learning and mastering it, intonation was ignored for many years in language teaching. However, it is now slowly gaining recognition as an integral part of language fluency, competence, and proficiency.

There appear to be several reasons for the current heightened interest in intonation. First, in theoretical linguistics, there have been important new advances in the theory of intonation and its representation aided by the growing accessibility of computational acoustic speech analysis. Second, the expansion of the analytical domains of traditional linguistics from sounds (phonetics and phonology), words (morphology), and sentences (syntax) to larger units of inquiry such as entire texts/discourses has given rise to such subfields as pragmatics, text linguistics, discourse analysis, and conversation analysis. In addition, the realization that language and language acquisition are influenced by other disciplines has strengthened the interdisciplinary focus of linguistics

to include psycholinguistic and sociolinguistic perspectives. Third, applied linguistics has evolved to emphasize *communicative function* rather than *linguistic form*. With particular regard to the teaching and learning of pronunciation, there has been a trend to adopt a top-down approach, i.e., to focus more on global meaning and communication (via context- and discourse-based instruction), rather than to adhere to the traditional bottom-up, phonemic-based approach (based on isolated or contrasted sounds).

In accordance with these major shifts in emphasis in both general and applied linguistics, this book about intonation has two main goals: (1) to show the practical importance of intonation, particularly discourse intonation (i.e., intonation in contexts/discourses beyond isolated sentences), which is acknowledged in theoretical linguistics as an indispensable component of language and communication and which thus ought to be taught to L2 learners, and (2) to integrate theory and practice, i.e., to draw upon research both in theoretical linguistics as well as applied linguistics in order to (a) render intonation accessible to teachers and pedagogues, and (b) suggest a pedagogical approach to the teaching of discourse-based phonology and intonation in L2 instruction.

Part I of the book presents an overview of the linguistic theories of intonation. In Chapter 1, the main components of intonation, including fundamental frequency (or pitch), intensity (or loudness), duration (or length), stress, accent, and rhythm are defined and also discussed in terms of how they are perceived by hearers. In Chapter 2, the main formal models of intonation are described, starting with the British contour approach, followed by the American phonemic (or levels) approach, and including Bolinger's theory of pitch accent. The current generative model is also summarized, and the chapter concludes with a discussion of discourse intonation. The purpose of the chapter is to make the concepts and representation of intonation accessible to non-linguists and language teachers. Chapter 3 discusses the meanings and functions of intonation. Although there are varying taxonomies, I suggest four main functional categories: first, the *grammatical* functions of distinguishing statements from questions, marking prominence or stress, and segmenting stretches of speech into words, phrases, and sentences; second, the *attitudinal* or *affective* functions of using intonation to express emotions, attitudes, and intentions; third, the *discourse* functions of expressing pragmatic meanings, achieving cohesion within a discourse, signaling turn-taking cues and distinguishing different discourse modes; and fourth, the *sociolinguistic* functions of intonation, employing it, e.g., to indicate role or status of the speaker or to

signal politeness. Important as all these categories are, it is the discourse function that researchers in pragmatics, discourse analysis, and conversation analysis as well as in second and foreign language acquisition are increasingly focusing their attention on.

Part II examines applied linguistic research during the last thirty years on pronunciation and intonation in L2 learning contexts. Chapter 4 reviews the past areas of research and traditional rationales and goals for teaching L2 pronunciation; until around the mid-1980s, pronunciation pedagogy generally adhered to the traditional phonemic-based, structuralist approach of teaching individual sounds via articulatory descriptions and contrastive minimal-pair exercises. Factors in L2 phonological acquisition are also discussed, among them constraints on pronunciation mastery, the relationship between production and perception, the role of feedback, issues related to transfer from L1, and the effect of different types of pronunciation training. Since past work fell short, both theoretically and methodologically, of describing and teaching suprasegmentals, Chapter 5 discusses the research agenda for the present and future. Recent work in second language acquisition and applied linguistics, along with the trends in English teaching toward a discourse-based, top-down approach to teaching pronunciation and toward communicative proficiency in language pedagogy, strongly support (at least in theory) the teaching of discourse intonation. Finally, inherent problems in teaching intonation in the classroom are also noted in the introduction to Part III.

To see how this pragmatic challenge has been met in the past and to help correct problems in teaching, the introduction to Part III first briefly surveys the treatment of intonation in textbooks and handbooks for L2 learning and teaching in the last several decades, contrasting traditional methods of teaching pronunciation with the innovative ones which include intonation. In addition, it provides suggestions for a framework for practical applications of the linguistic and applied linguistic theories of intonation. An initial goal is to sensitize language instructors, many of whom may not themselves be *consciously* aware of how intonation functions in language, to the importance of including this aspect of pronunciation in their syllabi. In its most modern framing, teaching discourse intonation means teaching aspects of pronunciation beyond segmentals in order to make intonation serve larger goals within discourse. The first goal concerns the *production* of suprasegmentals in L2 and has a communicative emphasis. It is not sufficient just to get the individual sounds of L2 right, it is also necessary (1) to keep thought units or idea units together when speaking, (2) to provide markers for focus, given vs. new

information, "propositional attitude," and turn-taking cues in discourse, and (3) by providing such markers, to help the hearer parse, process, and understand what is said. The second goal thus involves developing students' *perception* or listening comprehension, i.e., enabling speakers (1) to become aware of how they are "coming across" to others and (2) to become attuned to listening for suprasegmental "cues" as produced by native speakers instead of trying just to understand words (individual lexical items) or sentences (syntax). To help teachers achieve these goals, Chapter 6 suggests exercises for practicing rhythmic patterning and stress placement in English at all levels (syllable stress, word stress, sentence stress, and discourse emphasis). Chapter 7 focuses on ways of practicing intonation, first for *grammatical* purposes (ranging from one-word sentences to complex sentences), then to convey *attitudinal* functions (various emotions and attitudes), and, finally, to signal different kinds of *discourse*, *pragmatic,* and *sociolinguistic* functions. These functions correspond to the elements of communicative competence that are currently being emphasized in second language acquisition.

In the last two decades, much has been published on intonation theory, on models of intonation for many languages and on the acoustic measurement of intonation and its various components. In addition, there are many handbooks on teaching pronunciation in L2, several of which address suprasegmentals and suggest ways of teaching them (for English, cf. Avery & Ehrlich, 1992; Bradford, 1988; Brazil, 1994; Celce-Murcia et al., 1996; Dalton & Seidlhofer, 1994; Dauer, 1993; English, 1988; Gilbert, 1984, 1993; Hagen & Grogan, 1992; Hahn & Dickerson, 1999; Henrichsen et al., 1999; Hewings & Goldstein, 1999; Kenworthy, 1987; Morley, 1992; Orion, 1988; Pavlik, 1986; Rogerson & Gilbert, 1990; Sheeler & Markley, 1986; Wong, 1987). However, while there is an ever-growing number of theoretical linguistic books and articles on discourse intonation (cf. Bolinger, 1989; Brazil, 1975, 1978; Brown et al., 1980; Couper-Kuhlen, 1986; Couper-Kuhlen & Selting, 1996a; Ford & Thompson, 1996; Johns-Lewis, 1986a; Liberman & Pierrehumbert, 1984; Pierrehumbert & Hirschberg, 1990), there are as yet no books on the theory of discourse intonation as it interfaces with applied linguistic theory and the pedagogical implications — with one exception: Brazil et al. (1980). To supplement Brazil et al. (1980), this book will explain both the theoretical linguistic bases of discourse intonation and their application to second and foreign language acquisition theory and also provide practical applications for language teachers who are eager to help their students gain greater oral competence and proficiency as speakers and hearers of other languages.

I propose viewing intonation not from a structuralist or generativist standpoint, merely using minimal pairs to show dichotomous contrasts of phonemes, but rather basing pedagogical models of intonation on naturally and authentically occurring speech, interactions, and conversations. There is no one-to-one correspondence between form and function; rather, intonation must be viewed and interpreted from the context in which it occurs, i.e., is spoken. It has a signaling function within discourse and is part of a real-time, ongoing process of interaction in which speakers react to the way their interlocutor is using intonation (pitch, rhythm, timing) and "conform" to it or "break away" from it. Speakers and recipients cooperate to avoid conflict and to resolve conflict when necessary. With the fields of both foreign language pedagogy and prosodic linguistics moving in parallel with regard to communicative/interactional competence and the interactive functions of intonation, our concern should be to integrate the main functions ascribed theoretically to intonation with the main elements of communicative and interactive competence being addressed in L2 instructional materials and curricula.

Part I

Linguistic theory: Intonation in L1

Over the last twenty-five years, the study of intonation has enjoyed a solid revival, and great strides have been made in both the theory and measurement of intonation (see Botinis et al., 2001, for a summary of recent developments and paradigms in intonation research and Hirst and Di Cristo, 1998, for an excellent survey of the intonation systems of twenty languages). The purpose of Part I of this book is to provide a concise summary of the underlying linguistic foundations of intonation for the non-specialist in order to form the basis from which applied linguists and language teachers can proceed, i.e., to make it possible to understand how speech "melody" functions in human communication and why it should be taught to language learners.

A cogent overview of what a theory of intonation should contain is provided by 't Hart et al. (1990) [my italics]: "…intonation can be approached from a variety of angles, all of which are equally indispensable if one wishes ultimately to understand how speech melody functions in human communication. Ideally, a theory of intonation should comprise a *phonetic* and a *linguistic* component. The phonetic part of the theory should account for the *physiological*, *acoustic,* and *perceptual* aspects of intonation, and elucidate the relation between them. The linguistic component of the theory should aim at a *phonological* interpretation of the phonetic facts and at a *pragmatic* explanation of how intonation functions in the communicative interaction between speaker and listener. Finally, the theory should comprise a natural link between the linguistic and phonetic components: it should clarify how the melodic performance of the language user results from the interaction between his communicative intent and the peripheral means of his vocal and perceptual apparatus" (p. 2).

Chapter 1 thus describes the phonetic component and is divided into two sections: the first defines the technical linguistic terminology used in discussions and descriptions of intonation, specifically, the *phonetic* and *acoustic* components. The second section reviews briefly some of the *perceptual* aspects of intonation, namely the relationship between the physical or acoustic parameters and how these parameters are perceived by listeners. Chapter 2 then

describes the linguistic component by presenting the major schools of thought and theories of intonation, i.e., the *phonological* interpretations of the phonetic "facts." Chapter 3 links the linguistic and phonetic components by examining the *functions* of intonation. First, the *grammatical* functions ascribed to intonation are discussed; then the *attitudinal* and *affective* functions of intonation are described. Next, the *discourse* and *pragmatic* functions of intonation are examined. These functions refer to the communicatively relevant information not explicitly contained in the lexical and syntactic elements. Lastly, the *sociolinguistic* functions of intonation are discussed. These four categories of functions correspond to the elements of competence being cultivated in second language learners — issues to be taken up in Part II of the book.

Chapter 1

Intonation, suprasegmentals, prosody

Definition of terms

Intonation is a term often used interchangeably with several other terms, in particular *prosody* and *suprasegmentals*. Intonation is often referred to in everyday language as *speech melody* or *sentence melody*, terms that focus on pitch variations and modulations. These and related terms will be defined in this chapter. (See also Hirst & Di Cristo, 1998, Chapter 1 for a concise discussion of terminology.)

Suprasegmentals is perhaps the broadest term. *Suprasegmental features* are phonological units that stand in contrast to so-called *segmental features* or simple sounds, i.e., consonants and vowels. Whereas each segmental feature is considered to be an entity in itself and sequences of them are strung together in an utterance, a suprasegmental feature typically extends over more than one sound segment in an utterance, over longer stretches of speech. For example, features such as pitch, tempo, and rhythm extend beyond a single vowel or consonant to syllables, words, and entire sentences.

Prosody has been defined as a "continuum of functions and effects, ranging from the nonlinguistic or extralinguistic at one end, through the paralinguistic, to the essentially linguistic. At the nonlinguistic end, for example, are features of voice quality that reflect the nature of the speaker's larynx and the vocal tract; at the linguistic end are features such as stress and tone, which are functional within specific linguistic systems" and often vary widely in function and use from language to language (cf. Clark & Yallop, 1990, p. 276). However, there is a gray area between these two extremes; for example, the paralinguistic features of a particular speaking style may be due to the speaker's physiology or anatomy as well as to a conscious and deliberate (and therefore communicative) effort by the speaker to convey a certain personality, affect, or attitude. Bolinger (1985) suggests that intonation is inherently iconic and is part of a gestural complex that includes facial gestures such as raised eyebrows, widened or narrowed eyes, and smiles or scowls as well as other bodily gestures such as raised shoulders, hand movements, and clenched fists. In this chapter, I focus

on the linguistic end of the continuum, i.e., on systems such as intonation, stress, and tone manifested in such features as pitch, loudness, and duration. In later chapters, the focus will be on the communicative and discourse functions of prosody.

Although there is not complete agreement as to the precise domains of *suprasegmentals* and *prosody*, in this book the two terms will be used interchangeably as more general cover terms for the non-segmental components of speech, i.e., those over and above the segmentals (vowels and consonants). The term *intonation* has been defined as referring to "the distinctive use of patterns of pitch, or melody" (cf. Crystal, 1985, p. 162). This narrowing of the definition or scope of intonation to encompass those aspects of the speech signal that involve pitch or fundamental frequency is widely accepted and is adopted in this book.

The phonetic and acoustic basis of suprasegmentals

The three features that are most consistently used for linguistic analysis and description of prosody are *pitch, length,* and *loudness.* The physical properties of these features are termed *frequency, duration,* and *intensity*, respectively. Both the physiological (production-related) and the acoustic (both production- and perception-related) correlates of these features will be examined below. The perceptual aspects involve human psychological responses to acoustic stimuli.

Pitch

The term *pitch* refers, as in everyday language, to the varying level or height of the sounds produced in speech. Pitch is produced or created physiologically by the vibration of the vocal folds (vocal cords) within the larynx. This frequency of vibration, i.e., the number of times that the vocal cords completely open and close in a given period of time, can be measured acoustically. The acoustic (or physical) correlate of vocal fold vibration is the *fundamental frequency* or F_0 of the sound wave generated at the glottis, measured in hertz (Hz) and also referred to as "cycles per second." The fundamental frequency among male speakers on average varies between 60 Hz and 240 Hz; among female speakers, the range is between 180 Hz and 400 Hz. The average fundamental frequency for men is approximately 120 Hz, for women 225 Hz, and for children 265 Hz (cf. Cruttenden, 1997, p. 3).

While fundamental frequency involves acoustic measurement of what is produced physiologically by speakers, pitch usually refers to how fundamental

frequency is perceived by listeners, i.e., whether a sound is "high" or "low," whether one sound is "higher" or "lower" than another and by how much, and whether the voice is going "up" or "down." Various factors influence the perception of pitch, including loudness and duration. Pitch may also be altered by phonetic or segmental factors such as vowel quality or the nature of the surrounding consonants. Only voiced sounds have a repetitive waveform and hence may be perceived as having pitch. Voiceless consonants have no fundamental frequency, but they do have an effect on adjacent voiced sounds; for example, vowels have a higher fundamental frequency when preceded by voiceless consonants than when preceded by voiced consonants. It is also the case that different types of vowels have inherently higher and lower fundamental frequencies; open vowels, for example, tend to have a lower fundamental frequency than close vowels. However, in terms of pitch differences that are perceived by hearers, it is only duration and loudness, not vowel type or adjacent consonant, that carry intonational functions and thus influence perception.

Briefly, one of the main functions of pitch, at least in English, is to mark prominence. That is, when a syllable or word is perceived as "stressed" or "emphasized," it is pitch height or a change in pitch, more than length or loudness, that is likely to be mainly responsible (cf. Fry, 1958; Fudge, 1984; Gimson, 1980, pp. 222–226; Lehiste, 1976). It is generally agreed that the three features of pitch, length, and loudness form a scale of importance in bringing syllables into prominence, with pitch being the most significant, duration next, and loudness the least important factor (cf. Cruttenden, 1997, p. 13).

Length

Phonetically, *length* refers to the physical *duration* of a sound, but phonologically, it refers to the **relative** duration of sounds and syllables when these are linguistically contrastive. Depending on the language, both vowels and consonants can be phonologically "long" or "short." That is, in absolute terms, for most languages, there is no set duration or range of duration for a particular sound, e.g., one cannot say that in English, the long-*a* sound is 50 milliseconds (ms) while the short-*a* is only 25 ms in duration.

Acoustically, the duration of individual sound segments can be measured on a waveform or a spectrogram, but the criteria are not always simple for determining precisely where one sound segment ends and where the next begins. Similarly, how length or duration is perceived by listeners can also be quite complex. For the purposes of this book, however, specific details on determining the precise duration of sounds are unnecessary because our con-

cern is not the acoustic or phonetic properties of segmental duration but rather the suprasegmental functions of relative length. In addition, there are other influences on the absolute duration of a segment or syllable, e.g., whether there is an "innate" length for vowels in a given language, and the effect of so-called pre-pausal lengthening, i.e., the phenomenon by which the last syllable before a pause is lengthened. In other words, the same sound or syllable can vary in length depending on the neighboring sounds, whether the syllable a sound is in is stressed, or whether the syllable occurs immediately before a pause. As noted above, it is generally agreed that duration plays a somewhat greater role in signaling stress or accent than loudness, which is discussed below.

Loudness

The everyday meaning of the term *loudness* extends to the linguistic use of the term. It is an auditory phonetic feature whose acoustic correlate is *intensity*, which is a physical characteristic of sound that is related to the amount of energy present in the production of a sound. This acoustic feature, measured in decibels (dB), is based on the size of the vibrations of the vocal cords as a result of variations in the pressure of air coming from the lungs. However, there is no direct or linear correlation between loudness (or volume) and intensity: factors other than intensity may affect a hearer's perception of loudness. Although loudness is the subjective property of a sound that is most directly related to intensity, intensity also depends on the fundamental frequency and the duration of the sound as well as on its spectral characteristics. For example, open vowels are acoustically of greater intensity than close vowels. As mentioned earlier, although loudness contributes to the perception of stress or emphasis, it is probably the least salient feature of the three involved.

Other dimensions of prosody

Stress, accent, and prominence

In the previous sub-sections on pitch, length, and loudness, the articulatory and acoustic correlates of these three phonetic features were discussed, particularly with regard to how they function to make some syllables more prominent in words and in sentences. Such prominence is often referred to as *stress.* In a language like English, where words have their own stress pattern, the term *stress* is often used to mean *word stress* or *lexical stress* (e.g., word stress on the first or second syllable of *content* differentiates the minimal pair *cóntent,* the noun meaning "that which is contained," from *contént,* the adjective meaning

"satisfied"). The other term often used interchangeably with *stress* is *accent.* A brief (and perhaps oversimplified) summary of common definitions of these terms, as well as how they are used in this book, now follows.

The term *stress* has been used mainly in two different ways in the literature. First, it has been used to refer to actual syllables (or vowels) in words or sentences that are made prominent for communicative purposes. Second, it has been used to refer to the degree of force used in producing a syllable or loudness. In other words, on the one hand, it refers to the entity that is stressed (syllable or vowel), and on the other, to the means of making a syllable prominent (popularly termed *loudness*). However, this second sense is misleading if loudness is the only factor considered, since loudness (as discussed above) has been shown to play only a minor role in producing prominence, whereas pitch and length play greater roles. In the cases in which *accent* refers to the prominence given to certain syllables, such prominence is principally associated with pitch and is thus designated by many as *pitch accent* (cf. Cruttenden, 1997, p. 13; Crystal, 1985, pp. 288–9).

In this book, I adopt the usage in Cruttenden (1997) by which *stress* in the more general, less specified way refers to prominence in which a change in pitch is not the primary factor. *Accent* will refer to prominences in which pitch is involved, making it equivalent to *pitch accent.* Although these definitions might suggest simple, straightforward explanations, these terms are notoriously ambiguous, and in Chapter 2 on the different theories of intonation, it will become evident that the complexities of the various proposed intonational systems make it difficult to set forth simple definitions. For the moment, however, I will proceed to define the other basic components of prosody.

Rhythm

As suggested above, patterns of stress are important in languages like English because — first, they distinguish one word from another, and second because the *rhythm* of the language is determined largely by strong "beats" falling on the stressed syllables of phrases and sentences. The most general definition of rhythm is thus the perceived regularity of prominent units in speech. Although these regularities are sometimes described in terms of patterns of syllable *length* (long vs. short) or *pitch* (high vs. low), rhythm is most commonly discussed in terms of patterns of *stressed* vs. *unstressed* syllables (cf. Crystal, 1985, p. 266). To take the example of *content* again: the noun *cóntent* consists of a strong, or stressed syllable followed by a weak, or unstressed syllable. The adjective *contént,* on the other hand, consists of a weak (w) syllable followed by a strong

(s) syllable. Further examples include pairs such as *ímport/impórt, éxport/expórt, próduce/produce*, where the first word in each pair is a noun and the second a verb. Words in English are built from metrical *feet*, and a metrical *foot* is a grouping of one strong syllable followed by any number of weak syllables Utterances are said to consist of a number of these feet or rhythmic units, each of which is dominated by the beat of the stressed syllable (cf. Clark & Yallop, 1990, p. 287).

A typical spoken utterance of English thus consists of a number of *feet*, each of which is dominated by the beat of its stressed syllable. In speech at a normal tempo, unstressed syllables are greatly reduced and form a tail of varying length in each foot. In English, there is a tendency to equalize the duration of each foot. This means that each foot, regardless of whether it consists of a single syllable or several, will tend to take roughly the same amount of time to produce. Thus, for words consisting of a single foot, production of a one-syllable word such as *cheer* is approximately the same duration as that of the two- or three-syllable words *cheerful* and *cheerfully*. This so-called *isochrony* (equal timing) based on stress has led to the frequent description of English as a *stress-timed* language. That is, in English (and in German, Russian, Arabic, and many other languages as well), as a general rule, speakers take an equal amount of time moving from one stressed syllable to the next. In other words, in stress-timed languages, stressed syllables are said to occur at approximately regular intervals of time, regardless of the number of intervening unstressed syllables. However, isochrony in absolute terms has not been experimentally verified.

By contrast, many of the world's languages — among them French, Spanish, Japanese, Italian, Turkish, and Yoruba — are *syllable-timed*. In syllable-timed languages, it is the syllables (as opposed to the stresses) that occur at regular intervals. That is, unlike in stress-timed languages, in syllable-timed languages there is no strong pattern of stress, at least not to the extent that unstressed syllables are markedly reduced or compressed. In such languages, the total duration of a word or an utterance is therefore dependent more on the number of syllables it happens to contain than on the number and position of stressed syllables (cf. Clark & Yallop, 1990, p. 287). For example, as will be shown in Chapter 6, individual syllables are longer in Spanish words than in English or German words (see Figure 16 comparing the duration of the word meaning "chocolate" in these languages).

Sentence stress, emphatic stress, contrastive stress

Under discussion in the previous sub-section was the rhythm of speech, i.e., the stress patterns of individual words and larger units of utterances. In examining the prosodic or intonational patterning of spoken language as a whole, it is necessary to go beyond the word, beyond word or lexical stress, to phrases and sentences. To be sure, in order to describe intonation, it is first necessary to know which syllables in words are stressed in order to know which syllables are potentially stressable in utterances. In addition, however, it is necessary to know which syllables are stressed in utterances because such syllables form the framework for intonational patterns. In English, for example, the phrases *the table* or *a party* or *leave it* normally have the stress pattern of single words, with only one prominent syllable. The prominent syllables are found in the so-called content words (i.e., nouns, verbs, adjectives, and adverbs), whereas the so-called function words are characteristically unstressed. Although there are usually a limited number of function words in a given language, they are generally of very high frequency, including articles, prepositions, pronouns, and conjunctions, such as (in English) *the, a* (articles); *at, for,* and *to* (prepositions); *he, her,* and *them* (pronouns); *and, than,* and *that* (conjunctions).

However, these normally unstressed words may be stressed in certain contexts, and the stress is then meaningful — in contrast with the normal or unmarked pronunciation — as in the following example, where the word *was* is unstressed in the first but stressed in the second utterance (cf. Clark & Yallop, 1990, p. 296):

ˊ ˇ ˊ ˇ Joe was angry (two feet, each with stress on the first beat, i.e., on 'Joe' and on 'an-')

ˊ ˊ ˊ ˇ Joe WAS angry (three feet, with each of the first two syllables constituting a separate foot).

The second reading signals that the speaker is contradicting a previous statement or implying that Joe was angry but no longer is.

What the above pair of utterances demonstrates is simply that inherent lexical or word stresses, i.e., the normal rhythmic or accentual pattern of a word, may be overridden by what are called *sentence stress, emphatic stress,* and *contrastive stress.* (As noted earlier, some use the term *accent* to refer to relative prominence within longer utterances, but I have chosen to use *accent* synonymously with *pitch accent* and to use *stress* to refer to prominence at any level, i.e., word-, sentence-, or discourse-level.) From a phonological standpoint, the main function of stress is to provide a means of distinguishing degrees of

emphasis or contrast in sentences. This overriding of lexical or word stress can occur in one of two ways and for one of two reasons: (1) a stressed syllable, already prominent within the normal rhythmic pattern determined by word stresses, can be further accentuated relative to other stressed syllables for *emphatic* purposes, or (2) as seen in the example above with the normally unstressed word *was*, although the characteristic pitch prominence in English utterances usually falls on syllables that are potentially stressable by virtue of word stress patterns, in certain cases or contexts, a syllable or word that does not normally receive lexical stress may be stressed for *contrastive* purposes (see Chapters 2 and 3 for further explanations of the meaning and function of these different types of stress).

The perceptual aspects of intonation

The preceding sub-sections summarized the phonetic and acoustic aspects of prosody. At various points above, a third aspect, perception, was also touched upon. In dealing with the perceptual aspects of intonation from a phonetic, i.e., articulatory (physiological) or acoustic, perspective, two main aspects are relevant: first, the human ability to perceive the physical properties of frequency, duration, and intensity, and second, the psychological response to various acoustic stimuli. Almost independent of the linguistic theories of prosody, there has been a long history of the study of sound perception in the field of "psychophysics of sound." While psychoacoustic and psychophysical research and approaches focus on the parameters and limits of human perception, linguistic approaches focus on meaningful units that listeners are able to perceive or learn to perceive.

Research has found that human listeners, starting virtually from birth, attend and react to speech sounds differently than to other acoustic signals. It is suggested that the human brain has a number of devices or property detectors that respond selectively to particular types of acoustic signals and perceive speech signals as a mode different from other sounds. Many of these neural property detectors respond in particular to signals that the human vocal tract makes (cf. Lieberman & Blumstein, 1988, pp. 148–150).

In brief, starting with pitch or frequency, the range of human hearing is generally postulated to be between 20–20,000 Hz, and for language in particular between approximately 40–16,000 Hz. For the perception of intensity, the range is approximately between 0–140 dB; whispered speech is generally be-

tween 30–40 dB, "normal speech" between 40–80 dB, and loud rock music about 100 dB. "Differential thresholds" for frequency, intensity, and duration have also been studied, i.e., the smallest changes that can be detected by human hearing. For frequency, 0.1% changes in frequency can be heard, e.g., with synthetic speech, for a tone around 1,000 Hz, listeners can detect a 1–2 Hz difference. In a study on intensity using pure tones, it was found that listeners could detect a difference of 0.5–1.0 dB, no matter what the frequency. As for duration, the smallest gap between two sounds that could be detected was 2 ms.

Lifschitz (1933) found that human judgments of loudness are a function of the duration and amplitude of sound, not solely of the physical intensity of the speech signal. Similarly, other psychoacoustic tests cited by Lieberman and Blumstein (1988, p. 154) have shown that the factors that influence how human listeners make stress distinctions include the total fundamental frequency (pitch) contour of the utterance, the amplitude (loudness) of the syllabic "peaks," and the relative durations of segments of the utterance.

Although pitch, length, and loudness are all important components of prosody in speech, their perception does not correspond linearly to how each is physically produced or realized. For example, there is not a one-to-one relationship between frequency and pitch, i.e., a tone that is judged to be twice as high as another tone does not necessarily have twice the Hz value. Neither is there a one-to-one mapping between intensity and the perception of loudness. For example, a sound that is twice as strong in intensity as another sound is not necessarily perceived as being twice as loud. In addition, the perception of pitch is influenced by intensity in a nonlinear way: sounds at different frequencies with the same intensity are often perceived as having different loudness. With respect to the interaction of duration and intensity, it was found that longer sounds are not only easier to detect than shorter sounds but will also be heard as louder (cf. H. Gopal, personal communication, 1994).

In sum, although some previous psychophysical research has been able to isolate the individual elements of fundamental frequency, duration, and amplitude and to determine listeners' abilities to perceive them, psychoacoustic research on the suprasegmentals pitch, length, and loudness has suggested that listeners use a combination of these components in order to make optimal perceptual judgments. As mentioned earlier, Cruttenden (1997) found that "Perceptual experiments have clearly shown that, in English at any rate, the three features [pitch, duration, and loudness] form a scale of importance in bringing syllables into prominence, pitch being the most efficacious, and loudness the least so" (p. 13).

The importance of the perceptual aspects of intonation in speech is often overlooked, in part because in studying intonation one seems to be faced with the dilemma of either choosing a linguistic, i.e., phonological, approach at the risk of overlooking phonetically important features, or, if opting for an instrumental-phonetic perspective, of perhaps missing communicatively relevant features. 't Hart et al. (1990), in an attempt to solve this dilemma, focused on perception instead of production and proposed a "model of the listener" to answer the general question "What does the listener make of pitch in speech?" That is, they investigated the questions of "which melodic units the listener distinguishes, how he structures them to the overall percept of a pitch contour, how he relates perceived contours to more abstract melodic entities (intonation patterns), how he integrates melodic and textual information into one linguistic message" (p. 4).

This is not to deny the importance of speech production and acoustics, but rather to suggest that auditory impressions should be studied systematically as they are related to the properties of the acoustic signal and particularly to those that result from voluntary physiological action on the part of the speaker. Thus, the central question is not so much "What melody do we perceive?", but rather "Which properties of the acoustic signal are relevant for our perception of speech melody?"

't Hart et al. (1990) concluded that communicative *function* is of paramount importance in imposing structure on the phenomenon of pitch cues in speech — that it is more important than *form* as in traditional structural or phonological approaches (p. 69). This chapter will now be concluded by summarizing two studies on listeners' abilities to perceive and discriminate different intonational *forms* using synthesized speech; studies of the different *functions* of intonation that listeners are able to perceive and ascribe to particular acoustic components of natural speech will be discussed in Chapter 3 on the meaning and function of intonation.

Gårding and Abramson (1965) sought to find some American English intonation contours that could be identified by native speakers on the basis of fundamental frequency alone. They tried to determine the extent to which the pitch-curve of a given contour could be synthetically changed before being perceived as different. To this end, the fundamental frequency movements of five American English intonation contours were varied in a systematic fashion, and subjects had to decide whether pairs of contours were "the same" or "different." The results of these tests showed that "each contour has a considerable margin within which changes can be made without any effect

on perception, as long as these changes do not disturb a certain pattern" (p. 75).

In another study using synthesized tone sequences modeled after the intonation contour of a short sentence, Scherer (1979) found that listeners attribute emotional meaning to (synthesized) auditory stimuli based on characteristic patterns of acoustic cues. He found that the parameters that seem to have the most influence on listeners' ratings are tempo (speed, rate) and pitch variation (p. 106):

> Moderate pitch variation leads to ratings of generally unpleasant emotions, like sadness, fear, disgust, and boredom… Extreme pitch variation and up contours produce ratings of highly pleasant, active, and potent emotions, such as happiness, interest, surprise, and also fear. Down contours have similar effects but do not seem to contain elements of surprise or uncertainty. Fast tempo leads to an attribution of high activity and potency as in the emotions of interest, fear, happiness, anger, and surprise. Slow tempo is seen as indicative of sadness, disgust, and boredom.

These representative studies of the perception of intonational contours suggest that fundamental frequency, or pitch, plays an important role in how hearers perceive prosody in speech and, in moving toward the linguistic end of the continuum, also how listeners impute speaker emotions on the basis of different acoustic cues.

In the next chapter, I discuss the linguistic component of intonation, moving from the traditional domain of phonetics (acoustics and perception as presented in this chapter) to the domain of phonology, that is, phonological interpretations of phonetic facts as set forth in various theories of intonation.

Chapter 2

Phonological organization of prosody

Theories of intonation

There are a number of theories about how the fundamental components of intonation should be described, whether as discrete elements or variable ones, wholes (tonal contours) or parts (tonal sequences). For English, historically, the main approaches, while both being form-based and generally restricted to sentence-level phenomena, have differed depending on which side of the Atlantic the research was being conducted. In Europe and particularly England, the contour analysis has been preferred: pitch patterns are described in terms of their overall shape, as *tone units, tone groups,* or *sense groups*, analyzed further into a head and a nucleus, both of which have a particular pitch contour. These approaches have typically been more "impressionistic," i.e., judgments about rising or falling contours have been made by listening to speech data. On the other hand, in the United States, so-called levels analyses, with increasing emphasis on instrumental analysis, have been favored. In many of these analyses, pitch patterns are analyzed in terms of sequences of tones or levels of pitch, i.e., as pitch phonemes and morphemes. A notable exception is the American Bolinger's (1951) theory of pitch accent, which resembles the British prosodic approach to some extent but differs in that it ascribes a major role to stress (or *prominence*). In recent years the focus has been on generative models of intonation, also form-based, that rely on metrical representations of text, indications of stressed vs. unstressed syllables, and on phrasal tunes represented by strings of two discrete tones, high (H) and low (L).

In contrast to the traditional form-based, sentence-level approaches, recent research in the U. S. and elsewhere has been taking a more functional, discourse-based approach, focusing on the role of intonation in natural communication and interaction based on the *intonation unit* (IU) as the basic linguistic unit (cf. Chafe, 1993; Couper-Kuhlen & Selting, 1996b; DuBois et al., 1992, 1993). In addition to current interest in function-based approaches, another major focus in recent years has been the attempt to integrate the more impressionistic, or *auditory*, approaches to intonation with instrumental, or

acoustic, approaches. An example of an acoustic approach is the so-called *psychophonetic* research that considers both physical or acoustic evidence in conjunction with perceptual judgments of linguistic relevance. Although the traditional literature tended to concentrate on generalized abstractions (for example, about typical intonation melodies or the functions of tones), advances in technology have enabled researchers to give much closer scrutiny to phonetic details (cf. Clark & Yallop, 1990, p. 278). In this chapter, the major theories of intonation will be summarized, starting with traditional, form-based tune and tone theories, moving to generative models (several of them integrated with instrumental or acoustic phonetic analyses), and concluding with discourse-based approaches. (For the history of studies in English intonation, cf. Couper-Kuhlen, 1986; Cruttenden, 1997; Crystal, 1969.) Based on a critical analysis of underlying theories, the final section of this chapter proposes a theoretical model upon which to base an integration of theory with language instruction, which in turn will be expanded upon in the third part of this book.

Prosodic or contour approach

The prosodic approach generally preferred by British linguists to describe British English can be divided into two sub-groups: (1) *tune analysis* or the *whole tune approach* dating back to Jones (1909, 1918) and Armstrong and Ward (1926) and (2) *tonetic analysis* or the so-called *nuclear approach,* which began with Palmer (1922) and analyzes tunes further into shorter prosodic segments. In other words, the first focuses on overall shape and the second on local changes.

Tune analysis posits a set of contrasting *tunes* or pitch configurations to describe intonation. The domain of these tunes, which are also called *tones*, is a *sense group* consisting of a word or a closely-knit group of words that express a thought. Sense groups are usually separated from each other by pauses and contain a single *nucleus*. The nucleus is the element in a group of words that receives the greatest stress or is given the most prominence; this is usually marked by the convergence of several suprasegmental features, generally pitch, intensity, and duration. Every sense group, no matter how long or short it is, has a tune pattern, which starts from the stressed syllable of the last prominent word in the sense group and extends over the remainder of the sense group. In tune analysis, two other elements associated with the intonation system are

pitch range and relative pitch height. Sweet (1890), in his analysis, proposed the following 5 tones, and their designations suggest that the pitch trajectory is the most important feature (p. 3):

> The tones are level (–); rising (´), as in ´<u>whot</u> 'what?'; falling (`), as in `<u>now</u> 'no!'; falling-rising or compound rise, as in <u>teik ˇkea</u> 'take care!'; rising-falling or compound fall, as in ^<u>ou</u> 'Oh!' as an expression of sarcasm. The tone-marks are put before the word they modify; if they modify a whole sentence, they are put at the end of it.

Although similar in many ways to the tune approach, tonetic analysis goes a step further and analyzes tunes into smaller segments. The term *tonetic* generally refers to the notational technique of writing tone diacritics in the line of text, before each stressed syllable, to indicate the type of pitch movement beginning on that syllable. Perhaps the most important characteristic of the tonetic approach is that the basic unit, the *tone group*, is sub-divided into two distinct categories, a *head* and a *nucleus*, both of which are characterized in terms of pitch contours. In the tune approach, the entire tone group or sense group consists of one tune. By contrast, tonetic approaches maintain that the head, which consists of all the syllables that precede the nucleus, should be characterized by a different kind of patterning than that of the nucleus.

Following the tonetic approach, O'Connor and Arnold (1961) distinguish further between a *phonetic* and a *phonological* level of analysis for intonation. The tunes, which are described according to an essentially phonetic system, are grouped into equivalence (or phonological) classes on the basis of how they function. These equivalence classes are called tone groups. A tone group is defined as "a grouping of tunes all conveying the same attitude on the part of the speaker" (p. 29). Thus, for O'Connor and Arnold, the main contribution of intonation is to express the speaker's attitude. However, many of the tone groupings show considerable redundancy, and there is no one-to-one correspondence between a tone group and a specific meaning, which demonstrates the difficulty of matching attitudinal labels with intonation contours. Of particular note here is that the number of tunes or tone groups is of secondary importance. What is rather of primary importance are the functions assigned or attributed to the different intonation patterns.

Below are six examples of sense groups in British English appearing in O'Connor and Arnold (1961), each containing the prominent word (nucleus) "two" followed by other words that are *not* prominent (p. 14). The important idea is that all tunes must end in one of the ways shown below. No matter how

long or short the sense group is, no matter how many prominent words it contains, the pattern of its tune from the stressed syllable of the *last* prominent word onwards will correspond to one or other of the six general patterns. Note that two types of notation are included: first, the traditional tonetic diacritics immediately before the nucleus "two," and second, O'Connor and Arnold's symbols of a large dot for the nucleus and small dots for the unstressed syllables — all shown graphically between two horizontal lines representing the normal high, and low limits of the voice.

Low Fall: ˏTwo, ˌPeter.

Hlgh Fall: ˥Two, you ˌsilly ˌchap.

Rise-Fall: ˄˥Two, inˌdeed!

Low Rise: ˏTwo, I ˈthink.

High Rise: ˊTwo, did you ˈsay?

Fall-Rise: ˬ˥Two, you ˌmean.

The tunes are described as follows (p. 7):

Low Fall: the voice falls during the word from a medium to a very low pitch.

High Fall: the voice falls during the word from a high to a very low pitch.

Rise-Fall: the voice first rises from a fairly low to a high pitch, and then quickly falls to a very low pitch.

Low Rise: the voice rises during the word from a low to a medium pitch or a little above.
High Rise: the voice rises *during the word* from a medium to a high pitch.
Fall-Rise: the voice first falls from a fairly high to a rather low pitch, and then, *still within the word*, rises to a medium pitch.

In contrast to O'Connor and Arnold, who group tones according to the attitude being conveyed by the speaker, Halliday (1963, 1967a) attempts to establish phonological categories on the basis of meaningful grammatical contrasts. The choice of pitch movement and other suprasegmentals is not only related to other grammatical choices, such as word order, use of verb tense or use of a negative morpheme, but the intonational contrasts themselves are at the same time distinct grammatical choices in their own right. Thus, according to Halliday, since *I'm GOing* and *I'M going* differ in meaning due to different words being stressed, the difference in intonation should be classified as a grammatical distinction comparable to the difference between *I'm going* and *I'm not going*.

Theoretically, there is no limit to the number of different pitch contours that can be produced, but not all the pitch variations in a language are significant. Halliday (1967a) lists five significant intonation choices and postulates a basic significance for the falling and rising tone movements (see Table 5.1 below). Falling tone 1 means that the polarity (truth value or validity) is known and stated, i.e., there is certainty with regard to *yes* or *no*; rising tone 2 implies that the polarity is unknown and information is desired. Falling-rising tone 4 means that something is known, but there is some doubt or reservation, and the opposite rising-falling tone 5 raises or notes doubt in order to dismiss it. The low-rising tone 3 implies uncertainty and that the assertion is conditional on something else.

Table 5.1 Significant intonation choices

Tone	*Symbol*	*Tonic movement*	*Terminal pitch tendency*
1	`	falling	low
2	´	rising	high
	ˇ	falling-rising	high
3	¯´	rising	mid
4	˘	(rising)-falling-rising	mid
5	ˆ	(falling)-rising-falling	low

(Halliday 1967a)

In addition to the general meanings outlined above, the following example illustrates how the tones can also be ascribed specific, at times even attitudinal, significance similar to the attitudinal meanings ascribed to intonation by O'Connor and Arnold. The tones can additionally provide speakers with five major options to indicate degree of involvement with the information and/or the interlocutor. The double lines // indicate the boundaries of the tone group, the number at the beginning of each line indicates the tone, and the syllable in italics is the tonic (cf. Coulthard, 1977, p. 100; Halliday, 1967a, p. 37):

// 1 I saw him *yes*terday // (neutral)
// 2 I saw him *yes*terday // (contradictory, challenging)
// 3 I saw him *yes*terday // (non-committal, disengaged)
// 4 I saw him *yes*terday // (reservation)
// 5 I saw him *yes*terday // (committed, involved)

The addition of attitudinal meanings has great potential for discourse analysis in that it consists of a small and therefore powerful set of categories linked to a set of general meanings that have to do in part with speaker involvement with the discourse partner(s) (cf. the section in this chapter on Discourse Intonation). These general meanings, when combined with individual clauses and lexical items, generate more specific meanings. However, one of the main problems with these types of contour analyses is simply the question of how many different pitch contours are significant in a given language. For example, distinctions are made between high rises and low rises, but how many types of rise (or fall) should be distinguished? Furthermore, how are contours actually associated with or mapped onto the words and syllables within the contour? For purposes of learning about intonation in either an L1 or an L2, will learners be able to decipher notational systems containing only symbols and not graphical representations of pitch curves? These questions are problematic both for theoretical accounts of intonation as well as for pedagogical applications to teaching intonation. One promising avenue of research is the acoustic analysis of intonation, which can be applied in several ways. First, quantitative measurement of phonetic tunes and tones in terms of pitch movement and pitch range could aid in their phonological classification, e.g., in determining how many types of rise or fall can be distinguished for a given language. Second, in terms of listener perception, listeners could be presented with different tunes and asked to classify them according to meaning, again contributing to determination of phonological groups for a given language. Third, graphic representations of acoustic data,

e.g., pitch curves, could help L2 learners to "see" the intonation patterns they produce and might be more useful than diacritics marking tunes or descriptions such as "low rise" or "fall-rise."

Bolinger's theory of pitch accent

If there is a common factor uniting the theories and descriptions of English intonation, it is their reliance on pitch as the distinguishing feature of intonational systems as opposed to other prosodic systems such as those of stress, rhythm, or juncture. However, the focus on this single feature is not as simple and straightforward as the British prosodic or contour approaches (discussed in the previous section) or the phonemic approaches (discussed in the next section) might suggest. Proponents of both the prosodic and the phonemic approaches acknowledge that some aspects of pitch patterning cross the boundaries between the traditional intonational systems and the stress or accent systems. Bolinger, for example, believed strongly that the prosodic systems are closely interrelated and was among the first to attempt quantitative acoustic experiments (cf. Bolinger, 1957a, 1958). Although Bolinger's (1951) theory of *pitch accent* for American English bears a number of resemblances to the British prosodic approach (e.g., he posited that pitch *configurations* are more meaningful than pitch-*level* sequences), it is usually considered apart from other theories, because stress (or prominence) plays a major role in his perspective, and I allot Bolinger a separate section in my discussion as well.

The term *pitch accent* refers to a prosodic element that is simultaneously a marker of prominence and a building block of intonation contours. Bolinger proposed that (1) pitch and stress are **not** phonemically independent, i.e., pitch (or pitch prominence) is the main cue to stress, and (2) since changes in stress affect intonational contours, intonational morphemes should be defined in terms of both pitch and stress, thus "pitch accents," rather than just in terms of pitch. When used alone, the term *stress*, in his view, should refer to the domain of word stress.

Bolinger (1958) distinguishes three kinds of pitch prominence or accents for American English intonation, labeled A, B, and C (p. 112):

> …it is not pitch *rise*, but rather pitch *prominence* that is essential to what we react to as stress. By prominence I mean a rapid and relatively wide departure from a smooth or undulating contour. A rise is only one kind of pitch prominence, though it is certainly the commonest kind.

Accent is said to be an "all-or-none" contrast; that is, if a subject were asked whether a pitch obtrusion or pitch prominence can be perceived, the reply would be "yes" or "no." Intonation, on the other hand, is a "more-or-less" phenomenon that encompasses the non-accentual aspects of pitch patterning, i.e., the kind of patterns that are said to be gradient rather than discrete in structure and use. Examples of gradient intonational patterning are: steep falls as opposed to gradual falls, the relative height of the accent peaks or troughs and terminal junctures, which also involve pitch movements. Thus, the primary cue of what is usually termed *stress* in an utterance is pitch *prominence*. In addition, length, loudness, rhythm, and vowel quality also contribute to manifestations of accent, though intensity is found to be negligible both as a determinative and as a qualitative factor in accent. Furthermore, delayed release of consonants (a form of extra lengthening) and breathiness also contribute to achieving and/or perceiving prominence.

Some thirty years after his 1958 paper, Bolinger summarized his work spanning more than four decades in two volumes, *Intonation and its Parts* (1986) and *Intonation and its Uses* (1989). He retains his three major accent types, now called *profiles*:"…shapes determined by how the pitch jump cueing the accent is realized, will be referred to as the *profile* of the accent" (1986, p. 139). His original Accent A, which is now Profile A, the most frequent shape, reflects the same general intonational shape as originally in 1958: "an intonational configuration whose distinguishing feature is an abrupt fall *in* or *from* the syllable that is made to stand out by the fall" (1989, p. 3). Similarly, Profile B, like Accent B, almost equally as frequent as A, is characterized by upward motion, being "marked by a jump up *to* the syllable that is made to stand out by the jump, with any following unaccented syllables usually continuing with a gradual rise but often staying level or even falling slightly" (1989, p. 3). Finally, Profile C, like Accent C, the least frequent of the major types, is the opposite of Profile A, "a kind of anti-accent A" (1958, p. 143); it is "marked by *down to* rather than *down from*" (1989, p. 4).

There are several additional profiles. Examples from Bolinger (1986) follow, illustrating his unique notation system of showing pitch patterns with the printed words themselves (p. 141):

> "Take the word *John.* The commonest shape is a relatively high pitch followed by a quick drop:
>
> ```
>          Jó
> It's        hn.          (Profile A)
> ```
>
> An almost equally common shape starts higher than a preceding pitch and does not fall:
>
> ```
>          Jóhn.           (Profile B)
> It's
> ```
>
> A less common shape is approached from above, and does not fall:
>
> ```
> It's
>          Jóhn.           (Profile C)
> ```
>
> A still less common shape starts at a relatively low pitch, goes up, and abruptly comes down again:
>
> ```
>             h
> It's       ´             (Profile CA)
>          Jo
>             n.
> ```
>
> A somewhat commoner one is like the first one above (Profile A) but ends in a rise:
>
> ```
>          Jó
> It's                     (Profile AC)
>            hn.
> ```
>
> The least frequent of all is a sort of combination of CA and AC, and is the limit of what one normally encounters on a single syllable.[1] It has the shape of a tilde; the *o* is repeated to show the display:
>
> ```
>            o
> It's                     (Profile CAC)
>          Jó  hn.
>             o
> ```

In keeping with his penchant for nontechnical simplicity and his trademark of illustrating pitch movement with the actual words of an utterance, Bolinger (1986) uses as a model for the shape of a "normal" utterance the conversational dyad, in which one person asks a question and another answers it (p. 46):

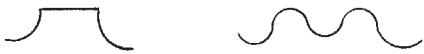
```
                it?
           like          lóve
Do you                I
                             it.
```

For longer utterances, approximately the same shape can be found where the second clause of a sentence "answers" the "question" in the first clause:

```
                it
          líke          trý
If you             then
                           it.
```

The first part is called the *theme* and the second part the *rheme.* This two-accent shape is "so typical that it has been designated the "hat pattern" (Cohen & 't Hart, 1967) or the "suspension bridge" (Bolinger, 1989) — shapes that usually

turn out to be lopsided in practice but nevertheless "exhibit the two corners of the hat or the two towers of the bridge" (Bolinger, 1986, p. 47):



In sum, what distinguishes Bolinger's theory from prosodic and phonemic approaches to intonation is that it posits the interdependence of pitch and stress and asserts that pitch is the main cue to stress. Although his theory of *pitch accent* involves pitch contours as in the prosodic approaches rather than pitch levels or phonemes as in the phonemic approaches (see following section), it differs from the prosodic approach in that it considers stress directly intertwined with pitch (cf. also Pierrehumbert's (1980) theory which posits *pitch accents* based on metrical phonology).

Phonemic or levels approach

With rare exceptions, the phonemic approach to intonation has been confined to American linguistics. This approach was modeled after the structural linguists' phonemic theories of segmentals with the goal of devising a system to describe intonation by a small number of apparently discrete pitch levels that would be tied to a system of stress levels as well as to junctures. Most American analyses have involved four pitch levels (1–4) and three terminal junctures (the pitch direction on the last syllable of an intonation-group) falling, rising, and level. Historically, two different perspectives can be identified: the early phonemic analyses, including that of Bloomfield and the pitch height or pitch level approach of Pike.

Bloomfield (1933) employed the concept of the phoneme when describing the use of pitch in language. He did not explicitly use the term intonation, but stated that "differences of *pitch* … are used in English, and perhaps in most languages, as secondary phonemes. The actual acoustic forms are highly variable" (p. 114). He used fairly primitive orthographic means to represent intonation (conventional punctuation marks such as periods, question marks, and commas), noting only the sentence-final changes in pitch trajectories. He distinguished five *phonemes*, which are pitch contours rather than pitch levels and are in some ways akin to the British tune approaches: (1) fall, marked by a period and used for statements; (2) rise, marked by a question mark and used for yes-no questions; (3) lesser rise, marked by an upside-down question mark and

used for wh-questions; (4) exclamatory pitch, marked by an exclamation point and used when a sentence is distorted as to pitch and also as to stress, such as in the intonation of anger, surprise, sneering, and the like; and (5) pause-pitch or suspension-pitch, marked by a comma and consisting of a rise in pitch before a pause within a sentence to show that the sentence is not ending.

Pike's (1945) pitch phoneme theory became the hallmark of American intonation analysis for the next two decades and beyond. Three of the most important characteristics of his theory are: (1) its use of pitch heights or pitch phonemes as the basic elements for characterizing intonation contours; (2) its use of a relatively systematic set of functions pertaining to speaker attitude; and (3) its recognition of the interdependent systems that coexist and influence intonation, namely stress, quantity, tempo, rhythm, and voice quality.

Pike (1945) divides intonation characteristics into several types, the two main ones being "colorless" contours, which give the listener no implication of the speaker's attitude or feeling, and those contours which do reflect the speaker's attitude or the attitude that the speaker expects the hearer to perceive. These contours are "affected or caused by the individual's physiological state — anger, happiness, excitement, age, sex, and so on" (pp. 20–21).

Pike found the descriptions rising, falling or falling-rising in the prosodic approaches insufficient. Rather, he found it necessary to specify the internal structures of these contours and to distinguish among variations within the same type of contour. He analyzed contours as sequences of four pitch height variants or pitch phonemes. These phonemes or levels are defined relatively, not absolutely, i.e., their significance is determined by their height relative to each another. The distance between the four levels is neither fixed, uniform, nor predictable; rather, it varies from individual to individual, and an individual varies her/his own intervals from occasion to occasion.

The four pitch levels are numbered from 1 to 4: 1 = extra high, 2 = high, 3 = mid, 4 = low. In addition, the symbol [°] is used to mark the beginning point of a primary contour. Primary contours are those contours with the strongest meanings, generally occurring at the ends of sentences. A stressed syllable constitutes the beginning point for every primary contour; there are no primary contours without a stressed syllable, and every heavily stressed syllable begins a new contour. Pitch levels are almost meaningless in themselves; rather, "it is the intonation contour as a whole which carries the meaning while the pitch levels contribute end points, beginning points or direction change points to the contours — and as such are basic building blocks which contribute to the contours and hence contribute to the meaning" (p. 26).

In Pike's transcription system, not every syllable is marked for pitch level — just those *contour points* that are crucial to establishing rises or falls. In any rising or falling contour, there are two contour points: the pitch level at its beginning and the pitch level at its end. In the first sentence below, which is an example of falling intonation, the first contour point is marked °2 and occurs on *want-* while the second contour point is marked 4 and occurs on *it*. In the second sentence, an example of rising intonation, the first contour point, marked °4, again occurs on *want-* and the intonation rises to level 1 on *it*.

He wanted to do it.
°2– –4 /

He wanted to do it?
3– °4– –1 /

With regard to the meaning of primary contours, Pike (1945) states "All falling contours have a meaning of *contrastive pointing*. The word or syllable which contains the beginning point of the primary contour is singled out as the *center of selective attention* of the speaker, or constitutes a demand by the speaker that the hearer focus his attention at that point… When placed before the final type pause, … contours falling to pitch four tend to have a further meaning of *finality*" (p. 44). This constitutes one of the chief differentiating characteristics between the group falling to pitch four and the group falling to pitch three. The group falling to pitch three has a nonfinal meaning. Compare the two sentences below: the second sounds incomplete, as if the person might later add, after a hesitation, "but couldn't":

He wanted to do it.
3– °2– –4 /

He wanted to do it (but couldn't.)
3– °2 –3 / 4– °2– –4 //

Pike's rising contours, on the other hand, generally imply that the speaker considers the utterances to be *incomplete* by themselves and that the speaker and possibly the hearer consider them to be in need of *supplementation* of some type by the speaker or by the hearer. For example, the first sentence below is a statement, to be supplemented by the second clause, and the second is a question to be supplemented by an answer from the hearer — or even by one from the speaker (p. 51).

He'll do it when he can.
4– °3–2 / 3– °3–4//

Should he do it?
4– °3 –2 /

There are various problems with this approach to describing intonation (cf. Cruttenden, 1997, pp. 38–39 for a discussion the main criticisms of the levels analysis). For linguistic theory in general and for language teachers in particular, the main problem is the fact that the different pitch levels are said to be *relative*. But the question then is "How relative?" If the levels are relative, then three or five levels would work equally well. If one assumes four levels, should the pitch range be divided up into four equal intervals? What are the phonetic details of going from one pitch level to another, e.g., where exactly does a rise in pitch from level 4 to level 1 take place? These difficulties represent one of the reasons for not adopting the phonemic approach in applying linguistic theory to the teaching of intonation. This is an area in which acoustic analyses, which provide quantitative measurements of pitch, may be useful in dealing with some of these problems.

Generative approaches

Parallel to the Chomskyan revolution against structuralist approaches to morphology and syntax, a generative approach was also introduced for phonology in contrast to the structuralist/phonemic approaches. Before discussing the generative approach to intonation in this section, the two main approaches to stress within generative phonology will be presented briefly, since, in the generative approach, stresses in words are converted into stresses/accents in utterances, and understanding the generation of patterns of sentence stress is necessary for discussing the generation of intonation patterns.

As with the principles of generative syntax, the basic assumption of generative phonology as set forth in Chomsky and Halle's *The Sound Pattern of English* (1968; SPE) is that stress need not be marked in the lexicon because it can be predicted by rule. The rules described in SPE are applied to the output of the syntactic component of a grammar, i.e., to a surface structure, to a sentence marked as to its constituents with labels such as N (Noun), V (Verb) or NP (Noun Phrase). The surface structure must then be divided into "phonological phrases," the equivalent of so-called "intonation groups." These phonological

phrases often coincide exactly with sentences. To oversimplify a bit, this model of *metrical phonology* to explain rhythm and stress is based on the idea of *metrical grids* (as opposed to that of *metrical trees* described below).

All segments (vowels and consonants) are initially marked [–stress]. At the word level, a rule then assigns [1 stress] or *primary stress* to particular vowels. Further conventions weaken previously assigned stresses and non-main stresses within a word. Above the simple word level, there are rules for compounds and the Nuclear Stress Rule for all constituents higher than the word. The Nuclear Stress Rule first assigns stress in the right element in phrases, then does so cyclically in all higher constituents.

For example, in the sentence [1]*Old* [1]*Tom* [1]*grows* [1]*roses*, each word will first get a primary stress as (now) marked. Then the noun phrase subject *Old Tom* and the verb phrase *grows roses* will each be reassigned a primary stress on their right element while the left element is downgraded to stress level 2, yielding [2]*Old* [1]*Tom* and [2]*grows* [1]*roses*. Finally, the noun phrase and the verb phrase are put back together to form the sentence, and stress is again (cyclically) reassigned, putting primary stress on the right element and downgrading the left elements to yield [3]*Old* [2]*Tom* [3]*grows* [1]*roses*.

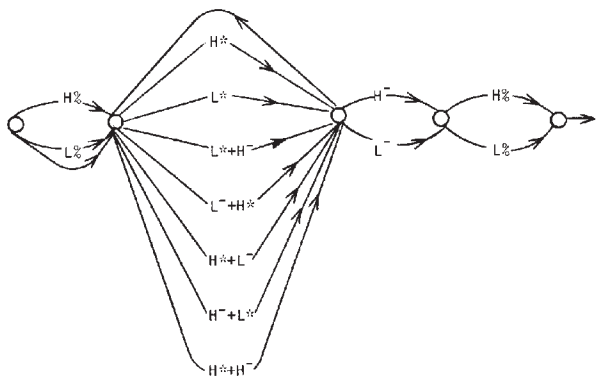
Liberman and Prince (1977) introduced an alternative approach to stress within generative phonology which forms the basis of generative models of intonation, such as the model of Pierrehumbert (1980). Liberman and Prince's theory of *metrical phonology* sought to eliminate the numbering of stress levels as in SPE and its inherent problem of "indefinite lowering," i.e., the problem that because previously assigned stresses are downgraded each time a new primary stress is assigned, there is in theory no limit to the number of possible stress levels. They propose instead a formal system in which stress is defined by *metrical tree* structures whose nodes are divided only *binarily* into **s** (strong) and **w** (weak) branches. As in SPE, the system can be applied at both the word and sentence levels. Also as in SPE, the assignment of strong and weak nodes is governed by two rules, the first for words and compounds (a Lexical Category Prominence Rule) and the second for the predicate and sentence levels (a Nuclear Stress Rule). One of the key notions in this theory is that stress levels are relative rather than absolute.

There is an extensive literature on metrical phonology that will not be taken into account here, in part because of a lack of consensus on which model to adopt, but mainly because a discussion here would unduly complicate matters for L2 instructors and learners. (Cf. Goldsmith, 1990, Chapter 4 for an introduction.)

Pierrehumbert's (1980) seminal monograph sets forth the now-standard generative model of intonation based firstly on the principles of metrical phonology (Liberman, 1975; Liberman & Prince, 1977), which label in an utterance the stressed vs. unstressed syllables, and secondly on a system of *tunes*, which are represented by a sequence of high (H) and low (L) *tones*. A key to both the theory of metrical phonology and Pierrehumbert's intonation theory is that they are both based on *binary* systems, **s**(trong) vs. **w**(eak) stresses and **h**(igh) vs. **l**(ow) tones. Pierrehumbert's structured strings or "tunes" consist of one or more *pitch accents* (which are aligned on the basis of the metrical pattern of the text) plus two additional tones which characterize the intonation of the end of the phrase, the *phrase accents* and the *boundary tones*. While the original model does not provide a detailed account of the meanings of tunes, it is concerned with specifying a set of rules that "translate" such tunes into actual pitch patterns. In other words, as with other generative models, the goal is to develop a system of underlying representation, in this case for English intonation, to link tone units to text according to metrical theory. Pierrehumbert (1980) asserts that the model "gives an account of what different tunes are possible and how they are aligned with different texts [and] characterizes the rules which map the underlying [phonological] representations into phonetic realization" (p. 2).

In other words, the goal is to formulate a grammar which generates the set of well-formed tonal sequences for an intonation phrase. The diagram below illustrates the different sequences possible for English (p. 29).

Boundary Tone	Pitch Accent	Phrase Accent	Boundary Tone



As shown in the above diagram, each intonational phrase consists of a sequence of H and L tones. There are four components of each intonational phrase or tonal sequence: (1) a boundary tone at the beginning, (2) a pitch accent, (3) a phrase accent, and (4) a boundary tone at the end. Pitch accents are realized by a single tone (marked H* and L*) or bitonally (marked H– + L*, L*+H–, H*+L–, L–+H*, and H*+H–). The starred tones H* and L* indicate the center of the accent, while L– and H– indicate "leading" or "trailing" tones. Phrase accents (also marked L– or H–) occur near the end of the word that contains the last pitch accent. They account for any movement in pitch immediately following the last pitch accent. Boundary tones (marked L% or H%) occur at the very beginning of a phrase and on the very last syllable, in the latter case taking care of any pitch movement on the final syllable of the phrase.

Simple examples of intonational phrases are taken from Pierrehumbert (1980, p. 25):

–What about Anna? Who did she come with?
–Anna % came with Manny.
| |
H* L– H% H* L– L%

–What about Manny? Who came with him?
–Anna % came with Manny.
| |
H* L– L% H* L– H%

Pierrehumbert, like Bolinger, uses the term *pitch accents* to indicate how pitch and stress are directly interrelated, and follows in the American tradition of describing intonation in terms of levels rather than tunes. Proponents believe that her model is preferable to the "traditional" model of Pike because it posits only two levels rather than four. With only two levels, it is said to avoid many of the problems encountered with more levels. For example, there is issue of relative vs. absolute pitch. In Pierrehumbert's system, pitch level or height is relative rather than absolute, as in the theory of metrical phonology of Prince and Liberman in which stress levels are not absolute but relative. The actual height of any H or L in Pierrehumbert's model is computed by reference to three things: its relationship to the baseline, the degree of prominence which the speaker opts to give it, and its relationship to preceding tones. This concept of relative pitch is felt to be preferable to a system with four levels of pitch because in such a system there are questions of how to divide the pitch range

of an intonation-group into four levels and whether these four levels then represent absolute pitch levels. Additionally, when dealing with absolute levels, there is the problem of how to explain the phenomenon that sometimes very small pitch movements convey significant differences in meaning, whereas in other cases larger pitch differences carry no unusual meaning. These problems are solved, according to Pierrehumbert, by her system of two levels of relative pitch.

There are, however, some theoretical linguistic problems with Pierrehumbert's model (cf. Cruttenden, 1997, pp. 64–66). One main criticism is that the model ignores the problem of intonation meaning but focuses on describing a system for generating typical contours that occur in English. (In later work, however, Pierrehumbert and Hirschberg (1990) address the meaning of intonational contours with regard to the interpretation of discourse — see next section). Another serious drawback of much of Pierrehumbert's work is that it relies on linguist-generated utterances and elicitations rather than on naturally occurring discourse or speech.

Taking Pierrehumbert's work for English as a point of departure, Féry (1993) gives a phonological account of German intonation, describing the tonal structure of German in a tone-sequence model. In addition, as is becoming increasingly important in intonation studies, she shows the influence of discourse factors on the tonal pattern of utterances, stating that "…the choice of the form of the realized accents, as well as the tonal melody associated with unaccented syllables, heavily depends on discourse structures like the focus-background division of the text […], topicalization […], modality […], and pragmatic factors" (p. 1). This inclusion of discourse factors in the study of intonation will be discussed in the following section.

Discourse intonation

A first step in discussing discourse intonation is to define the term *discourse*. Crystal (1985) defines discourse as "a continuous stretch of (especially spoken) language larger than a sentence… a set of utterances which constitute any recognisable speech event" (p. 96). Blakemore (1988) asserts that "The study of discourse belongs to the study of language in use" (p. 229), i.e., in order to understand fully what an utterance means, one must go beyond traditional sentence grammar and include pragmatic explanations. *Pragmatics* is the linguistic subdiscipline in which language is analyzed "from the point of view of

the users, especially of the choices they make, the constraints they encounter in using language in social interaction, and the effects their use of language has on the other participants in an act of communication" (Crystal, 1985, p. 240). Although much work has been done in the domains of both intonation and discourse, combining the study of intonation with the study of discourse has, until recently, been relatively rare — a circumstance that is one of the reasons for this book. In this section, the research to date on discourse intonation theories will be summarized.

Brazil (1975, 1978) was one of the first to use the term *discourse intonation*, and he, along with various of his British colleagues, went on to develop a theory of discourse intonation with reference to British English (cf. Coulthard, 1977; Coulthard & Brazil, 1981; Coulthard & Montgomery, 1981; Crombie, 1985, and others). Brazil (1975) introduces his theory of discourse intonation by stating that he remains "unashamedly concerned with function" and describes intonation in English as "a set of speaker-options formulated without explicit reference to grammar" (pp. 1–2). He believes that there is a small, finite number of functionally contrastive pitch configurations and that each of these configurations has its own meaning. "Meaning" in this context does not refer to attitudinal notions like "expectant" or "surprised" nor to grammatically derived concepts like "interrogative" or "declarative." Rather, what is important for Brazil is the continuous assessment of discourse by the speaker and a choice of one intonation pattern over another for the purpose of achieving coherence and cohesion in the discourse — in other words, the *interactional* significance of intonation. Brazil's theory thus differs from previous theories of intonation not in proposing a different set of components, but rather in ascribing different meanings and functions (ones that derive from usage in discourse) to more or less traditional components.

Brazil (1975) proposes no new phonological categories with regard to pitch but instead uses the five *tones* proposed by Halliday (1963) and a distinction of *key* that was first suggested by Sweet (1890). He offers a new interpretation of the significance of these features. Tone refers to the pitch *change* that characterizes the tonic segment of a tone group. Although Sweet used the term *key* to refer to the general pitch of the sentence or sentence group, Brazil employs the term *key* to refer to the pitch level of the tone group, a smaller unit than the sentence or sentence group. The five *tones* are: (1) falling-rising or *referring* tone **r**, (2) falling or *proclaiming* tone **p**, (3) rising or marked version of the fall-rise, **r+**, (4) rising-falling or marked version of the fall, **p+**, and (5) low rising.

An example is given by Brazil (1975) of two sentences that are syntactically identical on the surface, but that differ in subtle ways. The fall-rise tone in the first sentence shown below in (1.) marks the content of a tone group as part of the information shared by both participants; it is thus termed the *referring* tone, symbol **r**. The first sentence below would be addressed to someone who is expected to know already that the speaker is reading Middlemarch but does not know what the speaker's future intentions are. By contrast, the falling tone in (2.) marks the matter as new and is thus called the *proclaiming* tone, **p**. (Old information is "referred to" and new information is "proclaimed.") In the second sentence, the fact that the speaker wants to or will read Adam Bede is known information, and the new information is *when* it will be read (after finishing Middlemarch) (p. 6):

1. // when I've finished Middlemarch // I shall read Adam Bede //
2. // when I've finished Middlemarch // I shall read Adam Bede //

1. \/ (fall-rise) = **r** (referring tone)
2. \ (fall) = **p** (proclaiming tone)

For both of the above two tones, fall-rise and fall, there are "intensified alternatives," i.e., marked versions of the referring or proclaiming tones, **r**+ and **p**+, that signal an extra measure of involvement on the part of the speaker. The intensified referring tone is a rising tone. For example, as a reply to the question *Where are the glasses?*, the use of a simple referring (**r**) tone // **r** *in the CUPboard* implies "that's where they always are," whereas if the intensified **r**+ tone is used, as in // **r**+ *in the CUPboard,* the connotation is "why don't you ever remember…?"

1. // **r** in the CUPboard \/ (fall-rise) = **r** (referring tone)
2. // **r**+ in the CUPboard / (rise) = **r**+ (intensified referring tone)

In Brazil's notation, capitalization and underscores are used to designate *tonic syllables*, i.e., the syllable on which there is a major pitch movement.

The rising tone has traditionally been associated with questions or interrogative functions. In the two examples below, the first version, with fall-rise, is a straightforward or neutral question, while the second version, with rising intonation, has a more insistent or intensive connotation:

1. // **r** had he READ it? // \/ (fall-rise) = **r** (referring tone)
2. // **r**+ had he READ it? // / (rise) = **r**+ (intensified referring tone)

There is also an intensified proclaiming tone **p**+ that is realized by a rise-fall instead of the simple falling proclaiming tone **p**. By choosing the **p**+ tone, the speaker signals that new information is being added both to the common ground and to his/her own store of knowledge. In other words, the information is marked as doubly new: "I also didn't know" and thus "I'm surprised/disappointed/delighted." For example, in response to unexpected news, one might say "Really?" with a rising-falling tone:

// **p**+ REALly? // /\ (rise-fall) = **p**+ (intensified proclaiming tone)

The final tone, a low-rising tone, is used in neutral situations when the speaker does not want either to proclaim or to refer to anything. This tone has often been characterized as "uninvolved," "careful," and "patronizing." The speaker seems to withdraw or disengage him-/herself from the interactive situation.

In summary, what emerges are two different and independent dimensions of functional meaning expressed by tones: the *refer/proclaim* dimension and the *involved/uninvolved* dimension.

In addition to making choices in the tone system, a speaker must also select relative pitch or *key* for each tone unit. The category *key* is taken from Sweet (1890), who states that "each sentence or sentence group has a general key or pitch of its own." Three keys are postulated by Brazil — high, mid, and low — with the middle one generally left unmarked. Selection of high or low key involves raising or lowering the pitch of the whole tone group relative to a pitch that can be established as the norm for the speaker concerned.

With regard to key, speakers make a potentially meaningful choice to pitch each successive tone group *at, above,* or *below* the level which for them can be regarded as the baseline or norm. Any occurrence of a high-key tone group can be thought of as being phonetically bound to a succeeding mid-key tone group; it carries the implication "There is more to follow." On the other hand, any occurrence of a low-key tone group is bound to the preceding mid-key one, with the implication "This is said in a situation created by something that went immediately before" (cf. Brazil, 1975, p. 10). In a discourse, high key sets up expectations whereas low key implies prerequisites.

The three keys represent three options that a speaker has when embarking upon a new tone-group (i.e., the choice of key for the first stressed syllable of a tone group): the speaker can choose to keep it neutral (by using mid key) or can mark it "contrasting" (high key) or "equivalent" (low key). In general, high key functions to signal contrast, as in the example below where *Bognor* is in contrast to what the hearer might have expected (cf. Brazil, 1975, p. 13).

high BOGnor
mid // **p** we're going to MARgate this year // **r** not //
low

In addition, at the beginning of utterances, high key can function to mark off the ensuing part of the discourse from that which has preceded it. For example, high key can be used to start a new topic or to change the subject, both of which entail a type of contrast.

Low key, on the other hand, is used in situations in which a speaker assumes that there is "equivalence" between two successive tone groups in the discourse. Low-key tone groups that occur medially and finally in utterances are generally uninforming, as they often signal previously known information. In the example below, low key conveys the information that the speaker has only one set of neighbors, that *the Robinsons* and *our neighbours* are, in this context, synonymous (Brazil, 1975, pp. 15–16).

high // **r** eVENtually
mid // **p** we gave it to our NEIGHbours
low
high
mid
low // **p** the ROBinsons

In terms of the interactive functions of intonation, utterance-final key is of great significance because it sets up expectations and can thus influence the behavior of another party in an interaction. High key, for instance, can have the implication that there is more to follow or that the utterance is part of an incomplete structure, so the other party is thus constrained to respond in some way. Low key, on the other hand, in signaling previously known information or a potentially complete structure, can inhibit response because there is an implication of finality that has to be countered if the hearer is to continue the conversation. Mid key sets up no particular expectations. Brazil (1975) provides the following example (pp. 28–29):

A: mid // **p** where is he
low NOW //

high // **p** in BED//
B: mid

Here the high key of the response suggests that Speaker B presumes that Speaker A will consider the answer "in bed" a surprising, scandalous, thought-provoking, or otherwise comment-worthy piece of news. If Speaker B continued with an utterance in mid key,

	high	// **p** in BED//	
B:	mid		**r** if you ever HEARD of such a thing

the implication that A should respond would be removed. It is thus the final key of a tone unit that is significant. If low key were used for the original reply "in bed," further reaction would be discouraged because of the matter-of-fact finality implied. Thus, high key marks the response as "likely to be surprising" while low key would mark it as "only to be expected," suggesting that A probably knows or would guess the answer to the question. Similarly, if A's reaction to the response from B is itself in low key, this has the similar effect of inhibiting further response from B:

A:	mid	// **p**	where is he NOW //	
B:	mid	// **p**	in BED//	
A:	low			**p** I SEE //

In summary, although Brazil (1975) retains the major distinction between rising and falling tones found in the Hallidayan model (discussed above in the section "Prosodic or Contour Approach"), he interprets the tones differently and radically moves away from linking them to grammatical entities. Instead, he argues that intonational choices speakers make are motivated by their moment-to-moment, situationally-specific decisions to add meaning to particular words or groups of words. Thus, for example, intonation marks the prominence of situationally informative items. Tones mark pieces of information as to their status of how well-known they are: falling-rising or *referring* tones indicate that the information is common knowledge and can therefore serve as a basis for further discussion, whereas falling or *proclaiming* tones indicate that the information is new and therefore needs to be announced. Choices of high or low key signal contrastive and equivalent information, respectively. Choice of key at the ends and beginnings of tone units has interactive effects on the hearer: high terminations tend to constrain an interlocutor to respond, while mid termination does not set up any expectations and low termination, in effect, discourages or inhibits a reply.

Since Brazil's early work, a growing number of researchers have focused on discourse intonation; selected studies will be summarized below to provide an

overview of the issues involved in the study of discourse intonation. Johns-Lewis (1986a) is one of the earliest of the following books on the subject, a collection of papers primarily on British English intonation. The volume gives a sense of the breadth and focus of activity in the field, ranging across psychoacoustics, conversation analysis, discourse analysis, syntax and semantics, sociolinguistics, psycholinguistics, clinical linguistics, and social psychology. Both abstract model-building and measurement and experimentation are represented in the volume. The chief questions addressed in intonational model-building involve how to delimit the number and type of pitch categories. In Johns-Lewis' volume, measurement and experimentation are concerned primarily with the relationship between physical parameters and perceptual phenomena.

Some of the questions about the discourse functions of intonation whose answers could lead to clarification of theories of discourse intonation, as well as to issues for language instructors to consider, are: Which intonational cues mark prominence or stress? Which intonational cues mark *focus*, e.g., contrastive information and new information? Which intonational features signal speaker attitude and affective information? Which intonational features function as cues in the management of spoken interaction, e.g., interruptions, marking finality or non-finality of utterances and turns, asking for confirmation?

Some of the findings that provide initial answers relevant to the teaching of intonation include the following (cf. Johns-Lewis, 1986a, pp. xxi–xxii). Segmental lengthening is a marker not only of intonational prominence, but also of sentence finality, paragraph finality, and conversation-turn finality as well. Two other cues, *creak* (laryngealization) before a boundary and *pause length*, are associated with the perception of boundaries. As does lengthening, pitch phenomena (*range of pitch* and *pitch movement*) play a significant role not only in marking prominence but also in marking boundaries at the sentence level and at the discourse level. *Sentence declination,* the lowering of fundamental frequency or pitch across an utterance to mark the boundary of a sentence, has been studied in a number of languages. The *height* of tone units relative to each other (e.g., the *downstepped contour*) has been strongly associated with turn finality and may also be associated with topic finality. Features of pitch height, specifically relatively high *pitch peaks*, are proposed as markers of initiality in utterances, paragraphs, and topics. Fundamental frequency *range* is shown to be narrower in some kinds of speaking activity than others. For example, normal conversation is characterized by a narrower frequency range than in either reading aloud or oral delivery during acting; reading a dramatic dialogue aloud produces a wider frequency range than reading a written prose text

aloud. The important point here is that intonation has been shown to have definite functions at the discourse level — beyond the sentence level — and thus a theory of discourse intonation as well as its application in language teaching need to be articulated and should include such functions as signaling new topics, changing topic, and marking turn-finality or turn-beginning.

The study of intonation in discourse with regard to American English has recently flourished (cf. Hirschberg & Litman, 1987; Hirschberg & Pierrehumbert, 1986; Liberman & Pierrehumbert, 1984; Pierrehumbert & Hirschberg, 1990; Ward & Hirschberg, 1985). Hirschberg and Pierrehumbert (1986) propose that intonational features such as *phrasing, stress* or *accent placement, pitch range,* and *tune* provide important information about the *attentional* and *intentional* structures of discourse. *Phrasing* refers to how a complex utterance is divided up. *Stress* refers to the rhythmic pattern or relative prominence of syllables in an utterance. *Pitch range* is the distance between the highest point in the pitch contour and the baseline or lowest point. When a speaker's voice is raised, the overall pitch range is expanded. *Tune* is the abstract source of fundamental frequency patterns.

Pierrehumbert and Hirschberg (1990) expand the theory of metrical phonology with regard to the contribution that intonation makes to discourse interpretation. Their primary unit of meaning analysis is the *intonational contour* or *tune* (comparable to the *tone unit* of Crystal, 1969, the *intonation-group* of Cruttenden, 1997, or the *intonation unit (IU)* of DuBois et al., 1992). Pierrehumbert and Hirschberg propose that speakers choose particular tunes to convey relationships between the propositional content of a particular utterance and previous and subsequent utterances as well as relationships between the current utterance and knowledge and beliefs common to interlocutors in a conversation. The various intonational components convey how hearers are likely to interpret an utterance structurally and what the speaker believes to be believed by all parties in the discourse. They propose a *compositional theory* of tune interpretation: *tunes* are composed of *pitch accents, phrase accents,* and *boundary tones* (cf. discussion of Pierrehumbert, 1980 in an earlier section "Generative Approaches"). Pitch accents convey information about the status of discourse referents and of relationships specified by accented lexical items. Phrase accents convey information about the relatedness of intermediate phrases, particularly whether one intermediate phrase forms part of a larger interpretive unit. Boundary tones convey information about whether the current intonational contour is "forward-looking" or not (p. 308).

As with much of Pierrehumbert's work, it should be noted that most of this work is based on utterances derived and elicited by the researchers.

By contrast, other theoretical approaches designed to determine how intonation contributes to signaling meaning in units beyond the sentence, e.g., in conversations or conversational sequences, are based on natural speech — on actual, authentic conversational data. Selting (1987), for example, devises a system of descriptive categories for the analysis of German intonation in natural conversation. In general, Selting regards intonation as a signaling system that contributes to situating utterances in context. Within the system are two types of categories, local and global. Local categories, such as accents, tend to fulfill mainly semantic functions, while global categories, such as different contour types, tend to fulfill interactive functions in conversation (p. 777).

Selting (1988) analyzes natural conversations in German and the role of intonation with respect to the organization of conversational sequences. She shows that, in general, intonation serves as a contextualization cue co-occurring with specific syntactic, semantic, and discourse-organizational devices to signal the status of an utterance in conversation — and specifically that (1) intonation is systematically used as a type-distinctive device in the initiation of so-called repair or local problem-handling sequences; and (2) intonation is systematically used as a means to constitute and control participants' cooperation in a global problem-handling sequence. Thus, the functions of intonation go above and beyond the level of individual sentences into the larger domain of discourse, linking the utterances of all speakers in a conversation.

Yang (1995) describes how intonation plays a crucial role in communicating emotions and attitudes of interlocutors towards the subject matter in Mandarin Chinese conversations as well as aiding in topic development (e.g., signaling topics and subtopics) and discourse flow (e.g., signaling agreement, interrupting, establishing a common background). She also shows that discourse in Mandarin is interactionally and cooperatively organized and that intonation patterns help to signal these discourse structures. For example, steep drops in both amplitude and pitch contribute to definite and emphatic utterances. In connected sequences of utterances, topics start with high pitch, and the pitch peaks for succeeding subtopics tend to gradually diminish. But there are also cases in natural conversations in which the pitch peaks become gradually higher, and this is attributed to the cognitive building up, step by step, of the discourse structure. As for intonational signals of cooperation between speakers, her data show that speakers' pitch movements in contiguous

utterances mirror each other. This type of pitch concord or convergence has been noted by others (Coulthard & Brazil, 1981; Couper-Kuhlen & Selting, 1996a), as well.

Although DuBois et al. (1992) do not describe a theory of discourse intonation per se, they cite the growing importance of discourse analysis and the need for research tools. This monograph on discourse transcription sets out the substantive details of speech ranging from pause to prosody to discourse unit structure, and particularly those features that are functionally significant to the participants in a discourse. The description of intonation and the identification and classification of intonation units are central to the monograph and contribute to theories of discourse intonation in that the authors' attempts to systematize a general framework for discourse transcription are based on general theories of intonation and methods of discourse analysis.

Discourse (in this case, conversation) can be divided into units of various kinds and at various levels. DuBois and his co-authors describe units of conversation at the following levels: the *turn*, which is a fundamental unit of conversational discourse; the *intonation unit*; and the *word unit*. An intonation unit (IU) is defined as the fundamental unit of the discourse production process (cf. Chafe, 1993) and is "a stretch of speech uttered under a single coherent intonation contour. It tends to be marked by cues such as a pause and a shift upward in overall pitch level at its beginning, and a lengthening of its final syllable" (DuBois et al., 1992, p. 17). The IU is similar to the *tone group* of Halliday (1967a), the *tone unit* of Crystal (1969), and the *intonation-group* of Cruttenden (1997). Each major intonation unit is generally characterized by some kind of prominent pitch movement, which carries the most significant intonational information about that unit. This prominent pitch movement is in general located on the word that bears the primary accent in the sentence or clause and is thus similar to the *nuclear accent* in British theories.

Similarly to the British tradition, DuBois et al. (1992) adopt the term *tones* for the various distinctive intonational shapes that start in the syllable with primary accent and can spread across several words, often extending from the last primary accent to the end of the intonation unit. Although they note that there is not universal agreement as to the classification of tones, they find that the most straightforward approaches involve simply classifying the movements of pitch by using the terms rise, fall, rise-fall, fall-rise, and level tones, transcribed with the followed symbols, respectively: /, \, /\, \/, _ .

In addition to the traditional British five tones, they list five major prosodic cues that contribute to signaling the boundaries of intonation units (p. 100):

1. coherent contour: a unified intonation contour, i.e., one displaying overall gestalt unity
2. reset: a resetting of the baseline pitch level at the beginning of the unit
3. pause: a pause at the beginning of the unit (in effect, between two units)
4. anacrusis: a sequence of accelerated syllables at the beginning of the unit
5. lengthening: a prosodic lengthening of syllable(s) at the end of the unit (e.g., of the last syllable in the unit)

Couper-Kuhlen and Selting (1996b) discuss current problem areas in discourse-level prosodic research: (1) questions regarding *formal* issues, e.g., whether prosodic categories should be viewed as phonetic categories or phonological categories; (2) questions regarding *functional* issues, e.g., whether the functions of intonation are distinctive or not; and (3) questions regarding *methodological* issues, e.g., what types of data should be used. There is ample evidence that there is little or no constancy between intonational form and meaning, i.e., the same tune can be shown to "mean" something different with each different morphosyntactic carrier. In addition, much of the previous research suggests "intuitively pleasing" theories and models but is based on the analysts' own intuitions and on constructed and/or elicited data. Couper-Kuhlen and Selting therefore propose an interactional perspective on prosody based on a complex interaction of verbal forms with contextual and situational factors. The discourse functions of intonation are related to the kind of pragmatic meaning that is situated and inference-based rather than to the semantic meanings of decontextualized linguistic forms. "In an interactional perspective, analysts are consequently not looking for minimal pairs and distinctive functions. Instead they typically find that intonation and prosody have a *contextualizing* function" (p. 13).

The shift in emphasis proposed by Couper-Kuhlen and Selting is to view intonation not from a structuralist or generativist standpoint in which minimal pairs are used to show dichotomous contrasts or phonemes, but rather to base theoretical models on naturally and authentically occurring speech, interactions and conversations. There is no one-to-one correspondence between form and function; rather, intonation must be viewed and interpreted within the context in which it occurs, i.e., is spoken. It has a signaling function within the discourse or context and is part of a real-time, ongoing process of interaction, where speakers react to the way in which their interlocutor is using intonation (pitch, rhythm, timing) and "conform" to it or "break away" from it. Speakers and hearers cooperate to avoid conflict and to resolve conflict when necessary. The focus is on the "reconstruction of patterns as cognitively

and interactionally relevant categories which real-life interactants can be shown to orient to" (p. 46).

The phonetic features that Couper-Kuhlen (personal communication, 1998) believes to be salient for prosody in conversation and that need to be further investigated are: pitch level at the beginnings of intonation groups (termed *onset* or *head* in tonetic research and *key* in discourse analysis studies); pitch register (either *high* or *non-high,* i.e., the pitch level for the entire intonation phrase that is either higher or lower than the speaker's normal pitch range); and rhythm and timing (which have to do with the regular beat of conversations and the metrics that participants use to judge when to attempt to take the conversational floor).

To summarize the historical development of research relative to intonation, although traditional theoretical treatments, including the British contour or tune approach and the American levels or phonemic and generative approaches, tended to focus on sentence-level phenomena, linguistic research during the last two decades on both sides of the Atlantic has increasingly emphasized *pragmatic*, *discourse*-level phenomena (cf. Brazil et al., 1980; Couper-Kuhlen & Selting, 1996a; DuBois et al., 1992). The added dimension of interactional functions of intonation will be discussed in greater detail in the following chapter. In the next section, a proposed theoretical basis for teaching discourse intonation that draws upon previous linguistic theories of intonation will be articulated.

Integrating theory and practice: A model for teaching intonation

In an attempt to make the myriad of intonation theories accessible to language instructors, the following model does not claim to be exhaustive or complete and runs the risk of seeming to distort some of the theories discussed above. Of note is the observation that the various established theories use different terminology for what often appear to be the same or very similar concepts. The proposed framework is, therefore, a synthesis of the similarities in the major theories that attempts to capture the essence of descriptions of intonational patterns useful for language learning, particularly at the discourse level.

Although from a theoretical linguistic perspective the recent generative theory of intonation expanded by Pierrehumbert and Hirschberg (1990) to include discourse interpretation may be intuitively appealing and potentially useful for research on language acquisition (cf. Wennerstrom, 1994), in terms

of pedagogical applications, language instructors who attempt to apply this model to teaching intonation may be daunted by the terminology, notation and abstractness of the theory. In particular, they would have to learn the complex set of rules that needs to be applied to the strings of H's and L's (high tones and low tones) in order to arrive at the surface-level pitch contours that would then be taught to learners. Similarly, the phonemic approaches that describe pitch contours in terms of numbers, typically from 1–4, would seem to be too abstract and not readily decipherable for instructors who are not intonation specialists — and even less accessible to students.

Of the existing theories, particularly appealing are those of Bolinger (1986, 1989), with his emphasis on the interrelatedness of pitch and stress, and Brazil et al. (1980), particularly with their notion of *key* that is integrated into their system of five *tones*. Recent work on discourse intonation in American English by DuBois et al. (1992), who find that the most straightforward approach involves simply classifying the movements of pitch in terms of five (British-like) tones, is also appealing for its elegant simplicity. Finally, Couper-Kuhlen and Selting's (1996b) work is most appealing from a pedagogical standpoint because it based on naturally-occurring discourse. They propose that analysis of natural speech and interaction should form the foundation for theories and descriptions of intonation and incorporate the concepts of tone, key, and the results of acoustic analyses of fundamental frequency and rhythm into models of prosody.

Thus, a modified prosodic or tune approach with a focus on natural discourse is suggested here as the basis for teaching intonation to language learners. Pitch movement is critical at two points in an utterance or intonation unit (IU). First, the greatest pitch prominence is generally found at the final primary accent, where five basic tunes are commonly distinguished in many theories: *rise, fall, rise-fall, fall-rise,* and *level.* These tunes start in the syllable with primary accent and can be spread across several syllables or words, often extending from the last primary accent until the end of the IU. If the primary accent falls on the last word, the tune can also be realized on a single syllable if the word is monosyllabic.

The second crucial point in an intonation unit is the pitch movement at the end of the IU, i.e., at transition points from one IU to another in a discourse. There are three main types of transitional continuity best characterized by their function: *final, continuing,* and *appeal* (cf. DuBois et al., 1992, pp. 28–31). "Final" transitional continuity in English and many other languages is manifested primarily (but not exclusively) by a fall to low pitch at the end of an

IU. The class of intonation contours whose transitional continuity is generally interpreted as "continuing" in a given language is realized in American English in one of several ways: one type of continuing contour is realized by a terminal pitch that rises slightly from its beginning at low or mid level; another type remains level; another is realized by a terminal pitch that falls slightly, but not low enough to be considered final. The third class of intonation contours, that in which transitional continuity is regularly understood as an "appeal," is generally realized in English by a non-descending high rise in pitch at the end of the IU.

In addition, the concept of *key* is relevant, particularly in the domain of discourse, as the element that signals the relationship between the propositional content of an utterance and that which precedes or follows it. Selection of *high* or *low* key involves raising or lowering the pitch of the entire tone group relative to a pitch that can be established as the norm (*mid* key) for the speaker concerned. In general, high key functions to signal contrast, e.g., it can mark a word or utterance as contrasting to what the hearer might have expected. Low key, on the other hand, functions to mark known information or even to inhibit a response by signaling that no response is necessary or expected (see discussion of Brazil above).

As for a transcription system to be integrated with the underlying theoretical bases proposed here, a modified version of the system of DuBois et al. (1992) is adopted. As noted previously, DuBois et al. transcribe the *rise, fall, rise-fall, fall-rise,* and *level* tones with the followed symbols, respectively: /, \, /\, \/, _ . I will use their system in conjunction with graphic representations of acoustically generated (not stylized) pitch contours and recommend this notation for teaching intonation to learners of second or foreign languages. Preferable to symbols alone are graphic representations and notations that show pitch height, pitch movement, slope of pitch, and pitch range or key, as well as indications of which syllables are prominent or stressed. Bolinger's notation using the typed words themselves to depict pitch movement is excellent graphically but extremely difficult to realize. A system similar to or based on a musical staff has been employed by some researchers (e.g., O'Connor & Arnold, 1961; Lohnes & Strothmann, 1980). Most notational systems are inherently labor-intensive to illustrate, but this is due to the simple fact that intonation is non-linear. Chapter 7 of this book uses instrumental trackings of fundamental frequency (F_0) or pitch, which represent the most direct (and therefore accurate) manifestations of acoustic phenomena. In light of the fact

that software for pitch tracking is becoming increasing accessible, providing learners with these types of graphic representations, along with a notational system like that of DuBois et al. and indications of pitch range or key, will be the system employed here and recommended for teaching pronunciation in the classroom.

Chapter 3

Meaning and function of intonation

Introduction

As discussed at the end of the preceding chapter, theories of discourse intonation that have emerged in the 1980s and 1990s reflect the current view that intonation provides some measure of redundancy to linguistically coded messages but also supplies additional cues to express the full meaning that speakers wish to communicate. Many researchers would agree that sentence prosody is determined more by communicative intentions or *functions* than by *formal* rules as in traditional structural or phonological approaches. In other words, intonation cannot be exclusively accounted for with semantically or pragmatically "blind" phonological rules (cf. Lambrecht, 1994, p. 296; 't Hart et al., 1990, p. 189). Bolinger insisted for decades (as in Bolinger, 1958, p. 37) that the relationship between grammar and intonation is a *casual* rather than *causal* one. In this chapter, the various interpretations of the meaning and functions of intonation will be discussed, beginning with the traditional grammatical and attitudinal functions and leading to the presently accepted discourse and interactive functions.

Table 1 provides an overview of the functions of intonation, which I have grouped into four broad categories: *grammatical, attitudinal, discourse,* and *sociolinguistic.* These functional categories expand on the categories described by Couper-Kuhlen (1986) and are suggested as a means of organizing the diversity of previous work but are also based on the elements of communicative competence being emphasized in language learning curricula. The purpose of this chapter is to review the various accounts of the functions of intonation and, by categorizing them into four main groups, to show how these functions are precisely those that are currently being recommended for language learners to master.

In most accounts, intonation is found to be multi-functional, though different theories emphasize different aspects or functions. For example, Crystal (1985) lists several functions, the most important being to signal *grammatical* structure, which includes marking sentence, clause, and other boundaries,

and contrasting grammatical structures such as questions and statements (p. 162). Halliday (1967b) focuses on a second function, how intonation reveals *information structure*. A third function of intonation is to communicate *attitudinal* or affective meaning, e.g., sarcasm, puzzlement, anger, as emphasized by O'Connor and Arnold (1961) and Uldall (1964). More recent theories have tended to focus on *discourse-level* functions and phenomena. Brazil (1978) is particularly interested in the way participants in a conversation use intonation to control interactive structure. Brown et al. (1980) propose a system in which intonation marks whether a speaker is continuing with an already established topic or is instituting a new topic.

Gumperz (1982) describes in general how conversationalists use prosody to initiate and sustain verbal encounters. Prosody helps listeners to hear a stream of talk and group words into clauses or utterances as well as to distinguish main from qualifying phrases and parenthetical remarks. During a conversation, speakers and hearers have expectations about possible goals or outcomes for the interaction, about what information is new or important and how it is likely to be signaled. Successful communication requires that participants share these expectations, and this implies that they must also have a common system for signaling or negotiating shifts or transitions from one activity to another.

Couper-Kuhlen (1986) summarizes six functions that have been attributed to English intonation: (1) informational, (2) grammatical, (3) illocutionary, (4) attitudinal, (5) textual/discourse, and (6) indexical. (Cf. Table 1, which expands on Couper-Kuhlen and subdivides the textual/discourse function into textual/discourse and interactive/discourse functions.) She gives the following to illustrate how intonation functions to signal the *information structure* of utterances (pp. 111–113):

(a) I saw a MÀN in the garden
(b) I SÀW a man in the garden

where (a) with the stress on "man" could conceivably answer the question "What happened then?" or "Who did you see?" but (b) with the stress on "saw" could only answer a question such as "Did you hear a man in the garden?" In other words, intonation in these examples functions to mark the most newsworthy piece of information in an utterance.

As illustrations of the traditional *grammatical* function of intonation, i.e., the use of rising and falling contours to distinguish sentence types such as statements, questions, and commands, the following examples are provided (p. 111):

(a) |John's going HÒME ||
(b) |John's going HÓME ||
(c) |shut the DÒOR||
(d) |shut the DÓOR||

where a falling contour is used in (a) for a statement but a rising contour in (b) turns it into a question and the rising contour in (d) makes a request out of the command with falling contour in (c).

To illustrate *illocutionary* function, i.e., that intonation signals the intentional force of an utterance in a given context, Couper-Kuhlen cites the following example from Sag and Liberman (1975, p. 488):

(a) ^{h} | WHẏ don't you move to CaliFÓRnia ||
(b) ^{l} | why don't you move to Cali↑↑FÓRnia || (Sag/Liberman 1975)

where (a) spoken with high (h) pitch range is appropriate as a "real" question while (b) spoken with low (l) pitch range renders it a suggestion.

As an example of the *attitudinal* function of intonation, Couper-Kuhlen cites Gimson (1980, pp. 276f.) and describes differences in *pitch range,* as in (a) and (c) with a high (h) pitch range vs. (b) and (d) with a low (l) pitch range, and in *pitch direction,* as in (a) and (b) with falling pitch on "great" vs. (c) and (d) with a rising contour on "morn-". The combination of high pitch range and a falling contour in (a) could signal that the speaker is excited, happy, or pleased; low pitch range with the same falling contour in (b) might be used if the speaker were sad, reserved, or being ironic. In greeting (c), high pitch range and a rising contour might express that the speaker is cheerful and friendly, whereas (d) with low pitch range makes the greeting more routine or perfunctory.

(a) that's ^{h} | GRÈAT ||
(b) that's ^{l} | GRÈAT ||
(c) ^{h} |good MÓRNing ||
(d) ^{l} |good MÓRNing ||(Gimson, 1980, pp. 276f.)

To exemplify the *textual/discourse* function of intonation, where *discourse* refers to the interactive aspects of linguistic organization beyond the sentence, Couper-Kuhlen adduces an example from Brazil et al. (1980, p. 75). Example (a) is taken from an actual doctor-patient exchange. Couper-Kuhlen explains that "In (a) the patient demonstrates 'compliant' behaviour as a discourse participant by answering 'in key' to the doctor's question (i.e., by using a mid onset following the doctor's mid termination […] Example (b), which is

fictive, would be said to illustrate 'non-compliant' behaviour on the part of the patient (a high onset following mid termination)" (p. 113):

(a) Doctor: it's ^{h} |DR$\hat{Y}$ skin || ^{m} | ÌSn't it ||
Patient: ^{m} |M`M || (Brazil/Coulthard/Johns 1980: 75)

(b) Doctor: it's ^{h} |DRY$ skin || ^{m} | ÌSn't it ||
Patient: ^{h} |M'M ||

The sixth function of intonation posited by Couper-Kuhlen is said to be *indexical* in nature (termed *sociolinguistic* in Table 1) when it establishes contrasts that enable hearers to identify speakers as members of different social groups, with "social groups" being groups differentiated by any of the criteria gender, age, occupation, or socio-regional background, as well as to identify individuals on the basis of idiosyncratic features.

Couper-Kuhlen and Selting (1996a) describe how their views have evolved to their current proposal of a major shift in emphasis that focuses on the *interactive* functions of intonation in natural speech. They now consider intonation to fulfill primarily a signaling function within a specific context or interaction. The other functions of intonation (informational, grammatical, illocutionary, and attitudinal) are subsumed under this broader *contextualizing* function — that is, they now link intonation to functions which derive from the situated use of language to accomplish interactional goals. In the introduction to their volume, Couper-Kuhlen and Selting state their goal of investigating how intonation functions to make "social interaction more than the mere exchange of words, namely a real-time encounter between conversationalists who establish and negotiate units of talk as situated meaningful activity" (p. 1).

In summary, the preceding accounts of the functions of intonation illustrate the diversity of opinions as to what the main functions are and how these functions should be classified. In this book, the various meanings and functions of intonation are grouped into four basic categories: ***grammatical, attitudinal, discourse,*** and ***sociolinguistic.*** An important underlying reason for distinguishing these categories is to parallel the communicative functions of language that are increasingly being emphasized in language learning syllabi and curricula. For each of these broad categories, examples will be given in ensuing sub-sections of how different meanings and functions are classified in some of the different theories, and the differing classifications show that the labels are somewhat arbitrary and subjective. Characterizing the nature of

various functions can indeed be problematic. For example, there can be a fair amount of overlap between what counts as attitudinal and discourse functions as well as between grammatical and discourse functions. Specifically, some researchers classify using intonation to mark *new* vs. *old* information as a grammatical function, whereas others categorize it as a discourse function. In this book, I will consider the function of distinguishing statements from questions *grammatical* while characterizing the function of revealing information structure, controlling interactive structure, and introducing a new topic or continuing with an established topic as a *discourse* function. However, it is clear that the four categories used here are often interrelated and do entail at least some degree of overlap.

Grammatical functions

On the most basic level, the fundamental question of what is included in *grammar* is by no means universally agreed upon. Is *grammar* synonymous with *morphosyntax*? Is *grammatical form* synonymous with *syntactic form*? From as early as the 16th century, intonation and grammar have been said to go hand in hand. That is, sentence type (statement, question, exclamation) and tune (falling, rising) have frequently been linked. However, as Crystal (1969) noted, there is a wide range of opinion on the extent to which grammatical considerations should enter into intonation analyses (p. 253). Near one end of the continuum is Bolinger (1958), who feels that "the encounters between intonation and grammar are casual, not causal" (p. 37), whereas for Halliday (1964), at the opposite end of the continuum, the relationship is central: "Only those distinctions which are shown in the grammatical description to be meaningful are represented in the phonological analysis" (p. 169). As seen in Chapter 2, Halliday believes that the choice of pitch movement is not only related to other grammatical choices, such as word order, use of verb tense, or use of a negative morpheme, but that the intonational contrasts are distinct and grammatical choices themselves. In the case of accent placement, Bolinger speaks of "free choice" depending on pragmatic considerations, whereas Halliday and others posit (grammatical) "predictability" (cf. also Gussenhoven, 1987; Ladd, 1990).

Lambrecht (1994) considers grammar to be much broader than morphosyntax and sees it as central to language systems in general. He suggests that grammar and grammatical patterns are "the result of multiple language-specific dependencies" between the following four components: semantics,

information structure, morphosyntax, and prosody. Grammatical constructions may appear at different levels as lexical, phrasal, clausal, or sentential structures, and also, he proposes, at the level of prosody (p. 42).

Couper-Kuhlen (1986), after presenting both the case for the grammatical function of intonation and the case against it, concludes that a three-way distinction is necessary "between *intonational* shape — rising or falling tone, etc.; *grammatical* form — declarative, interrogative, imperative, etc.; and *pragmatic* function — question, statement, command, etc." (p. 139). In this section, the more traditional grammatical functions will be discussed, and the pragmatic functions, including illocutionary and discourse functions, will be treated in a later section.

I divide the broad category of the grammatical functions of intonation into the following subcategories: (1) aligning or associating tones with particular syntactic structures, such as sentence types, and (2) "chunking" the flow of speech into various types of units, i.e., segmenting functions.

Association of intonational structures with syntactic structures

The theories that emphasize the grammatical functions of intonation span a wide range. At one extreme, there is the main division into two tunes based on an opposition of question and statement types, while at the other extreme there is the detailed system of tonal contrasts within the context of Halliday's (1963) grammatical theory, in which the nucleus and its placement are central to marking grammatical functions (see Chapter 2). In between, other theories distinguish different kinds and degrees of grammatical relevance.

As for linking tones or tunes with larger syntactic structures, intonation patterns are typically classified into three or four major sentence-types: statements, two types of question (those beginning with an interrogative word and those requiring a yes/no answer), commands, and exclamations. In English, statements, wh-questions, commands, and exclamations traditionally are described as having falling intonational patterns and yes-no questions as having typically rising patterns.

Below the sentence level, smaller syntactic units have also been classified. Cruttenden (1997) discusses the way in which intonation groups align with various portions of utterances. Intonation groups can be aligned with syntactic constituents such as clauses, adverbials, noun-phrase subjects, topicalized subjects, parenthetical clauses, vocatives, nouns in apposition and pairs of clauses. Although there is a good deal of flexibility in the choice of intonational phras-

ing, intonation groups co-occur most commonly with clauses (on 40% of occasions by one count) under the traditional definition of clause involving the occurrence of a finite verb in a surface structure (p. 69).

One of the main difficulties in associating tunes with sentence types is that this is most successful with so-called "neutral" meanings, e.g., a falling tone could be considered the "unmarked" tone for a syntactic type such as "declarative" or "statement." However, what is problematic is that it is not always easy to determine what the most neutral meaning of an utterance is. With yes-no questions, for example, would the "polite" tone be considered more unmarked and thus more neutral than a "businesslike" tone? Similarly, studies that have actually counted the occurrence of rises and falls on yes-no questions in English have usually found that the occurrence of one or the other is heavily dependent on the type of situation involved. The grammatical approach to the analysis of the meanings of English nuclear tones is therefore felt by some to be difficult to justify, although it may be a necessary simplification in initial stages of teaching intonation in courses of English as a foreign language (cf. Cruttenden, 1997, pp. 88–89). In this book, "neutral," "unmarked" examples are first presented, followed by selected examples of what would generally be considered more marked or situation-specific cases.

Segmenting functions

Related to the function of aligning tones with syntactic structures is the role of intonation in segmenting utterances (or parts of utterances) — that is, of marking various types of boundaries within or between utterances. Crystal (1969) states that analyzing speech into tone units in effect means defining their boundaries. He finds that in normal (unhurried) speech, there are regular, definable phonological boundaries for tone units. Since each tone unit has one peak of prominence in the form of a nuclear pitch movement, a tone unit boundary occurs after each nuclear tone and is marked by two phonetic factors. The first phonetic factor is a perceivable pitch change, either rising or falling. The second is the presence of junctural features at the end of every tone unit. This usually takes the form of a very slight pause, but this pause is frequently accompanied by segmental phonetic modifications, such as variations in length and aspiration, that serve to reinforce it.

DuBois et al. (1993) discuss how to determine and mark boundaries between units of various kinds including the intonation unit and the word unit. Their description of intonation unit boundaries is similar to that of

Crystal's tone unit boundaries above: "Roughly speaking, an intonation unit is a stretch of speech uttered under a single coherent intonation contour. It tends to be marked by cues such as a pause and a shift upward in overall pitch level at its beginning, and a lengthening of its final syllable" (pp. 46–47).

In summary, there is no consensus about what should be subsumed under the *grammatical* functions of intonation. In particular, the grammatical or syntactic domains have traditionally included primarily sentence-level phenomena, whereas much recent work on intonation has focused on the discourse level, and the grammatical domain is currently often expanded to include discourse or pragmatic considerations. As a result, there is definite overlap between what might previously have been designated *grammatical* functions and what are now considered *discourse* functions.

Attitudinal/emotional functions

Couper-Kuhlen (1986) begins her chapter "Intonation and Attitude" by stating: "It is an undisputed fact that intonation has an important role to play in the expression of emotions and attitudes. The linguist's task therefore is not so much to determine *whether* intonation expresses a speaker's inner states or not but rather *how much* of this expression is indeed linguistic" and whether intonational features that express emotions are universal or language-specific (pp. 173–174). Lieb (1980) states "Wer wütend die Stimme hebt, tut dies weder auf Englisch noch auf Deutsch, sondern eben nur wütend" (p. 34) translated by Couper-Kuhlen (1986) as "An angry person does not raise his voice in English or in German but simply in anger," leading to the proposition that emotion might appear to have a language of its own. "In sum, we must distinguish an unmonitored, purely physiologically determined externalization of emotional state, presumably universal across linguistic communities, from a "cognitively" monitored expression of attitude, conventionalized and communicative in purpose" (p. 174).

The current section examines views on the attitudinal functions of intonation in the different approaches to or theories of intonation, starting with the prosodic approach. In one of the earliest approaches, Sweet (1890) describes the basic functions of his five proposed tones using a mixture of terms, some of which refer to attitudinal functions, e.g., doubt, caution, warning, dogmatic assertion, obstinacy, sarcasm, contempt, and some of which refer to what are being classified in this book as grammatical and/or discourse functions, such as

interrogative, contrastive, expectant (suspensive), affirmative or conclusive (pp. 32–33).

In the tonetic approach, O'Connor and Arnold (1961) state that the "contribution that intonation makes is to express, in addition to and beyond the bare words and grammatical constructions used, *the speaker's attitude to the situation in which he is placed*" (p. 2). In the phonemic approach, one of the most important characteristics of Pike's (1945) pitch phoneme theory is the idea that intonational meanings are superimposed upon intrinsic lexical meanings, i.e., extrinsic pitch contours provide a temporary addition to the basic form and meaning of words, according to the attitude of the speaker or as an indication of the attitude with which the speaker expects the hearer to react (p. 21). Brown et al. (1980) suggest that "there seems to be a small number of intonation patterns which are conventionally related to a set of attitudes" (p. 22). For example, final rises are characteristic of speakers wishing to be kindly and encouraging to their interlocutors, when a "kindly" voice quality is also used. However, rising intonation patterns accompanied by other types of voice quality are associated with other quite different attitudes, such as "politeness" and "hectoring." They suggest that "the manipulation of voice quality seems to be a much better indicator of attitude than intonation alone" (p. 22).

One of the primary proponents of the close, even interdependent relationship between intonation and emotion, i.e., that intonational features are beyond grammar and are directly linked to emotion, is Bolinger. The association of intonation and gesture, a physiological phenomenon, is in fact a cornerstone of Bolinger's (1986) theory of intonation: "The fact that intonational configurations are matched by configurations of facial expressions and bodily gestures, that the two operate much of the time in parallel, and that their similarities betoken similarities of function, points to the configurational approach as the most likely to succeed" (p. 337). Bolinger (1989) expands upon the link between intonation and emotion, thought by some to be his most controversial claim. His basic contention is that intonation is best described by contours and is closely tied to gestural phenomena, e.g., falling intonation conveys the notion of termination, finality, where rising intonation and its quality of "upness" carries the meanings of "up-in-the-airness" or "incompletion," and "keyed-upness" or high emotivity, such as excitement, anger, surprise, curiosity.

However, Bolinger has been challenged for his insistence that whatever abstract basic meanings are discovered for intonation, they (a) must be directly linked to the prelinguistic uses of pitch in expressing emotion and (b) must therefore be universal. Ladd (1990), for instance, although conceding that

intonation *is* remarkably similar in many languages, particularly in terms of intonational contours, contends that accent placement can be very different in different languages and maintains that until these differences can be described, Bolinger's case for a universal link between intonation and emotion remains incomplete. In general, however, Ladd (1990) acknowledges that "Few linguists would disagree with the proposition that intonation affects the interpretation of utterances through the interaction of very general meanings and broad principles of pragmatics" (p. 808). The next section will discuss the pragmatic and discourse functions of intonation.

Discourse functions

As stated generally throughout this book and throughout this chapter in particular, there are no firmly established or universally agreed upon principles for classifying the functions of intonation. The *discourse* functions of intonation discussed in the literature encompass a range of functions beyond the sentence level for the purpose of achieving continuity and coherence within a discourse, regardless of the length of the discourse. Some of the functions that have been suggested are to mark prominence, focus, or newsworthiness of a piece of information in a discourse; to mark boundaries in a discourse, e.g., boundaries between sentences, paragraphs, topics, and conversational turns; to control interactive structure, e.g., to constrain the hearer to reply or to discourage the hearer from replying; to continue an established topic or to signal a new topic. Here, these and the following discourse functions will be discussed: to express a speaker's intentions, to indicate the expectations a speaker has about a listener's reply, to facilitate cooperation between speakers in structuring a discourse, and to mark the shared mutual knowledge of a speaker and listener. It should be noted that some of these functions have been mentioned and discussed in previous sections, albeit in part as sentence-level phenomena rather than at the discourse level — another reflection of the overlap between categories of intonational function.

Clark and Yallop (1990) note that in recent years, researchers have increasingly turned their attention to the role of intonation in discourse, for example, Brazil et al. (1980), Brown et al. (1980), Hewings (1990), and Johns-Lewis (1986a) (p. 303), all of which were discussed in Chapter 2 on theories of intonation. As discussed in the preceding section, recurrent (and not uncontroversial) themes in Bolinger's work are his emphasis on the pragmatic bases of

intonation, e.g., the role of speakers' intentions and the newsworthiness of individual words, and his argument against any role for grammatical rules.

Cruttenden (1997), in discussing the functions of intonation, begins with grammatical meanings and continues: "Other descriptions have emphasised the attitudinal aspects of meanings or the discoursal aspects of meanings, the attitudinal approach being the older, the discoursal approach being the more recent" (p. 97). He then states, in discussing discoursal meaning, that "whether information was new or old or contrastive [is] seen to be the most obvious factor in decisions about nucleus placement," thus indicating his preference for classifying information structure as discoursal rather than grammatical. He also cautions that, in practice, it is not always easy to separate discoursal approaches to tones from attitudinal approaches, e.g., to determine how adjectives and labels such as "protesting," "detached," "interested," "impressed," and "encouraging" should be classified.

In this book, the discourse functions of intonation are grouped into the following subcategories: *information structure marking, illocutionary/speech act; textual/discourse; interactive/discourse* (cf. Table 1). It should be noted that there is unquestionably overlap across these subgroups, such as whether expressing a speaker's intentions or expectations about a hearer's reply should be classified as an *illocutionary* or an *interactive* function.

Marking information structure

The theory of *information structure* in sentences and texts stems from the work of the Prague School linguists and also the work of Halliday, who accords a central role to intonational choice. Information structure generally refers to the division of a message or utterance into chunks and the organization of these chunks in terms of given and new information. Halliday (1967b) proposes that in each message or unit of information, i.e., in each *tone-group*, one or two elements are selected as "points of prominence" within the message and form the *information focus* or *foci* of the unit. The functions of focus itself are concerned with the distinction between *new* vs. *old* (or *given*) information and *contrastivity* .

Once again, the difficult question of whether information structure "belongs" in the domain of grammar or discourse surfaces. In his attempt to define *information structure* and the status of this component of language in the overall system of grammar, Lambrecht (1994) argues that grammar cannot be separated from pragmatics and defines the concept more broadly to include

pragmatics and discourse-level components. One of his underlying ideas is that sentences cannot be fully understood without looking at the linguistic and extralinguistic contexts in which they are embedded. The extension of the context beyond the sentence to the discourse level entails overlap between the traditional domains of grammar and discourse. Information structure is said by Lambrecht to be concerned with the "discourse circumstances under which given pieces of propositional information are expressed via one rather than another possible morphosyntactic or prosodic form" (p. 6). In other words, information structure can be manifested in various ways, i.e., not exclusively by prosodic means, but also by special grammatical markers: in the form of certain syntactic constituents, in the position and ordering of such constituents in the sentence, and in the choice of lexical items. However, for the purposes of this book, the discussion of information structure will be confined to the use of prosody or intonation to mark the *information focus* or *foci* of a sentence as they are situated in a discourse or context and, in particular, to emphasize and contrast as well as to signal new vs. given information.

Given vs. new information

Brown et al. (1980) state that "One of the currently most discussed functions of intonation is the part that it plays in marking the information structure of discourse (see for example Halliday, 1967b; Chafe, 1972, 1974, 1976)" (p. 27). Halliday (1967b), although generally describing intonation in terms of grammatical units or structures, characterizes "given" information as "recoverable either anaphorically or situationally" (p. 211), whereas he considers "new" information to be focal "not in the sense that it cannot have been previously mentioned, although it is often the case that it has not been, but in the sense that the speaker presents it as not being recoverable from the preceding discourse" (p. 204).

Brown (1983) finds in an experimental study of intonation and information structure that when a speaker introduces brand-new information it is typically marked with high pitch, and that when inferable information is introduced, high pitch is also used. Although the information may be potentially "known" to the hearer, the speaker treats it as new, in Halliday's terms. "It is only the information which derives saliently from the context, or specifically from previous mention in the discourse, that the speaker treats as given, in Halliday's terms — on low pitch" (p. 77). Chafe (1974) also asserts that, in general, the given/new dichotomy in English is reflected intonationally by the use of low vs. high pitch, respectively.

Emphasis and contrast

Not surprisingly, in terms of signaling *emphasis* or *contrast*, some would argue for the grammatical nature of this function, while others contend it belongs in the realm of discourse because it relies by necessity on what was previously known or has occurred in the preceding discourse.

Halliday (1967b), for example, views new information as either "cumulative to or contrastive with what has preceded" (p. 211). In particular, he claims that new information is contrastive or merits emphasis when items that do not normally receive focus in fact form the information focus, e.g., pronouns or prepositions as in *I'M going* ("not someone else"), *I put it ON the table* ("not under it"). Similarly, in the sentence *Joe WAS angry*, the verb would not normally be given sentence stress (copulas usually are not), but if it is, then it would be contrastive or emphatic — the speaker might be contradicting a previous statement or implying that Joe was angry earlier but no longer is. In other words, the normal accentual pattern of a word may be overridden by the focus being placed elsewhere for contrastive or emphatic purposes.

Chafe (1976), however, contends that contrastive information is qualitatively different than new information and belongs in a category of its own. The examples he gives are the following hypothetical dialogues:

A: I didn't know Susie could cook so well.
B: She can't. RONald made the hamburgers.

In the above example, *RONald made the hamburgers* is contrastive, contradicting the assumption that Susie was a possible or likely candidate to have made the hamburgers. However, in the next dialogue, *Ronald* is new information but not contrastive:

A: Who made the hamburgers?
B: RONald made the hamburgers.

Both contrastive and new items may have high pitch but they may be distinguished intonationally by different overall pitch configurations or contours. Chafe (1974) illustrates this with an example of a sentence with two foci that can be pronounced with different pitch patterns. In the first case, the situation is that the speaker is visiting a family with a child named Matthew. The speaker has brought a book as a present and says to one of the parents:

I brought MATTHew a BOOK.

There is a non-contrastive version of this sentence, where both *Matthew* and *book* are relatively high-pitched, as they both convey new information. The high pitch on *Matthew* is sustained, i.e., it continues high throughout the word. In the other case, in a different situation, both *Matthew* and *book* might be contrastive, e.g., if the speaker had brought Matthew's sister a game and was contrastively emphasizing which child received which gift. In this case, *Matthew* might be pronounced with a partial drop in pitch, which is not raised until the word *book*. Chafe suggests that contrastive pitch is always falling (p. 119) and concludes that new information and contrastiveness should be treated as separate phenomena.

These examples serve to illustrate that it is both pitch range or key and pitch contours that signal focus and that, according to Crystal (1985) "From the viewpoint of phonology, the main function of stress is to provide a means of distinguishing degrees of emphasis or contrast in sentences. The very nature of these functions of focus in information structure, i.e., that focused elements are often emphasized or contrasted across, and not within, utterances, based on the previous discourse or prior knowledge, supports the view that they are discoursal functions and not grammatical functions" (p. 288).

Illocutionary functions of intonation

A fundamental principle of linguistic study — one first elucidated by Austin (1962) and Searle (1969) — is that speakers use sentences in given situations for specific purposes. Utterances are thus verbal actions and constitute *speech acts*, which can be characterized as *locutionary, illocutionary,* or *perlocutionary*. It is the relationship between intonation and *illocutionary* acts that will be discussed in this section. Searle established a taxonomy of speech acts and grouped them in five major classes: representatives, directives, commissives, expressives, and declaratives. The table below summarizes the taxonomy:

Speech Act	**Examples**
representatives	asserting, describing, stating, complaining, concluding
directives	requesting, commanding, asking, entreating
commissives	promising, threatening, pledging
expressives	congratulating, welcoming, thanking, apologizing
declaratives	resigning, firing someone, appointing someone, declaring (e.g., war)

Simple examples of illocutionary functions of intonation involve using intonation to express the speaker's intentional or illocutionary force. For example, in

uttering a question such as "Why don't you move to California?" (Sag & Liberman, 1975), the intonation chosen will indicate whether the speaker is making a "genuine" inquiry (high key), a suggestion (low key) or an exhortation (level intonation rising slightly at the end). If someone says "It's hot in here," the intonation used helps to express whether the speaker is making a simple statement, is grumbling or complaining, or is making an indirect request that someone open a window.

Brown et al. (1980) discuss the speech function or illocutionary force of intonation, noting that a common claim is that "there is a close relationship between the form of an intonation pattern, in particular whether its end-point rises or falls, and the illocutionary force of an utterance, whether it functions as a statement, question, or command" — that is, intonation correlates with illocutionary force rather than syntactic structure (p. 30). But they note that one can go a step further and associate low terminals "with the ends of topics, with the end of a turn when a speaker has no more to say on a topic, and with conducive questions where the speaker has a high expectation of the correctness of the assumptions that lie behind his question."

However, Couper-Kuhlen (1986), in a chapter on intonation and illocution, notes that although intonation has often been felt to be a marker of illocutionary force, few studies have dealt with the question of how the two are specifically linked (p. 163). The main difficulty, according to her, is establishing any kind of one-to-one relationship such that for every distinct illocution there is a distinct intonational marking.

Textual/discourse functions

The *textual/discourse* role of intonation is concerned with stretches of speech larger than one utterance. For the purposes of this book, these textual/discourse functions of intonation are **related to the content of the speech or discourse itself**, as opposed to the speaker's and hearer's conversational perspective, which will be discussed in the next sub-section on *interactive/discourse* functions. *Text* is defined by Halliday and Hasan (1976) as any passage, spoken or written, that forms a unified or organized whole (p. 1). Much of the research into the organization and structure of text focuses on the notion of *paragraph*. The conceptual paragraph is an important building block of texts, and texts consist of paragraph units that are organized around topics. In this book, the first function of intonation in the *textual/discourse* subcategory is to achieve coherence between and among paragraphs or propositions in a text or discourse.

Coherence between two or more tone units can be signaled by intonation, either with alternating intonation patterns or with successive patterns. For example, a rising tone unit may have the subtle, quasi-musical effect of demanding a "resolution," followed by a falling tone unit that provides the necessary resolution or termination. Or, a cohesive unit may be established by a succession of two or more identical or similar intonation patterns, e.g., a falling nuclear pitch movement can be repeated at different (gradually descending) pitch levels over a series of tone units, and this succession of repeated patterns makes the units cohere.

Two additional functions of intonation that fall in *textual/discourse* subcategory are to mark shared knowledge between speaker and hearer (old or given information) and to distinguish this type of information from new information. Yet another *textual/discourse* function is to mark prominence, focus, or newsworthiness of an item in general. The mechanisms for marking information structure involve pitch accents and intonation contours in addition to syntactic and lexical devices. The speaker must make assumptions about shared common knowledge in assessing how to signal the new or newsworthy piece of information, and the hearer must be aware of the preceding discourse in order to understand the significance of the speaker's intonational signals.

A further so-called *textual/discourse* function of intonation is to mark various types of boundaries within a discourse. Oreström (1983) studied turn-taking in (British) English conversation using the *tone unit* (TU) as the basic unit. He hypothesized that "the end of turn is positively correlated with the end of TU, which is a point of prosodic completion" (47). His data showed that the end of a speaking turn correlated highly with three primary factors: "prosody: end of a TU with a non-level nuclear tone; syntax: end of a syntactically completed sequence […]; and semantics: end of a fully comprehensible stretch. The joint completion of these three, present in about 95% of all cases [four English conversations from the London-Lund Corpus], constitutes a major juncture in English…" (p. 77). Loudness reduction and pause, two other features characterized by DuBois et al. (1992) as signaling juncture, appeared to be secondary cues.

DuBois et al. (1993) state that one aspect of intonational function is to mark *transitional continuity*, i.e., mark the degree of continuity that occurs at the transition point between one intonation unit and the next (p. 53). When a speaker comes to the end of an intonation unit, the intonation contour chosen usually gives a fairly clear indication of whether the topic or discourse at hand will be continued or has been completed. The question of what it is that is

being continued or finished is somewhat open-ended: a "final" contour may apply to the end of a sentence, the end of a turn, or the end of some other discourse unit. Although it may be possible to make finer distinctions in transitional continuity, the distinctions among "final," "continuing," and "appeal" seem to be basic.

DuBois et al. defined *final continuity* as a class of intonation contours regularly understood as *final* in a given language. "For [American] English and many other languages, this means primarily (but not exclusively) a fall to a low pitch at the end of an intonation unit." On the other hand, *continuing continuity* is seen as a class of intonation contours regularly understood as *continuing*. "The contour is often realized in English as a slight rise in pitch at the end of an intonation unit (beginning from a low or mid level), but it may have other realizations as well, each of which presumably has slightly different pragmatic implications. One type of continuing contour is realized by a terminal pitch that remains level; another by a terminal pitch that falls slightly, but not low enough to be considered final" (p. 54).

In the transcription system employed by DuBois et al. (1993), the question mark (?) indicates a class of intonation contours whose transitional continuity is usually understood as an *appeal*. For English, this is often realized as a marked *high rise* in pitch at the end of the intonation unit. "Appeal" here refers to when a speaker, in producing an utterance, seeks some sort of confirmation from a listener. The most common type of appeal in this sense is a *yes-no* question, but not all *yes-no* questions are said with the appeal contour (p. 55). The underlying purpose of these various types of transitions, e.g., expressing finality, continuation, or appeal, leads to the next subcategory of discourse functions of intonation, namely that of controlling the interactive structure between and among participants in a discourse, because while the transitions themselves help to mark boundaries, their functions can be said to help in controlling interactive structure.

Controlling interactive structure

This subcategory of the *interactive/discourse* functions of intonation is distinguished from the preceding *textual/discourse* subcategory (despite, of course, some overlap) in that instead of being concerned with the content of the text or discourse, it is **related more to the conversational structure of the discourse**, particularly the so-called **turn-taking functions,** and the speaker's and hearer's roles or perspectives in controlling the interactive structure. Included in this

subcategory are the following: using intonation (1) to continue with an established topic; (2) to start a new topic; (3) to constrain a hearer to reply; (4) to discourage a hearer from replying, and additionally (5) to express a speaker's expectations about the hearer's reply; (6) to show cooperation and/or compliance with the discourse partner; and (7) to facilitate repair in cases of breakdowns or misunderstandings.

The use of intonation to produce cooperation between and among speakers in structuring a discourse can be categorized as an interactive/discourse function. Brown et al. (1980) state that in British English, a speaker embarking on a new topic often begins the utterance in a relatively high pitch range (pp. 23–24). The second speaker, responding to the first utterance, also generally begins the response in a relatively high pitch range. In contrast, if a speaker asks a question on an already established topic, a low pitch range might be used, and the second speaker will likewise frequently reply to it in a low pitch range. This *pitch concord* or echoing by the second speaker is explained in terms of the second speaker cooperating with the first.

As for the function of intonation to conclude a topic, Brown et al. (1980) posit that "the end of the topic is typically marked by the repetition of lexical items already introduced" or by a fading away on expressions such as "and so on" or "that's how I see it," etc. Phonetically, this fading-away is realized by speakers "dropping low in their pitch ranges, fading away in amplitude, and leaving a long pause at the end of the turn" (p. 25). If, on the other hand, a speaker does not wish to terminate a topic, low terminals are frequently associated with non-finality in topic or in a turn, particularly when a speaker indicates that there is more to come on the same topic by some other means such as incomplete syntax.

Additional interactive/discourse functions of intonation include realizing a speaker's desire to dominate a listener and expressing a speaker's expectations about a discourse partner's reply (cf. Cruttenden, 1997). As an example of the latter function, Cruttenden describes a common type of tag question in English that is of the "reversed polarity kind," i.e., if the preceding clause is positive, the tag is negative, and vice versa. "These sorts of tag can typically take a tone which is high-falling, as in (a), or low-rising, as in (b)," e.g., (p. 89):

(a) He isn't `coming/ `is he?
(b) He isn't `coming/ ´is he?

The difference in meaning between the tags with the two intonations can be explained in terms of discourse: although both presuppose "no" for an answer,

the falling intonation in (a) does so much more strongly than the rising intonation in (b). Use of the fall in (a) presupposes only a slight possibility of "yes" while the rise in (b) presupposes a positive reply. At the same time, Cruttenden (1997) notes that one could also "give attitudinal labels to the tones, the high-fall being 'demanding' and the low-rise 'doubtful'" (p. 89).

The intonational cue for turn-yielding is described for American English as the use of any pitch level/terminal junction with a non-level tone at the end of a phonemic clause (cf. Duncan, 1972, 1973). Oreström (1983, pp. 31–32) and Brown et al. (1980, p. 24) note that, in British English, a drop in pitch and/or loudness may signal that the speaker is ready to give up his/her turn.[1]

With regard to using intonation as one of the cues that shows that a speaker is anxious to continue a turn in the conversation, Shapley (1989) finds that, in American English, various speaker intentions in discourse correlate with characteristic pitch levels: "The highest pitched and most likely to be stressed information occurs with interactive utterances, making sure the listener understands the speaker, has things correct, and follows the discourse, rather than on conceptual information" (p. xv). In addition, she found that discourse modes of narrating and conversing are differentiated by mean pitch values, with the pitch of conversing mode (interactive contexts) higher than that of narrating mode.

Cutler (1983) describes the interrelations between prosody and speech repair in order to illustrate speakers' conceptions of the functions of prosody. First, speakers use prosody to insure that errors that are likely to disrupt communication are decisively repaired: errors are corrected by using prosody to mark the error, i.e., to distinguish it prosodically from the original utterance and to direct the hearer's attention to it. Such corrections seem to be applied to errors at the lexical level or above. Second, prosodic repairs are used when the speaker fears the hearer will be misled into an incorrect interpretation of the utterance. "The function of prosody is thus seen to be, in the speaker's view, primarily concerned with the semantics or pragmatics of the utterance" (p. 91).

Most recently, Couper-Kuhlen (1997) expressed the view that there has been a major shift in emphasis with regard to the role or function of intonation in language that has resulted in three main strands of research: (1) intonation as part of grammar, which has expanded beyond the traditional domain of grammar to include not only syntactic types but also propositional attitudes as well, albeit context-independent; (2) intonation as related to information flow, providing a window into the speaker's consciousness; and (3) intonation as contextualization, which posits that prosodic features are linguistic signs that

are embedded in contexts. The interactive approach proposed by Couper-Kuhlen (1997) is substantially supported by the studies reported on in Couper-Kuhlen and Selting (1996a) and is based on the concept that intonation is linked to functions which derive from the situated use of language to accomplish interactional goals. "In an interactional perspective, analysts are consequently not looking for minimal pairs and distinctive functions. Instead they typically find that intonation and prosody have a *contextualizing* function" (p. 13). Prosodic features "function as part of a signaling system which — together with syntax, lexico-semantics, kinesics, and other contextualization cues — is used to construct and interpret turn-constructional units and turns-at-talk" (Couper-Kuhlen & Selting, 1996b, p. 25; cf. also Selting, 1992, 1995). The studies reported on in Couper-Kuhlen and Selting (1996a) include examinations of turn-continuations, of how listeners or recipients react to their interlocutors prosodically, of how prosody is used by speakers in repair situations (e.g., to contrast or contradict listeners' expectations), and of how prosodic aspects of speakers' repetition can convey meaning to hearers.

In summary, the importance of intonation in the organization of conversation and discourse began to emerge some two decades ago, due in large measure to Brazil's (1975) model of *discourse intonation* (which was an application of Sinclair and Coulthard's (1975) framework of *discourse analysis*). The work of other key figures such as Brown et al. (1980), Bolinger (1986, 1989), Cutler (1983), and Lambrecht (1994) have also been discussed in this chapter. For the most part, the aforementioned work dealt principally with linguist-generated and elicited utterances. The recent work of Chafe (1993), DuBois et al. (1992, 1993), and Couper-Kuhlen and Selting (1996a) is of critical importance as it represents an updated account of the functions of discourse intonation — ones based on data from natural conversations and authentic interactions.

Sociolinguistic functions

As noted in the introduction to this chapter, Couper-Kuhlen (1986) lists her sixth function of intonation as *indexical*, meaning that intonation establishes contrasts that allow hearers to identify speakers as members of different social groups, which include sex-groups, age-groups, socio-regional groups, and occupational groups. In addition, intonation is said by Couper-Kuhlen to identify individuals on the basis of idiosyncratic features. In this book, this indexical function is termed *sociolinguistic*.

To date there has been a limited amount of sociolinguistic research on intonation phenomena and dialects of English or of other languages. As Cruttenden (1997) laments, there is currently no book or article which compares in any detail the intonation of different varieties of English. The difficulties of such a comparison would be immense, including the problem of sorting through the numerous theories and descriptions of intonation that exist for individual dialects, e.g., RP in British English or American English, and also the fact that it has been impossible thus far to determine a one-to-one relationship between intonational form and function in any language or variety.[2] Cruttenden himself devotes a chapter to comparative intonation, where he briefly describes some of the differences among different regional varieties of English (e.g., dialects within England as well as varieties in Ireland, Scotland, Wales, Canada, Australia, New Zealand, the United States, the Caribbean and India). There are, however, very few studies of differences in intonation that focus on gender, on age/generational differences, or on socio-economic or occupational membership.[3]

In this book, I propose the sociolinguistic function of intonation to be one of the major categories since it parallels a major type of competence recognized by communicative approaches to L2 learning, and, similar to the applied linguistic and pedagogical literature, I use the term *sociolinguistic* functions rather than the Couper-Kuhlen's term *indexical* functions. Sociolinguistic and sociocultural competence are increasingly being emphasized as critical components of overall L2 competence, and in Chapters 5, 6, and 7, I show how the sociolinguistic functions of intonation can be attended to and taught. At the end of the next section on Perceptual Studies, I also include a brief discussion of studies conducted to date on the question of foreign accent in nonnative speech and how it is perceived by native speakers.

Perceptual studies

The studies discussed thus far have generally centered on functions of intonation from the perspective of the speaker. Before proceeding to the application of discourse intonation theory to the task of teaching intonation to language learners, one further theoretical aspect to be discussed is the function of intonation from the perspective of the hearer or perceiver. Intonation provides some redundancy to the linguistically coded messages and supplies extra cues to secure the full meaning that is being communicated. Selected studies on

how the functions of intonation are perceived by listeners or hearers are summarized below (cf. Cutler et al., 1997 for a literature review of prosody in the comprehension of spoken language).

Grammatical functions

Distinguishing sentence types

In studying how sentence intonation is perceived by hearers, Gårding and Abramson (1965) established three categories of intonation contours in American English that are perceived by hearers: "neutral statement," "yes-or-no question," and "counting in a series" (in addition to two that could be called attitudinal: "anger" and "delighted surprise"). Hadding-Koch and Studdert-Kennedy (1965) found for Swedish and American English three categories: "statement," "question," and "talking-to-yourself." Isačenko and Schädlich (1966) found three categories in German: "*nicht-Frage*" ("non-question"), "*progredient*" or "*weiterweisend*" ("progredient" or "pointing further"), and "*Frage*" ("question"). Johansson (1978) established three categories in Swedish and labeled them "declarative," "continuation," and "interrogative."

In a study of the perception of sentence intonation in Danish, Thorsen (1980) found that fundamental frequency (pitch) was closely correlated with subjects' judgments of utterances as declarative, nonfinal, or interrogative. The "most steeply falling intonation contours were identified as being declarative, the least falling ones as being interrogative and contours in the middle of the continuum as being nonfinal" (p. 1014). Thorsen also found that the two most influential parameters are the pitch levels of the last stressed syllable and the succeeding unstressed syllable, respectively, with the pitch change between the last stressed syllable of an utterance and the following unstressed syllable determining the steepness of the contour. These parameters are also determined to be most important in many of the theories of intonation that were discussed in Chapter 2.

Boundary marking in sentences

Berkovits (1984a) investigated fundamental frequency and durational measures in order to determine the acoustic features which signal finality in English and Hebrew. She found that these languages "make similar use of the various aspects of F_0 [fundamental frequency] in distinguishing between finished and unfinished utterances. Final segments in unfinished sentences of

both languages showed higher F_0 peaks, higher valleys and smaller F_0 falls relative to their finished counterparts. Unfinished sentences in both Hebrew and English exhibited continuation rises [… but] starting F_0 values did not differ in finished and unfinished sentences. Whereas segments were longer in sentence-final versus phrase-final position in English, finished and unfinished sentences in Hebrew revealed no durational differences" (p. 255).

Berkovits (1984b) then conducted a perceptual study of sentence-final intonation in English and Hebrew and found that listeners do perceive acoustic cues to sentence boundaries. Few errors were made by subjects in identifying sentences as "finished" or "unfinished." Although not all F_0 data could be correlated with how the utterances were perceived, one consistent result was that "the peak value of the final stressed syllable […] correlated significantly with subject responses, such that the lower the peak value the greater the tendency to identify the sentence as 'finished'" (pp. 302–303). An interesting finding was that the tendency to reprocess the acoustic cues in terms of syntax increased as less attention was directed to prosodic features.

In sum, a number of studies of English and other languages indicate that intonation contours and pitch height are cues for marking grammatical sentence types (e.g., high rising intonation for questions and steeply falling intonation for statements) and that intonation is also used to mark finished vs. unfinished sentences (i.e., at least in English, lower pitch, low-falling contours, and lengthening of final segments signal sentence finality).

Attitudinal functions

With regard to the perception of affective states in intonation contours, Scherer (1979) found strong agreement among subjects (native speakers of English) in assigning affective states to synthesized tonal sequences and speculated that the encoding and recognition of emotions in different tones, pitches, and rates might be innate. In a study with native English speakers and nonnative English speakers learning English, Luthy (1983) found that the control group of native speakers interpreted a set of "nonlexical intonation signals" (associated with expressions like *uh-oh* or *mm-hm* in English) in terms of rudeness or politeness, doubt or certainty, surprise or nonchalance consistently, while the nonnative speakers of varied L1 backgrounds often misinterpreted them. This suggests that the encoding and interpretation of attitude and emotion in pitch and intonation may also vary among languages.

Illocutionary functions

Geluykens (1987) experimented with the role of rising intonation for marking so-called "queclaratives" in British English (based on Halliday's system of tones). His perception experiment suggested that pragmatic factors, specifically Searle's (1969) "felicity conditions," play a decisive role in the recognition of an utterance with declarative form as a question, but that rising intonation by itself is virtually without impact. However, if pragmatic cues are not sufficiently strong to determine speech-act status, intonation can (but will not necessarily) act as a cue for determining question status. It should be noted that pitch range was not varied in this experiment, and Geluykens acknowledges that it might be worthwhile to investigate contours which start higher than the baseline.

Textual/discourse functions

Other perceptual studies of intonation have investigated both sentence and paragraph boundaries. Lehiste (1979), for example, examined what happens to a sentence when it is placed in different positions within a paragraph or uttered in isolation, particularly to see whether listeners are capable of deciding whether a sentence has been produced in isolation or as part of a larger structure. She concluded that listeners have certain expectations regarding the suprasegmental shape of a sentence as it might appear within various positions in the paragraph or in isolation. The first systematic phonetic cue is that a sentence is longer in isolation than when it is part of the paragraph. Sentences with the longest duration were most frequently (correctly) identified as having been produced in isolation. Another effective cue was the fundamental frequency on the first stressed word of a sentence. Listeners identified sentences with higher F_0 peaks on the first word as having been produced in the first position in a paragraph. This finding was replicated for Finland-Swedish by Enkvist and Nordström (1977), who concluded that the most conspicuous paragraph marker was the high pitch of the intonation peak on the paragraph-initial sentence.

In a second experiment, Lehiste (1979) found that "the most obvious difference between sentence boundaries and paragraph boundaries was the length of the pause that followed the boundary. Longer pauses were associated with perceived paragraph boundaries" (p. 106). She had hypothesized that "the difference between sentence and paragraph boundaries might lie in differences in the degree of pre-boundary lengthening," but the results of the study failed to confirm the hypothesis completely (p. 107). She concluded that "There is no

single phonetic marker, but fundamental frequency and timing both play a role. The beginning of the unit is signaled by a high pitch peak. The termination of the unit is signaled by a falling intonation contour that reaches the lowest level of the speaker's range (frequently dropping into laryngealization), and by a following pause that is longer, on the average, than a pause following a sentence within the unit. Further, and crucially, the termination of this unit is signaled by pre-boundary lengthening [...] Speakers use these cues, and listeners respond to these cues" (p. 108).

Kreiman (1982) repeated Lehiste's basic experiments with some modifications, e.g., using natural conversation and having subjects make sentence and paragraph responses during separate stimulus presentations. Subjects were presented with two versions (with and without semantic information) of a natural conversation and were asked to respond when they heard a sentence or paragraph boundary. When semantic information was present, the number of both very quick and very slow responses increased. Kreiman found in general that sentence boundaries are cued by non-level intonation contours (rises and falls), laryngealization (or "creaky voice"), pre-boundary lengthening, and the presence of and length of pauses — all unit-terminal phenomena (p. 163). Subjects generally agreed quite well as to where sentence boundaries occur; there was disagreement in cases in which terminal pitch contours did not match or coincide with grammatical or semantic information.

Kreiman found further that paragraph boundaries, on the other hand, are marked not just by terminal contours, as sentences generally are; rather, they are signaled by unit-initial cues, unit-terminal cues, or "cues which span boundaries and indicate both that one unit has ended and that another has already begun" (p. 163). She noted the complexity of interpreting paragraph results because in a conversation, participants do not wait until a new sentence has begun to see if their interlocutors have begun a new topic or not. However, subjects in Kreiman's studies did wait: they listened for cues at the end of one utterance and the beginning of the next, compared the two blocks of speech, and if the differences between them were great enough, they posited a paragraph boundary. She therefore concluded that there is a two-fold marking of paragraphs: "terminal contours suggest opportunity for change, and initial contours signal that change has actually occurred" (p. 172).

However, although these cues for paragraph boundaries can both precede and follow a change in topic, "either part of the composite signal can occur without the other. Topics may be changed without warning, or may be continued despite one party's wish to end them" (p. 172). This sort of optionality

with topic changes is probably also characteristic of turn-taking, i.e., "a conversant can be offered the speaking turn, or can take it without its being offered, but he cannot be compelled to take it (cf. Duncan & Fiske, 1977). Thus, one would expect any suprasegmental characteristics of turn boundaries to exhibit a structure similar to that of paragraph cues" (Kreiman, 1982, p. 174). In the following sub-section, the results of several studies on turn-taking cues are discussed.

In sum, perceptual studies of boundary marking at both the sentence and paragraph levels have revealed that several intonational cues function together in signaling these boundaries. For paragraph or discourse boundaries, in addition to pitch height and contours, laryngealization (or "creaky voice"), pre-boundary lengthening, and pauses longer than those between sentences are perceived as salient by hearers. However, unlike sentence boundaries, paragraph boundaries are marked not just by terminal contours. Rather, one of the most conspicuous paragraph markers was the high pitch of the intonation peak on the item-initial sentence and it thus seems that cues which span boundaries and indicate both that one unit has ended and that another has already begun are used by hearers to determine the macrostructure of a given discourse.

Interactive/discourse functions

Schaffer (1984) investigated the role of intonation as a perceptual cue to turn-taking in conversation. From a series of listening tests incorporating both face-to-face and non-face-to-face conversations, she isolated utterances and filtered them so that they were unintelligible but retained intonational information. She found no conclusive evidence that listeners have more auditory cues to work with in non-face-to-face conversations than in face-to-face conversations.

Subjects in the study were asked to identify the filtered utterances as either turn-initial or turn-final utterances. One finding was that turn-ends were correctly judged significantly more often than turn-beginnings, though not because of intonation cues. Rather, Schaffer hypothesized that "the answer lies not in the actual differences in the production of turn end vs. turn beginning cues (since I have found no evidence of such differences here), but in the listener's attention to and interpretation of these cues. Listeners may be more aware of cues to turn ends than turn beginnings because as polite conversationalists they will attach more importance to the end of the current speaker's turn, which will allow them to take over the floor without violating any rules of

conversation, than to the start of their own, which they are in control of in any event" (pp. 254–255).

The one intonational result that was significant in Schaffer's (1984) study was that rising F_0 and falling F_0 do not function in the same way as turn-end cues, counter to what Duncan and Fiske's (1977) "intonation-marked clause" cue suggested (p. 255). Rather, rising fundamental frequency was a stronger cue to the end of a turn than was falling F_0 in both face-to-face and non-face-to-face conditions. Based on her data, Schaffer concluded that syntactic and lexical characteristics appear to be used much more consistently as cues to turn status than intonation alone. However, she suggested that prosodic and verbal cues do interact with one another. At the same time, she conceded that the isolation of the test items from their surrounding conversational context may have been another factor contributing to the statistically insignificant results.

In a study of turn-taking cues in British English, Cutler and Pearson (1986) found the fundamental frequency contour of an utterance to be a major cue (p. 152). Their listeners found a downstep in pitch a good turn-yielding cue and a pitch upstep a good turn-holding cue. However, Cutler and Pearson caution that this perception may be ambiguous because many of the utterances the listeners found indeterminate also had upstepped or downstepped pitch. They suggest that other prosodic or vocal quality features that occurred in these utterances may also have functioned as effective turn-holding or turn-yielding cues.

Ford and Thompson (1996) studied listeners' ability to *project* the ends of "turn-constructional units" (TCU) in spontaneous speech, exploring the role not just of intonation but also of syntax and conversational pragmatics in the construction of turns. As their starting point, they used the *intonation unit* (IU), defined by DuBois et al. (1992) as "a stretch of speech uttered under a single coherent intonation contour."

Based on previous work by Crystal (1969), Cruttenden (1997), DuBois et al. (1992), and Schuetze-Coburn et al. (1991), Ford and Thompson (1996) characterized as influential in how pitch patterns are perceived both the degree and direction of pitch movement on a stressed syllable and a change in pitch relative to the speaker's preceding utterance (known as "pitch reset"). In addition, they also attributed a role in the perception of intonation units to such timing cues as an acceleration in tempo on initial unstressed syllables, prosodic lengthening of final syllables and a noticeable pause (0.3 second or greater) between intonational units.

One of Ford and Thompson's major findings was the high degree of coincidence among the syntactic, intonational, and pragmatic completion points. Of the three types of completion, syntactic completion points were the least reliable indicators of completion. Intonation, on the other hand, played a major role in determining *which* syntactically complete utterances were projected by hearers as complete units. In addition to intonation, pragmatics played a significant role in determining which syntactic completions were treated as complete by listeners. Ford and Thompson suggest that what listeners seem to do in projecting the ends of another's utterances is to pay attention to syntax, intonation, and the pragmatic content of those utterances — that is, to what the speaker is doing in the interactional context.

Ford and Thompson (1996) conclude that the three types of cues — intonational, pragmatic, and syntactic — converge to a great extent to define the turn-constructional units of conversations. The boundaries of these units are complex transition relevance places (CTRP), and in sequencing their turns, conversationalists orient themselves to these CTRPs. Intonational completion almost always involves syntactic completion and pragmatic completion; hence intonation units, as discussed in Chafe (1979, 1980, 1987, 1993) and DuBois et al. (1992), are a major component of CTRPs, and therefore of the turn-taking system itself. Projecting when a new turn could start must therefore centrally involve the perception of intonation units and pitch peaks within intonation units.

Another major finding of Ford and Thompson's study is the fact that speaker change correlates strongly with CTRPs. They interpret this as evidence that the units defined by the convergence of syntactic, intonational, and pragmatic completion are real for, i.e., are perceived by, conversationalists. Both speakers and hearers must orient to these units and design their own turns in response to them. It is at these CTRP points that participants in conversations typically start and yield turns.

Sociolinguistic functions

With regard to the sociolinguistic functions of intonation, a number of studies have dealt with the perception of "foreign accent." In a series of studies, Munro and Derwing investigated the role played by intonation in L2 speech production and found that (1) foreign-accented speech could be identified as spoken by a non-native speaker on the basis of nonsegmental (intonational) informa-

tion alone, even when utterances had been rendered unintelligible through low-pass filtering (Munro, 1995); (2) utterances by nonnative speakers required more processing time for comprehension by native speakers than the same utterances by native speakers (Munro & Derwing, 1995b); (3) "although the strength of foreign accent is correlated with perceived comprehensibility and intelligibility, a strong foreign accent does not necessarily reduce the comprehensibility or intelligibility of L2 speech" (Munro & Derwing, 1995a, p. 74); and (4) regarding the relationship among accent, perceived comprehensibility and intelligibility, accent ratings were generally harsher than perceived comprehensibility ratings, which in turn were harsher than actual intelligibility scores (Derwing & Munro, 1997). These results provide evidence in support of the claim that "although some features of accent may be highly salient, they do not necessarily interfere with intelligibility" (p. 11).

Anderson-Hsieh and Koehler (1988) found that prosodic deviance by nonnative speakers may affect comprehension more adversely than does segmental deviance. A subsequent study found that when nonnative speakers of English were rated by native speakers with regard to pronunciation, deviance in segmentals, prosody, and syllable structure all showed a significant influence on pronunciation ratings, but the prosodic variable proved to have the strongest effect (Anderson-Hsieh, Johnson & Koehler, 1992).

The above research is important for L2 teaching (and, of course, learning) because it suggests that prosodic deviance by language learners contributes significantly to the perception of foreign accent, affects comprehensibility and may affect intelligibility as well. Of interest to second language teachers is the implication that improvement in nonnative speaker comprehensibility is more likely to occur with improvement in prosodic (and grammatical) proficiency than with a sole focus on correction of phonemic errors.

In conclusion, this chapter has demonstrated that the functions of intonation cannot be divided into neat, clear-cut categories since they typically involve the grammatical, attitudinal, information-structural, illocutionary, pragmatic, and sociolinguistic domains of conversations and discourses with much potential overlap. These different types of functions were examined both from the perspective of the speaker (i.e., the production of intonational markers) as well as from the perspective of the listener (i.e., how intonational markers or acoustic cues are perceived). It is precisely because there is *not* a one-to-one correspondence between intonational *form* (e.g., particular pitch contours or pitch levels) and *function* and because intonation functions beyond the sentence level that

discourse intonation has become increasingly important in linguistic theory as well as in second language acquisition research, as will become evident in the next part of this book.

Notes

1. For German, Ehlich (1979) investigated the four distinctive phonological *tones* used with *hm* by a hearer to steer the other speaker's speech activity, and Winkler (1980) conducted a phonetic-paralinguistic analysis of conversational organization. Schmidt (1983) reported on an empirical study of "recipient signals" such as *mhm* and their various possible intonations in French and German. Gibbon and Selting (1983) studied the intonation of a stretch of German dialogue with respect to its function in discourse development. They showed that certain formal properties of intonation, e.g., pitch accents and their patterns, have strategic value for speakers and characterize both turn-taking processes and the semantic development of a discourse.

2. In their survey of the intonation systems of twenty languages, Hirst and Di Cristo (1998) devote one chapter each to American English and British English.

3. Examples of the very few studies on non-regional sociolinguistic aspects of intonation are Loveday (1981) and van Bezooijen (1995), who studied pitch use and gender differences in English and Japanese, and Japanese and Dutch, respectively.

Table 1. Functions of intonation

(1) **Grammatical**	(2) **Attitudinal**	(3a) **Discourse**
grammatical structure	affective, attitudinal meaning	information structure
Speaker's perspective	*Speaker's perspective, attitude toward partner and/or situation and toward content of proposition*	*Speaker's perspective, based on presuppositions about hearer and about hearer's sentence-level knowledge*
contrast sentence types: statements, questions, commands	express attitudes, emotions, e.g., sarcasm, puzzlement, anger, excitement, irony, happiness, sadness, animosity, friendliness, reservedness, cheerfulness, perfunctoriness	(1) provide sentence-level focus, emphasis; given vs. new information (2) mark sentence boundaries
Couper-Kuhlen, 1986; Crystal, 1985	Couper-Kuhlen, 1986; Gimson, 1980; Lambrecht, 1994; O'Connor & Arnold, 1961; Selkirk, 1984; Uldall, 1964	Couper-Kuhlen, 1986; Halliday, 1967a; Lambrecht, 1994; Selkirk, 1984

(3b) **Discourse**	(3c) **Discourse**	(3d) **Discourse**
illocutionary, speech act	textual/discourse	interactive/ discourse, contextualizing function
Speaker's perspective or intent; expectation toward partner and/or content	*Speaker's perspective, based on presuppositions about hearer's discourse-level knowledge*	*Speaker's intention, attitude, and reaction toward the conversation itself (e.g., turn-taking)*
convey intentional force, e.g., inquiry vs. suggestion; statement vs. complaint	(1) provide coherence between propositions (2) mark shared knowledge of discourse (old information) (3) mark discourse-level prominence, focus, newsworthiness (new information) (4) mark boundaries in a discourse (5) indicate expectations about hearer's reply	control interactive structure: (1) continue with topic (2) start new topic (3) constrain hearer to reply (4) discourage hearer from replying (5) show cooperation and/or compliance (6) facilitate repair
Couper-Kuhlen, 1986; Lambrecht, 1994; Liberman & Sag, 1974; Sag & Liberman, 1975	Brazil et al., 1980; Couper-Kuhlen, 1986	Brazil et al., 1980; Brown et al., 1980; Couper-Kuhlen, 1996a

Functions of intonation		
(4) Sociolinguistic		
sociolinguistic		
Hearer's and speaker's perspectives, based on knowledge about interlocutor		
allow hearers to identify speakers as members of different sociolinguistic groups, e.g., gender groups, age groups, socio-regional groups, occupational groups; signal that speaker belongs to a certain sociolinguistic group		
Couper-Kuhlen, 1986; Hurley, 1992; Loveday, 1981; Luthy, 1983; van Bezooijen, 1995		

Part II

Applied linguistic research: Intonation in L2

As discussed in Part I, linguistic theories of intonation have focused in recent years on the forms and functions of intonation above and beyond the sentence level. Even the generativist treatments of intonation have dealt with discourse-level phenomena, though the highly theoretical and rule-based nature of their descriptions make such treatments difficult to apply to L2 learning and teaching. This increasing emphasis on pragmatic analyses and on empirical studies of intonation in natural conversation, i.e., into how intonation is used to convey the communicative intentions of speakers, parallels the current focus in applied linguistic and second language acquisition (SLA) research on communicative competence and proficiency. Part II of the book thus examines the basic shift in applied linguistics towards the overall goal of more communicative, discourse-level competence and the renewed interest and emphasis on pronunciation with an increasing focus on intonation. The goals of pronunciation teaching are revisited, and it is suggested that the importance of intonation be noted in standard proficiency guidelines.

Chapter 4 reviews the research agenda of the past with regard to the teaching of intonation, examining the various factors involved in the learning of pronunciation in a second or foreign language and noting that the traditional teaching approach based on structural linguistics is found to be restricted to the word and sentence levels. Chapter 5 examines more recent research with its shift in emphasis towards discourse-level pronunciation teaching. The field of ESL/EFL has taken the lead in teaching pronunciation at the discourse level, and the increasing focus on communicative proficiency in second and foreign language methodology shows promise that intonation will receive more attention in the near future.

Chapter 4

Research agenda of the past

Structuralism and the first attempts to teach intonation

In the 1940s, 1950s, and into the 1960s, pronunciation was viewed as an important component in both the audiolingual methods developed in the U. S. and the British system of situational language teaching. Morley (1991) explains that language was generally viewed in both schools of language teaching as:

> consisting of hierarchies of structurally related items for encoding meaning. Language learning was viewed as mastering these forms, the building blocks of the language, along with the combining rules for phonemes, morphemes, words, phrases, sentences. The pronunciation class in this view was one that gave primary attention to phonemes and their meaningful contrasts, environmental allophonic variations, and combinatory phonotactic rules, along with structurally based attention to stress, rhythm, and intonation. (pp. 484–485)

In the decades from the mid-1960s to the mid-1980s, attention to pronunciation waned, with two general changes in the approaches to teaching pronunciation. The first was to teach articulatory phonetics in an essentially structuralist fashion, i.e., sounds to be learned were isolated, they were contrasted with the L1 sounds, and exercises for practicing them were provided. The second, a fairly common response from the late 1960s, through the 1970s, and into the early 1980s, was simply to avoid or ignore pronunciation altogether.[1] In a state-of-the-art paper on second language pronunciation learning and teaching, Leather (1983) suggested two reasons pronunciation had been ignored and had often been thought to be a dispensable component of second- or foreign-language learning. First, pedagogical concern had shifted from linguistic form to communicative function and focused on "getting the message across" rather than on "getting the sounds correct." Second, theoretical linguistics had not provided helpful input for teaching applications: the phonemic approach of structuralist linguistics had not proven to be enormously successful, and generative linguistics had offered little in the practical realm. Morley (1991) suggested that

both the process and the product were seen as flawed: "The process, viewed as meaningless noncommunicative drill-and-exercise gambits, lost its appeal; likewise, the product, that is, the success ratio for the time and energy expended, was found wanting" (p. 486). There were questions about whether pronunciation could be taught directly (and learned) at all. In addition, there were dogmatic claims that adults were simply unable to acquire L2 phonetics (cf. Scovel, 1988).

However, in the decade and a half since Leather's article, from the mid-1980s to the present, some inroads have been made. As previously suggested, the main aims of this book are to build on the positive trends of the 1980s and 1990s towards the revival of teaching pronunciation. I will demonstrate how correct and appropriate use of discourse intonation greatly enhances the communicative functions of language and that it should therefore be taught rather than ignored. In addition, I will attempt to bridge the gap between linguistic theories of intonation, second language acquisition (SLA) theories, and practical applications to the teaching of pronunciation and intonation. This chapter will examine closely the areas of research suggested by Leather (1983) that address the rationale for L2 pronunciation teaching in general. The specific focus will be on the problems and shortcomings, both theoretical and methodological, of describing and teaching intonation. In the next chapter, recent state-of-the-art references will form the basis of an updated rationale for discourse intonation pedagogy and directions for future research.

Goals of pronunciation teaching

A first research question is "What are our pronunciation goals?" We must decide whether we are striving for "comfortably intelligible" pronunciation, "near-native accents," or "socially acceptable" pronunciation.[2] In the past, it seems that the first option, "comfortably intelligible" pronunciation, was all that was striven for by language teachers, particularly because accent-free speech by adult learners seemed to be nearly impossible or at least very rare. Although "intelligible pronunciation" seemed to mean having a potentially near-native accent, learners were only taught how to pronounce individual segmental sounds (vowels and consonants). Therefore, even if learners did master all of the individual sounds of an L2, their speech melody and rhythm, or intonation, were often transferred from their L1–meaning that they would still speak the L2 with "an accent." If our goals in teaching pronunciation are

that students attain "near-native" and "socially acceptable" pronunciation, however, then teaching suprasegmentals (stress, rhythm and intonation) will be an important complement to the teaching of segmentals (the consonants and vowels of old-style pronunciation teaching).

In order for learners to communicate in a socially acceptable way, they need to know such interactional conventions as how to ask questions politely and how to interrupt more than they need to know how to pronounce individual sounds perfectly. Although, in the United States, the *ACTFL Guidelines* (American Council for the Teaching of Foreign Languages) for foreign language learning do not address pronunciation at great length, the goals of the *Guidelines* are divided equally among *linguistic, sociolinguistic, discourse,* and *strategic* competence — but the *Guidelines* prescribe more than what is conventionally taught (cf. Byrnes et al., 1989; Omaggio Hadley, 1993, pp. 5–8). As with the *ACTFL Oral Proficiency Interview*, in both the standardized TSE (Test of Spoken English, 1992) and the SPEAK (Speaking Proficiency English Assessment Kit) Test, ESL/EFL speakers are evaluated with regard to categories of pronunciation, grammar, fluency, and comprehensibility. However, no explicit attention is paid to intonation, nor are speakers given the opportunity to demonstrate *sociolinguistic, discourse,* or *strategic* competence.

One crucial sociolinguistic question in this context is what "accent" or style of pronunciation should be taught to L2 learners. Bailey (1978) suggested that more than one "style" of pronunciation should be taught. Beebe (1980) found that learners are able to acquire style-shifting or register-shifting ability in L2, an ability in which suprasegmentals play a role insofar as rhythm and reduction of vowels and syllables are concerned. In many languages, unemphasized words become more and more reduced as speech becomes more rapid or more informal. As an example of syllable reduction in English, Gilbert (1984) noted that auxiliaries (*is, will, have*, etc.) are "*not* contracted in formal written English, but they usually *are* contracted in the spoken language and in informal writing," so that the question of what students should be taught needs to be considered at the start (pp. 31–32).

In sum, if our goal is both "near-native" and "socially acceptable" speech, if students are to develop comprehensive language competence and proficiency, i.e., linguistic as well as sociolinguistic, discourse, and strategic competence, then learning appropriate pronunciation registers can no longer be considered "icing on the cake" but must be made an integral part of teaching and learning a language and striving for comprehensive language competence. Students must be taught how to express nuances of meaning beyond the face

value of individual words or phrases (e.g., intent, politeness, agreement, or disagreement) and how to argue or to keep the floor — all aspects of communication in which intonation plays a crucial role.

Phonetic and linguistic foundations

The traditional theoretical linguistic basis for the learning and teaching of pronunciation was a focus on the *segmentals*, i.e., the articulatory phonetics of *individual sounds*. Earlier, intonation had not been extensively researched either theoretically or acoustically and was considered a luxury in terms of teaching. However, since the early 1980s, there have been a number of appeals to reverse the emphasis from segmentals to prosodic patterns (cf. Chun, 1988a, 1988b; Leather, 1983; Morley, 1991; Pennington & Richards, 1986; Yule, 1989), and these new directions will be taken up in the next chapter.

Linguistic research on intonation, including acoustic phonetic, phonological, sociolinguistic, and pragmatic research began in the late 1970s (see Chapters 2 and 3) and has continued to flourish ever since. However, practical applications for teaching intonation to language learners have not emerged as rapidly. Although mention of prosody, suprasegmentals, and intonation has not been totally absent in either second/foreign language research literature or in pronunciation textbooks or handbooks, intonation is only slowly gaining in status as an integral component of pronunciation and overall language competence.

Historically, there have been strong interrelations between L2 pronunciation teaching, phonemic theory, and the development of transcription systems as outlined by James and Westney (1981). Typically, the sounds (segmentals) of a language were categorized into a set of phonemes, and articulatory descriptions of each sound informed learners as to where and how these sounds were to be produced. Reference to a standard transcription system such as the International Phonetic Alphabet made it possible for learners of any L2, in principle, to contrast phonemes in L2 and L1 as well as similar phonemes within L2 using minimal pairs.

In terms of the suprasegmentals, there were generally two main schools of thought until the 1980s. One of these was described by Leather (1983) "… most pitch training in non-tone languages until recently was limited to associating pitch contours and sequences with sentence types (declarative, interrogative, etc.) and attitudes like 'impatience' and 'surprise'…" (p. 200). As discussed in Chapter 2, the American structuralists were the ones to associate pitch se-

quences with sentence types: they attempted to analyze stress and intonation in terms of phonemes. For example, for English (and German), four phonemic pitch levels were proposed, numbered 1–4 with 1 being the highest and 4 being the lowest pitch and with the symbol ° indicating the most prominent syllable in the sentence. Intonation was then described in terms of sequences of pitch levels, e.g., 3 °2 4 with falling terminal pitch representing the intonation pattern for neutral declarative sentences, while 3 °4 1 with rising terminal pitch represented a surprised question, and questions with a 3 °2 2 sequence and rising terminal pitch requested corroboration (see sub-section in Chapter 2 "Phonemic or Levels Approach").

3 °2 1	3 °2 2
She lives here.	She lives here?

Similarly, for stress, three or four phonemic levels of stress were postulated, i.e., syntactic, primary, secondary, and weak stress.

The British, on the other hand, preferred the so-called tune or prosodic approach (as opposed to the phonemic approach) and associated pitch contours with both sentence types and attitudes or emotions. They listed a finite inventory of tunes for a given language and proposed various notation and transcription systems (e.g., O'Connor & Arnold, 1961 for British English and Fox, 1984 for German).

Generative phonology, which has flourished since Chomsky and Halle (1968), as a whole was not intended to nor has it in fact had direct applications to pronunciation teaching. However, it has made an indirect contribution in that the phonetic features and feature sets that have been proposed have been used to capture regularities wherever they might appear, though they have not been confined in principled fashion to the articulatory, perceptual or acoustic level as they would be in theoretical phonological theory (Leather 1983, p. 201). In addition, although increasingly more is being done along generativist lines to account for the pitch patterns of intonation (cf. Pierrehumbert, 1980 and others cited in Chapter 2), applications to the teaching of intonation have not been forthcoming. Furthermore, the somewhat daunting-looking notations of the feature and rule calculus have doubtless discouraged many a teacher from attempting to understand it, much less teach it.

The most recent approach to teaching pronunciation, and one that has received increasing attention in research, is *discourse intonation*, i.e., how intonation is used beyond marking sentence types or expressing attitudes to serve pragmatic functions in discourse (see Chapter 3).

Briefly, intonation can function in discourse to mark such boundaries as between sentences, paragraphs, topics, conversational turns; to control interactive structure or organize conversational exchange, e.g., to constrain a hearer to reply or to discourage the hearer from replying; to continue an established topic or to signal a new topic; to indicate the expectations a speaker has about a listener's reply, to facilitate cooperation between and among speakers in structuring a discourse, to mark the shared mutual knowledge of a speaker and listener. The task for linguists is to describe the intonational features that fulfill these functions. Recent studies of discourse intonation emphasize the need to use authentic conversations and to describe utterances and their accompanying intonational features in the contexts in which they occur (cf. Couper-Kuhlen & Selting, 1996a).

Altogether, neither structuralist theories of phonetics and phonology nor generative phonology have provided adequate or effective bases for the teaching of pronunciation to L2 learners. What began to emerge in the mid-1970s and then with greater acceleration in the 1980s was research into the pragmatic and discourse aspects of intonation. In the next chapter, I will draw on the pragmatic and discourse sub-disciplines of linguistics in combination with more recent phonetic and phonological theories to suggest new directions for applied linguistic research on intonation and applications to L2 pedagogy.

Transfer from L1

One factor in L2 acquisition that results in a so-called "foreign accent" in L2 is the transfer of patterns of pronunciation from L1. Studies of both segmentals, such as Bieritz (1974), Flege (1981), Flege and Port (1981), and Politzer and Weiss (1969), and prosodic phenomena, e.g., Pürschel (1975) and White (1981), have shown that some recurrent phonetic patterns in learners' productions can be accounted for by "interference" or "negative transfer" from L1.

There are different degrees and consequences of transfer. Inevitably, transfer of sounds and/or intonation from L1 to L2 will result in a "foreign accent," which need not hamper understanding. More serious, though, would be problems of comprehension or misunderstandings that arise from so-called "negative transfer." Nash (1972) suggested that whereas transfer at the level of the phoneme is "self-limiting," transfer at the prosodic level is more serious because it is "cumulative." Loveday's (1981) study made clear that prosodic transfer at the level of politeness-related and sex-related pitch patterns may

have direct sociocultural consequences. Specifically, Loveday's Japanese female subjects adopted an extremely high pitch to express politeness in Japanese, clearly separating themselves acoustically from Japanese males in the same circumstances. On the other hand, the pitch of males speaking English as an L1 was considerably less differentiated from L1 English-speaking females. Loveday thus concluded that pitch and intonation are employed for distinctly different sociosemiotic functions in different language communities.

Edmondson et al. (1984) reported that German speakers of English tended to mix different kinds of messages and intonation contours, e.g., they used the "uncertain" rising intonation contour in declarative sentence types and produced a so-called "contradictory communicative effect" because they did not intend to indicate uncertainty, but their delivery was so perceived by native speakers of English (p. 121). Similarly, Gumperz (1982) reported that Indian and Pakistani servers in a British cafeteria irritated native speakers by saying "*Gravy*" with falling intonation when offering gravy, as opposed to the native-speaker servers who encoded polite uncertainty by saying "*Gravy?*" with rising intonation.

As these findings suggest, there is increasing awareness that more emphasis must be placed on suprasegmentals than on the segmentals — or that they should at least be given equal attention — if we are to help learners most closely approximate a native-like accent and, more importantly, help students avoid pragmatic and sociolinguistic types of misunderstandings due to transfer from L1. In the next chapter and in Part III, I will address the issue of how to make inroads in this area.

Constraints on pronunciation mastery

Whatever one's views or opinions of the "critical period hypothesis," the hypothesis that after a certain age language can no longer be acquired as easily or as "perfectly" as one acquires one's first language (cf. Lenneberg, 1967), it is critical to distinguish between acquisition of *language* and acquisition of *speech*. Many studies of the so-called critical period lump together the acquisition of phonology, morphology, lexicon, and syntax. Moreover, the inability of adults to acquire native-like phonology, especially the complex suprasegmental features that identify native speech, is often ignored. Nevertheless, most researchers would agree that there are a number of factors that apparently constrain learners' achievements in second-language pronunciation. Such constraints on

pronunciation mastery would include factors like age; sex; personal variations in aptitude, motivation, and attitude; personality; individual rather than developmental differences, e.g., patterns of hemispheric specialization in the brain; psychological variables like "empathy," "intuition," "self-esteem," and "flexibility of ego boundaries" (cf. Conrad, 1991; Moyer, 1999).

Although studies such as Olsen and Samuels (1973) and Snow and Hoefnagel-Höhle (1977) have concluded that there is no evidence for a critical period even at the phonological level, intonation and other suprasegmental information have **not** generally been studied. One of the few studies on the acquisition of phonology that **did** include the acquisition of suprasegmentals (albeit indirectly) is Neufeld (1978), in which the data seem to refute the critical period hypothesis. In that study, Anglophone subjects were taught to reproduce lengthy sound sequences in three non-Indo-European languages, one of which, Chinese, was a tone language. The results suggested that adults appear to be able to acquire native or near-native pronunciation proficiency in the sound patterns of new languages — that is, they have not entirely lost their ability to perceive and produce novel sounds. However, in this particular study, subjects were given neither rules of the language nor meaning of the target utterances. It is thus quite possible that the subjects were processing only the phonetic material and were therefore imitating sounds without attaching any meaning to them or using them to communicate — making the validity of the findings questionable.

Since its formulation by Lenneberg (1967) and Krashen's (1985) reformulation, many applied linguists — though certainly not all — have ascribed to the critical period hypothesis. Interestingly enough, in a later article, Neufeld (1980) found that age is not a factor in the acquisition of linguistic competence **except in** the domain of L2 phonology or pronunciation (p. 296):

> Insofar as the adult's acquisition of linguistic competence is concerned, I see little evidence to support neurophysiologically-induced language learning disability. The only consistent indicator of adult inferiority that I can find is the average learner's inability to get rid of a foreign accent when speaking L2. In other words, the disability, if one exists at all, may be less psycholinguistic than psychomotor in nature. While adult students may know what their second language should sound like, many may find it difficult to get their vocal apparatus to obey cerebral instructions.

Others, such as Scovel (1988), firmly believe that accents stem from biological constraints. He reviewed experimental studies and found that if one picked up a second language after about the age of 10 to 12, one ended up "easily

identified as a nonnative speaker of that language" (p. 123). Adams (1979), on the other hand, attempted to identify the rhythmic and stress difficulties faced by foreign learners of English but was unable to discover an anatomical or acoustic correlate of the rhythm and stress differences manifest in the speech of native and nonnative English speakers.

In sum, it is precisely because of the fact that pronunciation is one of the most difficult facets of language in which to achieve native or near-native competence that we must seek ways to help learners overcome some of the most challenging aspects of pronunciation. Very few studies have been done on the acquisition of suprasegmentals, and more work is needed to gain insights into why prosody is one of the first components of language acquired by children but one of the last that adult language learners can potentially master. Those insights into the acquisition of suprasegmentals that have been made will be discussed in the next section and the next chapter.

Factors in L2 phonological acquisition

In discussing the factors involved in L2 phonological acquisition, Leather (1983) suggested that, as with other areas of language pedagogy, "advances in pronunciation teaching might well depend upon better understanding of pronunciation learning" (p. 202). He included among the factors in achieving a better understanding of pronunciation learning: (1) the relationship between learners' perception and production, (2) the fact that different measures of articulatory and perceptual training may be called for because acquisition of intonation may be different from acquisition of vowels and consonants, and (3) the importance of auditory feedback. In this section, the general factors in L2 phonological acquisition are summarized with particular focus on the acquisition of *prosodic* sound patterns.

Relationship between perception and production

Although the relation between speech production and speech perception is not fully understood (cf. Conrad, 1991 and Sanders, 1977), it has usually been tacitly assumed in pronunciation teaching to be direct and simple. Neufeld (1978), for example, had proposed that correct imitation of target sounds presupposes correct perception of those sounds. However, Neufeld (1980) cautioned that a distinction must be made between acquisition phenomena

and performance phenomena because "there is frequent asymmetry in the adult's receptive and productive performance in L2 at the phonological level" — that is, there may be a discrepancy between what an L2 learner can produce and what that same learner knows or can perceive (p. 295).

The question of the relationship between production and perception has yet to be answered adequately. Specifically with regard to second or foreign language learning, the question remains as to the exact nature of the cross-linguistic issues in perceiving the L2 sound system (including intonational patterns) and in contrasting it with the L1.

Berkovits (1984b), in a perceptual study of sentence-final intonation in English and Hebrew, provided support for perceptual effects of prosodic differences between English and Hebrew. Both native speakers of English and native speakers of Hebrew made more errors in identifying "unfinished" sentences in Hebrew than in English. Berkovits suggested that this was due to the fact that English uses both duration and fundamental frequency (F_0) in distinguishing between finished and unfinished sentences, whereas Hebrew makes use of F_0 alone. With fewer prosodic cues used in Hebrew, it was more difficult for either group of subjects to correctly identify unfinished Hebrew sentences. In the case of finished sentences, on the other hand, the two groups responded similarly to English sentences, but the native speakers of English (English-dominant subjects) made more errors on Hebrew sentences than did the native speakers of Hebrew (Hebrew-dominant subjects). The English-dominant subjects were more likely to identify a finished Hebrew sentence as "unfinished," which may again have been an effect of prosodic differences between the two languages. The English-dominant subjects may have expected final lengthening in finished sentences, and when this acoustic cue was absent, a sentence might have therefore been perceived as "unfinished." Berkovits concluded on p. 305:

> It thus appears that English-dominant listeners are affected by the absence of final lengthening in Hebrew to a greater extent than are Hebrew-dominant listeners, suggesting that perception of intonation in a weaker language may reflect acoustic properties of the dominant language.

Clearly, more cross-linguistic research of this nature is needed: first, on the perception of intonation patterns in L2; second, on whether correct perception must always precede correct production (or whether accurate perception and production can be achieved at about the same time or in either order); and third, whether perceptual training improves perception and/or production.

The next section summarizes some of the recent research on perceptual, auditory, and visual training of language learners, which will ultimately have implications for how intonation can and should be taught in the classroom.

Perceptual and articulatory training

There are various types of training in L2 pronunciation, including perceptual, articulatory and some combination of both. The theoretical justifications for perceptual training range from the weak claim — that through attention to the new phonetic contrasts of L2, learners will subsequently be able to acquire the necessary set of new phonetic categories for perception of L2 speech — to the strong claim that a phonetic contrast in the target language must be perceived before it can be mastered in production.

On the basis of perceptual (discrimination) experiments, Henning (1966) suggested that learners achieve better pronunciation in a foreign language after finishing a sound discrimination course. However, the course described in the study dealt only with segmentals, i.e., vowels and consonants. Feldman (1973, 1977) used electronic synthesis to simplify several prosodic features that were then presented to hearers for discrimination. After this initial training, natural speech models that contained the same prosodic contrasts were presented, and positive results in discrimination were obtained.

't Hart and Collier (1975) sought to provide Dutch learners of English with information about English intonation with the goal of influencing their intonation in English. They found that "four aspects of pitch change are perceptually relevant: (1) direction of pitch change (rise, fall, or level), (2) range of pitch change (difference between high and low levels), (3) speed of pitch change (how abruptly or gradually the change happens) and (4) place of pitch change (in sentence, word, or syllable)." These four relevant aspects of intonation were then explained and demonstrated to the subjects in a 12-minute tape, using single tones and speech samples generated by a computer. The different pitch changes were also represented in the tape. Though the tape itself was only concerned with the perception of intonation, it caused a significant improvement in the learners' production of intonation.

However, although *perceptual* training can influence phonetic perception (and possibly production) in the short run, Leather (1983) suggested that it does not always permanently "re-tune" the perceptual system. Similarly, one cannot automatically assume, with regard to *articulatory* instruction, that the relation between speakers' articulatory activity and their acoustic output is a

simple one, such that fine tuning of the acoustic output can be effected through small adjustments in the vocal tract. Stevens (1972) and Stevens and Perkell (1977) found rather that the relation between articulation and output is not generally linear, but "quantal": within a particular range of an articulatory parameter, "the acoustic attribute is relatively insensitive to perturbations in the position of the relevant articulatory structure." They found specifically that speakers were able to learn to modify their articulatory output to within a good approximation of a target sound, but that fine tuning to achieve the exact sound was much more difficult (though it might possibly be facilitated by perceptual training).

Leather (1983) asserted that different classes of sounds are acquired differently. Articulatory instruction depends upon *orosensory* perceptions (sensory perceptions related to the mouth/oral cavity), but the sensory feedback that speakers "receive" and perceive of their own speech is much greater for consonantal articulations (e.g., "the <th> sounds" in English [θ] and [ð]) than for vocalic ones (e.g., <ö> umlauts [ø] and [œ]), and even less for a prosodic feature like voice pitch. Specifically, speakers can more readily perceive that their tongue is between their teeth in the production of <th> than that their tongue is in the correct fronted position for the <ö> umlaut. Very difficult for the great majority of speakers is perceiving whether they are producing a pitch pattern correctly. Mallard et al. (1978), for instance, found that laryngeal sensory perception played only a very small part in the production of voice pitch. Auditory information instead appeared to be of primary relevance in monitoring pitch production in a tone-matching exercise, though this was not the case for intonational patterns. They concluded that intonation patterns appear to be independent of peripheral monitoring.

Other studies have also questioned the usefulness of using an auditory perceptual-training approach in the classroom. For example, de Bot's (1981a) study found that learners who heard a playback of their productions appeared to perform worse in subsequent attempts to improve their intonation. By contrast, however, de Bot and Mailfert (1982) showed that improved perception of intonation due to training led to improved production. They found that a 45-minute training session in the perception of intonation did result in a statistically significant improvement in the production of English intonation patterns by both Dutch and French students in pre- and post-tests. However, questions remain about the long-term effects of such brief training as well as about the effects of more extensive training.

A year later, de Bot (1983) demonstrated that explicit instruction may be effective in helping L2 learners master some of the prosodic aspects of language. Specifically, he found that visual and audio-visual feedback provided by computers was helpful to L2 learners and was more effective in promoting intonation learning than auditory feedback alone. Subjects who viewed the pitch contour of sentences on a screen and heard the sentences spoken performed better on imitation post-tests than subjects who only heard the sentences but did not view the pitch contours.

A number of other studies have been conducted to evaluate discrimination training, articulatory instruction and exercises in imitation. These different types of training were used alone and in varying combinations, and the results were also varied. Lane and Schneider (1963), for example, found that discrimination training alone did not bring about better production of Thai lexical tones by non-native learners. One possible reason for the divergence of the results may be that there is perhaps no simple and symmetrical relation between a learner's perceptual ability and her/his productive skill, and that these abilities may not progress in parallel.

In sum, although the learning of suprasegmentals might be enhanced by different types of training, it is not yet known what the ideal combination might be. More studies are needed to determine whether an ideal combination of training types exists and to determine whether perceptual and productive abilities in language learners are or are not symmetric and do or do not progress in parallel. This issue is addressed in more detail in Chapters 5 and 7.

Importance of feedback

Leather (1983) summarized the research on feedback with regard to segmentals during speech production (pp. 202–3):

> One focus of both theoretical and applied research has been the part played in learning by self-perception based on feedback during speech production. Three sources of control feedback in speech are usually recognised: auditory, tactile (from the surface of the articulators) and proprioceptive (from receptors embedded in the muscles). […] Vowel quality appears to be determined more by auditory feedback, while consonant characteristics are determined more by sensory feedback.

In terms of research on feedback with regard to suprasegmentals, de Bot (1980, 1981a) cautioned against assuming that any type of feedback, no matter in

what form, would have a positive effect on the acquisition of intonation or on learners' attempts to improve intonation. Despite characterizing experimental or instrumental measurements of intonation as an improvement over the impressionistic basis of traditional intonation teaching methods and handbooks, he warned that simply being able to demonstrate aspects of intonation physically and perceptually does not necessarily render the process useful from a teaching point of view.

Subsequent studies by de Bot (1983) using visual and audio-visual feedback with computers showed that audio-visual feedback is more effective in intonation learning than auditory feedback. Learners who listened to English sentences and then imitated them and were able to view the pitch (F_0) contours of both the native speaker model and their own imitation made greater improvements in their own production of English sentences than did learners who only listened to the model sentences and their imitations but did not receive graphic representations of the F_0 curves. In addition, the type of feedback received had a strong impact on the learning behavior of subjects: those in the audio-visual feedback groups **repeated** the target sentences more often than the subjects who received only auditory feedback, whereas the latter **listened** more often to their own imitations than did the former (p. 347). The motivational impact of learners attempting more repetition may prove to be significant in the long-term. Weltens and de Bot (1984) showed that "feedback delay is not a critical factor when using a pitch visualizer for intonation teaching, but that the nature of the speech material [voiceless vs. voiced consonants, neutral vs. contrastive intonation…] does dramatically affect the quality of the visual feedback" (p. 79). In other words, the fact that feedback was not provided in "real-time" (i.e., with a minimum of delay between the production of a speech signal and its visualization) did not adversely affect the effectiveness of the visualizer. But if sentences contained many unvoiced sounds, the quality of the pitch curve was significantly poorer than sentences with few unvoiced sounds. This underscores the importance of having pedagogically meaningful feedback, i.e., feedback from computer software that explains the significance and detail in visual representations of pitch curves.

Thus, in sum, more research is definitely needed with regard to both the relationship between perception and production and the effect of feedback, particularly with regard to intonation (cf. also the next section on "Instrumental Aids") as well as with regard to speech in discourse — that is, speech beyond the sentence level. As will be seen in the next chapter, this type of research will

overlap with related disciplines, particularly psychology and educational psychology, in the area of learning theory.

Instrumental aids

The evidence cited in the immediately preceding section on training suggests that, as Leather (1983) asserts, "the learner makes use of auditory and sensory feedback to adjust his or her motor programs in accordance with the auditory-perceptual 'target' s/he has adopted, and possibly also in progressively redefining the target itself. Learners who acquire good L2 pronunciation in naturalistic conditions (i.e., without formal instruction) presumably make much constructive use of feedback" (p. 210). There are two main types of exercises that rely on constructive use of feedback. The first type, the traditional record-replay exercises in language laboratories that have existed for decades and that in principle enable learners to listen critically to replays of their production, aims to exploit auditory self-perception. In practice, however, this type of exercise requires learners to monitor themselves even though they generally lack the phonetic criteria for critical listening, since "the phonetic criteria which the learner needs [...] are precisely those which the exercise in question aims to develop" (p. 211). Instead, it is essential that the learner be helped to perceive the salient features of the L2 sound contrasts so that accurate production targets can more readily be established.

This dilemma can be remedied via the second main type of exercise: providing real-time feedback to the learner in the form of visual displays about aspects that would otherwise remain below the threshold of discriminability, as for example in the case of contrasts that are significant in L2 but not in L1. Various devices have been produced that provide real-time visual displays of articulatory activity and/or its acoustic consequences in speech output (cf. Abberton & Fourcin, 1975 and Fourcin & Abberton, 1971, who used output from a laryngograph to derive visual displays of vocal fold vibration as feedback to deaf speakers; Lindblom & Sundberg, 1971 on visual presentation of vocal tract information on vowel quality for teaching pronunciation).

In the past, costly and often one-of-a-kind devices were used to obtain pitch contour displays in an attempt to facilitate the learning of L2 intonation. Almost four decades ago, Vardanian (1964) reported on teaching English intonation through oscilloscope displays; Bluhme and Burr (1971) gave learn-

ers of Chinese visual representation of individual tones "with minimum delay"; Leon and Martin (1972) used a pitch analyzer for visual presentation of intonation; James (1976, 1977, 1979) developed a visualizer to study the acquisition of and to teach second language intonation, stress, and rhythm; and de Bot and Mailfert (1982), discussed above, used a computer to generate tonal speech samples demonstrating four aspects of pitch change that had been found to be perceptually relevant (direction, range, speed, and place of pitch change). Although several studies report positive effects of the use of visual displays of intonation among language learners (de Bot, 1983; Hengstenberg, 1980; Lane & Buiten, 1966), other studies found no such effects (Vardanian, 1964; Wichern & Boves, 1980), and none of these display systems came into widespread use.

In addition to the costliness and the technical limitations of earlier hardware and software, other problems in the studies in the preceding section derived from pedagogical issues related to the underlying assumptions of the respective software's content and design. Subjects in an experiment by Wichern (1979), for example, reported that they had considerable difficulties in relating visual and auditive signals. De Bot (1981b) also noted the problem in the application of visual feedback that the visual signal does not indicate what parts of the signal are perceptually relevant and what parts are irrelevant from a student's point of view (pp. 38–39). Only those parts that are perceptually relevant should be paid attention to and imitated as exactly as possible; students must be able to make these judgments themselves, or it must be possible to set or program the device to evaluate student performance as acceptable or unacceptable. However, this is actually quite a difficult task because first, there is generally **not** a one-to-one correspondence between a given sentence and its intonation pattern (because there is not always "one and only one" correct or appropriate intonation (or meaning) for a given utterance), and second, there is a very great variation in pitch and intonation that occurs within and among speakers of a given language.

De Bot (1981b) cautioned further that "Methods making use of visualisation as an aid in intonation learning wrongly imply that visualisation by itself is useful in teaching. Experimental phonetic research has shown that there is no empirical basis for a number of assumptions made in these methods. Especially, the tacit assumption that the pupil is an unbiased perceiver of whatever the teacher is presenting leads to many problems, because the perceptual abilities of the pupils in general have not been developed sufficiently and specifically enough for the perception of intonation" (p. 39).

In 1983, Leather discussed the existing technological applications in pronunciation training at that time (visual displays of speech parameters) and described future "instrumental aids" as consisting preferably of (p. 212):

> computer-managed pronunciation training which makes use of synthetic as well as natural speech models, which processes learners' productions to provide visual displays of salient features together with an assessment of phonetic accuracy, and which leads the individual learner through a series of perception and production training activities selected according to ongoing performance, while simultaneously compiling a detailed record of progress for teacher supervision.

He cautioned, however, that before applying any new technology in practice, one must first have a better understanding of L2 phonological acquisition and pronunciation teaching theory in order to know which natural learning processes can or should be facilitated. Both technological and pedagogical advances in the area of instrumental aids for the teaching and testing of pronunciation since the early 1980s will be examined in detail in the next chapter.

Conclusion

The foregoing discussion has been concerned with the principles in L2 pronunciation teaching as set forth in Leather's 1983 article on the state of the art of teaching pronunciation. While pronunciation teaching had traditionally been linked to classic articulatory phonetics and structural phonology, which generally restricted their scope to segmentals, an examination of more recent research has shown that (1) the scope is being broadened to include subdisciplines such as psycholinguistics, sociolinguistics, pragmatics, and discourse analysis as well as the broad field of speech science, (2) there is a shift in emphasis from individual sounds to entire utterances as they occur in context or in discourse, and (3) computer technology may play an increasingly important role in providing objective and individualized feedback to learners. However, Leather also warned that "As the focus in L2 pedagogy generally has moved towards more learner-centered designs, it has been realised that too little is known about pronunciation **learning** to advocate dogmatically any particular design for pronunciation **teaching**" (p. 198). In other words, along with the teaching of intonation, more research into the learning process must be done in order to promote the effectiveness of both segmental and suprasegmental training.

In the next chapter, I examine the most recent research and the definite shift in emphasis towards discourse-based pronunciation teaching. The ultimate goal is to lay firmer theoretical and empirical foundations so that classroom time spent on pronunciation will be better used and new technological aids can be adopted or rejected on the basis of clear theoretical arguments and empirical evidence.

Notes

1. Cf. Yule, 1990, p. 107 and Morley, 1987, preface. There are welcome exceptions, of course (e.g., Lantolf, 1976 for American English; Cook, 1968 and Currie & Yule, 1982 for British English; Gutknecht, 1975 and Kelz, 1977 on teaching English intonation to German learners; Preu, 1973 for teaching intonation to foreigners learning German).

2. Leather (1983) states that "What makes a non-native sound acceptable to a native has yet to be understood; it seems to involve not only how native-like his or her pronunciation is, but what variety of L2 it most approximates to" (p. 198). These social psychological and sociolinguistic aspects will not be dealt with here, but the reader is referred to Leather's (1983) summary of the literature. More recent studies investigating "foreign accent" are summarized in Chapter 5.

Chapter 5

Research agenda for the present and future

Communicative proficiency and discourse intonation

Until the mid-1980s, both research on the acquisition of phonology and materials for teaching pronunciation traditionally concentrated on lower-level speech phenomena, primarily the phoneme, the distinctive feature, and the syllable. During the period that the audiolingual method was widely used (1950s-1970s), pronunciation teaching was deemed important, but in the early phases of the cognitive and communicative approaches (1980s), attention to pronunciation seemed to diminish. Many teachers who adopted a communicative approach believed that getting one's message across (in whatever form) was the goal of second and foreign language pedagogy, as communication was more important than competence in any particular sub-component of language (e.g., phonology or grammar). Although the proficiency approach contains a strong focus on speaking competence, no specific guidelines have been given for how pronunciation is to be taught. Recent research in applied linguistics has exhibited a renewed interest in the acquisition of second language phonology in addition to emphasizing the communicative functions of language. As a result of, and as a reflection of, these changing research focuses, instructional materials (including computer software) are gradually also placing more emphasis on pronunciation and intonation. Leather and James (1991) described the resurgent interest in the acquisition of second language (L2) speech over the previous decade and called for more attention to higher-level patterning in the discourse domain, beyond the sentence level, in order to more fully understand L2 speech acquisition (p. 332).

To clarify the implications of these simple-sounding shifts in emphases, this chapter will focus on two main themes. First, the recent focuses in linguistic theory on intonation, pragmatics, discourse analysis, and conversation analysis as discussed in Chapters 2 and 3 will be linked to second language acquisition (SLA) research and to the changing emphases in applied linguistics and L2 pedagogy toward meaning- and context-centered instruction, i.e.,

toward communicative competence and proficiency. It will be shown that the English-teaching profession (both English as a Second Language (ESL) and English as a Foreign Language, (EFL)) has made more rapid and more extensive advances than its "counterpart" profession (the teaching of languages other than English, in many cases "foreign languages"), both in applied linguistics research and in the teaching of contextualized pronunciation and discourse intonation. One obvious reason for this is that ESL or second language learners have practical, real language demands and challenges in the environment in which the L2 is spoken whereas foreign language students generally have far fewer opportunities to use the FL with native speakers of that language.

However, in the cases of teaching both second and foreign languages, there are compelling reasons for advancing the current trend toward communicative proficiency. In order for language learners to become more proficient communicators and comprehenders, they must be taught how to use and perceive discourse intonation. The second theme in this chapter is a model for doing just that. The final sections of the chapter discuss how to incorporate these changing emphases and goals into the national (U. S.) *ACTFL Proficiency Guidelines* (American Council on the Teaching of Foreign Languages) as well as to integrate them into computer-based materials for teaching pronunciation and intonation.

Second language acquisition (SLA) research

As discussed in Chapter 2, the sub-disciplines of linguistics that have attracted increasing attention in recent decades include the study of intonation, pragmatics, discourse analysis, and conversation analysis. In addition, the field of second language acquisition (SLA) has emerged as a sub-discipline firmly based in fundamental, if not universal, theories of how second languages are acquired. This section contains a brief review of some of the more important studies of the last two decades that reflect the growing trend towards more pragmatic, discourse-oriented approaches to the acquisition of phonology, particularly the acquisition of suprasegmentals. Yet, despite the increasing interest in intonation, there is relatively little applied linguistic research devoted to the teaching and learning of intonation.

James and Leather (1987) attributed the genesis of their collection of papers, *Sound Patterns in Second Language Acquisition,* to three developments

since Chomsky and Halle's *The Sound Pattern of English* (1968): (1) the proliferation of a number of (fairly radical) theories of phonology beyond "generative" phonology; (2) the fact that the study of second language acquisition had gained some measure of autonomy as a research field; and (3) a growing recognition of some commonality of goals and data between the fields of SLA and post-generative phonology. The twelve papers exhibited a diversity of frameworks and paradigms but all derived their evidence from L2 speech perceptions and productions. The scope was limited to research motivated primarily by theoretical rather than pedagogical concerns, meaning that although some analyses and discussions may be applicable to the teaching of second language pronunciation, their first aim was to elucidate the psycholinguistic processes of the learner. A general question that is addressed in many studies of L2 phonological acquisition is: when a new sound system is learned, to what extent are deviations from L2 due to universal constraints on the development of sound structures and to what extent are they due to prior habitual use of the patterns of L1?

Three of the twelve papers in James and Leather (1987) dealt with suprasegmentals or intonation, which is a significantly larger proportion than would have been found in previous decades. The paper by Leather investigated how adult speakers of non-tonal languages (English and Dutch) perceive and learn to distinguish tonal patterns in an L2 with lexical tone (Chinese). The paper by Dechert and Raupach was concerned with the prosodic components of so-called "proceduralized speech" and suggested that the processing of L2 speech may be greatly aided by prosody. The paper by Cruz-Ferreira dealt with the perceptual processing of L2; she believed that misperceptions and misinterpretations due to defective comprehension of intonation often go unrecognized and therefore also uncorrected. The results of her experimental study of non-native intonational competence with English and Portuguese speakers indicated that listeners use various strategies to interpret intonational meaning. The transfer strategy can be either positive or negative, depending on whether a given intonational pattern is used in both languages. A lexico-syntactic strategy is used when the non-native listener cannot interpret the foreign intonational pattern, disregards intonation, and simply interprets the words of a sentence or chunk of discourse.

Similarly, Lepetit (1989) found in a study of the acquisition of French intonation by native speakers of Canadian English and Japanese that cross-linguistic influence in intonation is of central importance in the learner's acquisition of the target system and that the degree of complexity of this

influence should not be underestimated. In his study, intonation was limited to the domain of phonosyntax, where intonational cues are correlated with syntactic units.

In another collection of papers, Ioup and Weinberger (1987) specifically investigated interlanguage phonology and how the sound system of a second language is acquired. The study of "interlanguage" is a subset of the research field of second language acquisition; "interlanguage" is a learner's "principled, orderly, if complex, evolution of an approximating system" (cf. Selinker, 1972); it is a dynamic, ever-changing phenomenon. Three of the twenty-five chapters of Ioup and Weinberger's volume were devoted specifically to suprasegmentals. The Neufeld article was a reprint of his 1978 study "On the acquisition of prosodic and articulatory features in adult language learning" (discussed in Chapter 4). The study reported that adult learners could imitate sound sequences in unknown languages quite well though they were given neither the rules of the language nor the meaning of the imitated utterances. The paper by Ioup and Tansomboon was on the question of age differences in second language acquisition, specifically the ability to acquire tone in Thai. The paper by Broselow, Hurtig, and Ringen addressed the role of transfer from L1 in the perception of L2 prosody.

In investigating the question of transfer and suprasegmentals, Broselow et al. (1987) looked at how intonation is perceived by second language learners and presented two types of evidence that transfer plays a role in the *perception* of Chinese tone by English speakers. The first involved how easily each of the four Mandarin tones was perceived in different positions in strings of two and three syllables. Only one of the tones, the falling tone, seemed to be affected by position. This is the only tone that is markedly similar in its acoustic properties to a common English intonation contour, i.e., the falling contour that normally occurs in sentence-final position in declarative sentences. The falling tone in Mandarin was perceived significantly better when it occurred in final position — that is, in the position in which subjects are accustomed to hearing it in their native language — and it was more difficult to perceive when it occurred in an unfamiliar context (p. 351). This suggests that the L2 learner perceives or "tunes into" familiar pitch patterns automatically (as a matter of transfer), but only when they occur in a comparable position as in the learner's L1.

The second type of evidence of transfer involved the errors made by English speakers when they **mis**identified the Mandarin falling tone. Again, this misidentification seemed to be affected by the position of the contour in

the string: in sentence-final position, it was often thought to be a high level tone instead of a falling tone. The explanation for this was that the hearers analyzed only part of the tonal contour (the high level part) as having lexical significance; the rest of the contour (the falling portion) was presumably dismissed as part of the sentence intonation pattern. In both cases, the indication was that the English speakers were perceiving the second language strings in terms of their native language phonological system (p. 351). This result again suggested that phonological transfer from L1 to L2 was made by some L2 learners, if only in the perception of tones and intonation.

The studies discussed above reaffirmed the growing importance of prosody and intonation in current research into the acquisition of phonological proficiency, particularly with regard to the perception of intonational patterns. They suggest that, for L2 learners, it is not sufficient just to be able to produce appropriate patterns; learners also need to be able to correctly decode patterns they hear.

The seminal article by Leather and James (1991) described the state of the art with regard to the acquisition of second language speech but contained limited references to intonation and prosody. However, in the conclusions it was acknowledged, first, that current theoretical concern with suprasegmental issues (e.g., research on stress, prominence, and intonation in metrical and autosegmental theory) "can be expected to provide more differentiated and, potentially, more explanatory frameworks for L2 speech analysis," (p. 331) and second, that

> it is very noticeable that, although some research has been devoted to (sentence) intonation, most work has hitherto concentrated on what traditionally have been considered lower-level speech phenomena — the phoneme, the distinctive feature, the syllable, and so on. For a fuller understanding of L2 speech acquisition, this concentration will need in the future to be balanced by closer attention to higher-level patterning in the discourse domain. (p. 332)

A subsequent volume edited by James and Leather (1997) was concerned with the acquisition of second language (L2) speech. Only two of the seventeen papers focused on suprasegmental phonology, with one investigating the acquisition of lexical tones in Chinese by Dutch speakers. The paper by Grosser began "Perhaps one of the most neglected areas of L2 phonological acquisition is that of intonational structure" and described L2 intonational structures of Austrian learners of English, specifically the tonal and sentence accentual properties of these learners, which were more similar to their L1 systems at the end of two years of L2 study than at the beginning.

Of further note is the fact that in one of the primary journals for SLA research, *Studies in Second Language Acquisition,* only two research articles whose main focus was intonation have appeared during the twenty-three years of its existence, from 1978 to 2000 (cf. Gutknecht, 1978, with an applied focus and Wennerstrom, 1998, on discourse intonation). During the same period, there were thirty-one articles on phonology, pronunciation, and/or accent, and three "research notes," one of which, Lepetit (1985), dealt with intonation and fossilization in interlanguage.[1] In a study on the critical factors of age, motivation, and instruction in the attainment in L2 phonology, Moyer (1999) found that learners of German who received both suprasegmental and segmental feedback on their pronunciation were judged to have more native-like pronunciation than those who did not. In her conclusion, Moyer stated that SLA phonological research should expand to explore further the process of second language suprasegmental acquisition and the effect of incorporating suprasegmental training into overt instruction. Other examples of articles that have considered prosody and intonation (without, however, applications to L2 teaching or learning) include Munro (1995), who reported on a study of suprasegmental factors that mark foreign accents using filtered speech; Bongaerts et al. (1997), who studied factors such as age in investigating ultimate attainment in the pronunciation of a foreign language; and Derwing and Munro (1997), who studied "goodness of prosody" as one of several factors in L2 accent.

Similarly, the journal *Language Learning* has featured a relatively large number of articles on phonetics and phonology during the last twenty-three years (1978–2000) but only a few that specifically address the acquisition of suprasegmentals (e.g., Anderson-Hsieh et al., 1992, on "the relationship between native speaker judgments of non-native pronunciation and deviance in segmentals, prosody and syllable structure"; Lepetit, 1989, on the acquisition of French intonation; and Schmid, 1986, on a comparison of children's and adults' acquisition of tone accents in Swedish). In recent years the focus in SLA phonological research has been on foreign accent, comprehensibility, intelligibility, and age of acquisition, with some attention given to prosody. Munro and Derwing (1995a), for example, found that intonation is an important factor in listener judgments of comprehension and accent (cf. also Anderson-Hsieh & Koehler, 1988; Anderson-Hsieh et al., 1992; Griffiths, 1990; Koster & Koet, 1993; Munro & Derwing, 1995b, 1998; Purcell & Suter, 1980; Thompson, 1991).

In addition, SLA research on the role of the learner's awareness in learning has also studied speech elements that L2 learners attend to or perceive in the stream of speech. Kim (1995) looked specifically at prosodic characteristics of

speech elements that affect perception and the developmental phases of speech perception and found that noticed elements were characterized as having phonetic or prosodic prominence (e.g., primary stress in an intonational phrase with one intonation contour). But her results regarding the effects of shorter phonological phrases and slower tempo of speech on learner processing were inconclusive. Further research into relationships between speech intake, particularly its prosodic characteristics, and developmental phrases of aural processing is needed.

Pennington and Ellis (2000) reported on two experimental studies of Cantonese speakers' memory for English sentences with prosodic cues and found that their memory was generally poor, both when the contrastive focus was implicit in the experimental task and when it was the explicit focus of attention. Learners' performance was improved significantly when their attention had been explicitly directed to intonation, but only for sentences in which prosody cued a marked informational focus ("contrastive stress") versus an unmarked one ("neutral" sentence intonation). The authors suggested the importance of raising awareness of prosody in a second language.

Pennington (1998), after summarizing the research on the teachability of L2 phonology, concluded that physiological, psychological, and sociocultural factors are mutually reinforcing in the development of phonological competence and in the retention of a distinctive accent in adulthood. As a result, she asserts that both language learners and language teachers must accept the fact that the acquisition of phonology beyond childhood is a gradual and extended process (p. 338).

Applied linguistic research

While SLA research tends to focus on underlying theoretical constructs, applications of such models and paradigms to teaching are addressed in other areas of research in applied linguistics. As a means of summarizing the general directions in which applied linguistics research is moving, Morley (1991) listed the following significant changes in theoretical paradigms, first in learning models and then in linguistic models (pp. 483–484):

> – From a language learning perspective […] a changed concept of language acquisition that views the learner as the active prime mover in the learning process (Corder, 1967), and an emerging paradigm shift in which learners are seen as active creators, not as passive recipients, in a process which is cognitively driven.

- Following from this altered conceptualization of the learning process, a movement from a focus on the group, to an increasing focus on individual learner differences and individual learning styles and strategies […]
- From a focus on language as simply a formal system, to a focus on language as both a formal system and a functional system, one that exists to satisfy the communicative needs of its users.
- From linguistic preoccupation with sentence-level grammar, to widening interest in semantics, pragmatics, discourse, and speech act theory.

Recent research that brings together the "new" emphases on both discourse-level pragmatics and intonation includes Pennington and Richards (1986), who, in re-examining the status of pronunciation in language teaching, proposed replacing the traditional phoneme-based view of pronunciation with a broader discourse-based view comprising segmental, voice-setting, and prosodic features. Morley's (1987) collection of papers focused (1) on "working with pronunciation as an integral part of, not apart from, oral communication" and (2) on "the primary importance of suprasegmentals […] and how they are used to communicate meaning, with a secondary importance assigned to segmentals […]"(Preface).

Hurley (1992) asserted that "pragmatics, prosody, and non-verbal communication are important components of face-to-face interaction that are often overlooked in L2 teaching programs" (p. 259). Furthermore, he acknowledged that "our knowledge of the prosodic and non-verbal contribution to conversational management and conversational support — particularly as these differ cross-culturally — is at its beginning stages" (p. 276) and stressed the importance of addressing these issues in order to help learners develop full pragmatic competence in their L2.

Van Els and de Bot (1987) noted that "both in the teaching of pronunciation and in research concerning foreign accents, too much attention is paid to the segmental and too little to suprasegmental aspects." The goal of their research was to determine the part played by intonation in a foreign accent. By monotonizing the speech signal, low-pass filtering it, and then presenting it to listeners along with the unaltered signals, they determined that pitch variation does play a part in a foreign accent. Similarly, Scovel (1988) suggested that the answer to the question "What makes a foreign accent foreign?" is a combination of segmental, suprasegmental, and voice setting characteristics that collectively mark a speaker phonologically with a kind of *gestalt*, or overall profile (p. 165).

Kelm (1987), in acknowledging that correct intonation is a vital part of being understood, focused on the different ways of expressing contrastive

emphasis in Spanish and English and investigated acoustically whether the ranges of pitch of non-native Spanish speakers differed from those of native Spanish speakers. Bowen (1975) had found earlier that unusual intonation in moments of high emotion might cause the non-native speaker of Spanish to sound angry or disgusted (p. 33). Kelm found that the native Spanish-speaking group clearly varied less in pitch than the two American groups — that native speakers of English use pitch and intensity to contrast words in their native language and transfer this intonation when speaking Spanish. Although the results showed a difference between native and non-native Spanish intonation in contrasts, they did not show the degree to which those differences affect or interfere with communication.

As another example of sociocultural misunderstanding caused by differences in intonation, Hurley (1992) pointed out that while drops in loudness and pitch are turn-relinquishing signals in English, Arabic speakers of English often use non-native-like increases in loudness that could seem like or be misinterpreted as efforts to hold the floor (pp. 272–3). Similarly, in a study of politeness with Japanese and English speakers, Loveday (1981) found more sharply defined differences in both absolute pitch and within-utterance pitch variation between Japanese males and females in uttering Japanese politeness formulas than he did between English males and females in uttering English formulas. In addition, the Japanese subjects transferred their lower native language pitch ranges when uttering the English formulas. Low intonation contours are judged by native speakers of English to indicate boredom and detachment, and if male Japanese speakers transfer their low contours from Japanese to English when trying to be polite, this could result in misunderstandings by native English speakers. Van Bezooijen (1995) corroborated earlier studies and found Japanese women to have higher pitches than Dutch women.

As evidence for culture-specificity with regard to the encoding and perception of affective states in intonation contours, Luthy (1983) reported that although a number of "nonlexical intonation signals" (associated with expressions like *uh-oh* or *mm-hm* in English) were interpreted consistently by a control group of English native speakers, non-native speakers of varied L1 backgrounds frequently misinterpreted them. He concluded that many foreign students appear to have difficulty understanding the intended meanings of some "backchanneling" signals in English and that these nuances are not being explicitly taught.

Related research on suprasegmentals in languages other than English includes: Tarone (1980), who described foreigner-talk as having such features as longer pauses, extra volume, and exaggerated intonation — features that teachers tend to identify as typical of non-native speech (p. 423); Edmondson et al. (1984), who studied German speakers learning English and found, for example, that learners combined syntax and intonation incompatibly, focusing instead on learning the pragmatics of discourse in non-intonational areas; Hieke (1984, 1985), who looked at speech rate and linking (liaison, consonant attraction) as markers of fluent speech and showed how divergent the views and definitions of oral fluency are; Scanlan (1987), who investigated native pause behavior in French with the intent of improving fluency in spoken French; Keller (1989), who investigated the question of whether perceptual accuracy and production accuracy in French-speaking adults' acquisition of Chinese tonal contours proceed in parallel or whether one precedes the other; and Chun (1988a, 1988b), who echoed Pennington, Richards, van Els, de Bot and the researchers cited above in emphasizing the theoretical principles underlying the need to include discourse intonation in language teaching and in deploring the neglect of intonation in syllabi and materials.

Morley (1991) summarized the recent significant changes in second and foreign language instructional models (p. 484):

- From an instructional focus on linguistic form and correct usage to one on function and communicatively appropriate use.
- From an orientation on linguistic competence to one on communicative competence.
- From a global competence concept to detailed competency specifications and the introduction of an especially useful model that brings together a number of viewpoints in one linguistically oriented and pedagogically useful framework: grammatical competence, sociolinguistic competence, discourse competence, and strategic competence.

In applied linguistics research, discourse-level phenomena essential to aspects of communicative competence particularly in the area of phonology and pronunciation had not, until recently, been widely addressed. Morley's (1991) review of the changing patterns in pronunciation teaching in ESL recognized that there is currently a renewed professional commitment to enabling students to become effective communicators and that it is therefore imperative to start incorporating pronunciation teaching into instruction because "intelligible pronunciation is an essential component of communicative competence"

(p. 488). She asserted that the teaching of pronunciation should not be narrowly regarded as one-dimensional, but rather should be approached as a multidimensional process — that pronunciation must not be isolated, conceptually or practically, from language learning or teaching, but rather must be viewed and taught as an integral part of oral communication. Furthermore, in redefining learner goals, the focus should be on *functional intelligibility* and *functional communicability*. Likewise, Gilbert (1984), in her handbook on teaching pronunciation and listening comprehension in English, also stated simply that "there are two main reasons to teach pronunciation: Students need to understand and to be understood," alluding to the important link between pronunciation and comprehension rather than the earlier emphasis on "pronouncing the sounds correctly" (p. 1).

In summary, prior to the 1980s, most theoretical studies of intonation had not incorporated the changing emphases on linguistic pragmatics into their work. Thus, despite the relatively extensive literature on intonation in the past twenty-five years as described in Chapters 2 and 3, it was not until the 1980s that the role of intonation in discourse and conversation analysis emerged as a research topic in linguistics as well as in applied linguistics.

As will be seen in the next two sections, pedagogical applications for discourse intonation in language syllabi have differed somewhat in the fields of ESL and the teaching of languages other than English. While the ESL profession has made a notable effort in the last decade to incorporate the teaching of suprasegmentals into their materials and curricula, the FL community, at least in the U. S., has been moving towards proficiency-oriented methodologies and meaning- and comprehension-based approaches but has only recently begun to recognize intonation as an important component of linguistic and communicative competence.

Recent revival of interest in pronunciation in ESL/EFL

Morley (1991) discussed the major influences on the changing patterns of pronunciation teaching in ESL in the U. S. and cited two main catalysts in bringing about change: (1) the urgent needs of ESL learners whose pronunciation difficulties may result in professional or social disadvantages, and (2) "a number of emerging principles that seem to reflect an underlying belief system shared by many new pronunciation programs" (pp. 489–490). Included in her list of principles were the growing trend towards communicative approaches

to teaching pronunciation; a reordering of priorities to focus on the critical importance of suprasegmentals (stress, rhythm, intonation, etc.) as well as of voice quality, paralinguistic features, kinesics, and proxemics; and a focus on meaningful practice, particularly speech activity experiences adapted to the communication styles and needs of learners' real-life situations.

Research by Pennington and Richards (1986) and Pennington (1989b) emphasized the need to teach pronunciation from a "top–down" approach, i.e., focusing on the rhythm and sentence melody of entire phrases or sentences as opposed to the "bottom–up" approach of mastering single sounds or words. With these and other studies, discourse intonation was beginning to be seen as a fundamental component of both listening comprehension and oral proficiency that serves to mark thought units and salient or new information as well as to signal different types of conversational strategies. When this perspective is applied in a classroom, students must be made aware of how intonation functions in language and must be given opportunities both to hear and to practice intonational patterns so that they will comprehend more fully and be better communicators themselves.

In many language learning methodologies and approaches of the past, intonation was not considered a critical component of language learning as it was thought to contribute little, if anything, to the production of correct forms or sentences. Moreover, since in many cases intonation is redundant relative to the syntax or to the surface semantic meaning of an utterance, teachers were able to consider teaching it as "icing on the cake." Morley (1991), however, included in her list of changing principles and priorities the emerging focus on the link between perception and production — that is, between listening and pronouncing/speaking (p. 494). In other words, the minor role previously attributed to intonation was based on the faulty assumptions that mastery of syntactic structures automatically insured that a sentence would be understood as a statement or a question and that properly chosen lexical items would adequately signal a particular attitude. However, as modern linguistic research has shown (cf. Couper-Kuhlen & Selting, 1996a), the redundancy argument is flawed because first, intonation can be shown to provide meaning not coded in any other form, and, second, the language learner needs not only to make her/himself understood, but also to be able to understand intonational contrasts made by native speakers.

As noted in the previous chapter, from the 1960s to the early 1980s, pronunciation was often ignored or treated minimally in L2 teaching. In the 1980s, the field of ESL/EFL took the lead in reviving interest in pronunciation

and in promoting the teaching of stress, rhythm, and intonation, whereas research in the teaching of languages other than English seemed to be lagging behind. For example, for the 20-year period 1980–2000, there were only nine articles dealing with pronunciation in the *Modern Language Journal*,[2] four in *Foreign Language Annals*,[3] two in the German pedagogical journal *Die Unterrichtspraxis*,[4] six in *The French Review*,[5] and five in *Hispania*.[6] Of those twenty-six articles, only seven dealt with intonation (or suprasegmentals).[7] Similarly, the 1999 *ACTFL Proficiency Guidelines for Speaking* published in Breiner-Sanders et al. (1999), the standard for the FL profession in the United States, described in detail below, make virtually no mention of intonation. Surveys of elementary language textbooks of the last two decades for languages other than English also reveal very little concern with intonation, and one can deduce that it is not being taught explicitly in the classroom (see the Introduction to Part III of this book).

In contrast, however, to the dearth of both research and instructional materials for intonation in languages other than English, there is considerable research interest and many materials available for ESL/EFL. In the same 20-year period, for example, there were eighteen articles in the *TESOL Quarterly* alone that dealt with some aspect of pronunciation; of the eighteen, seven dealt specifically with intonation.[8] Interestingly, despite the greater attention being paid to intonation and advances in both theoretical and applied research on intonation, Levis (1999) laments the fact that there has been a lack of innovation in intonation teaching materials.

In addition, the following studies in other journals deal with suprasegmentals: Anderson-Hsieh (1990) on teaching suprasegmentals to international Teaching Assistants and Boyle (1987) on teaching English stress and intonation to speakers of Chinese. Anderson-Hsieh found in an informal, impressionistic evaluation that at the end of the semester "students were using intonation more effectively to highlight contrasts and new information, and they were pausing less often and more appropriately" (p. 210). Wennerstrom (1994) studied intonational meaning in discourse and found that native speakers of English made significant use of pitch contrasts to signal discourse meaning but that the speakers of Spanish, Japanese and Thai in her study did not consistently use pitch to signal meaningful contrasts in many of the same environments.

Wennerstrom (1998) proposed that certain intonation components help to achieve coherence in discourse. She measured four aspects of intonation: (1) the pitch difference between newly introduced content words and function words; (2) the use of high pitch at phrase boundaries to link related constitu-

ents; (3) the use of pitch to distinguish contrasting items from given items; and (4) the increased pitch range at rhetorical junctures to signal topic shift. These measures were chosen for their contribution to the cohesion of the English lectures given by Chinese speakers, and it was found that macro-level intonation — the use of increased pitch range to mark topic shifts — makes the strongest contribution to scores of effectiveness in lecturing and to communicative competence in general.

In addition to journal articles, recent handbooks for teaching English pronunciation all include at least a partial focus on the phenomena of stress and intonation.[9] There is, however, one possible lacuna in the field of ESL/EFL: although the handbooks and textbooks contain pedagogical material on intonation, there are no standard guidelines for whether it should be taught, how it might be taught, or what the goals and results of teaching it should be.

Despite the fact that relatively little is being done on intonation per se in foreign language pedagogy, either theoretically or practically, the 1989 *ACTFL Oral Proficiency Interview Tester Training Manual* and the work by Byrnes (1987), Kramsch (1986), and Lantolf and Frawley (1985), to name just a few, reflect in principle a recognition of the need for more research on pronunciation skills that the language learner should learn and how they can best be taught. Based on a critical analysis of research on the teachability of phonology in a second language, Pennington (1998) argued that phonology both can and should be taught to adult learners. In her brief overview of studies on the effectiveness of pronunciation instruction, she found mixed results and suggested that both the nature of the training and the nature of the learners must be considered as a possible explanatory factor. In addition, she cited the difficulty of evaluating pronunciation ability or improvement. The next two sections discuss these issues of describing pronunciation standards and incorporating intonation into these standards.

Communicative competence and proficiency models for L2 and FL learning

As discussed at the end of Chapter 2 and in the first section of this chapter, recent research on discourse, both in theoretical linguistics and in second and foreign language pedagogy, has focused on describing rules for the comprehension and production of coherent verbal interaction. The contributing elements of coherent interaction are not simply the rules of grammaticality in

any given language, but include also the pragmatic or functional rules that govern the use of utterances in spoken discourse. Yule (1989), in characterizing prosodic elements as key factors in pronunciation, supported the view that teaching the spoken language should be concerned with general communicative effectiveness and should concentrate on activities with the goals of using language for its transactional ("getting one's message across") and interactional functions. However, many applied linguists in language pedagogy have largely overlooked intonation when describing the functions of natural language, i.e., what is actually *done* by language rather than what is *said*, and in turn, language learners have generally not been taught how to use intonation to signal discourse strategies. One notable exception is the work of Brazil, Coulthard, and Johns (1980). This section shows how intonation has not up until now received much attention either in the communicative competence or the proficiency approaches, despite the fact that it is a important tool for negotiating meaning, managing interaction, and achieving discourse coherence.[10] In a typical textbook exercise, Kramsch (1983), for example, did not explicitly discuss intonation but stated merely that the discourse strategies presented should be taught "with the proper intonation" (p. 182).

The first caveat in positing discourse functions of intonation is that the broad communicative value of an utterance, including not only syntax and lexical choice but also intonation, cannot be determined by examining isolated utterances. The interpretation of an utterance is dependent on the many interactive functions of that utterance within a given discourse. A second caveat is that intonation should be viewed as a powerful and as yet untapped discourse tool that should be developed as part of the communicative competence of the second or foreign language (L2) student.

To be sure, the language teaching profession is moving in the direction of this larger picture of language as communication, but intonation is only slowly finding its place. At present, the most widespread standard of oral competence for foreign languages in the United States is the *ACTFL Proficiency Guidelines,* the revised versions of which (1986, 1989, and 1999) focused, to a greater degree than do the *ACTFL Provisional Proficiency Guidelines* (1982), on interactional and discourse competence of language learners. The revisions of the *Guidelines* followed in the direction prescribed by many researchers to emphasize the interactive nature of communication and language competence. Savignon (1985), for example, had criticized the *Provisional Guidelines* for not addressing the issues of sociolinguistic, discourse, and strategic competence. Lantolf and Frawley (1985) had warned that a distinction needed to be

made between "displaying linguistic ability and ability to engage in a real communicative interaction" and that "we cannot ignore the fact that in normal verbal interaction the factor of 'maximal interlocutor cooperation' comes into play" (p. 339).[11] They had also noted further that as Bachman and Palmer (1984) correctly observed, "many educators and researchers have lost sight of the fact that communication involves two parties, and success in communicative performance will always be dependent upon the abilities of two people," to speak coherently as well as to understand what the other is saying.

However, a close examination of the stated components of both communicative competence and proficiency reveals that: (1) only the *expression* of meaning is fully dealt with, while the *interpretation* and *negotiation* of meaning have not been integrated into proficiency in particular, and (2) neither the *use* of intonation nor its *role* in language competence is addressed in any detail. These deficiencies also suggest that the concept of language as an *interactive social skill* has not yet been fully incorporated into communicative or proficiency guidelines.

Let us examine the communicative competence (CC) and proficiency models to see how they accommodate the phonological component of intonation. The basic principles underlying the communicative competence model include, first of all, the view of language as process rather than product, which is essential to Savignon's perspective on communicative competence, i.e., the notion of the competence of language learners in the "*expression, interpretation,* and *negotiation* of meaning involving interaction between two or more persons."

A second principle of the CC model is that communicative competence comprises four skill areas: *grammatical, sociolinguistic, discourse, and strategic* competence. Canale and Swain (1980) provided a thorough review of the various theoretical models of communicative competence but acknowledged that there is as yet no *formal* consensus about what communicative competence is. Nonetheless, Savignon (1983) and Omaggio (1986) indicated that most researchers agree about the four components of CC.

Grammatical competence is the ability to produce accurate phonological, morphological, syntactic, and lexical forms. Sociolinguistic competence includes the extent to which a speaker uses and understands grammatical forms appropriately "in various contexts to convey specific communicative functions," such as persuading, giving commands, warning, and being polite. Discourse competence involves "the ability to combine ideas to achieve cohesion in form and coherence in thought." (Though not explicitly stated, another

function of this type of competence would also be that of signaling, for example, whether an utterance is intended to invoke a response or whether it is intended to mark a boundary in discourse.) Strategic competence involves the use of "verbal and non-verbal communication strategies" to express one's meaning and convey information to a hearer, particularly in the event of a breakdown in communication or a misunderstanding (Omaggio, 1986, pp. 7–8).

In his modified model, Canale (1983) used the term *communicative competence* to refer to both underlying *knowledge* about language and to the *use* of this knowledge. Omaggio (1986) stated that "this conceptualization of the term is very similar to the notion of *proficiency*" (p. 8). According to Omaggio, in contrast to the four components of communicative competence, there are three interrelated criteria underlying the proficiency descriptions: *context, function, and accuracy*. Specifically, "the term *proficiency* includes specifications about the *levels* of competence attained in terms of the *functions* performed, the *contexts* in which the language user can function and the *accuracy* with which the language is used. Thus, the notion of *proficiency* enables us to broaden our understanding of *communicative competence* to include more than the 'threshold level' needed to simply get one's message across."

Overall, ACTFL has offered language teachers in the U. S. a synthesis of communicative competence and proficiency — Bachman and Savignon (1986) suggested the term "communicative language proficiency" (p. 381) — at least in theory. However, with specific regard to intonation, neither theory nor practice adequately reflects the integration of teaching discourse intonation into language learning syllabi.

As implied above, underlying both the communicative competence and the proficiency models is the recognition that speakers must not only make themselves understood but must also understand their interlocutors. As Kramsch (1981), for one, suggested, strategic competence should also involve such receptive and comprehension strategies as "turn-taking, linking and expanding, negotiation, and repair" (p. 18). Although ACTFL's speaking proficiency descriptions do include reference to interactive skills and to whether or not a speaker can or will be understood, the productive and the receptive skills are *not* tested to the same extent in a typical oral proficiency interview (OPI). Nor does the OPI include the bilateral negotiation of meaning facilitated (in part) by intonation.

Omaggio (1986) discussed a number of "ideal formats and elicitation techniques" for testing oral skills but also stopped short of testing receptive skills in. In the first part of most oral interviews, the interviewer does all the

questioning and follows up on the student's progress by linking up and expanding on what was said. The second part of the interview consists typically of a "conversational exchange" in which the student asks questions of the interviewer based either on a "topic card" or on a role-play situation. This task is intended to determine whether the student can initiate conversation but, in fact, the student is provided with a host of possible questions (in English) and begins by rendering these questions into the target language. Although the student is then encouraged to come up with additional questions, the conversation has, from the very outset, been largely predetermined. Since in both parts the student is given a role, i.e., either responder or questioner, opportunity for spontaneous interaction, for "turn-taking, linking and expanding, or negotiating" on the part of the student is limited, with the result that intonation can scarcely play a critical role in the interaction, either from a perceptual or a productive perspective.

In sum, the theoretical arguments for viewing language as an interactive process that potentially involves intonation — a process in which information and intent are both transmitted as well as received — were present in the research of the 1980s and into the 1990s. The logical implications were that models of both communicative competence and proficiency models should include the ability to use intonation as one of the tools in managing discourse. In the classroom, at the very least, the proficiency-oriented teacher should attempt to foster communicative language proficiency by making learners explicitly aware of the role that intonation plays in successful and coherent communication and begin to incorporate the teaching of intonation into the second and foreign language curriculum. If, and more importantly, even before this attention to intonation is effected, one would hope that intonation would find its place in future proficiency guidelines.

Incorporating intonation into guidelines of the profession

One of the most serious critiques from the perspective of this book and the research it reviews is the fact that the *ACTFL Proficiency Guidelines* in the U. S., standard ESL measures like the SPEAK test and the Test of Spoken English (TSE), and the descriptions of the most widely used language teaching models and methods — virtually all fail to include prosody (which includes suprasegmentals such as pitch, stress, pause) as a vital component of proficiency and of what students should be taught in the classroom. Yet, in the

context of discourse analysis and its importance to the goals of communicative competence and proficiency, current research is clearly indicating that intonation is a fundamental component of the communicative process. As a welcome development, many ESL and EFL textbooks have begun to include chapters on stress and intonation (e.g., Celce-Murcia et al., 1996; Dalton & Seidlhofer, 1994).

In the descriptions of speaking competence in the most recent, revised *ACTFL Proficiency Guidelines for Speaking* (1999), for example, pronunciation and intonation are given scant attention. Only the briefest mention is made of *pronunciation* in the different levels of proficiency. *Intonation* is only mentioned in two of the ten levels. Moreover, from both a phonological and discourse point of view, there is little consistency in the vague descriptions provided for the raters. Specifically, intonation is mentioned only in the highest two levels, as shown below, and these descriptions are so general as to be of questionable value to raters. Savignon (1985) also noted that of all the features in the original FSI scale, *accent* has the lowest weighting (p. 131).

Below are all of the direct references to intonation in the *ACTFL Proficiency Guidelines.*

- Novice-Low: none (one reference to pronunciation: "Speakers at the Novice-Low level have no real functional ability and, because of their pronunciation, they may be unintelligible").
- Novice-Mid: none.
- Novice-High: "These speakers' first language may strongly influence their pronunciation […]"
- Intermediate-Low: "Their pronunciation, vocabulary, and syntax are strongly influenced by their first language […]"
- Intermediate-Mid: "Because of inaccuracies in their vocabulary and/or pronunciation and/or grammar and/or syntax, misunderstandings can occur […]"
- Intermediate-High: none.
- Advanced Low: none.
- Advanced Mid: none.
- Advanced High: "They use precise vocabulary and intonation to express meaning and often show great fluency and ease of speech."
- Superior: "Superior speakers command a variety of interactive and discourse strategies, […] as well as intonational features such as pitch, stress, and tone."

The concept of discourse intonation, which has become an important issue in research (as discussed in Chapters 2, 3, and 5) is thus not mentioned until the very highest level of proficiency. However, recent research in intonation supports my contention that such types of critical discourse and strategic skills should be considered part of language competence and thus should be taught at a much earlier level. Since the *Guidelines* only describe how learners typically perform in a foreign language — and do not prescribe what they ought to be able to do — it is not surprising that little mention is made of intonation. Until now, students have not been taught how to use intonation for discourse purposes, and the typical beginning language learner is simply not even aware of its importance. Raters are not evaluating it with discrimination, and teachers are left to teach it on their own, at a rudimentary level, if at all.[12]

The outlook for including intonation in oral proficiency descriptions is, however, not completely bleak: in the 1989 *ACTFL Oral Proficiency Interview Tester Training Manual*, the emphasis is on *natural* conversation, and much attention is given to the pragmatic and discourse aspects of speech, e.g., the negotiation of meaning, recognition of the importance of context, cultural, and sociolinguistic competence, and task- or goal-oriented activities. In addition, the *OPI* manual outlines *interviewer* strategies for eliciting speech, e.g., rephrased questions, hypothetical questions, prelude questions, polite requests, and so-called intonation questions. It is this attention to the pragmatics of natural conversation — to initiating and managing conversation — that students, the *interviewees*, should also be taught.[13] Guidelines in the future should address and describe sentence intonation and meaning instead of only "pronunciation" — a focus on the vowels and consonants.

Looking ahead: Integrating technology into research and instruction

In addition to increased recognition and acknowledgement of the importance of teaching prosody, one exciting prospect for the 21st century with regard to implementing intonation instruction is that of taking advantage of technological advances, particularly in *multimedia* computer-based materials, for production, perception, and comprehension purposes. This line of research would correspond to other needs suggested by Morley (1991) in looking toward the future of pronunciation teaching (p. 511):

- A continuing need for development of pronunciation/speech activities, tasks, materials, methodologies, and techniques across the spectrum of imitative, rehearsed, and extemporaneous speaking practice experiences […] (One tool now becoming an economic and practical possibility is self-study computer programming both for student practice and for assessment through the use of visual displays of speech parameters […] [since] laboratory and speech analysis and synthesis capabilities have become more accessible for instructional uses.)
- [T]he need for more definitive evaluative measures and methods to quantify changes and improvements in the learner's intelligibility and communicability.
- A need for controlled studies of changes in learner pronunciation patterns as the result of specific instructional procedures…

Esling (1992) also concluded that "it is increasingly necessary to train language specialists in applied linguistics programs to become familiar with the use of technological systems for the recording, storage, and delivery of speech sound information" (p. 267). He stressed the importance of methods of instruction in phonetics courses as well as that of software design being consistent with principles of second language acquisition theory and with language teaching methodologies. His own *Phonetic Database* system for microcomputers emphasized the use of the computer as a tool instead of a tutor, for cross-referencing several forms of written and auditory information in order to build associations between sound and symbol, for analyzing prosody holistically as a learning exercise as opposed to a point-by-point didactic style, and for observing language from the prosodic and indexical perspective before concentrating on segmental phonology.

In recent work, Pennington and Esling (1996) provided a basic technological account of computer-aided pronunciation software, and Pennington (1999) presented an overview of the promise and limitations of using computers to improve pronunciation in a second language. Two of her ten principles for software design were to "link pronunciation to other learning and communicative goals" and to "raise awareness of contrast with L1 and range of targets for L2" (p. 433). More specifically, Pennington (2000) reported on a study of the training of intonation on the computer comparing seven different pedagogical orientations. The study underscored the need to explicitly focus language learners' attention on prosody and also served as a basis for pedagogical recommendations. One of the recommendations was that software should provide support using both visual and auditory models.

I would suggest four specific functions for technology that can be integrated into instruction and research. Computers and computer software can

be used: (1) to provide learners with visualizations of their intonational patterns and with specific feedback to help them perceive the meaningful contrasts between L1 and L2 so that they can improve their speech production; (2) to provide learners with authentic and extensive speech and cultural input and in turn to hone learners' perceptual abilities; (3) to facilitate, record, and analyze interactions between and among speakers; (4) to build tools for research purposes, e.g., data collection tools to record student performance, progress, and steps toward self-correction (cf. Chun, 1998).

Provide learners with visualization of their intonational patterns and specific contrastive feedback

Although computers have been utilized for providing learners with visualizations of their intonational patterns since the 1970s, relevant hardware and software for computers (particularly microcomputers) since the mid-1980s have become increasingly more accessible in the form of sound and speech digitizers, pitch trackers to produce displays of intonation curves and computer-assisted language learning (CALL) software, including pronunciation tutors with audio and graphic components.[14]

A very popular and successful product from Kay Elemetrics called *Visi-Pitch* has been available for a number of years for IBM personal computers (PCs) and compatibles (cf. Abberton & Fourcin, 1975; James, 1976, 1979). With *Visi-Pitch* students are able to see a native speaker's pitch curve and their own simultaneously. Fischer (1986a, 1986b) reported on the use of *Visi-Pitch* for teaching Chinese tones and French intonation (see Figures 1 and 2 below). Molholt (1988) used *Visi-Pitch* together with a Speech Spectrographic Display for improving both the segmental and the suprasegmental phonology of Chinese speakers of English. In this constellation, students first spoke a sentence into a microphone; their utterance was then digitized and pitch-tracked, and they saw a display of their pitch curve directly under a native speaker's pitch curve of the same sentence. Anderson-Hsieh (1992) reported on using *Visi-Pitch* to teach English suprasegmentals to international teaching assistants and found that "the major benefit of electronic visual feedback for teaching suprasegmentals is that it provides the students with an accurate visual representation of suprasegmentals in real time paired with the normal auditory feedback that occurs during speech. Students can thus more easily replicate native suprasegmental targets using both the target form and the visual feedback from their own speech to guide them" (p. 61).

The figures below were produced by *Visi-Pitch* (cf. Fischer, 1986a, 1986b). In the first figure, the top half shows the pitch curves of a native Chinese speaker pronouncing three words that should have the so-called Tone 1 (high level pitch), as represented by the relatively straight lines. The bottom half shows the pitch curves produced by a learner, exhibiting erroneously falling pitch contours. In the second figure, the top half shows the pitch contours of the French question *Qu'est-ce qu'il fait?* "What is he doing?" as spoken by a native speaker. The bottom half shows how an American learner produced the same question with English-like intonation contours, i.e., pitch peaks on the stressed syllables rather than continuously falling sentence intonation.

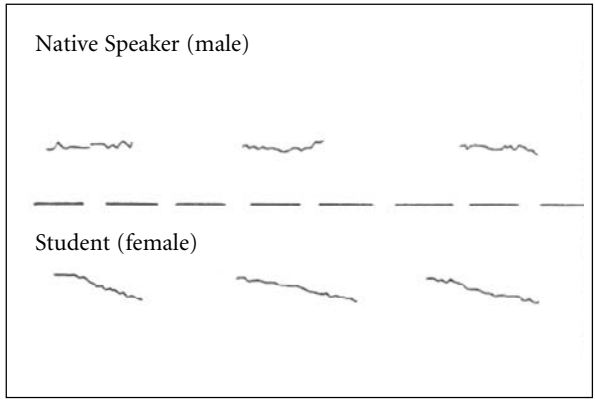


Figure 1. Chinese Tone 1

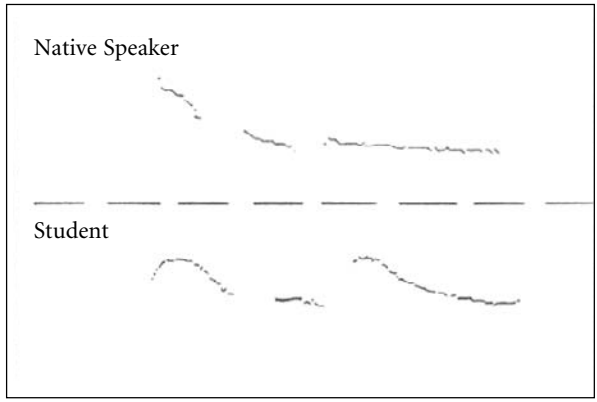


Figure 2. French sentence *Qu'est-ce qu'il fait?*

In conjunction with Speech Technology Research (STR) of Victoria, B. C., Canada, the upgraded, next-generation system developed by Kay Elemetrics is *CSL (Computer Speech Lab)*.[15] A spin-off of the *CSL* is *Visi-Pitch II*, which is designed more as a speech/voice therapy tool. Another speech analysis system that has been used for both pedagogical and research purposes is the *Micro Speech Lab (MSL)*, developed by the Centre for Speech Technology Research in cooperation with the Department of Linguistics at the University of Victoria, B. C., Canada, reported on in Esling (1992). Various components of the program allow for speech editing (*MSLEDIT*), manipulatory exercises (*MSLSORT*) and displaying pitch and other acoustic (e.g., spectrographic) information (*MSLPITCH* and *MSLSPECT*).[16]

In addition to his observations on the importance of training applied linguists to use technology, Esling (1992) also emphasized the underlying pedagogical rationales for using such speech technology systems in instruction, describing different types of exercises that address perceptual skills; manipulation-oriented tasks which allow for active learner participation; the importance of providing extensive, easily accessible examples to represent the diversity of speech sounds and the great variation that exists within a given language; and the need to hear authentic illustrations of sounds and intonations within a format that promotes practice, feedback, and even collaborative learning.

One of the problems with some of the earlier software programs was the lack of feedback processing, i.e., pitch could be measured and directly fed back to the learner, but interruptions in the intonation contour during unvoiced parts of the utterance, on the one hand, and inclusion of perceptually irrelevant pitch variations, on the other, made it difficult for the learner to interpret the feedback. Spaai and Hermes (1993) therefore devised a visual intonation-display system called *Intonation Meter* that presents visual feedback of the intonation as a continuous representation of the pitch contour and contains only the perceptually relevant aspects of the intonation pattern (see Figures 3 and 4 below). Figure 3 displays an "unprocessed" fundamental frequency or pitch track: it is "unprocessed" because the breaks in the pitch curve represent points in the utterance where voiceless sounds occurred. Figure 4 displays the same underlying pitch track which is overlaid by a stylized, continuous line that is meant to help learners focus on the main pitch patterns of rises and falls. In the actual *Intonation Meter* program, the pitch contours produced by the foreign language teacher are displayed on the upper part of a computer screen and the student's version or "imitation" appears on the lower part of the screen, both in stylized form (that is, with a continuous (interpolated) pitch

"curve" which eliminates the irrelevant pitch variations). The program contains three types of exercises. The first set consists of auditory discrimination exercises, the second of "imitation" exercises, and the third comprises "production on demand" exercises, whereby spontaneous production from the learner is elicited by means of visual cues (e.g., sentences, parts of stories, or dialogues) (p. 26–27).

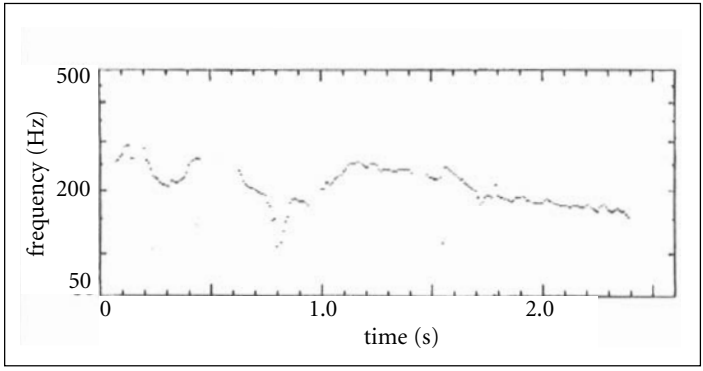


Figure 3. "Unprocessed" fundamental frequency (pitch) measurements

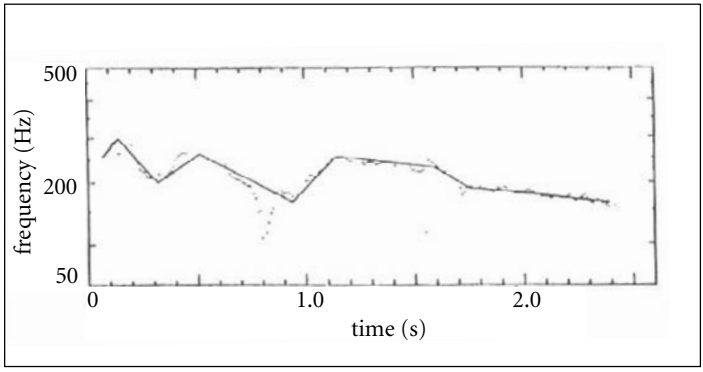


Figure 4. Unprocessed pitch contour (dots) and stylized pitch contour (straight line)

Anderson-Hsieh (1994) also discusses some of the technical problems with existing software and provides learners with information on interpreting visual feedback graphs of intonation contours produced by computer programs, in particular those generated by *Visi-Pitch*. In addressing the problem noted by Spaai and Hermes (1993), she compared pitch contours for sentences with

mainly voiced sounds (more or less continuous pitch curves) with sentences containing many voiceless sounds (broken or truncated pitch curves). In the example shown below from Anderson-Hsieh (Figure 5), the final syllable of both sentences is accented, but in the sentence with the word *nine,* which contains only vowels and nasals (all voiced), the final accent is shown as a long fall, while in the sentence with the word *six,* the voiceless consonants at word-beginning and word-end cause the final accent to be cut off and thus look like a truncated fall. This would be misleading to a learner, and more importantly, deep falls and truncated falls convey very different meanings in English: As seen in Chapter 3, Bolinger (1986) attributed the meaning of finality or assertiveness to deep falls and a sense of offhandedness or tentativeness to truncated falls. The solution to these inherent problems of voiceless consonants being incapable of carrying pitch is, at least initially, to use model sentences in training that contain mainly voiced sounds (especially sonorants) particularly at utterance-end, in accented syllables and wherever it is imperative for learners to be able to see the movement of pitch.

In a recent European Union project named *Spell*, several intonation training strategies were investigated. One problem that was addressed was what amount of variability in student production was deemed "acceptable" or "correct." Their solution was to heavily smooth the student's F_0 curve, normalizing it to the model utterance and allowing some variation with a so-called "pitch tunnel" between the "pitch anchor points" of the F_0 curve (see Rooney et al., 1992). Determining the pitch anchor points in the student utterances requires segmentation by a speech recognition algorithm and is thus a multi-step process.

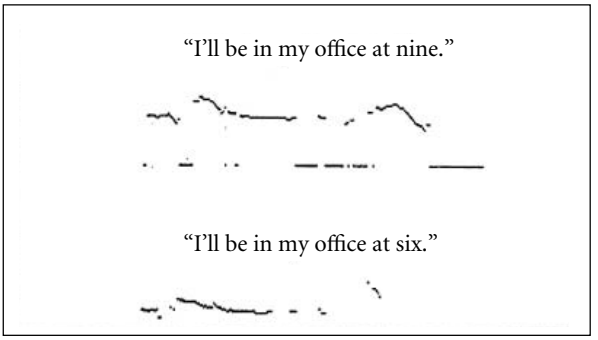


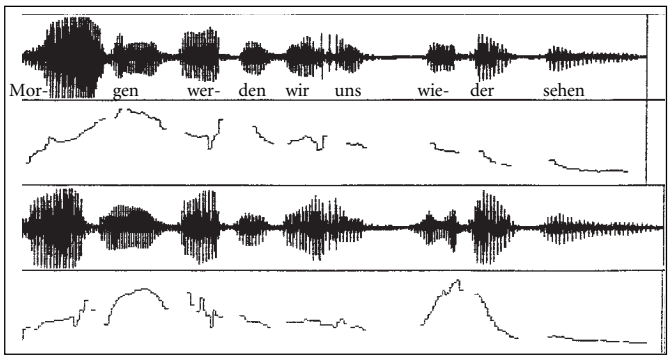
Figure 5. Sentences with sonorant vs. non-sonorant final words

For the Macintosh computer, any sound digitizer (e.g., MacRecorder, Digidesign's Audiomedia board or the internal digitizer on most Macintosh computers sold today) can be used in conjunction with a program by Keller (1988) called *Signalyze*© to produce and store intonation curves of native speakers and learners as well as to display them next to each other (cf. Chun, 1989 and Figure 6 below). A program called *SpeechLab Authoring Tool*©, based on *Signalyze,* has been developed to automatically display first the native speaker's utterance and then a student's version (cf. Fidelman & Keller, 1994). The purported benefits are that students thus get immediate visual and audio feedback of the otherwise elusive dimension of pitch. However, the tool is limited in its pitch-tracking ability and, as with the earlier software for PCs and Windows-based computers, in not providing pedagogical analysis of the feedback. Other programs, *MacCECIL*© *(Computerized Extraction of Components of Intonation in Language),* developed by SIL (Summer Institute of Linguistics) for the Macintosh, along with its counterpart for Windows *WinCECIL*© and an expanded program *Speech Analyzer*© (http://www.sil.org/computing/catalog/speech analyzer.html), are very easy to use, generate very good pitch contours and were employed for most of the figures generated in Chapter 6 of this book. It can only be hoped that a pedagogical application of this software for language learning will be developed, based on the principles discussed above for interpreting visual feedback and facilitating L2 intonation learning.

The example in Figure 6 below from Chun (1991) shows the intonation curves for two versions of a German sentence. In the first case, the word at the beginning of the sentence (*Morgen* 'tomorrow') is stressed, and in the second case, the word *wiedersehen* 'to see again' receives sentence stress. The different pitch curves would give the students a visual reinforcement of the acoustic distinction that has been contextualized (by the different placement of sentence stress) and thus shown to be functionally significant. In instruction, students could be given different contexts in which these two versions might be used and then asked to replicate the native speaker's pitch curves.

In sum, past use of software to help learners with visualizations of their intonation patterns was restricted by the relative inaccessibility of hardware and software, technical limitations and a lack of capacity to give comprehensive pedagogical feedback. In addition, the focus was typically on sentence-level intonation, specifically on contrasting different sentence (syntactic or *grammatical*) types, e.g., declarative statements, yes-no questions, wh-questions, and exclamations. Pitch-tracking software can certainly be used to teach these basic intonation contours, but for the future, in accordance with the

"Tomorrow we'll see each other again."



MOR gen wer den wir uns wie der sehen.
Mor gen wer den wir uns WIE der sehen.
(Tomorrow we'll see each other again.)

Figure 6. Pitch curves created by *Signalyze*

current emphasis on *communicative* and *sociocultural* competence, more attention should be paid to discourse-level communication and to cross-cultural differences in pitch patterns. Examples will be presented in Chapters 6 and 7, but for the purposes of this chapter, I suggest that software programs must:

– distinguish meaningful intonational features with regard to four aspects of pitch change: (a) direction (rise, fall, or level), (2) range (difference between high and low levels), (3) speed (how abruptly or gradually the change occurs), and (4) place (which syllable(s) in an utterance is/are affected)
– go beyond the sentence level and address the multiple levels of communicative competence: *grammatical*, *attitudinal*, *discourse,* and *sociolinguistic* competence
– start with utterances containing a minimum of voiceless sounds in order to insure decipherable pitch curves, particularly for the initial exercises, where learners must familiarize themselves with the visual representations of the pitch curves.

Provide learners with authentic and extensive speech and cultural input to hone learners' perceptual abilities

The second function for technology that can be integrated into intonation instruction is to provide learners with authentic and extensive speech and

cultural input and in so doing to hone learners' perceptual abilities. As advocated throughout this book, authentic speech should be used as input whenever possible. For example, short conversations between native speakers can be taken from corpora such as the CSAE (Corpus of Spoken American English)[17] then be pitch-tracked and presented to learners. To as great an extent as feasible, conversations with as few voiceless segmentals as possible should be selected. Learners would first do a very close listening and analysis, i.e., they would listen to the conversation (or selected part thereof) multiple times as well as view the intonation curves. They would be told which acoustic intonational features to focus on, e.g., male-female differences in pitch patterns or pitch range. Alternatively, they would be presented with utterances and asked to determine the nuance or attitude being expressed. Following the perception activities, they would be asked to practice these utterances by recording themselves and comparing their pitch tracks to those of the native speakers. Finally, they could be asked to role-play a conversation or an interaction with a partner on a similar topic. They would record themselves, pitch-track their utterances and note the direction of pitch changes, the syllable(s) on which pitch changes occurred, the steepness of the pitch falls and rises, and the overall pitch range used. If the learners' own natural speech contained a high number of voiceless sounds, the pitch curves might contain many breaks unless a system such as the one described by Spaai and Hermes (1993) were being used. If they did not have access to such a system, they would have to have their instructor listen to their conversation and provide feedback and guidance. Although there are limitations in current speech analysis systems, natural discourse should still be encouraged, and instructors will have to be prepared to provide the definitive feedback.

In addition to computer-based applications, so-called "lower-tech" multimedia solutions (traditional audiotapes and videotapes) can be utilized until such time that many of these are digitized. Since language and communication are multi-faceted in the sense that they involve the visual, auditory, and kinesic channels, it stands to reason that multimedia should allow for the presentation of authentic linguistic and cultural materials presented as entire "discourses" not only via the printed word in textbooks, but more importantly through (1) other types of visual input (e.g., videos of actual communicative interactions or other types of speech "events"); (2) audiotapes and videotapes of authentic conversations, e.g., television news reports, interviews, talk shows; (3) not only verbal descriptions of cultural traditions or gestural phenomena, but the full range and sequence of events accompanying them (e.g., SCOLA broadcasts

and even entire courses based on videotaped materials such as *French in Action* by Capretz (1987) or Fidelman's (1992) *In the French Body* and *In the German Body*, described below).

As an example of computer-based multimedia, Fidelman (1992) has produced the interactive videodisc learning materials *In the French Body* and *In the German Body,* which emphasize oral comprehension, oral production and the nonverbal characteristics of language. Both packages consist of a computer-controlled videodisc with unscripted conversations between pairs of native speakers for the respective languages as well as student exercises ranging from listening comprehension tasks to speaking exercises that focus on intonation, pronunciation, gestures, and body language. These packages embody a holistic concept of language instruction and use a "method acting" approach (originally conceived by Wylie, 1985) that encourages students to memorize and replicate the verbal and nonverbal aspects of native-speaker conversations. The software allows students and teachers to catalogue, in minute detail, the behavior and speech exhibited by native speakers. A speech analysis feature of the software is designed to aid students in obtaining more native-like intonation patterns.[18]

Design software to facilitate, record, and analyze interactions between and among speakers

The third function for technology that can be integrated into intonation instruction is in providing opportunities for interactions between speakers. Pennington (1989a), who advocated meaningful instruction in speaking and listening, pointed out that computers can both provide training in production and perception of speech and create environments that facilitate interaction. Citing the work of both Chaudron (1985) and Richards (1986), Pennington stressed the need for language learning software to move to skill-based and task-based learning activities that not only offer users practice in listening comprehension but also elicit and practice specific types of interactions, language forms, sound contrasts, or nuances of meaning signaled by intonation (cf. also Piper, 1986; Young, 1988). In other words, she extended the scope of intonation practice to include context and transactions, not only sentences. This sentiment that the segmental pronunciation teaching must decisively be left behind is one with which I wholeheartedly concur.

As suggested in the sub-section above, due to the importance of having natural discourse as the target input and output, software should first of all

present authentic speech samples in their cultural contexts and should call learners' attention to important intonational features. The second component of the software should contain activities requiring learners to interact and converse in pairs. Thirdly, the software should provide tools to enable the learners to record their utterances, pitch-track them, and view visual representations of their pitch curves. Such features are admittedly quite extensive and would require sophisticated implementation, but they address the deficiencies in previous software and would also take into practical account all the recent research on language pedagogy focusing on the discourse-level linguistic and cultural aspects of language learning.

Build research tools into the software

Besides providing learners with auditory and visual feedback of their intonation patterns and having them engage in dialogue with each other, technology can be integrated into intonation instruction as a research tool. Tracking tools can be built into the software so that the software can serve simultaneously as a pedagogical tool, a data collector, and a testing instrument. For example, a database for each speaker — assuming there is adequate hard drive space or storage capacity on peripheral devices — could be compiled. Utterances produced by learners could then be ordered chronologically and compared for progress over time. Research studies could be designed to evaluate a number of questions such as: (1) what are the effects of providing only visual feedback, only audio feedback, only descriptive contrastive feedback, or different combinations of the above three types of feedback? (2) what is the relationship between perception and production? (scores from perception exercises could be compared with performance on production tasks), and (3) what are the long-term effects of intonation training using different types of feedback and using both perception and production exercises?

Specifically, computer-based materials for the training of speech perception might contain listening comprehension tests as well as tests that focus on listening acuity and perceptual accuracy. Such tests focusing on perception rather than production could also increasingly be computerized to track and assess student performance. Automatic recordkeeping and scoring, for instance, would allow for individualized branching and pacing of a test for more differentiated testing. This recordkeeping capability of the computer would make possible detailed tracking of the stages which learners pass through in acquiring oral and listening skills, e.g., by measuring the gradual changes in

students' perception of intonation contours and in their production of intonation contours over time. Providing individualized feedback and micro-tutors would allow a much more sophisticated language learning pedagogy to emerge, which should ultimately heighten proficiency. Until this technology is fully implemented and has widespread accessibility, however, teachers must first know what it is about intonation that students ought to learn, begin to teach it in the classroom and determine the kind of feedback that is most useful to learners.

In addition to studies reporting on research into the effectiveness of audio and/or visual feedback from the computer, other studies have examined the possibility of using computers to actually evaluate speaking proficiency and pronunciation and thereby reduce the role of the teacher in modeling contrived interchanges that may be gender- or socially inappropriate. Molholt and Presler (1986) conducted a pilot study of the feasibility of computer-assisted evaluation of pronunciation and found that machine-generated scores of speech samples ranked the samples in the same order as trained human raters. Clark (1986) reported on a tape-mediated "semi-direct" test of speaking proficiency for English-speaking learners of Chinese. The test was designed to be scored and interpreted according to the ACTFL/ETS/ILR guidelines for proficiency tests. A similar project, *TOPS* (*Test of Oral Proficiency Skills*), was reported on by Lowe and Hughes (1990).

Conclusions

We have seen that the fields of ESL/EFL and the teaching of languages other than English have proceeded at different paces in both recognizing the importance of discourse intonation and in implementing its teaching. This disparity can be attributed in part to the difference in goals and situational settings between second and foreign language learning. Nonetheless, the two professions are moving in the same direction, and a first conclusion in terms of the research agenda for the future is that research is needed to determine both (1) how intonation and particularly discourse intonation are best *learned*, drawing on learning theory and research, and (2) how they are best *taught*, utilizing empirical studies of efficacy of method and/or means (e.g., auditory, instrumental, or other). In any case, the two perspectives must go hand in hand.

The recent scholarship on the teaching of pronunciation also indicates a trend toward incorporating meaningful, contextualized practice of sounds,

rhythm, and intonation into the broader domain of oral proficiency, as described by Pennington and Richards (1986) and Pennington (1989b) for ESL/EFL and as implied in the 1989 *ACTFL Proficiency Interview Tester Training Manual* for FL learning in the United States. Learning to perceive and produce appropriate intonation is tied to comprehension and is needed to be able to achieve cohesion and coherence in discourse. While the emerging ESL/EFL instructional materials do reflect these changing emphases, materials for languages other than English are lagging somewhat behind in implementation. However, although not necessarily explicitly building intonation practice into textbooks, FL authors in the U. S. are increasingly basing their materials directly on the *ACTFL Guidelines* with the goal of providing students with opportunities to negotiate, initiate, and manage discourse as well as to comprehend input. Therefore, FL researchers and instructors need to integrate pronunciation practice of both suprasegmentals and segmentals into these types of discourse activities, and perhaps into the *Guidelines* themselves, together with a feasible rating system. Suggestions for implementing such activities will be presented in the last two chapters.

As a second conclusion concerning the research agenda for the future and the implementation of intonation instruction, one promising avenue for the early 21st century is to take advantage of technological advances — particularly in computer-based materials but also (for perception and comprehension purposes) in *multimedia* in its broadest sense. However, as Pennington (1989b, 1996, 1998, 1999, 2000) points out, there is still a need for further experimentation and research to determine the effectiveness of the various methods. One of the greatest potential advantages of using computer-assisted pronunciation and intonation tutors, for example, is that the computer could serve both as a medium of instruction and as a tool for research, i.e., while teaching pronunciation, a software program could simultaneously keep detailed and thorough records of student performance and progress. Consequently, the recent studies on the effectiveness of visual feedback in teaching intonation and on the use of technology to develop new ways to evaluate speaking proficiency represent the type of ongoing research that must be conducted in addition to developing better and more efficient computer software.

Recent (and increasingly rapid) advances in technology have made possible the use of computers for instructional, evaluative, and research purposes. One of the most serious deficiencies in earlier and existing computer programs is that they do not contain enough pedagogical explanations of how students should interpret the visual representations of their intonation patterns

(cf. Pennington, 1999). The work available from the 1990s, however, makes it possible for us to suggest realistic directions for research and the development of software packages in the near future (cf. Chun, 1998).

Thus it is likely that, in the foreseeable future, the computer may ideally be able to provide an almost self-contained system integrating training, testing, and research in the development of pronunciation and listening proficiency. "Perhaps the most exciting possibilities combining language training, assessment, and research involve two-person interactions which are both facilitated and analyzed by a computer" (Pennington, 1989a, p. 119). As this implies, only if the interactive or discourse functions are integrated into computer software will full pedagogical advantage be taken of this technology. And only when concomitant research on the effects of such technology on the learning process is complete can the role of intonation in oral proficiency be redefined in conjunction with linguistic and applied linguistic research (see also Pennington, 2000).

These conclusions are in line with Morley's (1991) desiderata regarding the future of pronunciation teaching into the 21st century in which she states that ESL (and EFL) teachers need to be equipped with knowledge of suprasegmentals and their functions in interactive discourse, that pronunciation teaching should be integrated with second language acquisition research, and that continuing research into the nature and development of learners' L2 phonological systems is essential (p. 511).

Notes

1. A special issue of *SSLA* on "Interlanguage Phonetics and Phonology" in June 1998 contains seven articles, but none of them deals with intonation.

2. Cf. Arteaga, 2000; Chun, 1988a; Elliott, 1995a, 1995b; McCandless & Winitz, 1986; McGinnis, 1997; Pennington & Ellis, 2000; Schairer, 1992; van Els & de Bot, 1987.

3. Cf. Ecklund & Wiese, 1981; Hieke, 1981; Mantini, 1980; Scanlan, 1987.

4. Cf. Chun, 1988b; Wipf, 1985.

5. Cf. Dansereau, 1995; Frommer & Weitz, 1981; Jourdain & Schuler, 1999; Landick, 1995; Shelly, 1998; Walz, 1980.

6. Cf. R. Brown, 1990; Elliott, 1997; Guzmán, 1992; Kelm, 1987; Simoes, 1996.

7. Cf. Chun, 1988a, 1988b; Guzmán, 1992; Kelm, 1987; Mantini, 1980; Pennington & Ellis, 2000; van Els & de Bot, 1987. In addition, from other journals, cf. Anderson, 1979 on some of the subtle differences between English and German intonation and how English speakers learning German could be made aware of these differences; Pegolo, 1985 on the role of rhythm and intonation in the silent reading of French as a foreign language; Pritchard, 1985

on teaching French intonation to native speakers of English.

8. Cf. Acton, 1984; Anderson-Hsieh & Venkatagiri, 1994 on syllable duration and pausing in the speech of Chinese ESL speakers; A. Brown, 1988; de Bot & Mailfert, 1982 on intonation; Esling & Wong, 1983 on teaching voice quality, a feature that is often characterized as a suprasegmental; Flege, 1981; Hill & Beebe, 1980; Leahy, 1980; Levis, 1999 on intonation in theory and practice; Molholt, 1988 on computer-assisted instruction of English segmentals and suprasegmentals to Chinese speakers; Morgan, 1997 on identity and intonation; Murphy, 1991, 1997; Neufeld, 1980; Pennington & Richards, 1986 on intonation; Riney et al., 2000; Temperly, 1983. Articles on pronunciation in other journals include Hieke, 1984, 1985; Tumposky, 1982.

9. Cf. Avery & Ehrlich, 1992; Dauer, 1993; English, 1988; Gilbert, 1984; Hahn & Dickerson, 1999; Henrichsen et al., 1999; Hewings & Goldstein, 1999; Morley, 1992; Orion, 1988; Pavlik, 1986; Sheeler & Markley, 1986; Wong, 1987 for American English, and Bradford, 1988; Brazil, 1994; Dalton & Seidlhofer, 1994; Kenworthy, 1987; Rogerson & Gilbert, 1990 for British English. Fifteen pronunciation textbooks were reviewed in the Book Notices section of *TESOL Quarterly*, 1994, pp. 757–776.

10. Although the term "communicative competence" was first used by Hymes (1967, 1972), it was not until the 1980s that the concept found widespread application to second language teaching and learning, cf. Canale & Swain, 1980; Canale, 1983; Savignon, 1972, 1983. The concept of "proficiency" began to be studied carefully by the language teaching profession in the late 1970s. The *ACTFL Provisional Proficiency Guidelines* appeared in 1983 and the revised *ACTFL Proficiency Guidelines* in 1986. Cf. Higgs, 1984; Liskin-Gasparro, 1982; Omaggio, 1986; Omaggio Hadley, 1993.

11. Lantolf and Frawley (1985) attribute the phrase "maximal interlocutor co-operation" to Spolsky (1984).

12. In the recent 1996 "Standards for Foreign Language Learning: Preparing for the 21st Century" for U.S. grades K-12, published by the National Standards in Foreign Language Education Collaborative Project, no mention at all is made of pronunciation or intonation.

13. Cf. textbooks such as Fischer and Richardson, 1989 for German, whose aims are to teach discourse strategy beginning at the elementary level. The topics they use all "grew out of the functional syllabus for the book, which is squarely based on the *ACTFL Proficiency Guidelines*" (W. Fischer, personal communication, 1990).

14. Computer programs for more traditional pronunciation exercises such as practicing vowels, consonants and common phrases or expressions are also currently available. Many of these include digitized sound, and some include perception tests and/or graphics of the mouth, face, and tongue.

15. The URL for Kay Elemetrics is: www.kayelemetrics.com. A low-cost, Windows-based speech analysis software program from Kay is *Multi-Speech*, but its inability to perform real-time analysis limits its usefulness for research or pedagogical applications. For Apple II computers, the *Video Voice Speech Training System,* which uses a graphic display of the human voice, is available from Micro Video Corporation, 210 Collingwood, Suite 100, P.O. Box 7357, Ann Arbor, MI 48107; (313) 996–0262 (URL: www.videovoice.com).

16. There are a number of speech analysis systems generally available, many of them designed primarily for acoustic phonetic research with potential for pedagogical application. For DOS- and Windows-based computers, cf. *CSRE* (Canadian Speech Research Environment), developed in Western Ontario (URL: www.icis.on.ca/homepages/avaaz/); *CSpeech*, developed at the University of Wisconsin, Madison (contact Paul Milenkovic: milenkovic@engr.wisc.edu); *WinPitch,* developed by Philippe Martin (URL: www.winpitch.com); and *WinSAL-V SPEECHLAB,* a two-part program developed in Trier, Germany (URL: www.media-enterprise.de/winsal/). *SPEECHLAB* is an acoustic phonetics analysis package; *WinSAL*, Speech Analysis under Windows, is an interactive multimedia program for teaching elementary phonetics, though not intonation per se. For UNIX environments, cf. *Speech Filing System*, developed at University College London (contact Mark Huckvale: SFS@phonetics.ucl.ac.uk) and *ESPS/Waves+* (Entropic Signal Processing System) by Entropic Research Laboratory (URL: www.entropic.com).

17. The CSAE (Corpus of Spoken American English) is a project of the Santa Barbara Center for the Study of Discourse, Department of Linguistics, University of California, Santa Barbara. See also their website: http://www.linguistics.ucsb.edu/research/sbcorpus/default.htm.

18. Cf. Johnston & Milne, 1995, who describe a multimedia tool, *Teacher's Partner,* that provides examples of authentic target-language discourse, i.e., communicative exchanges among native speakers, and was employed to improve classroom interaction patterns. Although this tool does not pay specific attention to pronunciation, it is a noteworthy instance of the benefits of using multimedia tools to provide authentic discourse to learners.

Part III

From theory to practice

Teaching discourse intonation

As was seen in the preceding two chapters, applied linguistic research on the teaching and learning of pronunciation has been growing steadily over the last two decades, particularly in the second half of the 1980s and throughout the 1990s. Applied linguists and language teachers are now stressing the importance of pronunciation not just in the interest of encouraging near-native accents but more importantly for communication, thus going beyond individual sounds and phonemes to sentence-level stress and intonation. A brief survey of materials available for teaching pronunciation will be given here as an introduction to Part III. Subsequently, some of the general problems endemic to teaching intonation in the classroom will be discussed. Chapters 6 and 7 then present some specific suggestions for incorporating discourse intonation instruction into the curriculum.

Traditional treatment of intonation in teaching materials

Since most language teachers and textbook authors themselves have learned intonation informally and indirectly and since linguistic accounts of intonation tend to be highly theoretical and not readily applicable to language learning situations, it is not surprising that while most instructors in the field might acknowledge the importance of intonation, they themselves personally tend to avoid teaching it explicitly in the classroom. Fortunately, this situation is gradually changing, as is reflected in recent teaching materials. Traditional programs of pronunciation teaching, however, frequently did not include attention to intonation, and if they did, then not to the discourse-level functions of intonation.

It may be somewhat surprising that treatments and mention of intonation can be found in nineteenth-century handbooks on pronunciation, e.g., for British English, Sweet (1890). During the twentieth century, some of the more

pedagogically oriented handbooks and manuals that focused specifically on intonation include, for British English, Palmer (1922); Armstrong and Ward (1926); Kingdon (1958); O'Connor and Arnold (1961, 1973); Cook (1968); Halliday (1970); Brown and Yule (1983) and for American English, Pike (1945); Fries (1953, 1954); Morley (1979); Prator and Robinett (1985).

Although many of these earlier handbooks on pronunciation did contain intonation components, they were for the most part designed as reference works, with very little emphasis being given to intonation in classroom instruction. As for textbooks, the great majority of those both for ESL/EFL and languages other than English contained minimal reference, if any, to intonation or how it should be taught and learned. Passing reference (or perhaps "sweeping dismissals") such as the following were typical: "Stress, rhythm, and intonation in the sentence are best learned by listening to the spoken language"[1]; for spoken exercises accompanied by a tape program, "pauses permit students to imitate the native pronunciation and intonation."[2] One Teacher's Edition suggested that the teacher "draw arrows on [the] board to indicate rising and falling intonation patterns" and that "many students are unaware of intonation patterns in their own language."[3] Thus, although many textbooks acknowledged that prosody should be learned, they gave no indication of what they meant by prosody or how it should be learned or practiced other than by imitation of tapes and through help from instructors.

Recent revival of teaching intonation

It was not until the mid-1980s that the subject of intonation seems to have been revived and brought to the forefront of pronunciation teaching — particularly with regard to the function of intonation in context, situations, and communicative activities. Since the mid-1980s, there have been a growing number of teacher resource books on teaching pronunciation for ESL/EFL, e.g., Avery and Ehrlich (1987), A. Brown (1991), Brown and Yule (1983), Bygate (1987), Morley (1987), and Swan and Smith (1987). In addition, there are a number of handbooks that specifically address the suprasegmentals and suggest ways of teaching them: Avery and Ehrlich (1992), Celce-Murcia et al. (1996), Dauer (1993), English (1988), Gilbert (1984, 2nd ed. 1993), Hahn and Dickerson (1999), Henrichsen et al. (1999), Morley (1992), Orion (1988), Pavlik (1986), Sheeler and Markley (1986), and Wong (1987) for American English and Bradford (1988), Brazil (1994), Dalton and Seidlhofer, 1994,

Kenworthy (1987), and Rogerson and Gilbert (1990) for British English.

Clennell (1997) argued for raising the pedagogical status of discourse teaching. However, Levis (1999) recently revisited the issue of intonation in theory and practice and asserted that "intonation as currently presented in North American textbooks bears a strong resemblance to textbook treatments from 30–50 years ago" (p. 37). He attributed the lack of innovation in adequate teaching materials to two factors: first, an overemphasis on the role of intonation in signaling grammatical relations and in conveying speakers' attitudes and emotions, and second, a lack of communicative purpose and a focus on uncontextualized, sentence-level practice of intonation forms. He thus proposed four principles for intonation teaching materials that echo those I suggested in Chapter 5: (a) intonation must be taught in context, (b) intonational meanings must be generalizable, (c) the teaching of intonation must be subordinate to larger communicative purposes, and (d) intonation should be taught with realistic language (p. 37).

Let us examine some of the existing materials to see whether Levis' criticisms are well-founded. In one recent teaching handbook on British English, *Intonation in Context: Intonation practice for upper-intermediate and advanced learners of English*, Bradford (1988) aimed "to help learners of English to perceive the system of intonation used by native speakers and ultimately to incorporate the system into their own performance, [...] to be used [...] as a supplement to a communicative coursebook" (p. 1). Based on the theories of Brazil (1975, 1978, 1985), intonation was viewed by Bradford primarily as a feature of discourse and is related not only to grammatical features or attitudes as had traditionally been the practice, but more importantly to its systematic use in interactions between and among speakers. The main functions of intonation were presented in eight units and include highlighting, telling and referring, marking roles and status of speakers, signaling "old" information, and contrasting. Activities were sequenced to have students become sensitized to or made aware of certain intonation patterns, read explanations of them, and then progress to imitating them, practicing them with a partner, and finally to using them in communicative situations. Bradford's emphasis on having learners practice intonation in contextualized situations for communicative purposes of interacting with their partners thus refutes Levis' critique that teaching materials lack communicative purpose.

In Gilbert's handbook on American English *Clear Speech* (1984, 2nd ed. 1993), the chapters progressed from the smallest to largest units, from individual sounds to words, sentences, conversations, and ultimately to "public

speaking." The unit "Stress: reduced vowels and the effect on rhythm" thus dealt not only with polysyllabic words in English and their stressed and unstressed syllables, but also with word pairs such as *can/can't*. The pairs were contrasted, starting with the unstressed forms, and it was then shown how the rhythm was affected when they were used in sentences, i.e., how the word *can* in *I can go* would usually not be stressed, whereas in the parallel *I can't go*, the negative *can't* would often be stressed with discernible rhythmic consequences. This type of pragmatic, discourse-level difference, i.e., the differing placement of sentence stress in the positive vs. the negative utterance, was often neglected in previous handbooks on pronunciation, with the focus typically on the syntactic, sentence-level features of negation.

A further example in Gilbert (1984) integrating the practice of individual sounds and syllables into larger sentential or discourse domains concerned the difference between *content* and *structure* words, where students were asked to identify each type in sentences before they practiced emphasizing the *content* words by lengthening the vowels in the stressed syllables and reducing the vowels in the unstressed syllables. Learners then progressed to discourse-level phenomena and learned that in English, at the beginning of a discourse, the last content word in a sentence is usually the focus of meaning, whereas after the beginning of the discourse, when the topic has already been established, any word can be a new focus. This emphasis by Gilbert on having learners understand the importance of a word in its surrounding context or discourse refutes Levis' claim that instructional materials have focused on uncontextualized, sentence-level practice.

In their ESL/EFL handbook for American English, Celce-Murcia and Goodwin (1991) emphasized that intonation needs to be brought to learners' conscious awareness and suggested four steps for doing this with dialogues: (1) "have students listen and mark stressed words," (2) "have students mark each line with an intonation contour (e.g., rise, fall, rise-fall)," (3) "read certain lines with various intonations and ask the students to decide which mood is being expressed (e.g., anger, sadness, amazement)," and (4) "ask students to read or act out the entire dialog in one particular mood and to note the variations in intonation patterns" (pp. 140–42). However, these suggestions were very general and no further details were given, e.g., as to how "angry" or "amazed" intonation could be identified or described in the L2, or what sorts of scenarios could lead to the "angry" or "amazed" emotions. Nevertheless, this handbook, as well as Bradford's, Gilbert's, and recent handbooks by Henrichsen et al. (1999) and Hahn and Dickerson (1999), did go beyond the traditional sen-

tence-level patterns and were concerned with dialogue-level, discourse-level intonation. Levis' critique is thus not entirely valid for at least those books discussed here.

In addition to addressing *production* of intonation patterns at the discourse level, most recent handbooks for teaching pronunciation give equal emphasis to *perception*, i.e., they include listening identification and comprehension exercises in addition to exercises for practicing sounds (segmentals) and sentence patterns. Many of the exercises are designed to be done by pairs of students, and students are asked both to produce the various sounds and to listen to their partners and determine whether they can distinguish the sounds or intonational patterns their partner is trying to make.

For example, for American English, Wong's (1987) *Teaching Pronunciation: Focus on English Rhythm and Intonation,* stressed the important preparatory steps leading up to pronunciation practice. She pointed out that learners must first acknowledge the psychological factors that inhibit them from wanting to change their pronunciation as well as how important it is to actually change it. According to Wong, a subsequent important component of teaching pronunciation is providing opportunities for focused listening as well as checking to see whether effective listening skills are being developed. As these emphases imply, it is not sufficient to provide exercises to practice the production of sounds or sentences, but learners must first be taught to listen for and discriminate actively between similar sounds or sentence intonations before they can be asked to imitate or produce them on their own.

For languages other than English, textbooks in the United States since the mid-1980s do reflect the changing emphases in language teaching toward communicative competence and proficiency but their treatment of pronunciation and intonation has not increased substantially, even though the approaches to language teaching and learning are more favorable to incorporating intonation into syllabi. For example, Fischer and Richardson (1989), in *Wie bitte?*, attempted to teach discourse strategies at the elementary level. Most activities in the text were structured around using language at a functional level for problem solving and for negotiating meaning in German culture. Informal, general intonation advice was included as a part of this discourse strategy, e.g., students were advised to use echo questions and the clarification question *Wie, bitte*? (literally "How, please?" meaning "What was that again, please?") as conversational stalls. Students were also told to create a question by repeating the last statement of the speaker with a questioning intonation (p. 15). Although there was little explanation or explicit notation on intonation in the

textbook, learners were encouraged to pay attention to intonation in communicative contexts.

For pedagogical handbooks on languages other than English, Benware's (1986) introduction to the phonetics and phonology of German included a chapter on stress, accent, and intonation. Three primary types of intonation patterns were listed for German (rise, fall, sustained), and the author discussed *unmarked* sentence accent ("where 'unmarked' refers to the accent pattern when the sentence is spoken in isolation or without any special contextual condition"), *contrastive,* and *emphatic* accent, the latter two allowing speakers to give prominence to syllables or words that would not be stressed in a more neutral situation or context (p. 115). However, Benware cautioned that "isolated sentences [...] can obscure the fact that prosody — intonation, sentence accent, duration, as well as tempo and pause — is closely linked to nonlinguistic factors, especially the situation. By 'situation' is meant the relationship between speaker and hearer, their shared knowledge, the social setting, and the disposition or mood of the speaker [...] The prosodic features of utterances play a significant role in our interpreting the speaker's attitude, but in addition, previous linguistic context, the social context, and the meaning of the utterance all make a contribution as well" (p. 113). However, as the purpose of Benware's book was not to provide pronunciation practice per se, no applications of these ideas were provided for learners to practice. In Chapter 7 of this book, suggestions for incorporating speaker attitudes as well as demonstrating the role of intonation in conveying contextual information will be presented.

In the field of *Deutsch als Fremdsprache* or Teaching German as a Foreign Language, some progress has been made in both research and pedagogical arenas in terms of recognizing the importance of intonation. Breitung (1994) found at a colloquium on the role of phonetics and intonation in instruction of German as a foreign language that these areas lacked emphasis in language programs, testing guidelines, and research publications. An analysis of sixty-six textbooks (fifty-one for language learning and fifteen specifically for teaching phonetics) published in Germany from 1980–1992 revealed that the inclusion of phonetics and intonation in instructional materials and curricula did not parallel their importance for the acquisition of linguistic and communicative competence (p. 7). For example, only 30% of the textbooks dealt explicitly with intonation.

A recently published CD-ROM for pronunciation training in German, *Phonothek interaktiv* by Hirschfeld and Stock (2000), included discussions of intonation (pitch, intensity, speech tempo, pausing, and tone of voice, as well

as contrastive and emotional accentuation) in its "Rules" sections as well as exercises to recognize and practice intonation and accentuation. A relatively small number of "emotional intonations" were included, e.g., neutral, emphatic, undecided, ironic, friendly, pleased, angry, sad, disappointed, anxious, and surprised. However, although sentence accentuation includes attention to theme-rheme structures, contrastive accents, and emotional accents, the rules and exercises are not taken to the discourse level.

There is another aspect to touch upon with regard to the renewed interest in pronunciation teaching, namely the sociolinguistic question of which pronunciation to teach. Dalton and Seidlhofer (1994) pointed out that pronunciation is never an end in itself but rather a means to negotiate meaning in discourse, and that teachers should accordingly be aware first and foremost of how pronunciation functions in language use. Furthermore, they discussed in the first chapter of their book how pronunciation is closely related to social and personal identity and asserted that "therefore careful consideration needs to be given to both the feasibility and the desirability of forcing learners into a mould of 'correct' pronunciation based on native-speaker norms" (p. 9). They thus recommended that teachers ask not what is *correct* in relation to a native-speaker norm (e.g. Received Pronunciation in British English), but rather what is *appropriate* and necessary to be able to communicate in specific situations.
In sum, great strides have been made in the last decade in ESL/EFL handbooks for teaching pronunciation and intonation. There has been a focus on listening and perceptual training, on the features of stress, intonation, and rhythm, and on contextualizing exercises to provide learners with more authentic situations for practicing purposeful communicative activities that require differentiated intonational patterns.

For languages other than English, with the focus on communicative competence and proficiency, textbooks from the late 1980s and through the 1990s were found to contain many more activities and exercises that promote conversation and active use of the language than had been the case earlier — meaning that the principles being espoused were in tune with contemporary pedagogical trends. However, although the very nature of communicative competence and proficiency goals suggests that improved pronunciation and intonation are of utmost importance, there has not been a commensurate inclusion of pronunciation exercises, and there are few detailed suggestions in textbooks and handbooks for the teaching of pronunciation or intonation in the classroom. Although neither teachers nor authors make overt reference to intonation, it is hoped that at least in the best-case scenarios, intonation is

being taught indirectly. That is, with all the emphasis on spoken input, communication, and contextualization, perhaps intonation is attended to and imitated despite not being directly addressed. Even then, more explicit attention and practical guidelines are surely needed.

Presented in the remaining chapters are concrete suggestions for practice and instruction in both listening/perception and in production as well as for teaching awareness of the use and function of intonation in discourse. Even if materials and curricula for languages other than English are still several years behind those of ESL/EFL, it is hoped that we may soon see more widespread attention devoted to teaching pronunciation, particularly among those materials employing discourse-centered, top-down methods.

Problems of teaching intonation in the classroom

As discussed in the preceding sections, teaching language and language use in context has become a priority in recent years. Prior to this new emphasis, a first limitation on integrating intonation into language instruction derived from the types of discourse actually prevalent in most classrooms. Much of what transpired (and to some extent still transpires) in typical classrooms did not consist of natural discourse, but instead either of "teacher-talk" or responses to questions or imperatives from the teacher — with few, if any, student-initiated utterances (cf. Kramsch, 1983, pp. 175–190). Students sometimes posed questions, but these questions were almost always directed at the teacher and very rarely at fellow classmates. Moreover, the very traditional textbooks tended to set up one-sided production or a stimulus-response structure, not true social interactions.

In the same vein, teachers traditionally used (and still use) a much wider range of communicative prerogatives than students; their "privileges" included addressing others, selecting the next speaker, choosing the topic, interrupting, asking for clarification, changing the subject, and concluding a discussion — none of them occurring with anything approaching a "real-world" frequency. Unlike in normal social settings, everything that the teacher said and did exercised control over others' behavior, thus managing both the classroom and its discourse (cf. Kramsch, 1981, pp. 13–17).

Furthermore, the intonation patterns used in the classroom are often not correlated naturally with the syntax used. Teachers' utterances are very likely to be interrogatives, though they are usually not genuine requests for informa-

tion. Rather, they are frequently designed to elicit a pre-determined response for the purpose of drilling a particular grammatical item or for checking comprehension and are thus not representative of an individual's free use of questions and answers. With each successive repetition of a question, its credibility as a true interrogative diminishes, and by the second or third round of questioning, the intonation used by the teacher is often statement-like, with falling or low termination. At the same time, both teacher and students focus on a discrete point of information and not on the by then highly artificial "discourse."[4] Conversely, students unsure of their responses often use high or rising terminal contours with their statements, as they are genuinely, if subconsciously, seeking confirmation or evaluation of their reply — thus "mixing messages" between syntax and intonation. In these ways, students thus are usually afforded minimal exposure to genuine discourse and appropriate intonational patterns.

An even more serious problem arises due to the fact that most of the exchanges in the classroom are unnatural from a discourse standpoint and unequal sociolinguistically. Learners lack the privilege of discourse management — they do not learn to steer conversations or elicit reactions. They thus often feel that their knowledge of grammar is inadequate if, in a genuine conversation with native speakers, the speakers do not understand them. The problem, however, may well lie elsewhere: their grammatical competence may be sufficient, but, in addition to faulty intonation, they may not have managed to make the necessary transitions in discourse, e.g., to interrupt in a way the native speaker recognizes as an attempt to take the floor or to ask for clarification of a specific item and not for repetition of an entire clause or phrase. In terms of comprehension, they may not know where the focus of a sentence might be in the L2 and may not be attending closely enough at the point where new information is commonly stressed — that is, they might misapprehend the native speaker's sentence stress.

In restructuring the syllabus and curriculum to facilitate interactive skills and intonational competence, then, the first priority must be to make discourse situations reflect real communication wherever and whenever possible, i.e., to decentralize the power of discourse management by decreasing the amount of teacher control. If the teacher's verbal control is diminished, the students will then have to develop their discourse management skills: learn to interrupt each other, take the floor, give an opinion, choose to avoid a topic, react, evaluate, agree or disagree, ask for clarification, etc. Such a move away from teacher-centered classrooms is already reflected in many of the more

recent textbooks and syllabi. The second priority in the classroom is for teachers not to treat the acquisition of intonation as automatic, but as something that must be learned actively in discourse situations that are as natural as possible. Students should be provided with systematic practice in intonational contrasts, both for production and perception. They should be presented with authentic discourse situations and activities in which they convey specific kinds of information via intonation as well as with authentic input for listening comprehension (e.g., television or radio broadcasts and interviews). Ways of implementing the second priority will be suggested in the remaining two chapters.

Teaching discourse intonation

In pronunciation teaching, a dual-focus program combining "a *microlevel* focus on *speech production* (i.e., a focus on discrete elements of pronunciation in a bottom-up sense) and a *macrolevel* focus on *speech performance* (i.e., a focus on general elements of communicability in a top-down sense)" presents the best approach (cf. Clennell, 1997; Morley, 1991, p. 497; Pennington & Richards, 1986; Pennington, 1989b, 1996, 1998).

Applying this same dichotomy to the teaching of intonation, I suggest that although one might initially teach stress and rhythm at the word and phrase level or *microlevel* (syllable structure, elisions, assimilations, reductions, contractions), the crucial next step is to progress to teach stress and intonation at the sentence, and more importantly, at the discourse level (*macrolevel*). There is some evidence that a "rich" context for teaching pronunciation distracts the learners from features of pronunciation (see Derwing et al., 1997). However, if instruction proceeds systematically from the word and phrase levels to the levels of sentence and discourse and begins with awareness training before requiring production, then the "distractions" may be reduced. At all levels, it is important to include perception/listening exercises as well as production/speaking activities. Chapter 6 thus contains suggestions for teaching and learning stress production and placement, including a focus on rhythm and unstressed syllables and on predicting placement of stress. Although some of the exercises described begin at the word and phrase level, in keeping with the purpose of this book, the goal is improving pronunciation of connected speech, and many of the exercises are therefore embedded in larger discourse frames. Chapter 7 focuses closely on intonation and pitch patterns at the

sentence and discourse level, emphasizing the need to go beyond the word level, and suggests ways of teaching intonation as well as how to transform existing materials into intonation and discourse-management practice.

Notes

1. Cf. Feld et al., 1973, p. 364.

2. Cf. Weiss, 1980, p. 14.

3. Cf. Moeller et al., 1986, pp. s31–32.

4. Cf. Hewings, 1987 regarding research on the role of intonation in the organization of classroom interaction, particularly the intonation used by the teacher when providing feedback in an EFL setting.

Chapter 6

Teaching stress and rhythm

Stress and rhythm are closely interrelated components of language. As suprasegmental features of language, i.e., patterns that exist on a level "above" that of individual vowels and consonants, they lend a distinct and often unmistakable character to the sound of a language. Infants learn to recognize the rhythm of their native language and apply it automatically and unconsciously to their own speech as soon as they begin to talk (cf. Lieberman, 1986). The tendency is for learners of a second or foreign language to subconsciously apply the rhythm of their native language to any language they are learning, unless they are explicitly made aware of differences between the two languages and have opportunities to practice the rhythm and stress patterns of the new language. Of particular importance to second language learning is the fact that rhythm provides a "systematic guide to the structure of information in the spoken message" (cf. G. Brown, 1977, p. 43). Stress and rhythm serve the dual function of (1) helping speakers to mark word stress as well as sentence stress and (2) helping listeners to identify both individual words and the most salient piece of information in an utterance.[1] One of the main purposes of this book is to heighten instructors' awareness of suprasegmentals and their function in language. Instructors must be able to listen to authentic speech in the target language, train their own ears to recognize stressed syllables and words, and then help make learners aware of how the target language (L2) realizes stress so that the learners can in turn perceive and comprehend better and ultimately produce more native-like rhythmic patterns. As mentioned in the introduction to Part III of this book, from a sociolinguistic viewpoint, teachers must decide whether nor not to target a particular native-speaker norm (e.g. Received Pronunciation in British English or Standard American English) and proceed from there in teaching word stress, sentence stress, and rhythm.

In this chapter, the definitions of relevant terms will be reviewed briefly and the various features of stress and rhythm to be taught and practiced will be discussed, first at the word level, then at the sentence level, and finally at the discourse level. To be examined is how pronunciation is affected by syllable structure, vowel length, reductions, assimilation, linking, and elision. These

phenomena are discussed first in terms of how they affect vowels and consonants within individual words and subsequently in terms of the overall rhythm of speech beyond the word level. In addition, suggestions for listening (perception) exercises will be offered for both the word and sentence levels, followed by suggestions for speaking (production) practice.

Definitions

Stress

As discussed in Chapter 1, *stress* is the linguistic means of marking syllables or words as *prominent* or perceptually salient in relation to others. For the purposes of this book, the term *stress* applies primarily at the *syllable* and *word* levels for any syllable or word that is prominent, no matter how prominence is achieved. Typically, language instructors unfamiliar with basic intonational knowledge explain stress as added emphasis, basically synonymous with loudness. However, acoustic phonetic and linguistic studies have shown that stress, when defined as that feature which makes a syllable prominent, is determined and marked by three suprasegmental features in the following order of importance (see Chapter 1; Bolinger, 1958; Fry, 1958) — three features increasingly discussed in pronunciation handbooks as well (cf. Hagen & Grogan, 1992):

– fundamental frequency, i.e., pitch change
– duration, i.e., length
– intensity, i.e., loudness or volume

Accent, differentiated in this book from *word stress*, is used here to refer to syllables or words that are stressed or made prominent at the *sentence* level or *discourse* level. The term is used synonymously with what some call *sentence stress* and others call *pitch accent* because the movement of pitch on the accented syllable is the crucial feature in marking prominence and in characterizing the intonational contour of the sentence or intonation-group. Duration also plays a very important role in stress, and intensity is again the least significant of the three features. For *accent* then, as for word stress, the three factors, in order of importance, are:

– fundamental frequency (pitch)
– duration (length)
– intensity (loudness)

Rhythm

As defined in Chapter 1, the rhythm of a language is largely determined by strong "beats" falling on certain syllables or words, i.e., on the so-called stressed syllables of words, with varying numbers and types of unstressed syllables in between depending on the language. Crystal (1985) characterizes rhythm as referring "to the perceived regularity of *prominent units* in speech. These regularities may be stated in terms of [the following] — or some combination of these variables" (p. 266):

– patterns of *stressed* vs. *unstressed* syllables
– syllable *length* (long *vs.* short)
– *pitch* (high *vs.* low)

That is, in addition to the pattern of strong and weak beats (stressed and unstressed syllables), the length of these syllables or beats is also relevant. It has been suggested that the key to the rhythmic system of English is syllable length (cf. Wong, 1987, p. 23), and this may well be true of all languages. Bell (1978) asserts as universal among languages the function of vowels as representing syllable peaks (p. 155). The fact that vowels are generally the "carriers" of stress and intonation — simply because when they are stressed, they are generally voiced and capable of having pitch and length — makes it true that for most (if not all) languages, vowel duration and syllable length are critical features of rhythm.[2]

Rhythm can be viewed and discussed at two different levels, that of the word (where syllable length and syllable timing are involved) and that of the sentence (where sentence emphasis or accent are involved). For example, in English, rhythm is based on varying syllable lengths but also on the effects of sentence emphasis; it is a so-called *stress-timed language* because the syllable lengths vary and the rhythm is determined by where the stresses in a word and in a sentence occur. The *Concise Oxford Dictionary of Linguistics* (1997, ed. by P. H. Matthews) defines a stress-timed language as one

> in which the intervals between stressed syllables in speech are either equal or at least more nearly equal than the intervals between the nucleus of each successive syllable and the next. Thus English, to the extent that, e.g. in a phrase such as incrédible explanátions, the interval in time between the two stressed syllables, [krɛ] and [nej], will tend, despite the number of intervening unstressed syllables, to equal that in phrases such as absúrd théories, where stressed [sɔ˞d] and [θij] are adjacent.

Stress-timed languages are also known as "irregular-syllable" languages because of the irregularity of syllable lengths. English is also said to be *isochronous,* because the stressed syllables tend to occur at regular intervals. Languages in which syllable length does not vary as greatly and in which most syllables receive approximately equal stress are termed *syllable-timed languages* or "regular-syllable" languages; the *Concise Oxford Dictionary of Linguistics* defines such languages as ones "in which the timing of syllables tends to be equal: e.g. Spanish." In syllable-timed languages, the rhythm of polysyllabic words and entire sentences is more regular, though sentence rhythm, much as in stress-timed languages, will also depend on the placement of sentence accents.

Listening practice

A common and advisable practice in teaching rhythm is to begin with the development of listening skills, to have learners focus perceptually on the placement of stresses and on the overall rhythm of the second or foreign language. In language teaching methods where a learner "silent period" characterizes the first few days or weeks of instruction, learners are asked simply to listen to and accustom themselves to the sounds and rhythm of the new language. However, even in programs that do not subscribe to this theory, exercises can be done to attune learners to listening for and actively perceiving features in the natural speech of native speakers as well as, subsequently, in their own speech.

It is not sufficient simply to expose students to hearing a language spoken; learners must be provided with focused practice in listening, e.g., with opportunities to focus consciously on which syllables receive stress and how the rhythm of a language can be described. A key principle expressed in this book is that different types of possible speech samples need to be provided, ranging from individual words and phrases to samples that fulfill the following *discourse-level* criteria (cf. Wong 1987, pp. 12–13):

- Samples should consist of texts or discourse in their entirety that go beyond the sentence level, e.g., conversations, story narrations, news reports.
- Samples should include different types of speaker roles and relationships, e.g., conversations or interactions among peers, speakers of the same or opposite sex, speakers of different ages and authority levels and with differing roles.
- Samples should consist of authentic speech wherever possible. If simplified or "staged" speech is used, speech should be as natural as possible (for example,

instead of slowing down the tempo, speakers should speak at normal speed but could insert longer pauses between sentences).

The following list provides a general overview of what students could be asked to do on a tapescript or transcription while listening to such speech samples (more specific examples will be provided later in the chapter):

- Indicate the number of syllables heard in a given word, phrase, or sentence.
- Mark the syllables that are perceived as being prominent or stressed in polysyllabic words or phrases.
- Mark the syllables that are perceived as being not pronounced at all, reduced, or "swallowed" in normal or rapidly spoken speech.
- Mark the words that are perceived as being prominent or accented in sentences or entire paragraphs.

The suggested procedure for such perception/listening exercises would be either (1) in the case of taped materials, to play a given passage several times and to have students listen for a single feature at a time, or (2) in the case of unrecorded, i.e., "live" speech, to have students focus on only one particular feature at a time. In general, training would typically progress from perception to production; students would first listen to spoken input until they are able to discern particular features, then do "active" listening exercises where they are asked to mark or code in writing what they are hearing, and finally practice pronouncing words, phrases, and sentences themselves. Specific examples will be provided in the sections below.

Speaking practice

In general, following listening and perception exercises, or conceivably introduced in conjunction with them, speaking exercises should be implemented in order for students to be able to apply what they have learned to listen for in the L2 to their own L2 speech.

Three integrated modes of practice for a pronunciation syllabus may be suggested, again with an ultimate emphasis on contextualized, discourse-level activities and progressing from imitating a model to guided practice without a model, to independent, self-generated speech (cf. Morley, 1991, pp. 509–510):

- imitative speaking practice
- rehearsed speaking practice
- extemporaneous speech practice

The purpose of imitative speaking practice is to focus on controlled production of selected pronunciation features; such practice should be kept to a minimum, if possible, and not used once the learner can produce the target feature(s) easily. It includes contextualized practice and may include self-study by individual students, pairs of students, or small groups outside of class as well as use of computerized speech-analysis systems that provide visual displays of user pronunciation and intonation.

Rehearsed speaking practice is an interim stage between imitative and extemporaneous speech, and its purpose is to work toward the stabilization of newly learned speech patterns so that the learner can manipulate them at will. Practice can include oral reading (the reading aloud) of a wide variety of scripts, pre-planned oral presentations, rehearsed performances for audio- and/or videotaping, and one-on-one sessions with the instructor or another native speaker.

For extemporaneous speech practice, the goal is to provide a wide variety of speaking tasks and activities simulating naturally occurring situations. It can include various in-class presentations, e.g., small-group panel discussions and follow-up interaction with the audience. In the following sections below, specific examples of how speaking practice can complement and expand on listening practice will be given for individual facets of teaching rhythm and stress, starting with stress and rhythm at the word level and progressing to stress and rhythm at the sentence and discourse levels.

Stress and rhythm at the word level

For rhythm at the word level, two main features come into play: (1) word stress, which in turn depends on vowel type and the syllable structure of the stressed syllable; and (2) polysyllabic rhythm, which is the pattern of stressed and unstressed syllables within a polysyllabic word.

Word stress

As defined above, the main factors that affect stress at the word level are, in decreasing order of importance, change in fundamental frequency or pitch, syllable duration, and loudness. Syllable length (or duration) is in turn affected by vowel type and syllable structure:

– vowel type: whether the vowel is long or short, full or reduced, tense or lax, or any combination of these qualities.[3]
– syllable structure: the phonotactic sequences or combinations of sounds that occur in a given language.

Vowel type

In English, there are two types of vowels — long and tense (e.g., in *gate, meat, goat, moot*) and short and lax (e.g., in *get, mitt, got, mutt*).[4] One way to start sensitizing hearers to listen for long vs. short vowels in English is to use minimal pairs of one-syllable words that contrast long vowels (and diphthongs) and short vowels, e.g.,

get [ɛ]	*gate* [e:j]
mitt [ɪ]	*meat* [i:j]
got [a]	*goat* [o:w]
mutt [ʌ]	*moot* [u:w]
did [ɪ]	*died* [a:j]

Learners can first be presented with pairs like these and asked to determine which word in the pair contains the longer (and tenser) vowel. As a supplement to the listening exercise, they could also be shown visual representations of actual acoustic measurements in order to confirm graphically the difference in length. Figure 1 below shows graphs of (1) the waveform and (2) the amplitude curves for the pairs *get/gate* and *did/died*; the waveform graph is a representation of the vibration of the vocal cords over time during speech and the amplitude curves represent the intensity or loudness of the speech. In addition, the duration of any segment of speech can be measured (typically in milliseconds, ms); as shown in these figures, this utterance of the word ***get*** is **363** ms in duration while that of ***gate*** is **503** ms, and ***did*** is **338** ms while ***died*** is **477** ms. Since acoustic phonetic research on stress has determined that duration is an important feature of stress, waveforms with the duration of words and syllables will be presented throughout this chapter.[5]

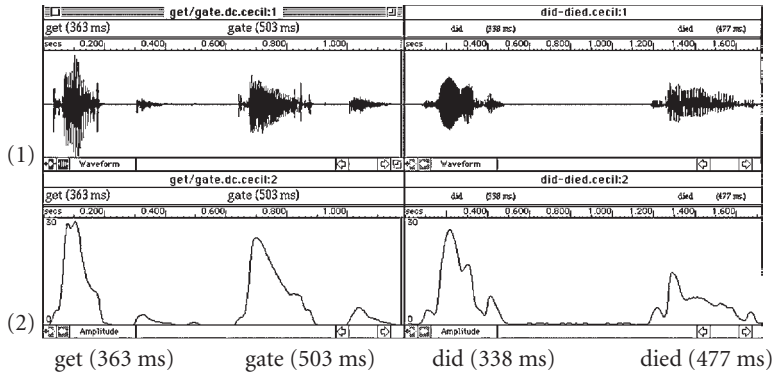


Figure 1. Waveforms and amplitude curves

As a next step, these words can be used in longer phrases or sentences, and the same procedure followed, whereby students listen to the entire sentence and determine whether the vowel is long or short (see Figures 2 and 3 below). This type of exercise goes beyond the word level to the sentence level, following the intention of this book to focus closely on sentence-level and discourse-level uses of suprasegmentals. Thus, in addition to being presented in isolation, these words can be embedded in sentences and learners can be asked to distinguish them in larger contexts, e.g.,

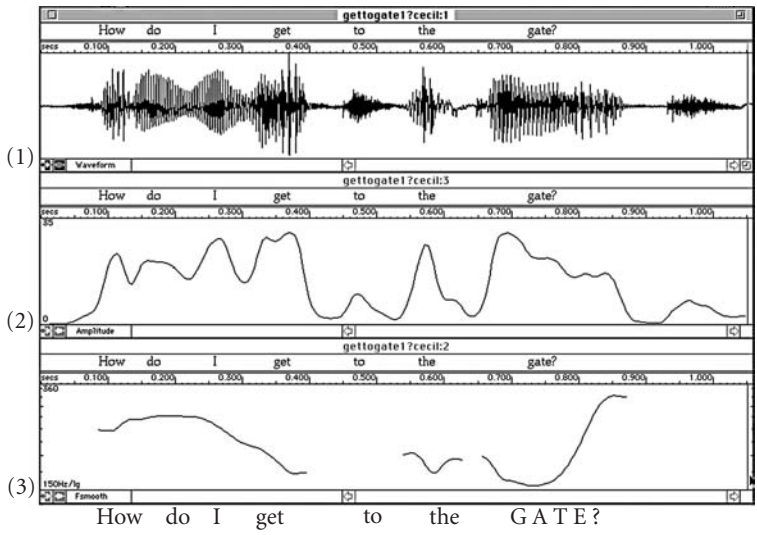


Figure 2. *How do I get to the GATE?*

In Figure 2 above, a third graph (3) shows the fundamental frequency or pitch curve of the utterance *How do I get to the gate?* The fundamental frequency curves presented throughout this chapter are intended to provide readers with initial information about how pitch is used to signal stress in individual words as well as in sentences. In Figure 2, for example, sentence stress is on the word *gate*, as shown by the steep rise in the pitch curve for this word. (Intonation curves and sentence intonation will be discussed more thoroughly in the next chapter.) One notes that the graphs or representations of these pitch curves are not always continuous lines. There are two reasons for this: (1) voiceless consonants cannot carry pitch, and therefore when there is a voiceless consonant in the word or sentence there is a "break" in the fundamental frequency or pitch curve (see the break in Figure 2 under the [t] sound of *get* and *to*), and (2) at the ends of utterances, when voice intensity drops too far, fundamental frequency cannot be calculated (see the end of the word *gate*). (Cf. Anderson-Hsieh,1994 for more detailed information about these shortcomings of computer software and ways to circumvent the resultant problems.)

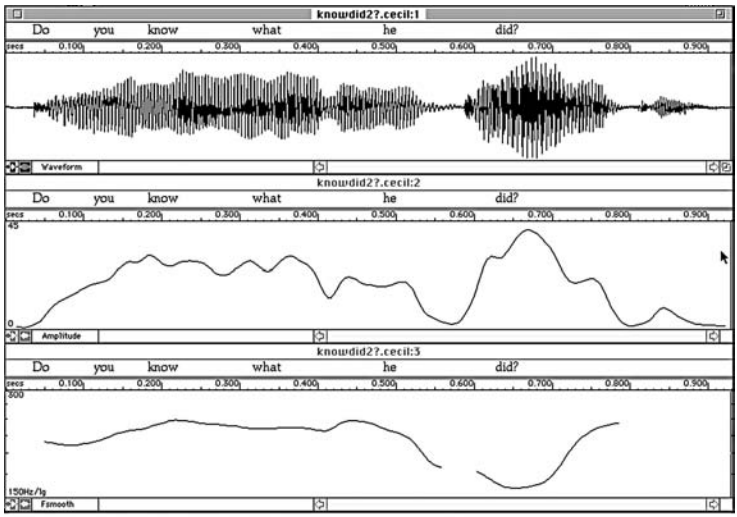


Figure 3. *Do you know what he DID?* (317 ms)

In Figures 3 and 4, the transcriptions of the speech included below the graphs capitalize the words that receive sentence stress, i.e., the words *DID* and *DIED* are most prominent. The word *died*, which contains a diphthong, is slightly longer in duration (379 ms) than *did* (317 ms), which contains a short vowel.

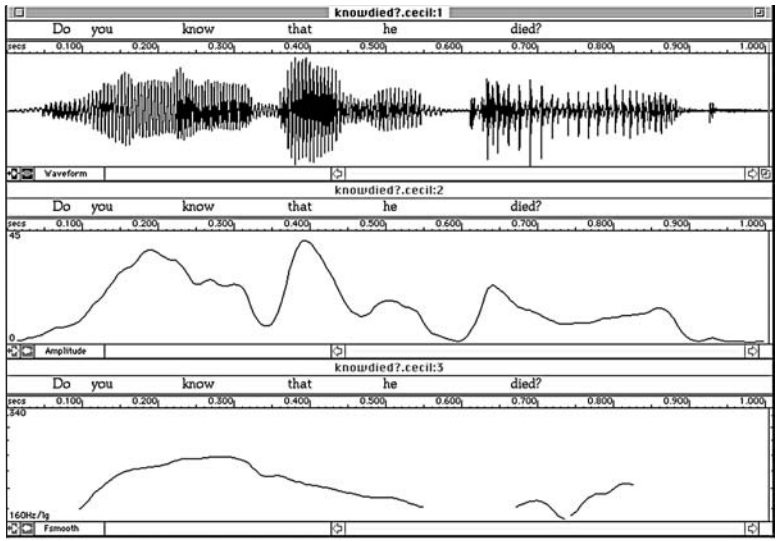


Figure 4. *Do you know that he DIED?* (379 ms)

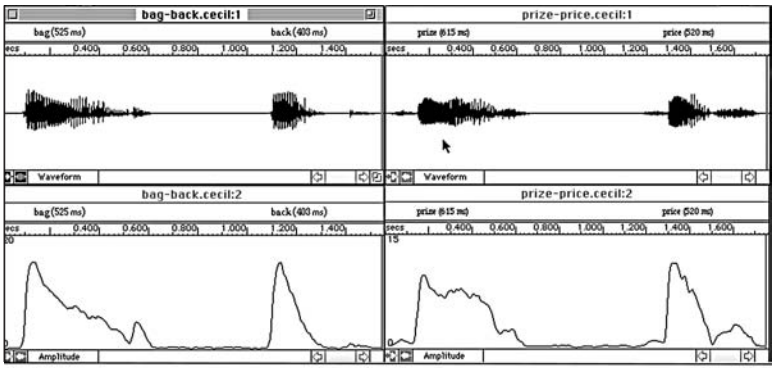
Syllable structure

In general, syllables that end in a consonant are called *closed* syllables, whereas those that end in a vowel are *open* syllables. In English, there are three different phonotactic environments in which vowels, both the long/tense and the short/lax vowels, may occur and which determine the relative length of the vowel. In order of relative length, vowels in closed syllables ending in a voiceless consonant are shortest (e.g., *goat, bet*), vowels in closed syllables ending in a voiced consonant are longer (e.g., *goad, bed*), and vowels in open syllables are the longest (e.g., *go*).[6]

– shortest: closed syllable ending in voiceless consonant (*goat, bet*)
– longer: closed syllable ending in voiced consonant (*goad, bed*)
– longest: open syllable (*go*)

Again, minimal pairs of one-syllable words can be presented to the learner to contrast long and short vowels and diphthongs in closed syllables before voiced vs. voiceless consonants, as shown in Figure 5 below (cf. Orion, 1988, p. 14):

cab	*cap*	*save*	*safe*
bed	*bet*	*bead*	*beat*
bag	*back*	*prize*	*price*

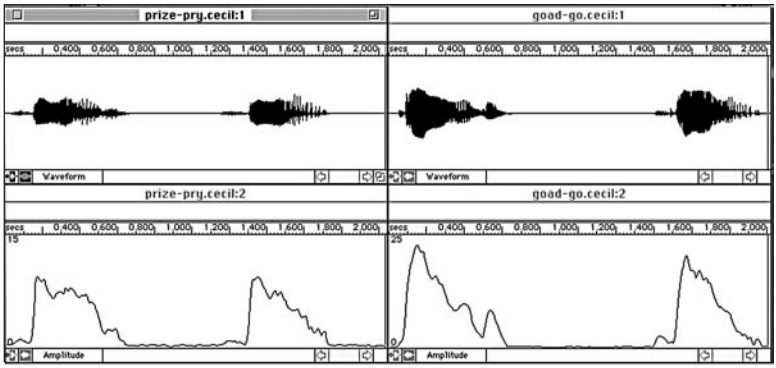


bag (525 ms); back (403 ms) prize (615 ms); price (520 ms)

Figure 5. Minimal pairs in closed syllables

and to contrast long vowels in closed vs. open syllables, as shown in Figure 6 below:

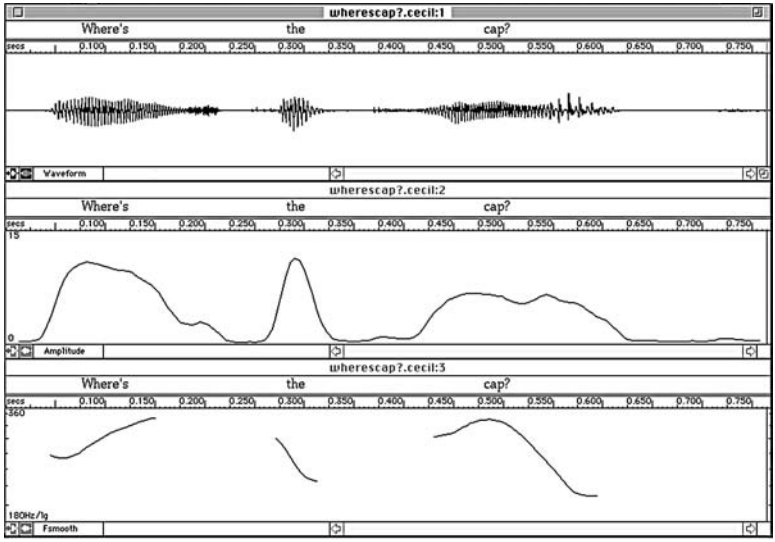
s<u>a</u>ve	*s<u>ay</u>*	*pr<u>i</u>ze*	*pr<u>y</u>*
g<u>oa</u>d	*g<u>o</u>*	*b<u>ea</u>d*	*b<u>e</u>*



[prai] in prize (474 ms) and pry (566 ms); [go] in goad (374 ms) and go (557 ms)

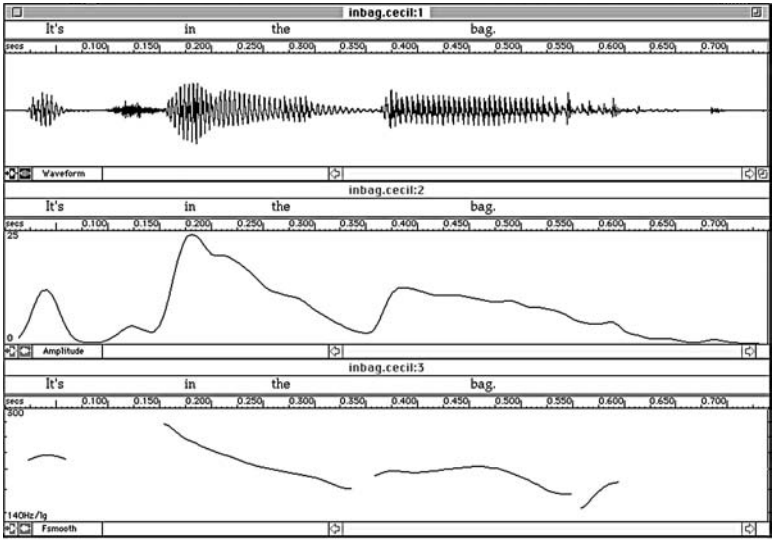
Figure 6. Minimal pairs in different syllable types

These words can also be embedded into question-answer sequences to go beyond the word level, as shown in Figures 7 and 8 below for the sequence (7) *Where's the cap?* (8) *It's in the bag* (potentially to be contrasted with the sequence *Where's the cab? It's in the back.*).



Where's the C A P ?

Figure 7. *Where's the CAP?* (247 ms)



It's in the B A G .

Figure 8. *It's in the BAG.* (340 ms)

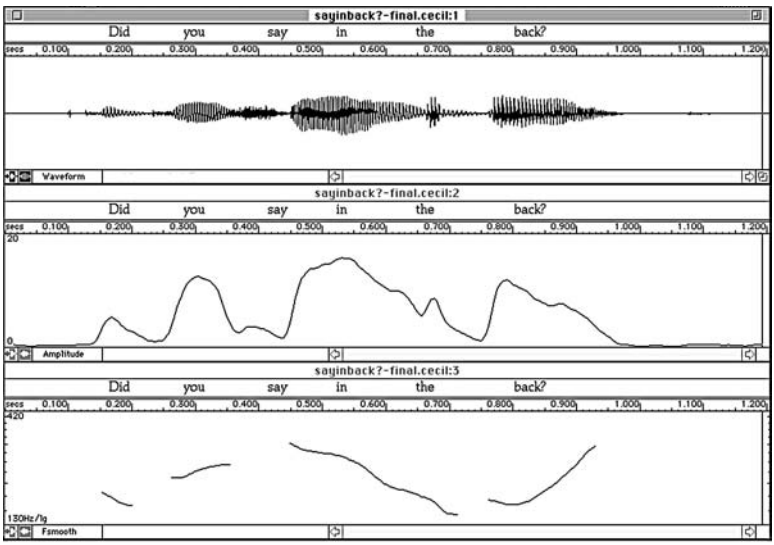
A further step is to use even longer sentences, since, typically, a vowel sound will be shorter when it occurs in the middle of a sentence than at the end, e.g.,

Where's the cab? It's in the back near the stable.
Where's the cap? It's in the bag near the table.

As a follow-up speaking exercise, students could work in pairs asking each other questions. The hearer would need to understand which word the questioner was saying and respond accordingly. The questioner would subsequently also decide from the reply whether the hearer had understood the original question, and if there had been a misunderstanding, the questioner would have to repeat the question, this time emphasizing and contrasting the two words. This type of activity helps learners to negotiate meaning and perform "repairs," as frequently necessary in everyday spoken language. At the same time, learners could practice emphatic and contrastive stress (e.g., see Figures 9 and 10 below, which also indicate the duration in milliseconds of the word being given emphatic or contrastive stress).

Speaker A:
(9) *Did you say in the BACK?*

Speaker B:
(10) *I said in the BAG!*



Did you say in the B A C K ?

Figure 9. *Did you say in the* BACK*?* (358 ms)

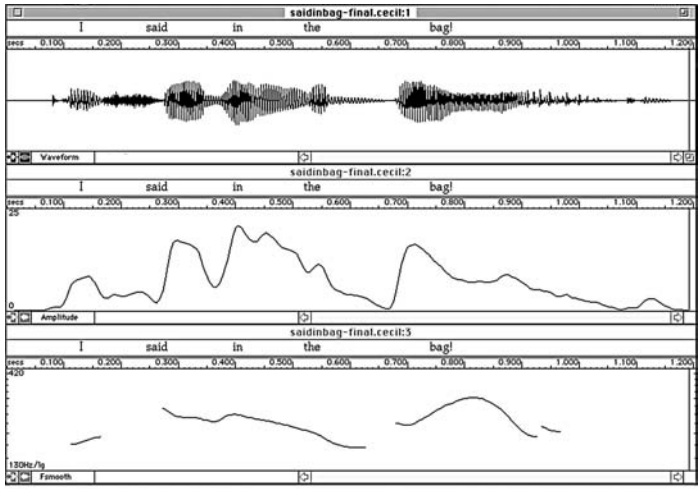


Figure 10. *I said in the* BAG*!* (486 ms)

Figures 11–14 below show additional examples of graded practice with the pairs *life* vs. *live* and *price* vs. *prize*, beginning at the word level, progressing to the sentence level, and then advancing to the paragraph level.

(11) *I like the price.*
(12) *I like the prize.*
(13) *That's life!*
(14) *That's live!*

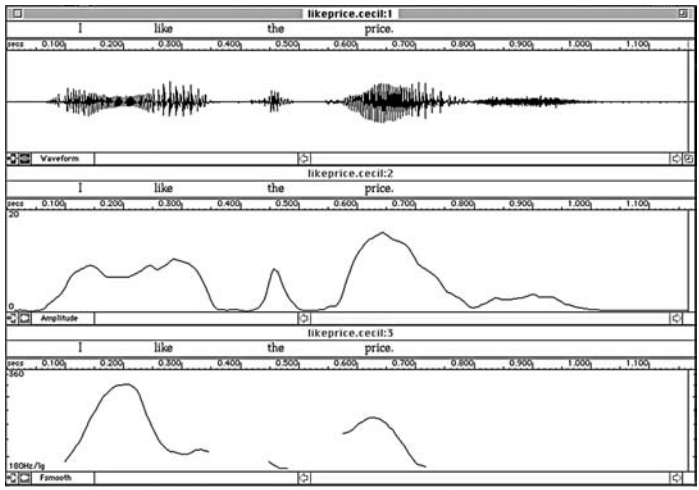


Figure 11. *I like the* PRICE. (515 ms.)

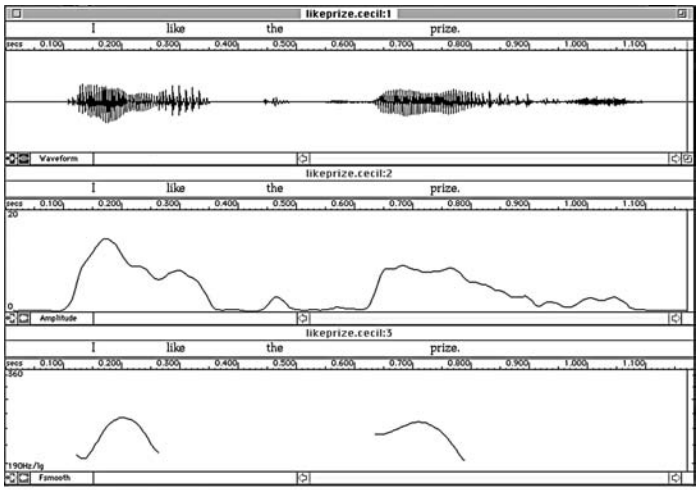


Figure 12. *I like the* PRIZE. (601 ms.)

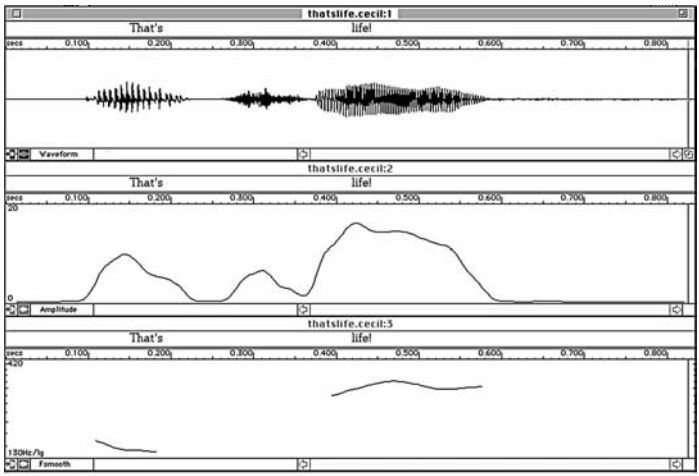


Figure 13. *That's* LIFE*!* (370 ms)

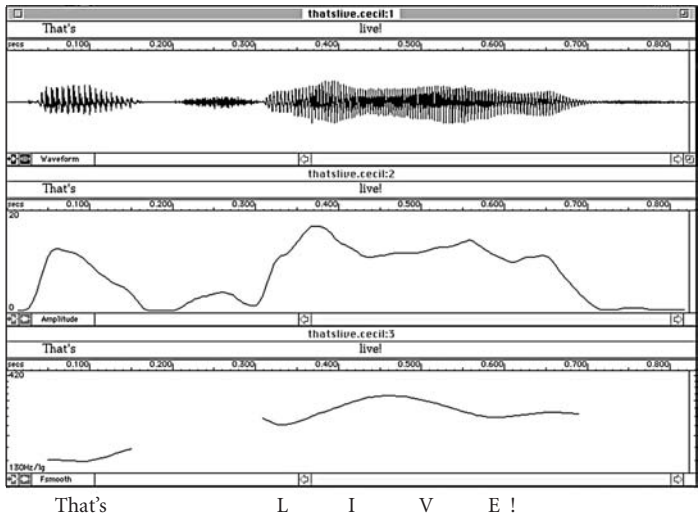


Figure 14. *That's* LIVE*!* (519 ms)

These short sentences could also be embedded in longer sentences or in longer paragraphs and discourses, e.g.,

> *That's live from New York/Hollywood/… (etc.)*
> *That's life in New York/Hollywood/… (etc.)*
>
> *I was once on a game show. The show was taped in New York, live. I actually won the game and a prize. The prize was nothing special, but being on television was. Getting there was expensive, but that's the price you pay, I guess. It was one of the more interesting things I've done in my life.*

For an initial listening exercise, students could be given the printed text and as the instructor reads the paragraph could mark the selected words as having longer or shorter durations. There are several cues: if the final consonant is voiced, then both the vowel and the consonant are longer than if the final consonant is voiceless. At the next level, they could be given (1) a list of the key words but not the text, and then asked to write down the words in the order that they hear them and indicate whether the final consonant is voiced or voiceless and therefore longer or shorter in duration than its counterpart. Alternatively, the students could be given (2) a cloze passage of the text with the key words left out and be asked to fill in the blanks with the correct words as they listen to the instructor read the passage.

I was once on a game show. The show was taped in New York, ____. I actually won the game and a _____. The _____ was nothing special, but being on television was. Getting there was expensive, but that's the _____ you pay, I guess. It was one of the more interesting things I've done in my ____.

For "rehearsed speaking practice," students could practice reading a text among themselves with various tasks. For example, they could work in pairs or small groups, with one student reading and the other(s) writing down the words they hear and the relative length of both the vowel and consonant produced by the reader — or students could be asked to tape themselves reading the passage as a sort of diagnostic quiz or speech sample to be evaluated and corrected by the instructor.

After learners have been introduced to the concept of stress and syllable length and have been exposed to such practical types of exercises, word stress can be explained from metacognitive or analytical perspectives. First, they can be given *rules* for the placement of word stress in the language to be learned.[7] Second, they can be taught the various ways in which stressed syllables are marked diacritically in different dictionaries so that they can then look up any word to learn which syllable to stress. And finally, in terms of production, students can explicitly be made aware of how the target language realizes stress acoustically, e.g., as has been shown for English, it is a combination of three suprasegmentals: pitch, duration, and loudness, in that order of importance. If learners have access to acoustic analysis software, they can record themselves saying words, phrases, and sentences. They can then take measurements of the length of various words or phrases they have uttered and compare the relative durations with those of their peers. Pronunciation software can and should provide guidelines for what is acceptable or near-nativelike in a given language as L2. In addition, as described more fully in the next chapter, learners can also examine their pitch contours and curves to see whether they have succeeded in using both duration as well as pitch change to signal stress and accent. Again, comparison of data among peers can prove to be both enlightening and motivational.

Word rhythm

The rhythm of a language is characterized by the timing pattern of successive syllables, and more specifically by the pattern of stressed and unstressed syllables. In many languages (but not English), every syllable is given approxi-

mately the same length. By contrast, English syllables vary in length; for example, a word may be composed of a sequence of a short unstressed syllable followed by a long stressed syllable and another short, unstressed syllable, as in the words *sucCESsive, conCLUsion, unFEELing.* This variation in syllable lengths is typical of English at both the word and the sentence level (cf. Wong, 1987, p. 22).

Stressed vs. unstressed syllables

The preceding suggestions for exercises for word stress dealt with monosyllabic words and vowel and consonant length. I now turn to polysyllabic words, phrases, and clauses and discuss how the rhythm of English is dependent on sequences of stressed and unstressed syllables at both the word and the sentence levels.

One of the most important features of English rhythm is that syllables are not equal in duration. As seen earlier, even comparing only stressed syllables, there are differences in vowel and consonant length that depend on syllable structure and on the position of the word in a sentence and its function in a longer discourse. In addition, in every word of two or more syllables in English, one syllable is stressed, i.e., is longer, is often at a higher pitch, and sometimes has more force than the other syllable(s) that is (or are) unstressed or weak(er). This pattern of stressed and unstressed syllables is characteristic of English and German, for example. However, in some other languages (such as Spanish, French, and Japanese), most syllables receive approximately equal stress. In English, this stressing and "unstressing" creates contrasts between and among strong and weak syllables, and these contrasts contribute to the rhythm of English.

Identifying stressed/long syllables

A type of introductory exercise to begin to sensitize students to listening for and perceiving rhythm is to have students first listen to polysyllabic words and to identify the stressed syllable. Orion (1988) suggests practicing word stress in English systematically by starting with one-syllable words, then progressing to two-, three-, and four-syllable words, etc. (p. 22). Within each group, subgroups would contain words with stresses on the first, second, or third syllable, e.g., for three-syllable words,

AC-ci-dent	*a-NOTH-er*	*dis-ap-POINT*
CIT-i-zen	*de-POS-it*	*guar-an-TEE*

This type of sensitizing exercise gives L2 students an opportunity to become acquainted with the various patterns in the L2.

There are variations in stress placement even among varieties and dialects of English. An example would be a word such as *laboratory*, which in American English receives word stress on the first syllable (and often reduces or eliminates the second syllable), whereas in British English it is the second syllable that is stressed, with the penultimate syllable often being elided. Other examples in which the penultimate syllable is elided include *secretary, strawberry, secondary.*

American English:	láboratory	[læːbɹətɔɹi]
British English:	labóratory	[labɔːɹətɹi]

Full and reduced vowels

Another preliminary exercise to do when teaching stress patterns in a second or foreign language is to make a list of cognates in the target language and the native language(s) of the students in order to compare cross-language syllable number as well as patterns of stressed and unstressed syllables (cf. Gilbert, 1984, p. 9). For example:

English	German	Spanish	Japanese
cocoa	*Kakao*	*cacao*	*kokoa*
- -	- --	- --	- - -
chocolate	*Schokolade*	*chocolate*	*chocolate*
- -	- - - -	- - - -	- - - -

Figure 15 below shows the waveform, amplitude, and fundamental frequency (pitch curve) for the words *cocoa/Kakao/cacao* in English, German, and Spanish, respectively. In English, the word stress is on the first syllable and the word contains two syllables, whereas in German and Spanish the word stress is on the second syllable which contains a diphthong. As seen in the waveform and amplitude curves, there are differences in the length of the stressed syllables. In German and Spanish, the stressed (second) syllable is significantly greater in duration than the stressed (first) syllable in the English word (498 ms and 522 ms, respectively, in German and Spanish for the stressed second syllable, and 185 ms for the stressed first syllable in English).

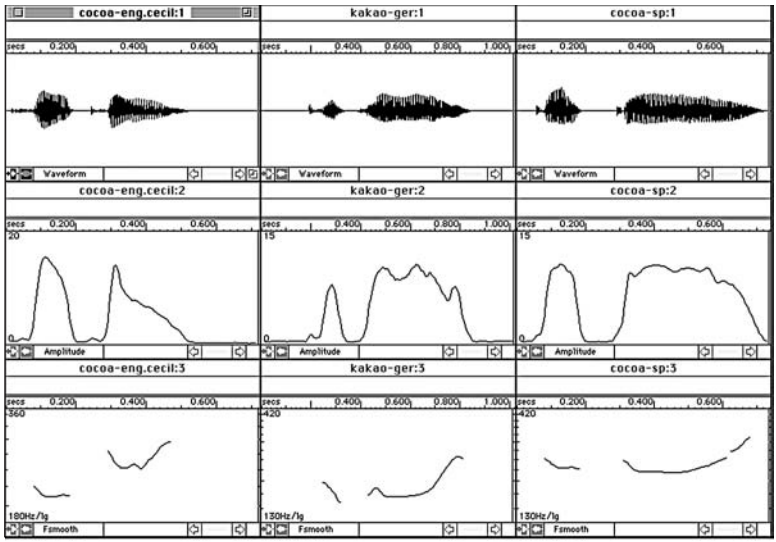


Figure 15. '*cocoa*' in English, German and Spanish

Similarly, in Figure 16 below, the word *chócolate* (two syllables) in English has word stress on the first syllable, but the corresponding words in German and Spanish, *Schokoláde* and *chocoláte,* respectively, are stressed on the third of four syllables. Again, the length of the stressed syllable in the German and Spanish words (306 ms and 262 ms, respectively, is greater than in the English word (219 ms for *cho-*). With regard to the pedagogical applications for such graphs, if language learners can visualize the intensity curves of what they produce as also showing the length of the syllables and compare them with those of native speakers, they might be better able to modify their production and achieve more native-like length of L2 syllables and words. At least some learners may also benefit from seeing their "performance" in "black and white" (concrete acoustic representations) and in repeated attempts to match native speaker examples. The visual "feedback" provided may help increase motivation as well.

Of these selected languages, English has the most unstressed or even completely elided ("swallowed") syllables while in the other languages each syllable tends to be pronounced. However, German is similar to English (and Russian and Arabic) in being a stress-timed language, as opposed to syllable-timed languages such as Spanish, French, Italian, and Japanese, where syllables are all of approximately the same length.[8]

Again, variation can be found among varieties and dialects of English.

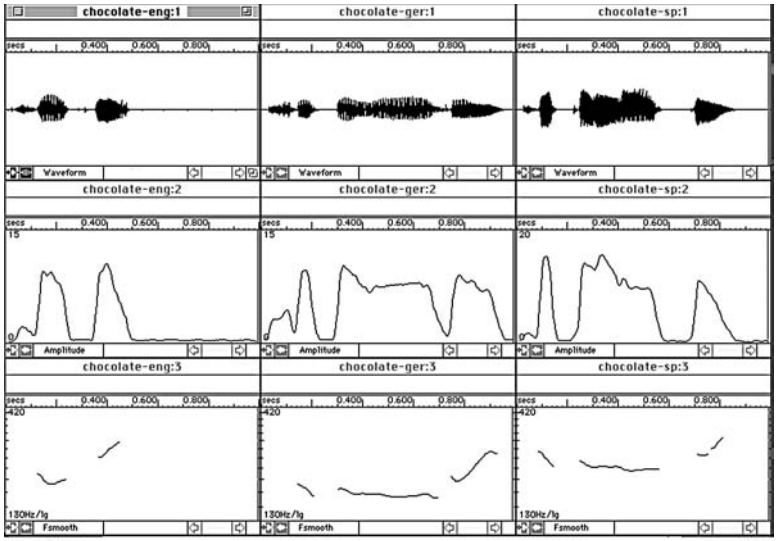


Figure 16. '*chocolate*' in English, German, and Spanish

Dalton and Seidlhofer (1994), for example, assert that Received Pronunciation (RP) in British English allows elision of at least one schwa per word, which explains why RP may sound "clipped" to speakers of varieties which do not follow this trend to such an extent (for instance, American English). As examples, they list the words *natural, penitentiary, exoticism, camera.*

As discussed, one of the most important features of English rhythm is that syllables are not equal in duration. In simple terms, Gilbert (1984, p. 21) posits three lengths for vowels in polysyllabic words (the underlined vowel in each example below):

– short: reduced (*SOf<u>a</u>*)
– half-long: full vowel unstressed (*<u>a</u>lteRAtion*)
– long: full vowel stressed (*alte<u>RA</u>tion*)

It is important to emphasize that vowel length is a relative phenomenon, and I echo Gilbert's (1984) sentiments: "These vowel lengths are not exact measurements, so there is no point in trying to reduce English to musical notation or in asking students to practice three-way or four-way timing contrasts. The purpose of this lesson is simply to increase their alertness to the irregularity of English syllables and to make them aware of the general principle that length adds emphasis" (p. 21).

In English words of two or more syllables, the unstressed vowel sound in both content and function words is often [ə] or [ɪ] (e.g., *b<u>a</u>nan<u>a</u>, fam<u>ou</u>s, sal<u>a</u>d, priv<u>a</u>te, tick<u>e</u>t; <u>a</u>bove, <u>a</u>side, <u>u</u>pon*). Because of the large number of unstressed syllables, in fact, [ə]/[ɪ] is the most common vowel sound in the language. Dauer (1993) presents a table of common unstressed suffixes in English with their reduced vowels (p. 63). In German, French, and Spanish, by contrast, while unstressed syllables tend to be shortened slightly, vowel quality is retained and vowels are not reduced to [ə]/[ɪ] to the extent as in English.

	English	German	Spanish
banana	[bəˈnæːnə]	[baˈnɑːnə]	[ˈplɑːtano]
computer	[kəmˈpjuːtəˑ]	[kɔmˈpjuːtɐ]	[komputaˈðoːra]
private	[ˈpɹaːɪvɪt]	[priˈvɑːt]	[priˈβɑːðo]

Wong (1987) suggests using common three-syllable words such as *bananas, pineapples, strawberries,* and *apricots,* or *computer, monitor, Macintosh,* and *IBM* and having students mark the short syllables with small dots and the longer ones with circles as they listen to the instructor read them (p. 26):

. o .	. o .
bananas	*computer*
o o .	o . .
pineapples	*monitor*
o . .	o . o
strawberries	*Macintosh*
o . o	o o o
apricots	*IBM*

Figure 17 below shows pronunciations of the words *compúter, Compúter, computadóra* (English, German, Spanish, respectively). Although the English and German words both have word stress on the second syllable, the first and third syllables in English are shorter, as shown on the waveform and the intensity curve, than the corresponding syllables in the German word. The Spanish word has two "extra" syllables and word stress on the fourth syllable. As in German, the unstressed syllables in Spanish are more fully enunciated than in English, as shown by the greater width of the intensity curves for each syllable, which indicates sustained length of vowels.

In choosing lists of words for classroom use, instructors should think of topics of particular interest and relevance to their students: words they need to use, words they will hear frequently, words in related lexical or semantic "families" (as illustrated in the examples above), and so forth.

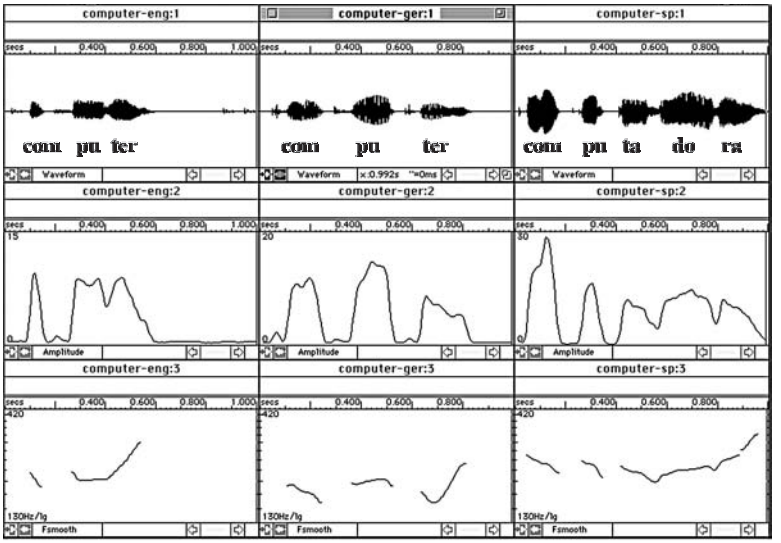


Figure 17. *compúter, Compúter, computadóra* (English, German, Spanish)

After students have been introduced to the idea of contrasting syllable length, they will need opportunities to listen to these contrasts in ordinary spoken language in contexts larger than just the word or the sentence. For example, one could take a text about computers in which commonly-used terminology appears repeatedly. The text can be read by the instructor and supplemented by having another native speaker read it on tape. On the first reading, students listen for the overall meaning. On the second reading, they are asked to count the number of times they hear the word *computer* or *computers* and to note the contrasting syllable lengths. On the third reading, they listen with the objective of being able to paraphrase the text to another student.

In summary, the intent here is to emphasize, first, that the rhythm of a language depends on the patterns of word stress and, second, that to complement a concerted effort to learn and practice the placement of word stress, words should subsequently be used in phrases or sentences. In phrases and sentences, it becomes important to note that syllable length and degree of stress are *relative*, depending on where in the sentence the word occurs. One important tool that should prove helpful for illustrating stress and its correlates (i.e., pitch, duration, and intensity) is computer software that allows for display of speech waveforms, intensity, and/or pitch curves. Representations of these

acoustic correlates can provide learners with concrete quantitative measures as well as qualitative graphic/visual impressions.

Stress and rhythm at the sentence level

In some of the preceding sections, brief references to and examples of stress at the sentence level were included as follow-ups on word-level activities. I now turn explicitly to the use of stress and rhythm at the sentence level and subdivide the discussion into three main parts, the first dealing with whether a language is a so-called stress-timed or syllable-timed language, the second dealing with sentence stress or *accent*, and the third dealing with the reduction and elision of words in connected speech.

Briefly, the main factors involved in stress and rhythm at the sentence level are:

- timing: whether the language is stress-timed or syllable-timed, that is, whether stressed syllables occur at regular intervals in connected speech with intervening unstressed syllables, or whether syllables tend to be equal in length with fewer intervening unstressed syllables in connected speech;
- accent: accented syllables, defined as those receiving sentence stress (as opposed to inherent word stress) are made prominent by the intonation or pitch contour of the sentence in which they occur and are also longer than unaccented syllables;
- reduced syllables: in English, these include syllables with reduced vowels, contiguous syllables that are linked or assimilated, and syllables that are lost or elided — at times beyond recognition

Stress-timed vs. syllable-timed languages

As described in the discussion of stress and rhythm at the word level, stress-timed languages are those in which stressed syllables tend to occur at approximately equal intervals. Unstressed syllables that occur between two stressed syllables tend to vary in length, depending on how many unstressed syllables there are. That is, no matter how many unstressed syllables there are, the interval between stressed syllables tends to be approximately the same length, so that if there are many unstressed syllables between two stressed syllables, each syllable will be even shorter than if there are fewer unstressed syllables. In

syllable-timed languages, on the other hand, syllables are roughly equal in length, so that the more syllables a sentence has, the longer it will be in duration.

Bolinger (1978) points out that English rhythm has a striking near-isochrony (i.e., the tendency to have equal intervals of time between accentual peaks) (p. 478).[9] Syllables with reduced vowels are said to "borrow time" from preceding syllables with "full" (unreduced) vowels, so that sentences such as the two in Figure 18 below are of nearly equal duration (only 69 ms difference, or less than a tenth of a second). The first sentence, *Marie drank tea*, consists of four syllables, while the second, *Maria drank some tea*, contains six syllables.

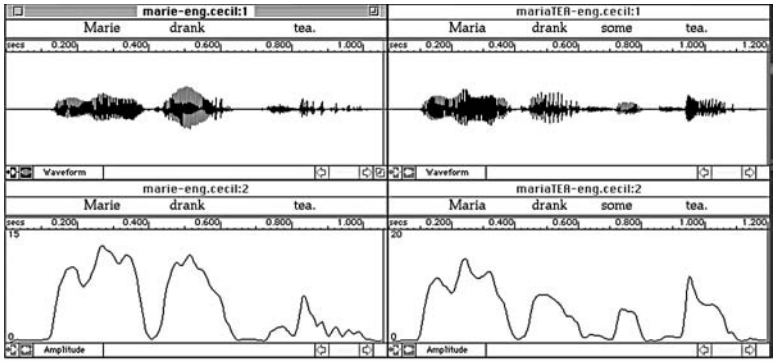


Figure 18. Duration of English sentences that differ by two syllables

The data in the table below confirm that German and Spanish tend not to reduce the length of unstressed syllables to the extent that English does. In the second sentence of the pair of English sentences, there were two additional unstressed syllables, but the 69-ms difference between the two sentences amounted to only 6%. In the second of the pair of German sentences, the three additional syllables added 238 ms (or 20%) to the duration, and in the Spanish pair, the second sentence also contained three additional unstressed syllables and was 176 ms (or 16%) longer in duration. Of note, however, is that a difference of 250 ms is only a quarter of a second. An interesting research question would be to determine whether listeners register such differences in length and thus to what extent such differences contribute to the perception of nonnative "accent."

Marie drank tea.	1,115 ms
Maria drank some tea. (+2 syll.; 69 ms/6% longer)	1,184 ms
Marie trank Tee.	1,196 ms
Maria trank einen Tee. (+3 syll.; 238 ms/20% longer)	1,434 ms
Mari bebió té.	1,120 ms
María bebió un té. (+3 syll.; 176 ms/16% longer)	1,296 ms

Figure 19 below shows the first sentence of these pairs in each of the three languages: *Marie drank tea* (4 syllables), *Marie trank Tee* (4 syllables), and *Mari bebió té* (6 syllables). Of note is that, as indicated by the display of amplitude, the intensity in English was great at the beginning of the utterance but decreased markedly at the end. In the other two languages, the intensity remained relatively constant in German and dropped slightly toward the end of the Spanish utterance.

Figure 20 below shows the second sentence in each language *Maria drank some [a cup of] tea* in English (six syllables), German (seven syllables), and Spanish (eight syllables). A feature to note here is that the intensity curves for the German and the Spanish sentences remain at roughly the same levels

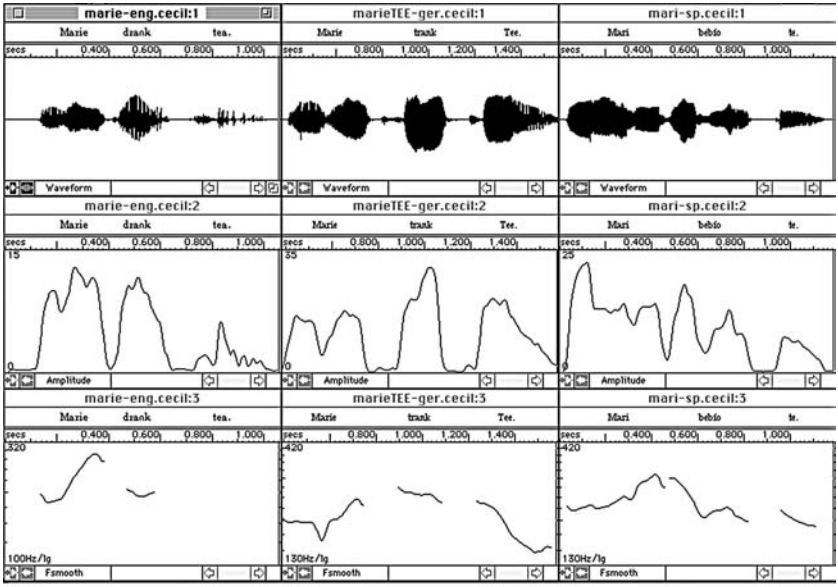
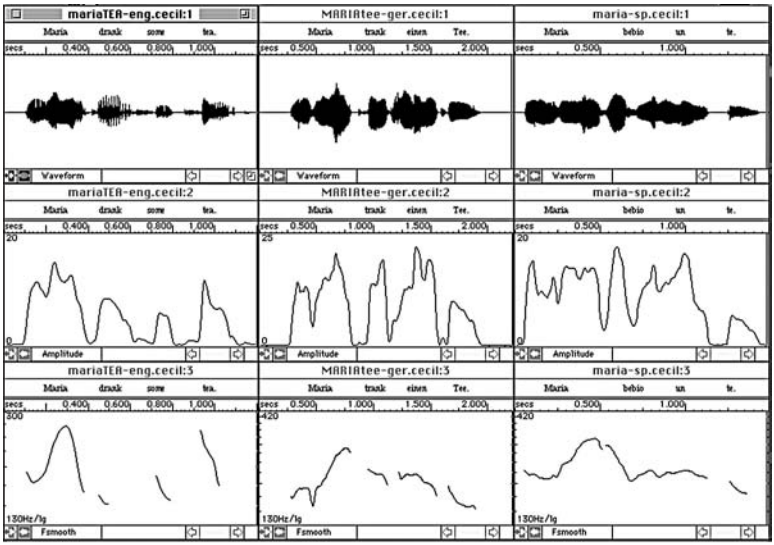


Figure 19. A comparison of English, German, and Spanish



Maria drank some tea. *Maria trank einen Tee.* *María bebió un té.*

Figure 20. Additional examples from English, German, and Spanish

throughout the utterance until the very end, when intensity drops dramatically on the final syllable. The English utterance, on the other hand, shows peaks in intensity on the first word, *Maria,* and then again on the last word, *tea,* but the unstressed syllables in between show lower intensity. In general, vowels in German and Spanish are pronounced with greater tenseness and intensity than vowels in English, and this is reflected in the amplitude waves. (Differences in the intonation or pitch curves will be discussed in the next chapter.)

Using *Visi-Pitch,* a software program discussed in Chapter 5, Fischer (1986b) compared the amplitude curves of English speakers learning French with those of native speakers of French and presented the example reproduced in Figure 21 below, where the top curve for each speaker represents the amplitude and the bottom curve the fundamental frequency or pitch. The differences to note between the native-speaker and the student pronunciations are (1) the native speaker has relatively constant amplitude (with the one drop in amplitude attributable to voiceless segments in the utterance), whereas the student shows more peaks and falls in amplitude; and (2) the native speaker starts with high pitch and then falls to and remains at a level pitch, whereas the student shows two peaks in pitch, a pattern typical of the stress-timing of English:

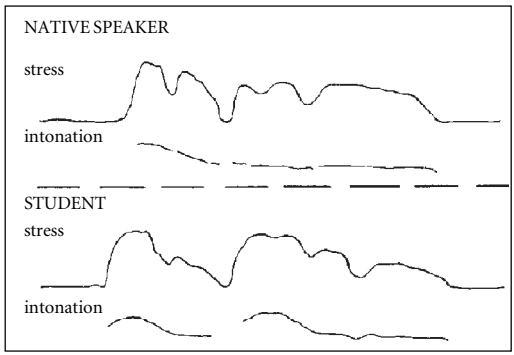


Figure 21. Native and non-native versions of the French sentence *Qu'est-ce qu'il fait?*

French is usually described as a language with fixed stress. Most traditional prosodic studies agree that French has a single rhythmic stress that is regularly assigned to the final full syllable (i.e. not containing a schwa) of the last lexical item of a stress group (Hirst & Di Cristo, 1998, p. 196). In the case of *wh*-questions, however, the pitch accent is placed on the stressed syllable of the question word, and following this initial prominence, the pitch drops regularly until the final syllable, as shown above (p. 205).

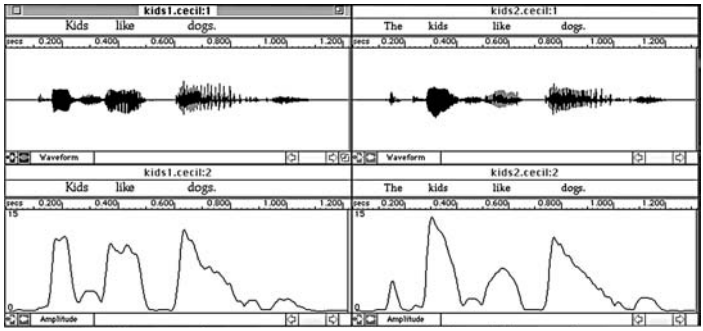
A type of exercise that can be done with L2 learners is suggested by Hagen and Grogan (1992), who present the following progression (p. 120):

(a)	*KIDS LIKE DOGS.*	1,062 ms
(b)	*The KIDS LIKE DOGS.*	1,120 ms (+5%)
(c)	*The KIDS LIKE the DOGS.*	1,265 ms (+13%)
(d)	*The KIDS might LIKE the DOGS.*	1,382 ms (+ 9%)
(e)	*The KIDS might have LIKED the DOGS.*	1,411 ms (+2%)

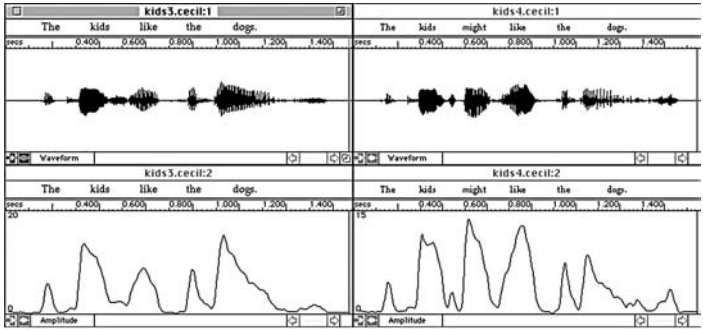
The capitalized words are content words, which are stressed or accented, while the uncapitalized words are function words that would not normally be stressed in a sentence. English, as a stress-timed language, tries to make stressed syllables occur at equal intervals, as it were. If there are unstressed syllables between the stressed syllables, they will be pronounced faster and reduced so that the speaker can reach the next "beat" on time. If, on the other hand, two stressed syllables are not separated by any unstressed syllables, they will often be stretched out longer in order to space them equally, e.g., compare *Kids* in (a) with *The kids* in (b) or *dogs* in (a) with *the dogs* in (c). In other words, the time it takes to say a sentence in English generally depends on the number of

stressed syllables, not on the total number of syllables. As is evident from the lengths of the sentences in milliseconds, unstressed syllables do increase the overall length, but only by relatively small increments, e.g., comparing the first two sentences, addition of the word "the" adds 58 ms or approximately one-twentieth of a second, but makes the second sentence only 5% longer in duration. The difference in length between the first sentence, with three syllables (all stressed), and the last sentence, which contains seven syllables (three stressed and four unstressed), is 349 ms, or less than half a second; it is, however, *not* twice as long, but rather 33% longer, even though it contains more than twice as many syllables.

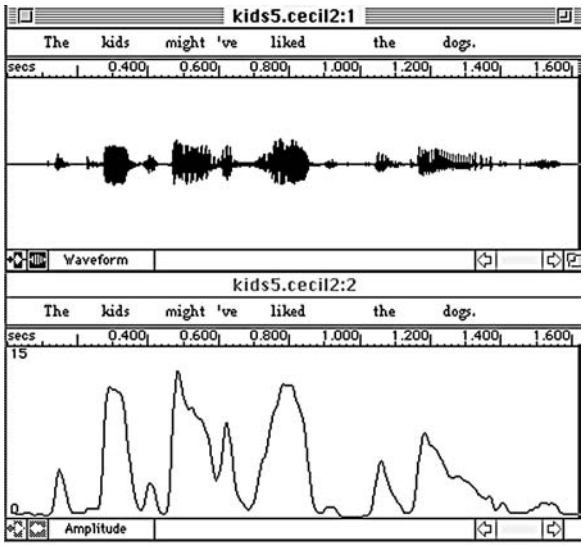
The figures below show the waveforms and the amplitude (intensity/loudness) curves for the five sentences. The content words exhibit greater intensity than the function words, which are unstressed. Interesting to note is that the intensity at the end of the sentence always decreases appreciably.



(a) *Kids like dogs.* (b) *The kids like dogs.*



(c) *The kids like the dogs.* (d) *The kids might like the dogs.*



(e) *The kids might have liked the dogs.*

Figure 22. Addition of unstressed syllables in (b)–(e)

By contrast, in languages that are syllable-timed, such as Spanish, French, Italian, and Japanese, the more syllables a sentence has, the more time it will take to say. Since the syllables are approximately equal in length, these languages sound as though they have a very even rhythm. For French, for example, Tranel (1987) asserts that "the basic rhythm is more dependent [than in English] on the enunciation of each syllable at a level of prominence which is practically speaking equal to that of its neighboring syllables" and that the prominence of stressed syllables is weaker than that of stressed syllables in English (p. 197), so that "from a practical standpoint, it is not unreasonable to aim for equally weighted syllables" (p. 200), in learning to pronounce French with a near-native accent.

For purposes of comparison (albeit with awkward word order in some cases), statements similar to Hagen and Grogan's in English were recorded by native speakers of German and Spanish in their L1. For the English example, successive sentences were increased by only one word/syllable. Due to the different syntactic structure needed to express these ideas in Spanish, more than one syllable may have been added to create the successive sentences.

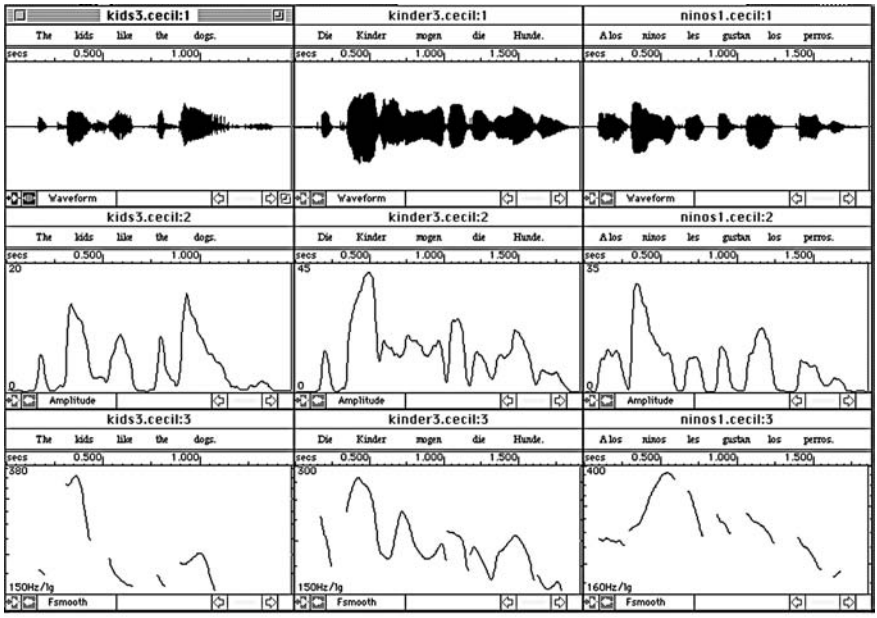
In the examples in English, successive sentences were increased by one syllable and averaged 87 ms more in duration per syllable. In German, succes-

sive sentences were increased by one or two syllables and averaged 218 ms more in duration per syllable. In Spanish, by adding more words (2–3 syllables per additional word), the sentence was lengthened by an additional 261 ms in duration per syllable. This is consistent with viewing English as a stress-timed language, where additional unstressed syllables do not increase sentence duration proportional to the number of syllables added. German, though also a stress-timed language, does not reduce unstressed vowels to the extent that English does, and therefore additional unstressed syllables are on average longer than their counterparts in English. Spanish, being a syllable-timed language, shows the longest average syllable length when additional syllables are added. (*In the Spanish sentences, (d) was compared with (a); **sentence (e) was compared with sentence (c).)

(a)	*Kinder mögen Hunde.*	1,236 ms
	(children like dogs)	
(b)	*Die Kinder mögen Hunde.*	1,498 ms
	(the children like dogs)	(+21%)
(c)	*Die Kinder mögen die Hunde.*	1,683 ms
	(the children like the dogs)	(+12%)
(d)	*Die Kinder mögen vielleicht die Hunde.*	2,090 ms
	(the children like perhaps the dogs)	(+24%)
(e)	*Die Kinder hätten vielleicht die Hunde gemocht.*	2,543 ms
	(the children would have perhaps liked the dogs)	(+22%)
(a)	*A los niños les gustan los perros.*	1,768 ms
	(the children like the dogs)	
(b)	*A los niños pequeños les gustan los perros.*	2,641 ms
	(the little children like the dogs)	(+49%)
(c)	*A los niños pequeños les gustan los perros grandes.*	3,454 ms
	(the little children like the big dogs)	(+31%)
(d)	*A los niños quizás les gusten los perros.*	2,374 ms
	(the children perhaps like the dogs)	(+34%)*
(e)	*A los niños pequeños quizás les gusten los perros grandes.*	3,593 ms
	(the little children perhaps like the big dogs)	(+4%)**

Figure 23 below shows phonetic displays of the three sentences *The kids like the dogs, Die Kinder mögen die Hunde,* and *A los niños les gustan los perros.* Because the German segmentals happen to be the most sonorant, i.e., the German words contain more nasals and voiced sounds than either the English or Spanish versions, the waveform for the German utterance shows heavy black

areas of voicing. Most noticeable is that the unstressed syllables in English are in fact very short in duration and low in intensity and that the intensity of syllables in English tends to drop sharply after the peak, whereas the intensity of each peak is slightly more sustained in both German and Spanish. In addition, the German and Spanish sentences show one intensity peak (on the grammatical subject, respectively *Kinder* and *niños*) with the remainder of the sentence having relatively similar intensity, whereas in English, the words *dogs* and *kids* (both direct object and subject) show high intensity peaks. It is these features that can be visualized with graphic representations such as the ones presented in this chapter that learners can be made aware of and in turn be enabled to use to sharpen their own pronunciation.



The kids like the dogs. *Die Kinder mögen die Hunde.* *A los niños les gustan los perros.*

Figure 23. Sentences in English, German, and Spanish

An exercise to practice "squeezing in" unstressed syllables is proposed by Hagen and Grogan (1992, p. 121):

Three Stressed Syllables	**Actual Syllables**
Sith is Thai.	3
Paul is French.	3
David is Czech.	4

Hilal is Turkish.	5
Fanny is Chinese.	5
Chuck-he is Korean.	6
Hendra is Indonesian.	7
Abdullah is Kuwaiti.	7
Hiroyuki is Japanese.	8
Jedediah is Canadian.	9

Although Hagen and Grogan posit that there are three stressed syllables in each sentence, I propose that there would normally be only two, unless the copula *is* were stressed to contradict a false assumption or idea. The two stressed syllables might be made evident visually if signal analysis software were used: learners could view a native speaker's intensity curves and would see only two peaks in intensity in non-contrastive contexts (but three for contradictions). In turn, learners could practice producing these and similar sentences and could check their intensity curves to insure that their unstressed syllables were reduced sufficiently in length and intensity.

Sentence stress or accent

As discussed at the beginning of this chapter, the three factors involved in *sentence stress* or *accent*, in order of importance, are:

– fundamental frequency (pitch)
– duration (length)
– intensity (loudness)

Syllable length — sentence level

Syllable length in English is in fact a *relative* phenomenon; that is, because it is influenced by a number of different factors, one cannot predict an absolute duration of a certain number of milliseconds for a given syllable in a given word (cf. Anderson-Hsieh, 1994 on the factors affecting syllable duration). A word spoken in isolation will have a different duration than when it is used unstressed in the middle of a sentence and yet another when it used stressed at the end of a sentence. For example, in answer to the question "Where are you going?" the following answers containing the closed syllable /hom/ are possible:

HOME. [home = 505 ms]
I'm HOMEward bound. [home = 165 ms]
I'm going HOME. ["Neutral" home = 337 ms]

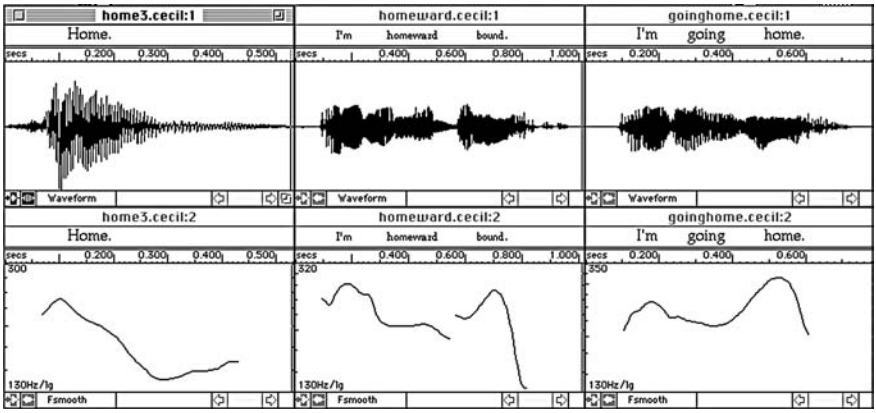


Figure 24. The syllable /hom/ in three different utterances

As shown in Figure 24 above, the first utterance, *HOME*, is a monosyllabic, unemotional response to the question "Where are you going?" containing a long vowel in a stressed syllable. In the second answer, if said in a matter-of-fact way, the syllable *HOME*, although it receives word stress, would probably not receive sentence stress, and the length of the syllable would be shorter than in either the first or the third instances. If the third answer were said neutrally — that is, with no particular emphasis — the syllable *HOME* would be shorter in duration than in the first case but longer that in the second because it occurs at the end of the utterance.

However, if the question had been a so-called echo question expressing surprise *You're going WHERE?,* then the answer might be *I SAID I'm going HOME!,* as shown below in Figure 25. If this answer were spoken in an exasperated tone, the added emphasis on *HOME* would cause the syllable to be lengthened considerably — to a duration perhaps approximately as great as when the word is spoken alone (in this example, 511 vs. 505 ms). One might also note the steeper falls in pitch in the curve of fundamental frequency appearing below the waveform.

The example in Figure 26 below shows the relative lengths of vowels in open and closed syllables. As the waveforms show, the vowel is longer at the

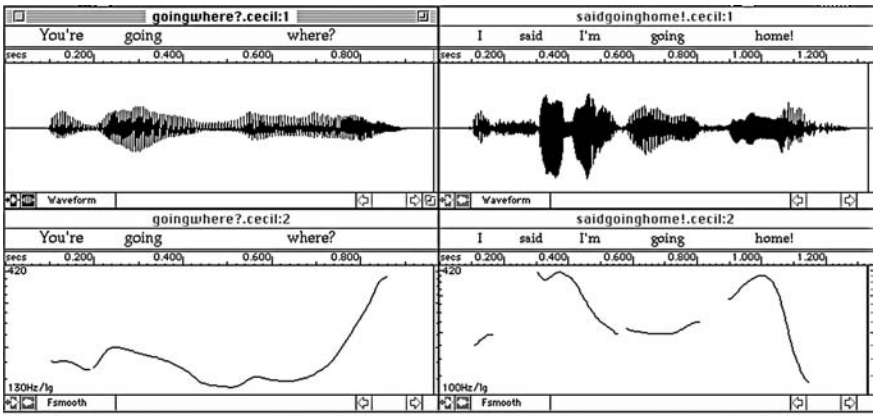


Figure 25. "Surprised" question and "exasperated" answer

end of sentences than in the middle regardless of whether the sentence-final syllables are open or closed (cf. Orion, 1988, p. 15):

	(g)o	*(h)o(me)*	*(sn)ow*
Let's GO.	235 ms		
Let's go HOME.	103 ms	193 ms	
Let's go home and play in the SNOW.	122 ms	102 ms	177 ms

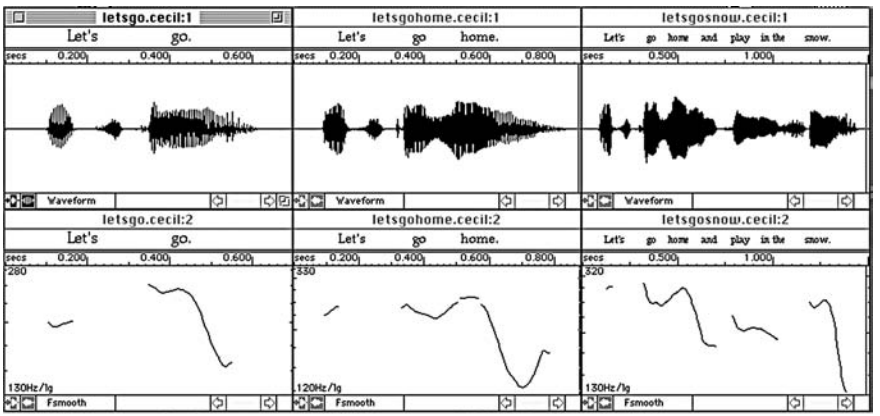


Figure 26. The vowel /o/ in open and closed syllables

As measured in these sample utterances, /o/ in the first sentence, *Let's go*, was 235 ms, while in the second and third sentences it was only 103 ms and 122 ms, respectively. The /o/ in the word *home* was 193 ms in the second sentence, where it occurred in utterance-final position, but only 102 ms in the third sentence, where it occurred in the middle of the utterance. In the utterance-final open syllable of the third sentence, the /o/ was 177 ms. These differences in duration are also attributable to information-structuring, that is, to the fact that a syllable representing the new or focal point of an utterance (or a discourse) will be stressed or accented and therefore longer in duration, a point elaborated in Chapter 7.

To provide a preliminary type of practice for discourse intonation, one might also begin to add attitudinal or pragmatic functions to these utterances; one could, for instance, imagine different situations and different speakers in different roles saying the above utterances and how the quality and length of the vowels would change accordingly. Consider the specific example of a parent trying to coax a reluctant child in a patient, non-threatening manner: the parent would say *Let's go* very differently from an angry, exasperated parent who is trying to get a slow child going with an abrupt and forceful *Let's go!* In the sample utterances shown below in Figure 27, the extended coaxing tone in the first utterance shows a long *go* of 865 ms as opposed to the short and abrupt *go* of 319 ms in the second utterance. To be sure, the pitch patterns of the word *go* (and the pitch patterns of the entire sentences) are also communicatively relevant, but for the purposes of the present discussion of stress and rhythm, duration, and intensity will be the main focus here.

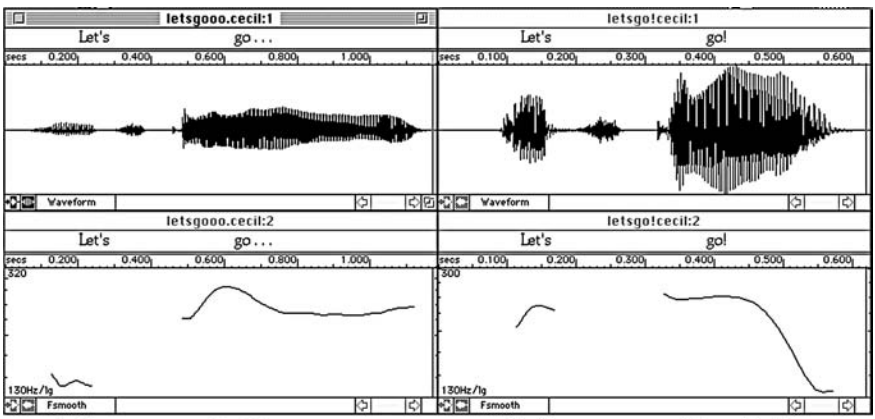


Figure 27. Two renditions of *Let's go.*

One of the first principles to observe when addressing the issue of sentence stress (accent) is that word stresses do not all retain the same strength in a phrase or a sentence. That is, overall sentence stress or accent will override individual word stress; word stress can be subordinated to and regulated by the syntactic structure of the phrase or sentence as well as by pragmatic considerations of contrast and emphasis.

In addition, although students have traditionally been taught about word stress, i.e., which syllable in a polysyllabic word to stress, they are often not instructed about which words in sentences tend to be stressed or where sentence stresses tend to occur in the second or foreign language.

In order to help L2 learners become aware of stress and rhythm in a new language, the instructor might start with songs or poems which have simple rhythmic patterns and have students listen for syllables and words that fall on the beat (cf. Wong 1987, pp. 24–25 regarding English; Tranel, 1987, p. 35, contrasting English and French). From there, one can progress to everyday expressions and have students identify where the heavy beats fall, e.g., in English they tend to fall towards the end of an utterance:

How's it GOing?
I'm doing GREAT.
Let's go out for DINner.

Another starting point for classroom work on sentence stress is to show how *content words* (nouns, main verbs, adverbs, adjectives, question words, demonstratives) tend to be accented whereas *function words* (pronouns, conjunctions, prepositions, articles, auxiliary verbs) are usually unstressed and do not normally receive sentence accent. (In particular contexts, particularly contrast, any word can be given sentence stress.) The most semantically salient word(s) in a sentence generally are content words. In introducing the concept of content words, Gilbert (1984) suggests asking students what kind of words they would put in a telegram. For example, students could be given a sentence and instructed to reduce the sentence to a specified number of words (shown in parentheses), such as (p. 26):

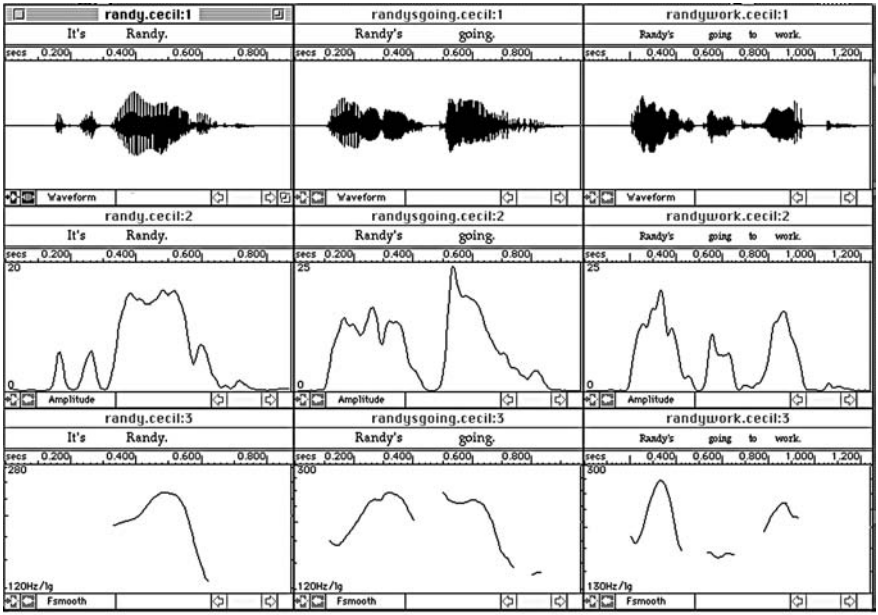
We need the report on Wednesday. (3) → *Need report Wednesday.*

Even in a "telegraphic" phrase such as the one above, where each word is stressed because of semantic content and thus informational importance, the heaviest stress would still tend to fall on the last word, in this case, because *Wednesday* might be the only bit of previously unknown information. Linguis-

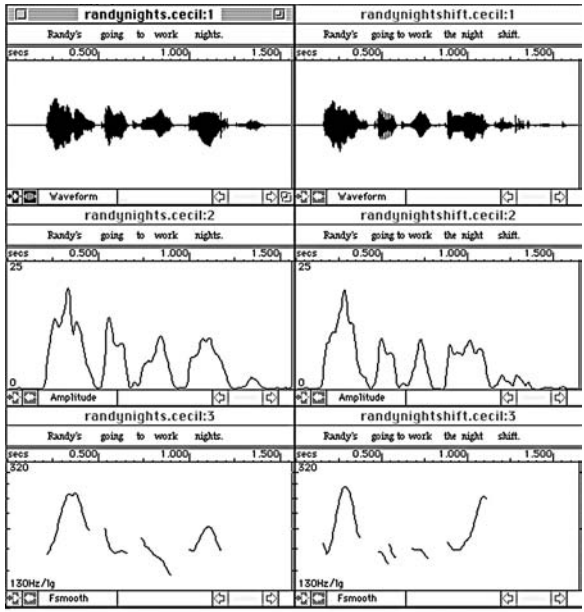
tic studies of sentence stress in English find that, in general, "the last accent in an utterance — the one that tends to gravitate toward final position — is usually regarded as the most important one. It is sometimes called the 'sentence accent,' 'sentence stress,' or 'nucleus'" (Bolinger, 1986, p. 49). Dauer (1993) illustrates in the following example how the so-called "neutral location of sentence stress" tends to occur toward the end of a sentence, typically on the last content word in the sentence (p. 222):

It's RANdy.
Randy's GOing.
Randy's going to WORK.
Randy's going to work NIGHTS.
Randy's going to work the NIGHT shift.

In Figure 28 below, the fundamental frequency (pitch) and the intensity curves of the utterances (b)–(e) generally exhibit high peaks at the beginning of the utterance and then another peak that is not quite as high as the first on the accented word at the end of the utterance. The reason for the high intensity and pitch peaks at the beginning of the sentences is that both *Randy* and the last content word toward the end of the utterance were stressed by the speaker.



(a) *It's RANdy.* (b) *Randy's GOing.* (c) *Randy's going to WORK.*



(d) *Randy's going to work NIGHTS.* (e) *Randy's going to work the NIGHT shift.*

Figure 28. Examples of "neutral location of sentence stress"

These sentences were read out of context, and both *Randy* and the last stressed word were considered to be new information (except in utterance (a) *It's RANdy*, where *Randy* was the only piece of new information). In an authentic conversation, if *Randy* were old or given information (e.g., if *Randy* were an established topic), then the main focus would probably have been on the piece of new information toward the end of the sentence (and the idea of *Randy* might have been expressed via a pronoun). What this underscores is that taking sentences out of context has its limitations, and learners must be made aware of how to signal stress or accent and in turn be enabled to base their decision as to which word to accent on the context and discourse situation.

On the question of accent placement, there is in fact an ongoing debate among linguists as to whether languages tend to have a "default" position. Bolinger feels strongly that the phenomenon of the sentence accent gravitating toward sentence-final position is simply a tendency and advocates "free choice" (cf. Gussenhoven, Bolinger, & Keijsper, 1987). Lambrecht (1994) gives numerous examples of how information structure relates pragmatic units such as *topic* and *focus* to sentence accent, suggesting that accent placement is therefore not

a matter of default. On the other hand, there are those who theorize about grammatical "predictability" and posit expected positions for accent placement (cf. Gussenhoven, 1987). This linguistic debate will not be elaborated upon here. The underlying principle is that in a given language there may be a tendency for sentence accent to occur in certain sentence positions, but that this "default" position may easily be overridden by pragmatic or discourse factors and that accent may thus occur in positions other than the expected position.[10] Although learners can be instructed as to the most common placement of sentence accent in the L2, they must also be made aware of other phenomena, such as pragmatic uses of emphasis or contrast that can override the more neutral (or "unmarked") placement of accent.

Words in connected speech: Reduced syllables

The phenomenon of reduction or loss of vowels (and consonants) in normal conversational speech has been covered thoroughly in many of the recent handbooks and courses on English pronunciation (cf. Avery & Ehrlich, 1992, pp. 84–89 on "Connected Speech"; Dauer, 1993, pp. 95–100 on "Pausing and Linking"; Gilbert, 1984, pp. 24–25; Morley, 1992, pp. 39–40 on "Elisions and Assimilations"; and Wong, 1987, pp. 49–53 on "Thought Units/Groups" and "Linking Sounds").[11]

As a way of introducing the concept of sound reduction in the L2 classroom, one could start with short utterances that are frequent in a given language, preferably from real discourse, or utterances that are likely to occur in natural conversation, e.g., greetings or formulaic utterances, including ones that are not complete sentences. One can point out to the students that such utterances tend to be made in informal situations, in which a less formal register is appropriate.

Thanks [a] bunch.	[ˈθæŋks ə ˈbʌntʃ]
See [you] later.	[ˈsi jə ˈlejtɚ]
Take [it] easy.	[ˈtejk ɪt ˈizi]
How's [it] going?	[ˈhawz ɪt ˈgoɪn]

As seen in the examples above, English has classes of words which typically bear secondary or weak stress rather than primary stress, i.e., the *function* words: articles, prepositions, pronouns, conjunctions, auxiliary verbs (*a, you, it,* and *is*, enclosed in square brackets above) as opposed to the *content* words: nouns, main verbs, adverbs, adjectives, question words, demonstratives. The vowels in

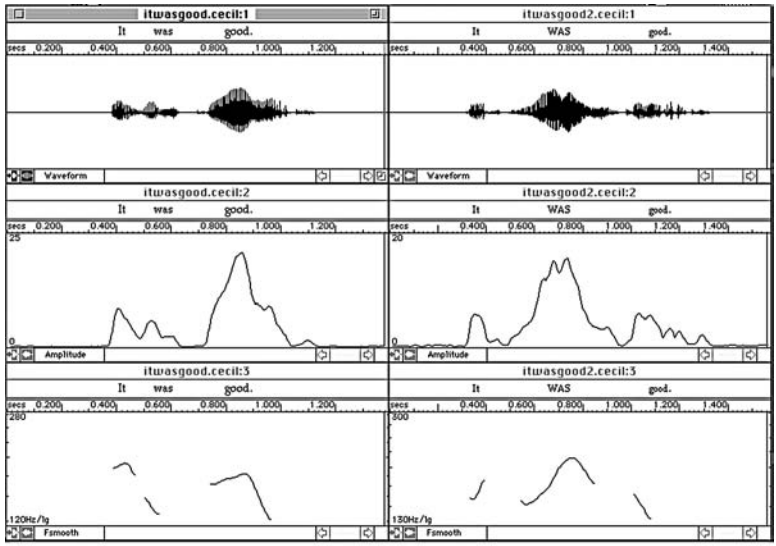
function words are often reduced to one of the centralized, lax vowels, the schwa [ə], and the high-central [ɪ] in English.

There are a variety of listening and speaking exercises for practicing stressing and unstressing function words with resultant differences in meaning (cf. Orion, 1988, pp. 37–44). In a "listen and repeat" exercise, students first hear a function word alone, spoken in its stressed form, and then unstressed in a sentence (here indicated by strikethrough).

an	*Have ~~an~~* [n̩] *orange.*
and, are	*Mary ~~and~~* [n̩] *June ~~are~~* [ɚ] *sisters.*

A follow-up exercise shows how meaning changes when function words are stressed, e.g., *Have AN orange* (not two or three). The examples below in Figures 29 and 30 also illustrate the discourse functions of stress. When spoken with reduced function words, the first sentence in each pair has normal or unmarked informational status, whereas the second sentence with stressed function words would normally be uttered to counter expectations or presuppositions.

It ~~was~~ good.	(simple statement of opinion)
It WAS good.	(speaker hadn't expected it to be good)



It was good. *It WAS good.*

Figure 29. Examples of discourse use of stress

I ~~can~~ DO it.	I am able to do it. (simple statement of fact)
I CAN do it.	Of course I can do it. (speaker contradicts assertion that s/he is not capable of doing it)

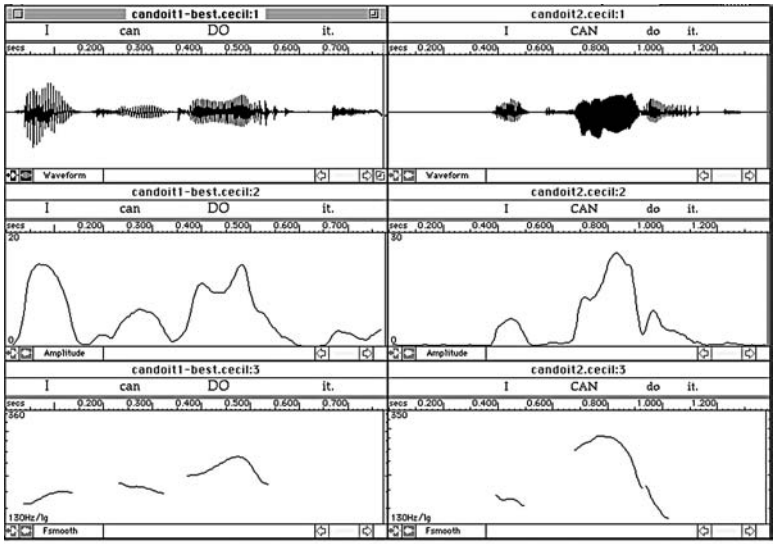


Figure 30. Additional examples of discourse use of stress

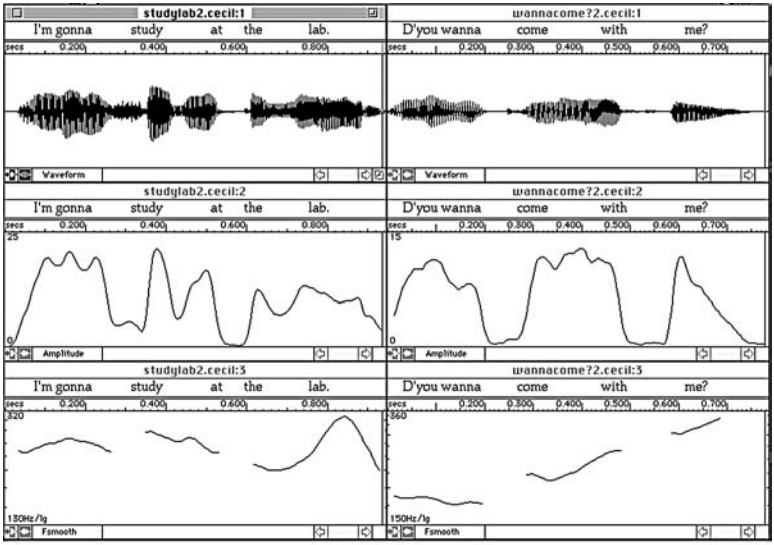
Loss or elision

In very fast speech and in informal registers, loss or elision of unstressed segments, particularly the schwa, often occurs, resulting in the usual contractions (cf. Dauer, 1993, p. 81 for a representative list of words with "disappearing syllables").

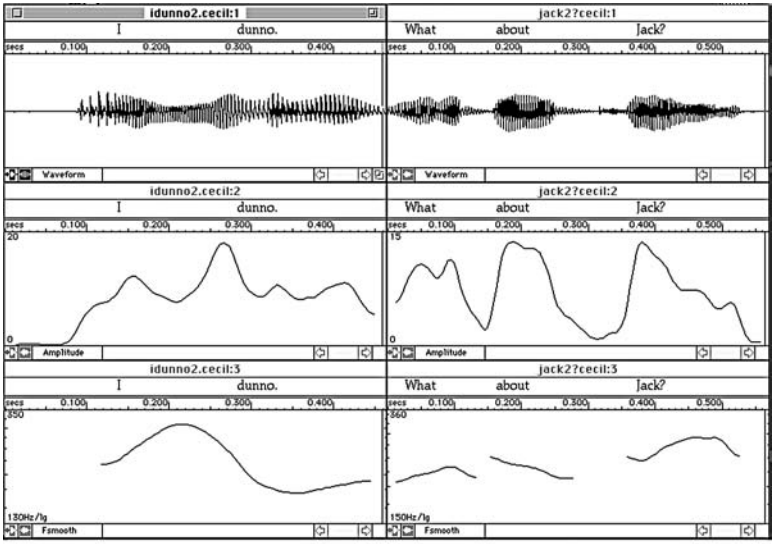
How're you?
How're ya doin'?
't's here.
I'm comin'.
't's time f'r lunch.

In addition, there are a number of other elided forms in which consonants are lost, e.g., in English the high-frequency phrases in informal speech (*wanna, gonna, dunno, doin'*). The dialogue below contains examples of some elided forms:

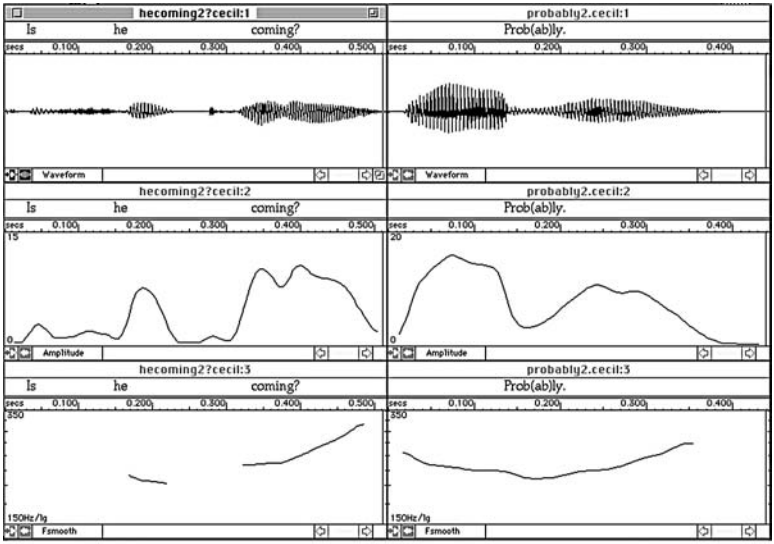
Speaker A: *I'm gonna study at the LAB. D'you wanna COME with me?*
Speaker B: *I dunNO. What about JACK? Is HE comin'?*
Speaker A: *PROB'ly. Y'know 'e has t' finish HIS homework too.*



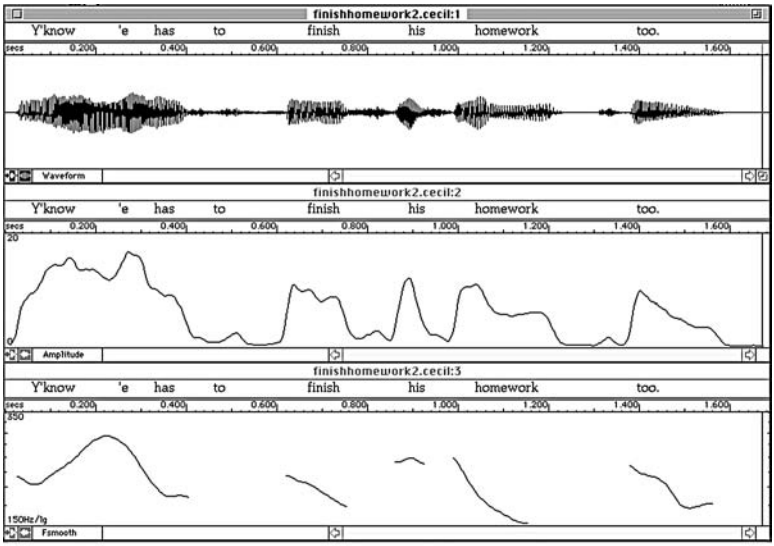
I'm gonna study at the LAB. D'you wanna COME with me?



I dunNO. What about JACK?



Speaker B: *Is HE coming?* Speaker A: *PROB'ly.*



Y'know 'e has t' finish HIS homework too.

Figure 31. Conversation with elided forms

One of the ways that these elided forms are readily apparent is that the fundamental frequency curves shown in Figure 31 above are continuous for the words *gonna, wanna*, and *dunno*. If the words were enunciated carefully as *going to*, *want to*, and *don't know*, they would contain the voiceless consonant [t] in *to* or *don't,* and this would appear as a break in the pitch curve. As the figure below shows, however, the portions of the pitch curves for *gonna, wanna*, and *dunno* are without breaks.

Displays such as the ones above can be incorporated into intonation practice and could be very useful in helping to provide learners with visualizations of their pitch patterns as well as to reflect the phonetic effects of various degrees of reduction and elision. Various scenarios could be presented to learners, e.g., ranging from formal to informal, and learners could then compare how such exchanges might be spoken depending on the speed and formality of speech appropriate to the situation.

One of the main tenets of this book is that exercises that go beyond the phrase and sentence level are always to be encouraged and implemented. This principle builds on the work of others who have stressed the importance of working with paragraph-length, passage-length, and discourse-length texts and conversations.

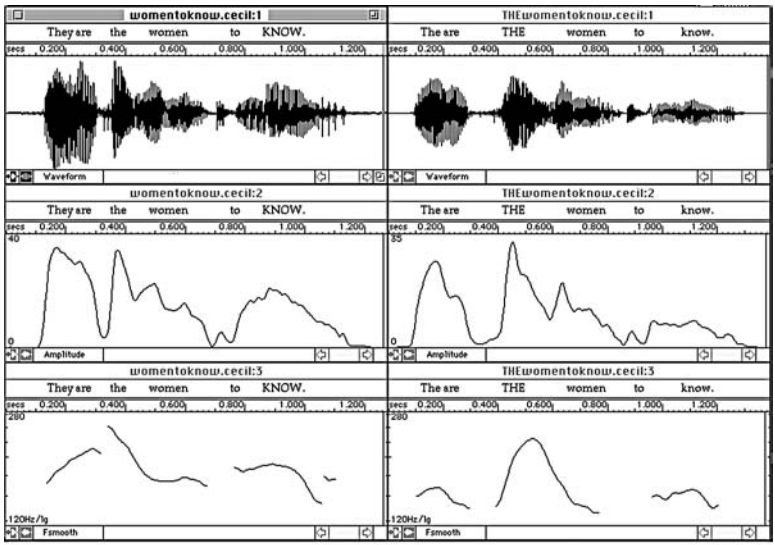
Stress and rhythm at the discourse level

Pragmatic stress or accent

There are exceptions to the very general rule of thumb that, in English, major sentence stress will fall on the stressed syllable of the final content word of a sentence. Most exceptions are due to *pragmatic* reasons, such as considerations of information structure and information focus, contrastive, or emphatic stress, the need to express a particular attitude or illocutionary purpose, and presuppositions on the part of the speaker about what is new vs. given information for the hearer (see Chapter 3 on the discourse functions of intonation). Pragmatic stress differs from grammatical stress or neutral (unmarked) sentence stress not only in its function, but also in its realization. Its function is basically to call the attention of the listener to a particular point of information with the purpose of contrasting it to another point of information or giving it added force or nuance of meaning. Where pragmatic stress occurs thus stands in direct relation to the semantics and pragmatics, rather than the syntax, of

speech. Its realization is also different. First, it is not bound to the usual placement principles. Second, its occurrence is generally more readily perceptible because it is usually given more articulatory prominence as seen in steeper pitch curves, longer duration, and sometimes greater intensity or loudness.

Some instances of the discourse functions of stress, i.e., as when syllables or words that would not generally be made prominent are given sentence stress or accent, have already been given above. For example, as seen in the preceding section, the function words *was* and *can* in the sentences *It WAS good* and *I CAN do it* were stressed to indicate contrast to or contradiction of what had been expected or assumed. Similarly, in the example in Figure 32 below, the function word *the* in *They are the women to KNOW* is unstressed, but stressed in *They are THE women to know*. The stress on *THE* not only changes the vowel from a schwa to a long, stressed [i:] but also indicates strong emphasis and an expansion in meaning of the indefinite article *the* to connote "the best" or "the (most) important." The duration of *the* in the first sentence of the pair is 113 ms, while the duration of *THE* in the second sentence is 264 ms, or more than twice that of the neutral occurrence. In addition to the high intensity of the opening word in both cases, there is added intensity on the stressed article in the second case, with very low intensity on the unstressed word *know* at the end of the second utterance.



They are the women to KNOW. *They are THE women to know.*

Figure 32. Example of neutral (unmarked) vs. emphatic stress

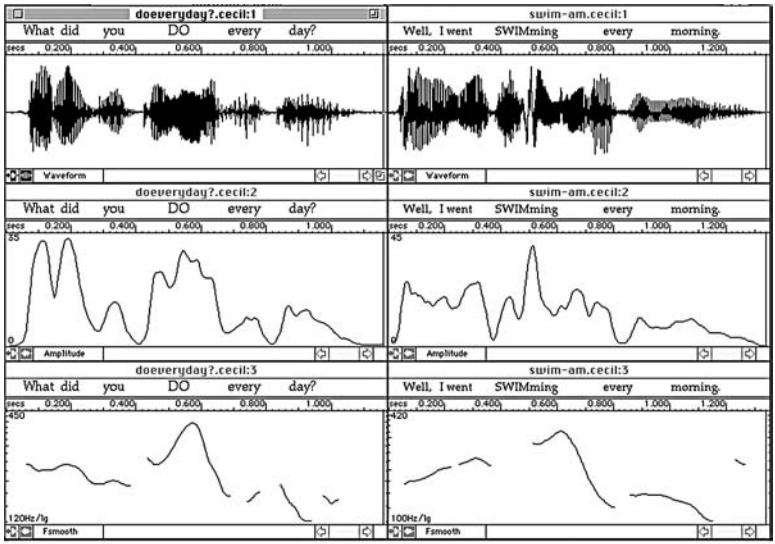
It might also be pointed out that function words can also receive sentence stress if they are contradictory, contrastive, or otherwise emphatic, as in the examples:

Are they FOR us or aGAINST us?
Did you do that TO her or FOR her?

When a speaker wishes to direct the listener's attention to a non-final content word in a statement, this word will receive major sentence stress, as shown in the exchange below where Speaker A and Speaker B are talking about Speaker B's vacation:

Speaker A: *What did you DO every day?*
Speaker B: *Well, I went SWIMming every morning.*

The utterance by Speaker B has sentence stress or accent on *SWIM-* because this is the element of the sentence that Speaker B is directing Speaker A's attention to. This element may be called the *information focus* of the sentence and is realized with added intensity and a rise in pitch, as shown in Figure 33. While the information focus most often occurs at the end of a sentence in English, it does not have to; in that case, the preceding discourse and knowl-



What did you DO every day? *Well, I went SWIMming every morning.*

Figure 33. Examples of accent before utterance-end

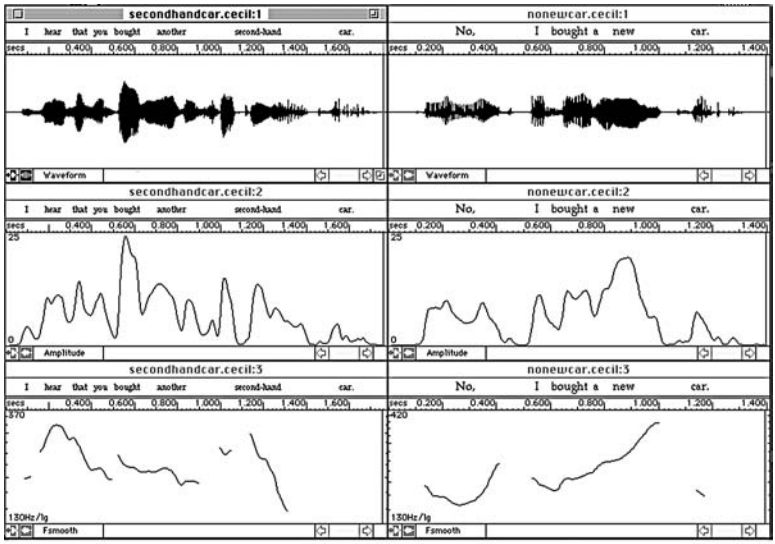
edge shared by speaker and listener will make it possible for the stress to occur elsewhere. The critical point here is that new information or that which is deemed most important will be accented and must not necessarily come at the end of the sentence.

Other examples of *discourse* stress would be cases in which stress is used *contrastively* or *emphatically*, as suggested by Avery and Ehrlich (1992) and exemplified as follows (p. 76):

> Speaker A: *I hear that you bought another SECond-hand car.*
> Speaker B: *No, I bought a NEW car.*

In Figure 34 below, the intensity of *new* is much greater than that of any other syllable. Note also that pitch reaches a peak on the word *new* and then drops steeply, though there is a break in the pitch curve due to very weak intensity after the word *new*.

In the exchange shown in Figure 34, the contrast takes the form of a contradiction. One might have expected the major sentence stress in Speaker B's responses to fall on *car* as the final content word. However, the idea of a car (and that of a car having been bought) were clearly already established and were shared by both A and B; it is the contrasted information in Speaker B's



I hear that you bought another SECond-hand car. No, I bought a NEW car.

Figure 34. Example of contrastive stress

responses, the fact that the car is new, that receives major sentence stress. "This contrastive stress can be even heavier and louder than the normal major sentence stress, particularly in sentences where a contradiction is being made" (cf. Avery & Ehrlich, 1992, p. 76). Note again, however, that speakers would normally replace the content word *car*, as being old information, with a pronoun such as *one*, with the result that *new* would be the last content word in the sentence after all.

Speaker B: *No, I bought a NEW one.*

Rhythm in conversation

For perceptual exercises to practice listening for discourse stress and rhythm, one could use recorded conversations that as realistically as possible represent language in use and illustrate the rhythmic features. Students could be asked to listen closely to the conversation, do a perceptual analysis of it, and identify the heavy beats in each sentence, particularly those that do not occur at the end of an utterance. Repeated listening and intense analysis will be necessary, and in small groups or as a class, the students can then discuss reasons for placement of stress and accent by examining the surrounding discourse.

For speaking practice, discourse intonation can be incorporated into communicative activities that involve such functions as negotiating, repeating, quoting, asking for clarification, and self-correcting or "self-repairing." Since most second and foreign language textbooks today already focus closely on using language to achieve certain functions and on providing students with opportunities to practice speaking in authentic contexts, adding prosodic practice can enhance and reinforce the learner's overall communicative competence.

In summary, the important principles to consider when teaching stress and rhythm are (1) materials must go beyond the word level to the phrase, sentence, and discourse levels, e.g., by embedding problematic or difficult words into larger contexts in both speaking and listening exercises, and (2) in parallel with the focus on communicative competence and using language for the negotiation of meaning, learners can first be made aware of how the language they are learning makes words more prominent, secondly how listening for such prominence will help them discern the information focus in a sentence or larger discourse, and thirdly how in their own speech they can (and should) try to produce phrases with stress on the content words to be accented (i.e., words which convey new, contrastive, or emphatic information). As a

final thought on rhythm, it might be noted that although sentence rhythm has been investigated extensively, particularly in the subfield of metrical phonology, rhythm in natural conversation and interaction has received less attention — but it is currently being studied and is a very rich topic for research (cf. Uhmann, 1996). Couper-Kuhlen (1996, 1997), one of the foremost researchers of prosody in conversation, has characterized rhythm in conversation as one of the three prosodic areas in need of further investigation, particularly with regard to whether there is a so-called "beat" in conversations and to the nature of metrical cues for when participants may "take the floor," "come in," or start a turn.

Notes

1. Levelt (1989) found that English speakers tend to store lexical items according to stress patterns, so that if a wrong pattern is perceived, listeners' comprehension is hindered because they may spend time searching for stored words in the wrong category (p. 373).

2. In addition, syllabic consonants, particularly liquids (/l/) and nasals (/m, n, ŋ/), may also be carriers of stress, pitch, and lengthening, but generally as such much less frequently than vowels (cf. Bell, 1978, p. 155).

3. Crothers (1978) states in his article on the universals of vowel systems that "Nearly half (45%) of the sample languages have contrasting long and short vowels. In most cases (70%) the vowels of the two systems are equal in number and arrangement, either identical in quality or showing minor differences. In another 19% the long vowel system is larger than the short vowel system, while 8% have more short than long vowels. By far the most commonly reported difference of quality between long and short vowels of corresponding positions is centralization (laxing) of the short high vowels, i.e., short /ɪ ʊ/ versus long /iː uː/, reported in 20% (19 languages) of the languages with long vowels" (p. 123).

4. The number and types of vowels differ among languages, of course. In German, for example, there are three main types of vowels: long and tense, short and tense, and short and lax. In Spanish, all vowels are generally tense and relatively similar in duration, i.e., they are relatively short, particularly in comparison with long vowels in other languages (cf. Barrutia & Terrell, 1982, p. 15; Quilis & Fernández, 1992). In French, much as in Spanish, all vowels are relatively equal in duration, and the basic principle is to pronounce each syllable with approximately equal strength and thus to maintain for each vowel its full, unreduced quality (cf. Tranel, 1987, p. 35). Moreover, Spanish and French (like other Romance languages) are generally characterized as syllable-timed.

5. Software for acoustic analyses is becoming increasingly accessible to the general public, and computer programs for pedagogically improving pronunciation and intonation are being developed at numerous institutions and by some software companies. The Learning Company, for example has a CD-ROM series "Learn to Speak X" for English, French,

German, and Spanish. Similarly, Transparent Language has a CD-ROM series "Language X Now!" for Arabic, Chinese, Dutch, English, French, German, Hebrew, Irish, Italian, Japanese, Latin, Polish, Portuguese, Russian, Spanish, and Swedish. Both the Learning Company and Transparent Language software make it possible for learners to record their own speech, play it back and compare it to the pronunciation of a native speaker. Transparent Language advertises "advanced pronunciation analysis" and the Learning Company advertises "advanced speech recognition," but neither provides the type of acoustic feedback that is useful to a beginning learner. There is thus a need for software that helps learners interpret the acoustic realizations of their utterances by explaining how they can change their pronunciation.

6. As pointed out by Wong (1987), "closed syllables that end in continuants (e.g., /m, n, f, v, w, r, s, z/) are longer than those that end in stops (e.g., /p, t, k, b, d, g/). For example, compare *glows* and *globe*" (p. 24). For current purposes, however, a three-way distinction in syllable structure and length will suffice.

7. For detailed accounts of the rules for stress placement in American English, cf., for example, Dickerson, 1989; for German, cf. Benware, 1986; Bleiching, 1992; Moulton, 1962; and Rausch & Rausch, 2000; for Spanish, cf. Dalbor, 1997; Cruttenden, 1997; for French, cf. Tranel, 1987. (In words of more than two syllables, weaker stresses may also be found "to the left" of the final stress, "usually distributed in every other syllable" (p. 195).)

8. However, Uhmann (1996) points out that "Although both English and German are said to be stress-timed, the unmarked degree of rhythmically integrated speech, i.e., *isochrony*, seems to be higher in British and American colloquial English than in German. This might be due to certain structural-phonological characteristics of German, namely the large number of secondary accents" (p. 326) — or, looking at it from the opposite direction, the relative high number of reduced syllables in English.

9. However, a number of studies provide counterevidence for the phenomenon of isochrony. Dauer (1983) concluded from her data analysis of English, Thai, Spanish, Italian, and Greek that intervals between stressed syllables were no more regular in English than in Spanish, a so-called syllable-timed language. However, the distinction is widely accepted and will be followed in the discussion that follows here. "Staccato" timing as in syllable-timed languages, when used in English, makes it seem that all words are being given equally high importance. Therefore, if known information in a stress-timed language is given stress, it might make the speaker sound arrogant (cf. Hatch, 1992, p. 279) or even accusatory (cf. also Borzone de Manrique & Signorini, 1983).

10. For French, Tranel (1987) proposes as a rule of thumb regarding sentence stress that "within each syntactically (and semantically) delimited portion of a sentence, it is normally the word furthest to the right which stands out, because of the presence of a primary grammatical stress which is not reduced (if the word occurs at the end of the sentence) or at any rate less reduced than those which precede (if the word is found at the end of a group that is not sentence-final)" (p. 199).

11. For German, cf. Benware, 1986, for a table which shows schematically how the stresses in compound words are altered depending on whether the word is spoken in isolation (dictionary entry), before a pause, or in slowed tempo (p. 109).

Chapter 7

Teaching discourse intonation

Almost two decades ago, de Bot and Mailfert (1982) lamented that "To our knowledge, a clear way to present a progression of intonation patterns to adult learners has not yet been worked out. Hanging patterns onto grammar/sentence function is not always valid, as Bolinger tells us; hanging intonation patterns onto attitudes and emotions gives only vague and unhelpful rules of use. Hanging intonation onto discourse seems to be promising, but has not yet been fully developed into teaching materials. It seems evident that more work must be done in this area" (p. 75). Fortunately, as summarized in the preceding chapters, there has been increasing interest in this area ever since, and this chapter will suggest ways in which learners can first be made aware of the discourse functions of intonation and can then practice recognizing and producing intonational patterns in a second or foreign language at both the sentence and discourse levels.

There are a growing number of textbooks and handbooks for teaching pronunciation which include sections or chapters on intonation, particularly for ESL (and EFL). Most of them acknowledge the importance of going beyond the word and phrase level to the sentence level and have begun to take the additional step of discussing the use and function of intonation at the discourse level. In teaching discourse intonation, the basic components of *rhythm* and *pitch* in the target language must first be presented (see preceding chapter on rhythm).[1] The plethora of existing theoretical and pedagogical treatments of intonation offer a wide range of both features to be described and notational systems for representing them. In this book, I have chosen a combination of systems in order to represent intonation as clearly as possible for the language teacher and learner. However, it is ultimately a matter of individual cognitive style and preference as to which system the user will find simplest to learn and comprehend.

Intonation notation

Most of the typical types of intonation notation can be associated with the major theory of intonation upon which they are based (as discussed in Chapter 2):

– the European tune and tonetic approaches which use diacritics to represent pitch movement, e.g., Halliday's (1967a) system exemplified in the two sentences below, where ` represents falling intonation, ´ rising intonation and ^ rising-falling intonation:

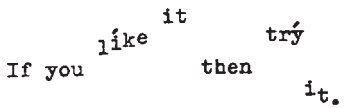
  it's `raining / ´isn't it?
  you ^know / ´do you?

– the American phonemic approach which designates pitch levels (or phonemes) by numbers, typically from 1–4, where 1 = highest pitch level and 4 = lowest pitch level (e.g., Pike, 1945), or by H(igh) and L(ow) in the case of the generative approach, along with other symbols for marking pitch accents (*) and boundary tones (%) (e.g., Pierrehumbert, 1980):

  I 'wanted to go but 'couldn't.
  3- °2- -4 4- °2- -4 //

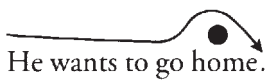
  – What about Anna? Who did she come with?
  – Anna % came with Manny.
    | |
    H* L- H% H* L- L%

– Bolinger's (1986) method of simulating a pitch curve by typing the words in the shape of the intonation curve:



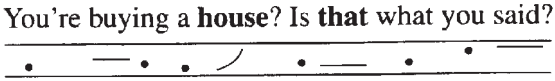
In pedagogical handbooks, the following practices are common:

– use of line drawings superimposing pitch/intonation curves over text, e.g., Avery and Ehrlich (1992, p. 222):



– use of stylized intonation "curves" drawn with straight lines above and below the words of a sentence, e.g., the German textbook by Neuner et al. (1979, p. 13):

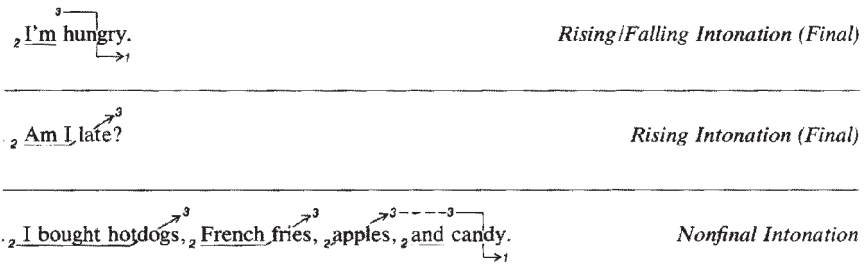
– use of dots for unstressed syllables and dashes or individual pitch curves for stressed syllables, e.g., O'Connor and Arnold (1961, Chapter 2), Dauer (1993, p. 222), Fox (1984), Lohnes and Strothmann (1980, p. 18).



– use of diacritics (including capitalization) in conjunction with raised (or lowered) typed lines to mark *key* in discourse intonation, e.g., Brazil et al. (1980, p. 28):

high WIFE
mid // *p* this is eLIZabeth // *p* peter's //
low

Various combinations of notations can also be found. For example, Morley (1991) combines Pike's method of numbering pitch levels from 1–3 and draws pitch lines above and below the text, including arrows to indicate pitch direction of stressed syllables (p. 34):



What needs to be taught: Components of intonation

The components of intonation that I feel are most crucial for language learners to be able to identify and practice are:

- *sentence stress (*or *accent)*: syllables or words that are most prominent because they represent the information focus or point of contrast or emphasis in a sentence; this stress is realized by pitch, duration, and intensity
- *terminal contour*: direction of pitch change, particularly at sentence end or at so-called "transition points"
- *key*: range of pitch used at points of transition (at both the beginning and end of an utterance) relative to preceding and succeeding utterances or parts of utterances

In this chapter, I will use a combination of notations, namely (1) graphs of actual pitch contours of utterances made possible by the accessibility of signal analysis software that can display fundamental frequency (pitch) contours and are perhaps preferable to the earlier more "stylized" representations, (2) the symbols used in the transcription system of DuBois et al. (1992) and (3) additional marking for *key*. As discussed in Chapter 2, DuBois et al. (1992) adopt the term *tones* for the various distinctive intonational shapes that start in the syllable with primary accent and can be spread across several words, often extending from the last primary accent to the end of the intonation unit. They classify the movements of pitch as *rise*, *fall*, *rise-fall*, *fall-rise,* and *level* tones, transcribed respectively by the followed symbols: /, \, /\, \/, _ . Due to the emphasis in this book on discourse intonation, one of whose earliest proponents was Brazil (1975), I adopt Brazil's concept of *key* and use the symbols *H, M, L* to designate the *range* of pitch (high, mid, or low). Capitalization of syllables or words that are the most heavily stressed (accented) will be used as an additional visual aid.

General principles for teaching listening and speaking

As discussed in Chapter 6, a typical progression for teaching intonation would include the following (cf. also Bradford, 1988; Pennington, 2000; Spaai & Hermes, 1993):

- sensitization (listening exercises)
- explanation (comparison with native language, if possible)
- imitation (*controlled* speaking exercises)
- practice activities (*rehearsed* speaking activities)
- communicative activities (*extemporaneous* speaking activities)

Listening or auditory discrimination exercises can precede or be used in con-

junction with speaking exercises. Since most textbooks are accompanied by tapes of dialogues read by native speakers, these tapes could initially be used. For intonation, the list of what students could be asked to listen for and do while listening to speech samples can include the following:

– Listen holistically to try to get an idea of the overall shape and character of the intonation before tracking its subparts.
– Mark points at which the speaker pauses with two periods (..) for a short pause and three periods (…) for longer pauses — to become aware of how native speakers (1) keep thought groups together, (2) convey information in "chunks" in the L2, (3) begin a new topic, and (4) add a parenthetical remark.
– Mark the words that are perceived as being prominent or accented in sentences and entire paragraphs.
– Mark pieces of information as "mentioned already" or "not known yet and thus new."
– Mark the movement of pitch at the sentence accent with symbols signifying rising, falling, rising-falling, falling-rising, and level tones: /, \, /\, \/, _.
– Mark each sentence at both the beginning and the end for high, mid, or low key (H, M, L).

A related listening exercise would be to have students match pairs of questions with replies (and vice versa). For example, one could ask students which of two questions is more likely to have been posed to elicit a particular reply. Alternatively, one could have students listen to the first utterance of a question-answer sequence, or the first few utterances of a conversation, and ask them to predict what the intonation contour for the next sentence(s) might be before listening to the actual dialogue on tape.

Another perception exercise would entail students guessing the mood, attitude, or intent of a speaker. The same sentence could be uttered with different emotions and embedded in larger contexts; students would first hear the utterance in isolation and make a guess before listening to the rest of the dialogue or conversation, and then stay with or change their original answer. Finding themselves changing their initial reaction would help them realize the importance of the surrounding discourse to accurately perceiving speaker affect.

As a follow-up to using the typically straightforward and formal renditions of dialogues read on instructional tapes, the instructor can suggest ways of "dressing up" the dialogue and discuss alternative scenarios with the class for production practice. For example, the characters in a dialogue can be ascribed

different personalities and moods; their relationships to each other can be made more explicit; their motivations and intentions can be determined for subsequent role-playing.

In accordance with recent linguistic research trends of basing theories of intonation on naturally-occurring speech data, language instructors should also incorporate authentic conversational data into their instructional materials. Examples should be taken from natural conversations and real interactions by native speakers of the target language, and learners could be asked to do a close analysis of the utterances, listening to conversations repeatedly (and possibly also observing them on videotape) and paying particular attention to the intonation patterns and the pitch ranges used.

As a general guideline for **speaking** exercises referred to above, I reiterate my emphasis on focusing on language in context, e.g., on using dialogues (answers *and* questions), conversations, interviews, and paragraph-length discourses (reporting, storytelling) rather than just isolated words or sentences (cf. Bradford, 1988 for examples of the types of dialogues and conversations that may be used as well as fully contextualized *communication activities*).

Another general principle is to *enrich the number and type of classroom discourse situations*. Students must be given opportunities to make choices, i.e., they must be provided with opportunities to encounter meaningful contrasts and to recognize and produce language in a wide range of communicative situations. This is based on Bolinger's tenet that one must look more at the influence or significance of the *function* of an utterance than at grammatical *form* in choosing appropriate intonation. That is, speakers in general do not assign intonation to an utterance on the basis of grammatical or syntactic structure or attitude alone. What is needed is a discourse context: the meaning of intonation is inseparable from context. Information structuring is derived from the speaker's step-by-step real-time decisions about the status of the discourse, e.g., which word(s) to stress or accent and which intonation contour to employ. This tenet is echoed in recent theoretical research on discourse intonation, in which Couper-Kuhlen and Selting (1996a) propose an *interactional* meaning and function of intonation. They ask the questions "How does [...] *interactional* meaning (as opposed to the *semantic* or *pragmatic* meaning of words and utterances) come about? What are the cues that help to make social interaction more than the mere exchange of words, namely a real-time encounter between conversationalists who establish and negotiate units of talk as situated meaningful activity?" (p. 1).

For example, as mentioned in Chapter 5, in traditional classrooms, the teacher is dominant, both socially and "politically," thus controlling and managing the classroom discourse. Teachers ask most of the questions, and students hear lopsided models of intonation, i.e., mainly question intonation. Students very rarely manage discourse, asking questions only when they need clarification. In addition, if questions and answers between instructor and students are the only types of discourse heard, students will not realize the totality of meaning that can be signaled by intonation. They thus need exposure to a variety of forms and contexts of speech. Intonation must therefore be taught and practiced with sufficiently rich verbal and nonverbal contexts, i.e., not only in terms of vocabulary and topics, but crucially also in terms of the *roles* and *status* of speakers in relation to each other. A variety of interaction types is needed in which the status of speakers varies along the dominant-equal continuum. A follow-up exercise to the close-listening activity suggested above would be for learners to use the real interactions that they analyzed closely as a model from which they would then construct their own interactions in the same style. Examples of specific exercises for practicing this are included below.

Exercises for practicing intonation

The main goal of this chapter is to suggest exercises for practicing intonation in different discourse types or genres and with different functions. As discussed in Chapter 3, I distinguish four main functions of intonation; in this chapter, I will present examples of how these functions are often realized intonationally and how they might be taught to L2 learners. The four basic functional categories — **grammatical, attitudinal, discourse,** and **sociolinguistic** — correlate with the main elements of communicative competence (discussed in Chapter 5), which the language teaching profession is striving to promote.

Beginning with the traditional *grammatical* functions of intonation, examples are presented in the following section of:

- one-word sentences and phrases
  - meaningful intonation within a word
- simple sentences
  - neutral declarative, interrogative, exclamatory sentences
- intonation in compound and complex sentences
  - keeping thought groups together

Secondly, for the *attitudinal* functions of intonation, examples are presented of:

- one-word sentences and phrases
  - meaningful intonation within a word
- simple sentences
  - sentences expressing emotion, intent, attitude (e.g., exasperated statement, incredulous question, impatient imperative)

The third category concerns the *discourse* functions of intonation, for which examples are presented of:

- conversations and longer discourses to express:
  - finality vs. non-finality
  - marking shared knowledge, presuppositions
  - turn-taking, interrupting
  - discourse management, e.g., constraining/discouraging reply, indicating expectations, facilitating cooperation
  - repair and negotiating meaning after misunderstanding
  - conversational support strategies
- different discourse modes
  - formal speech: reading aloud, news reporting, interview
  - informal speech: conversation, discussion

Fourthly, with regard to the *sociolinguistic* functions of intonation, examples given highlight:

- indicating role and status of speakers
- characterizing formal vs. informal speech
- signaling politeness, deference, etc.
- conveying gender differences, age differences
- conveying regional differences

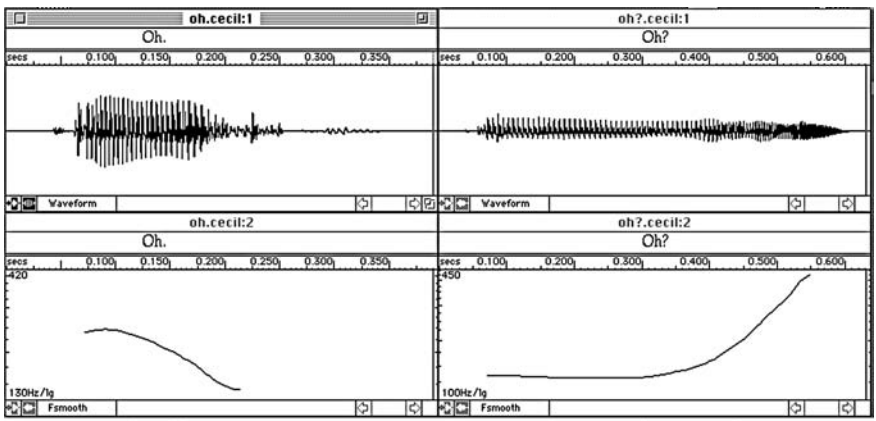
Grammatical functions of intonation

Words and phrases

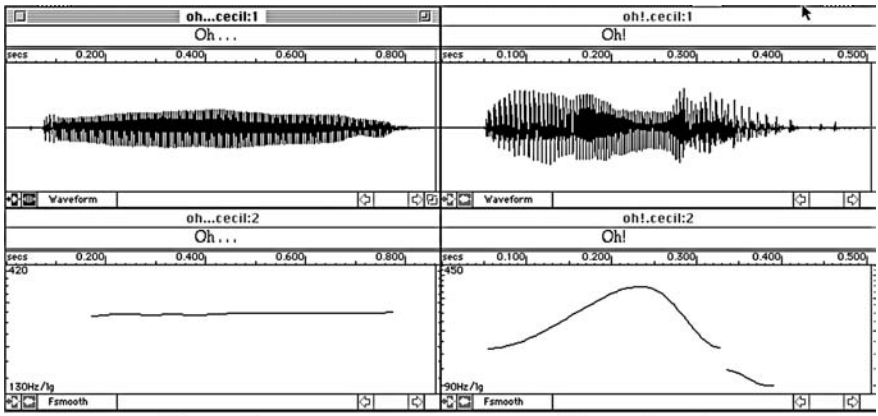
It is common to begin intonation practice (i.e., practice of pitch patterns) with single words and phrases before progressing to longer sentences. This type of practice does not need to be excluded even if the ultimate goal is working on

discourse intonation, because in conversation, many utterances (e.g., answers to questions, echo-questions or requests for confirmation, back-channel responses) do in fact consist of a single word or phrase. In traditional exercises, students are often asked to practice a list of single words almost random in nature; here, the individual words would eventually be practiced in several different contexts.

Figures 1(a)-1(d) below show four different examples of the word *oh* to illustrate some typical intonation patterns in English. (The waveforms and fundamental frequency or pitch curves in this chapter were all produced with the *MacCECIL*© program.) Figure (a) is a simple fall (\), typical of statements; (b) is a simple rise (/) typical of yes-no questions; (c) is a level contour (_), typical of an unfinished or uncertain utterance; (d) is an exclamatory rise-fall (/\). The exclamatory *Oh!* in (d) is spoken in high key [H], with a peak fundamental frequency of 346 Hz, as opposed to the more "neutral" *Oh* in (a) with a peak pitch of 246 Hz. As discussed in Chapter 2, key refers to the range of pitch used by a speaker. Speakers generally use mid key for unmarked or neutral intonation units and select high or low key to express "non-neutral" meaning. Use of high or low key involves raising or lowering the pitch of the whole intonation group relative to a pitch that can be established as the norm for the speaker concerned. For the speaker of the utterances represented below, mid key would be in the range 130–246 Hz for (a), whereas high key would be in the range 130–346 Hz as in (b).



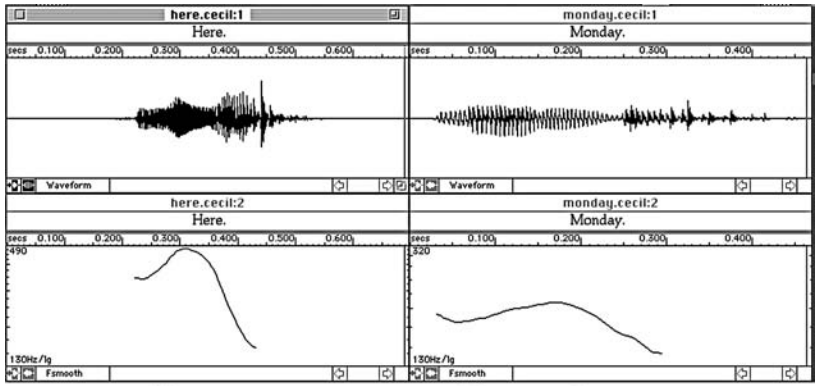
(a) *O h.* (b) *O h?/* [*H*]



(c) *O h… _* (d) *O h!* /\ [H]

Figure 1. Four renditions of *Oh.*

Figure 2 below shows one-word answers to questions such as *Where are we meeting?* and *When are we having our test?*, where the answers might be (a) *Here* and (b) *Monday*. Although both words are spoken with rising-falling intonation, the slope of the rise and the fall in (a) is markedly steeper than in (b). In addition, the pitch peak in (a) is nearly 500 Hz and the utterance is transcribed as having been spoken in high [H] key (with a pitch range from 130–490 Hz), whereas the pitch peak in (b) is 230 Hz (with a range from 130–230 Hz or mid key). The reason for these contrasts is that the reply in (a) was emphatic, as if to contradict the expectations of the interlocutor, whereas the reply in (b) was neutral and matter-of-fact.



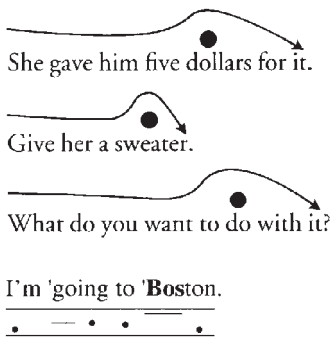
(a) *Here.* /\ [*H*] (b) *MONday.*/\

Figure 2. Answers to questions

Sentence level

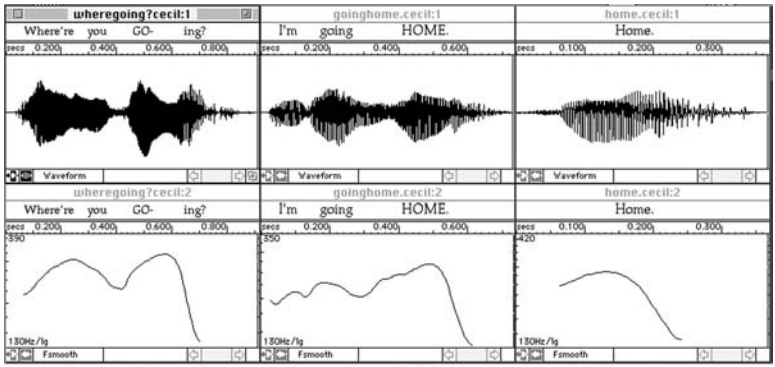
As discussed in Chapter 6, sentence stress or accent in English is traditionally described as occurring towards or at the end of a sentence and typically assigned to the final noun. Traditional accounts of sentence intonation describe what are thought of as typical or common intonational patterns; two types of *final intonation* are listed by Avery and Ehrlich (1992) and Dauer (1993). The first of these,

– **rising-falling** intonation (Avery & Ehrlich, p. 77), termed **fall** (high to low) by Dauer (p. 224) and **falling** by Brazil (1975, p. 6) is a very frequent pattern, characteristic of simple declarative sentences, commands, and questions that begin with a *wh*-word:



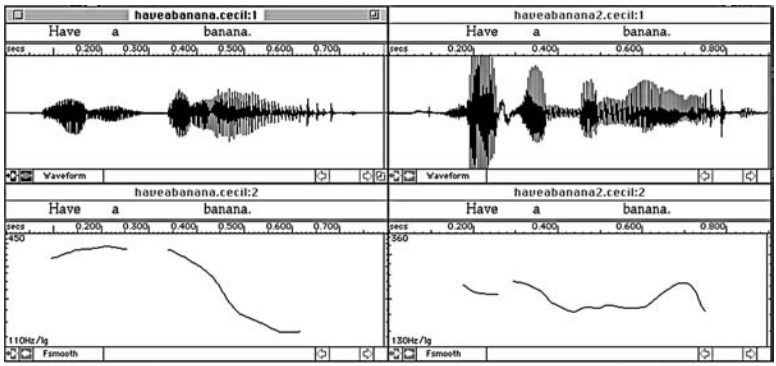
With their waveforms and pitch curves, Figures 1 and 2 show that both one-word statements and answers to questions can exhibit this *rising-falling* (1d, 2a) and *falling* (1a) intonation (see the displays for *Here, Monday* and *oh* above). In Figure 3 below are examples of *wh*-questions and corresponding answers with their intonation contours. The actual fundamental frequency curves show close similarity to the stylized notation used by Avery and Ehrlich (1992). The symbols adopted from the DuBois system (/\ and \) refer only to the final pitch contour.

The examples in Figure 4 below show two imperatives with slightly different stresses (or accents), but both with falling intonation. In (a), the stress is on the first word, as in [*Here,*] *HAVE a banana*, whereas in (b), the stress is on the word *banana* with the second syllable of *baNANa* receiving the sentence accent.[2]



Where're you GOing?/ *I'm going HOME./* *HOME.*

Figure 3. *Wh*-question and responses

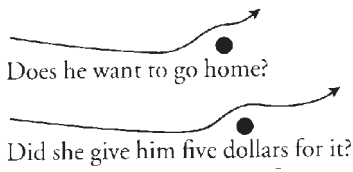


(a) *HAVE a banana.* (b) *Have a baNANa.*

Figure 4. Imperatives

The second type of final intonation in Avery and Ehrlich (1992) and Dauer (1993) is

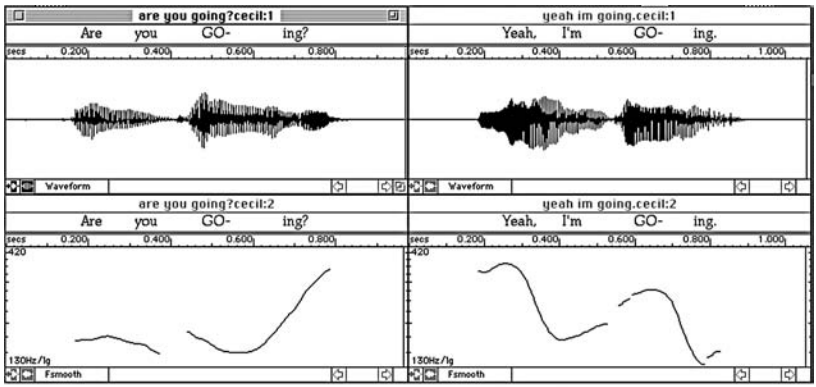
– **rising** intonation (Avery & Ehrlich, p. 78), also termed **rise** (low to high) by Dauer (p. 224) and Brazil (1975, p. 8), is characteristic of questions that require a yes-or-no answer.[3]



Are you 'going to 'Boston?

Figures 5 and 6 below show examples of yes-no questions (and replies) with their intonation contours.

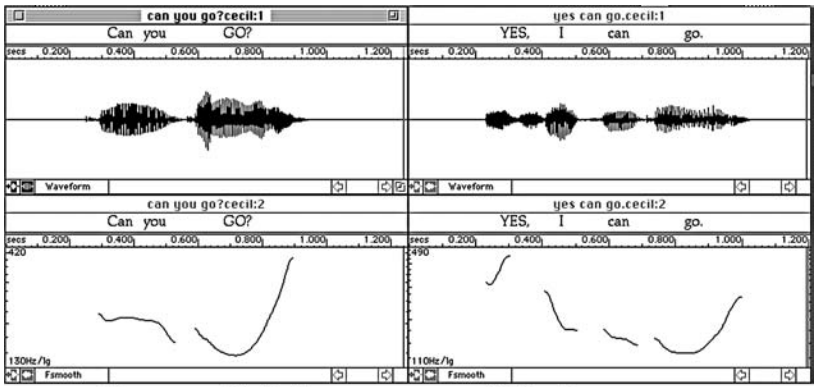
5(a): *Are you going?*
5(b): *Yeah, I'm going.*



(a) *Are you GO ing?/* (b) *YEAH, I'm GO ing.*

Figure 5. Yes-no question and reply

6(a): *Can you go?*
6(b): *Yes, I can go.*



(a) *Can you GO?/* (b) *YES, I can go.\/*

Figure 6. Yes-no question and reply

There are two types of *non-final intonation*, continuation rise and rising-falling intonation:

– **continuation rise** (Avery & Ehrlich, 1992, p. 79), termed **low-rise** (low to mid) by Dauer (1993, p. 225), is often used in lists and sometimes in the first clause of a complex sentence.

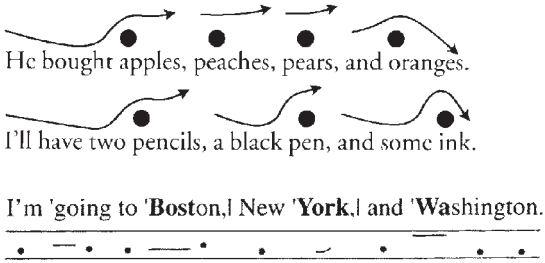
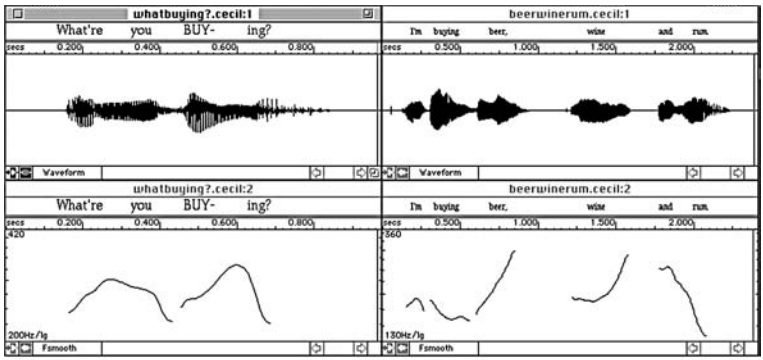


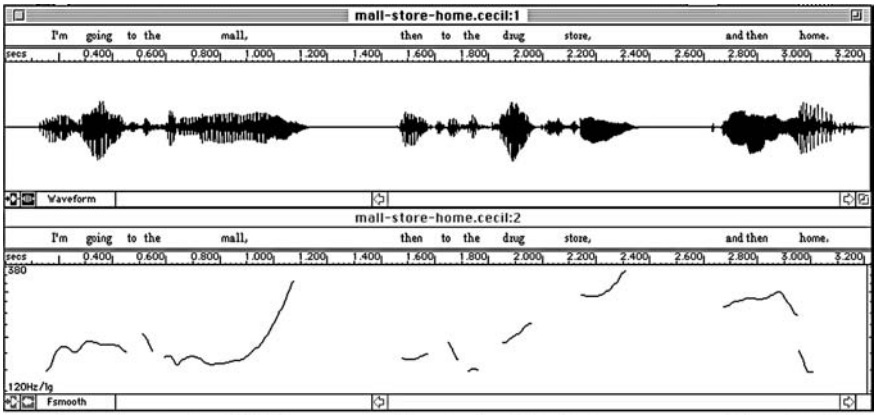
Figure 7 below shows the question *What are you buying?* and the reply *I'm buying beer, wine, and rum,* with rising intonation on the first two items of the list.



What are you BUY ing?/ *I'm buying BEER,/ WINE,/ and RUM.*

Figure 7. Question and answer series

Figure 8 below shows a sequential answer to the question *Where are you going?*: *I'm going to the mall, then to the drugstore, and then home.* The non-final items in this list have rising intonation ("continuation rises"), though this example does not exhibit "low rises" but rather rises to high pitch. The last item in the list has falling intonation at the end of the sentence. (Due to the voiceless consonants in the sentence, the intonation curves shown are not continuous



I'm going to the MALL,/ then to the DRUGstore,/ and then HOME.

Figure 8. Sequential list

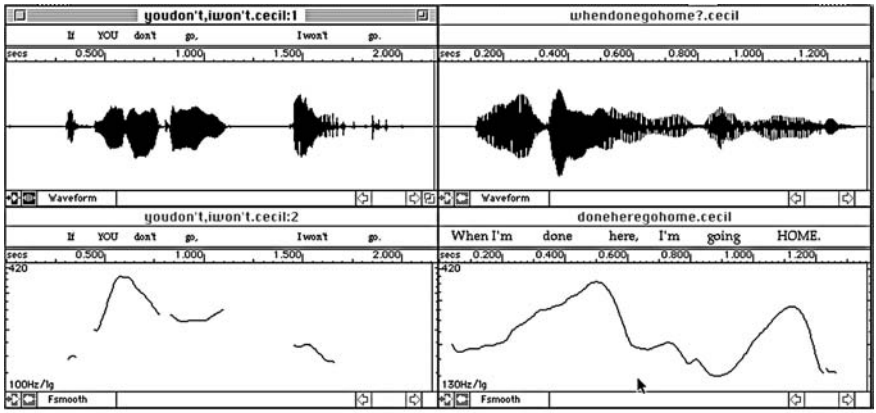
lines. As discussed in Chapter 5, this problem has led Spaai and Hermes (1993) to develop an intonation display system that superimposes a stylized pitch contour over the actual fundamental frequency measurements in order to help learners extrapolate a continuous pitch curve without breaks.)

The second type of non-final intonation is:

– **rising-falling** intonation (Avery & Ehrlich, 1992, p. 79), which is often used for the first clause of a complex sentence; pitch does not fall as far or steeply as in final rising-falling intonation.

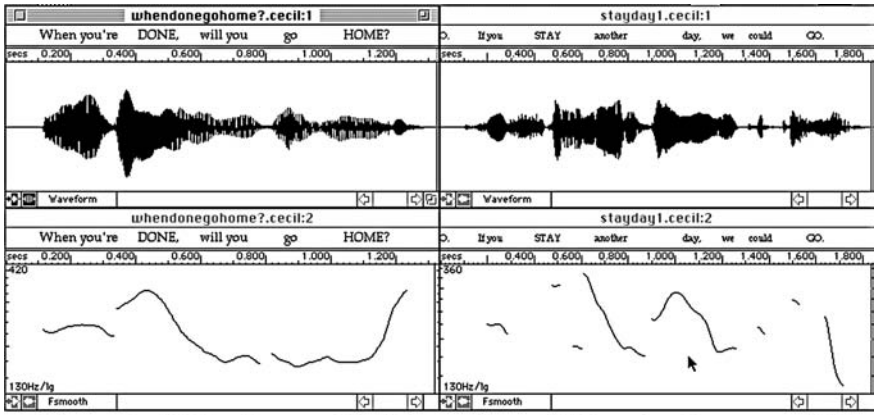
Figure 9 below shows two complex sentences that both begin with a subordinate clause. In the first sentence, *If you don't go, I won't go,* the subordinate clause ends with relatively level intonation on *go,* but in the second sentence, *When I'm done here, I'm going home,* there is rising-falling intonation (indicated by the arrow and described below) on the word *here,* though the pitch does not fall to as low a level as on the word *home* at the end of the sentence.

Other examples of rising-falling intonation with subordinate clauses are shown in Figure 10 below. In the first, *When you're done, will you go home?*, the rise-fall is on the word *done,* and in the second, *If you stay another day, we could go*, there is rising-falling pitch both on *stay* and *day.*



If YOU don't go,_ I won't go.\ When I'm done HERE,/\ I'm going HOME./

Figure 9. Complex sentences



When you're DONE,/\ will you go HOME?/ If you STAY another day,/\ we could GO.

Figure 10. More complex sentences

The patterns shown above are basic, neutral patterns. Practice with intensified versions, e.g., versions rendered with different *keys* or pitch ranges, should follow. These will be discussed further in a later section on the *discourse functions* of intonation.

Phrasing: Keeping thought groups together

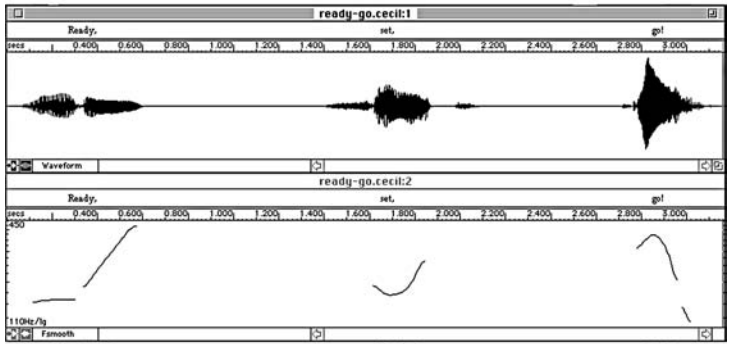
The chunking or demarcation of speech into thought groups (intonation units) is important both to listeners as an aid to comprehension and to speakers as a means of organizing their thoughts and keeping ideas separate. Languages can differ both in the way thought groups are marked intonationally (as well as morphosyntactically) and in the concept of what needs to be included in the group. Although Indo-European languages use pauses to mark groups, many languages (Cantonese, Korean, and others), use clause-final particles or affixes and do not rely on intonation or timing to indicate boundaries (cf. Gilbert, 1984, p. 32). Other languages (French, Spanish, Japanese, and others), although they use pauses for this purpose, put boundaries in different places than English (cf. Ballmer, 1980).

The first task for the teacher is to sensitize learners to listen for pauses, since pauses may be communicatively as important as the correct stress pattern of a word or correct accent in sentence focus.[4] As a very simple example, if telephone numbers are chunked or grouped in different ways from those they expect or are familiar with, foreign listeners may have difficulty knowing how many numbers will occur in a given group or chunk. In English, the boundaries of thought groups are marked by three basic phenomena:

– pause
– pitch movement or change in pitch
– lengthening of final stressed syllable

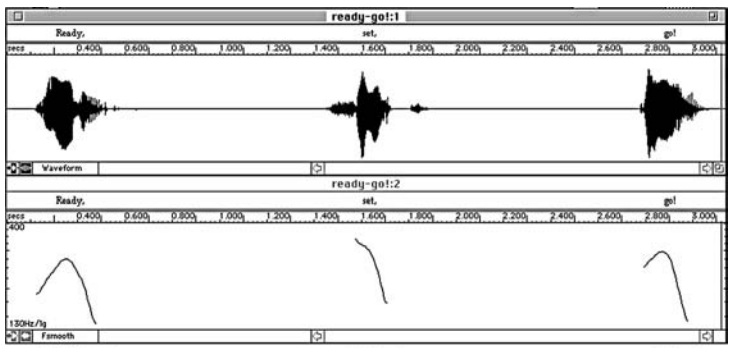
Pitch movement is interesting because in certain cases pitch rises before a pause, whereas in other cases, pre-pausal pitch falls. In lists, for example, all of the items except for the last one often have rising pitch in English, as shown in the examples earlier (see Figures 7 and 8) and in Figure 11(a) below. Listening exercises can include dictation of phone numbers, addresses, passport numbers, student identification numbers, social security numbers, etc. As an example of another type of list, an instructor could give a short "lecture" outlining a specified number of important points. "Today, I will talk about three main topics. First… Second… Third…" Students can be asked to listen for breaks between topics, that is, the end of one topic, the beginning of the next topic, etc.

For speaking or production practice, students can be asked to recite their own lists, e.g., *Ready, set, go!* and phone numbers: *three-oh-two, two-one-nine, eight-one-four-nine*, as shown in Figures 11 and 12 below. Shown here are two renditions of *Ready, set, go!* — the first, in Figure 11(a), with rising intonation



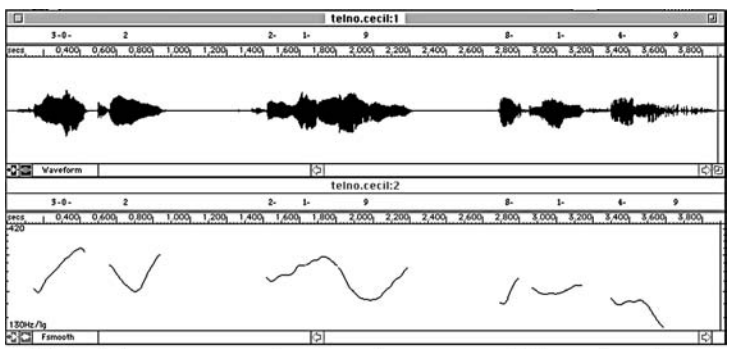
READy,/ *SET,/* *GO! *

Figure 11(a). Rising intonation on non-final items



READy, *SET,* *GO! *

Figure 11(b). Falling intonation on all items



three-oh-two,/ *two-one-nine,/* *eight-one-four-nine *

Figure 12. Rising intonation on non-final items

on the first two words and falling intonation on the last, and the second, in Figure 11(b), with falling intonation on all three items. For the telephone number in Figure 12, rising intonation for the first two groups of numbers was used and then falling intonation for the last group.

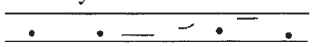
Further speaking practice could include commonly used games that entail building lists (e.g., "I'm going on a trip and I need to pack x, y, z…"), where each student in succession must add a new item to the list.

Special question types and patterns

In getting into the somewhat finer points of intonation, Dauer (1993) discusses certain special types of questions in English that can be used with more advanced students.

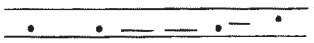
Choice questions: These are questions that give the listener two or more choices connected by the word *or*. When the listener is to choose only *one* of the alternatives, there is rising intonation on the first item and a fall on the second (so the question ends much like a *wh*-question), as in Figure 13(a) below: *Would you like tea or coffee?* ("Which one would you like?"). However, in the second example, Figure 13(b), the question is not really offering the listener a choice, but is rather a yes-no question meaning "Do you want anything to drink?", and thus ends in a *rise* (cf. also Prator & Robinett, 1985).

A: Would you like **tea**l or **c**offee? (Which one? - ends with a fall)



B: **C**offee, please. (or **Tea**, please.)

A: Would you like tea or **c**offee? (Anything? - ends with a rise)

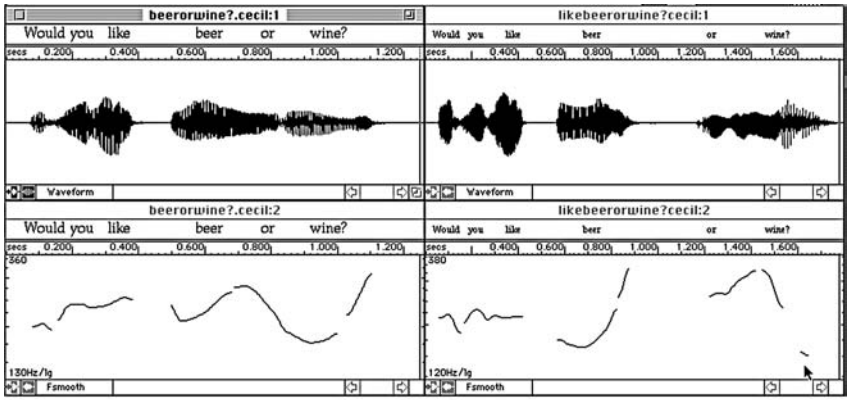


B: **No**, thank you. (or Yes, I **would**.)

(Dauer, 1993, pp. 238–239)

In Figure 13 below, both the yes-no question in 13(a): *Would you like beer or wine?* ("Would you like either/anything?") and the choice question in 13(b): *Would you like beer or wine?* ("Which would you like?") are shown, with rising intonation on *beer* and on *wine* at the end of the yes-no question, in contrast to the rise on *beer* but a fall on *wine* in the choice question. Note that the pitch on *wine* in the second instance, 13(b), actually falls quite low (as indicated by the

arrow), with the break in the curve being due to a drop in amplitude at the end of the sentence and the inability of the software to "track" the pitch.



(a) *Would you like beer or WINE?/* (b) *Would you like BEER/ or WINE?*

Figure 13. Yes-no and choice questions

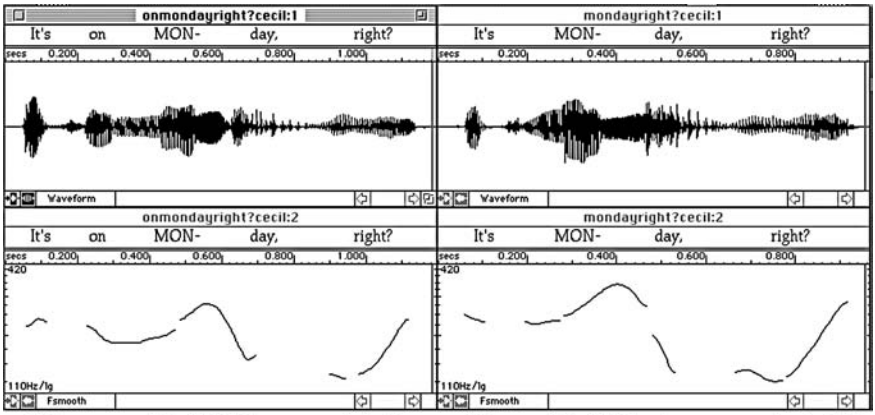
Tag questions*:* Tags can have either final (rising-)falling or final rising intonation in English, depending on the speaker's expectations about the answer.[5] A rising contour is used when the questioner really does not know the answer to the question, whereas with a rising-falling pattern, the questioner presumes to know the answer and is merely trying to confirm the presumption.[6]

A: You under**stand**,| **don't** you?
B: **Yes**, of **course**.
A: That's what I **thought**.

A: You under**stand**,| **don't** you?
B: **Well**, as a matter of fact, I **don't**.
A: **Oh**. Well let me ex**plain** it again.

Dauer (1993: 238–239)

Both of the examples in Figure 14 below show tag questions — 14(a) *It's on MONday, right?* and 14(b) *It's MONday, right?*–with rising intonation. Utterance 14(b) is said in a higher key, i.e., the overall pitch range is greater than in utterance 14(a), and the question sounds somewhat more insistent or in need of confirmation.



(a) *It's on MON day, right?/* (b) *It's MON day, right?/* [H]

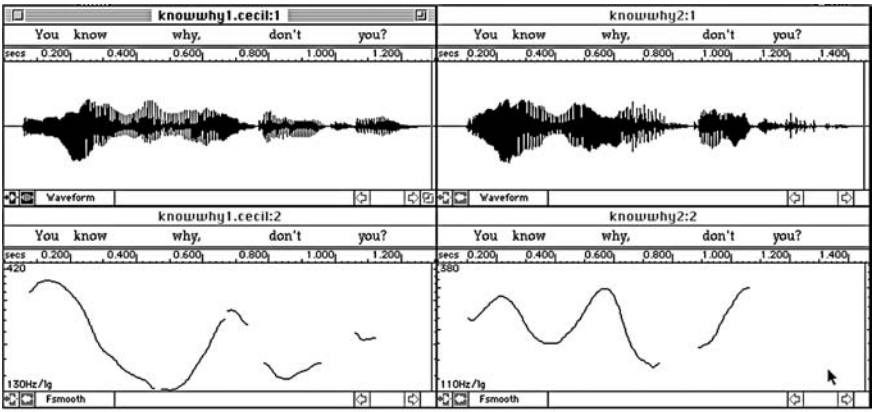
Figure 14. Tag questions

The examples in Figures 15(a)-15(d) below show four renditions of the question *You know why, don't you?* with different combinations of intonation patterns for each.

(a) *You know why,/ don't you?/*
(b) *You know why,\ don't you?*
(c) *You know why,/ don't you?*
(d) *You know why,\ don't you?/*

The first rendition, 15(a) *You know why,/ don't you?/,* shows rising intonation on both clauses, indicating a solicitous question in which the speaker, not having expectations about the answer, is asking the interlocutor a genuine question.

The second rendition, 15(b) *You know why,\ don't you?*, is said with falling intonation on both clauses, indicating statement-like intent whereby the questioner is expecting the hearer to know why and to confirm that this is the case. In the actual pitch curves below, only the fall on *why* is visible, and the fall on *you* was not detected by the signal analysis program because the intensity of the utterance had dropped too low at the end. The arrow points to where the curve would most likely have ended had the acoustic signal been stronger.

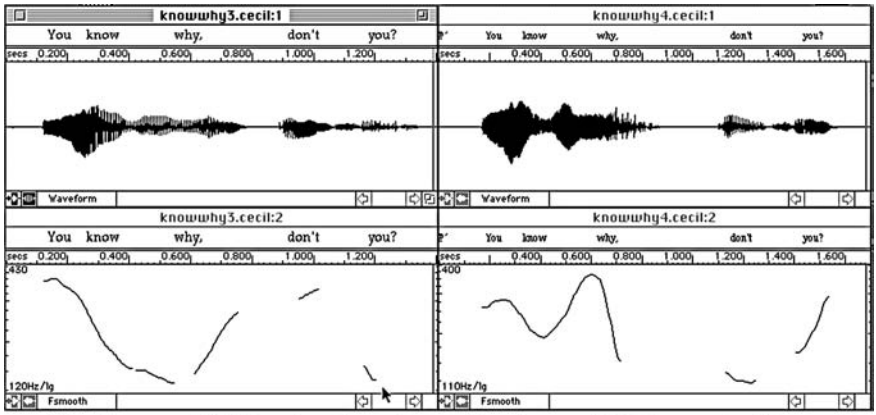


(a) *You know why,/ don't you?/* (b) *You know why,\ don't you?*

Figure 15. Different renditions of a tag question

The third rendition, 15(c) *You know why,/ don't you?*\, shows rising intonation on the first clause and falling intonation on the second, with the questioner again assuming that the listener does know why and wanting to confirm this fact. The rising intonation on the first clause may indicate a sort of reminder to the listener who may know why but may not be willing to accept the reason. Due once more to the very weak intensity of the speech signal, the falling pitch contour at utterance end is again not a continuous line, though a small portion of the end of the contour (pointed to by the arrow) is present.

The fourth rendition, 15(d) *You know why,\ don't you?/,* shows falling intonation on the first clause and rising intonation on the second. This pattern is used when the speaker thinks or expects that the listener knows why, but then has doubts and wishes to confirm this fact. These examples illustrate the contextualized nature of intonation in discourse: the utterances are marked intonationally not to signal that they are questions or statements, but rather to signal underlying assumptions and expectations about the response. Marking shared knowledge and presuppositions is a function of intonation at the discourse level. An important principle illustrated here is that there is *not* a one-to-one correspondence between syntactic type and intonation pattern, but that different combinations of these structures occur, depending on pragmatic intentions and meanings. In reality, i.e., in actual conversations, there might be very few "neutral" utterances; rather, utterances are closely bound to the context in which they are said.



(c) *You know why,/ don't you?* (d) *You know why,\ don't you?/*

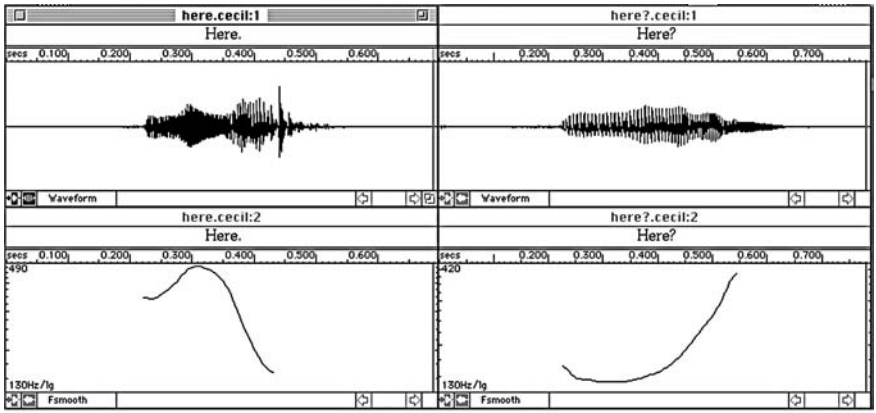
Figure 15. Different renditions of a tag question

Attitudinal functions of intonation

In the preceding sub-sections, the so-called *grammatical functions* of intonation and neutral intonation patterns were presented. In order to practice *attitudinal functions* of intonation, the same sentences or dialogues that were used to practice neutral intonation can be used but "spiced up," by assuming different contexts or creating new situations. For example, students can be instructed to turn relatively emotion-less questions into echo-questions requesting confirmation, or expressing incredulity, e.g., *WHEN did you say the test is?* (requesting confirmation) and *The test is on MONday?* (incredulous). Similarly, students can practice statements or exclamations that express attitudinal nuances by responding with exasperated or emphatic intonation, e.g., *No, it's NOT going to be on Tuesday. I told you it's going to be on MONday!*

Exercises for one-syllable or one-word utterances with different pitch patterns can be expanded for discourse practice by adding the dimension of *key* to signal additional meaning. For example, the replies *yes* and *no*, *here*, *Monday*, etc. could be given with one person asking yes-no or *wh*-questions and the other person replying with particular attitudinal nuances. For instance, following practice of neutral or unemotional responses, emphatic, surprised, exasperated, angry, or impatient connotations could be attempted. Students can subsequently progress to initiating their own statements, such as *We're not meeting at the university, we're meeting HERE* or *We're having an exam on*

MONday, to which other students would respond in high key as though surprised or incredulous, *HERE?, MONday?*. Figure 16 below shows intonation curves for the word *Here* spoken first with rising-falling intonation (in 16(a)) and then with rising intonation (16(b)). In both cases, the pitch range is quite wide, approximately 140–450 Hz, which contributes to making the first an exclamatory utterance and the second a surprised or astonished question — both using high [*H*] key.

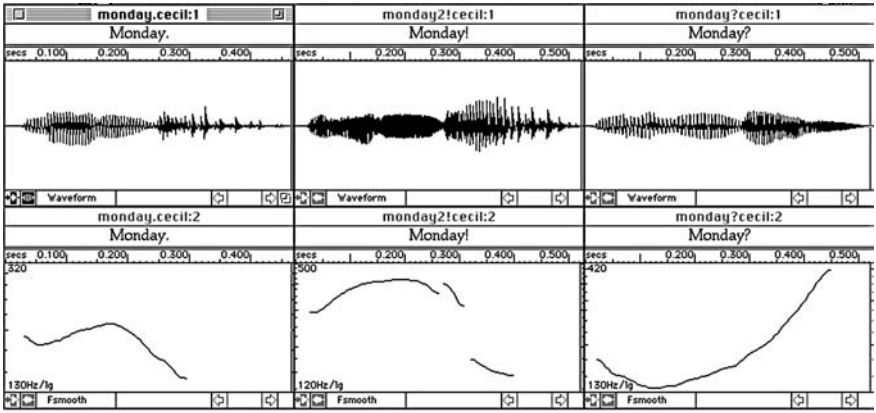


(a) [H] *Here*!/\ (b) *Here*?/ [H]

Figure 16. Two renditions of *Here*

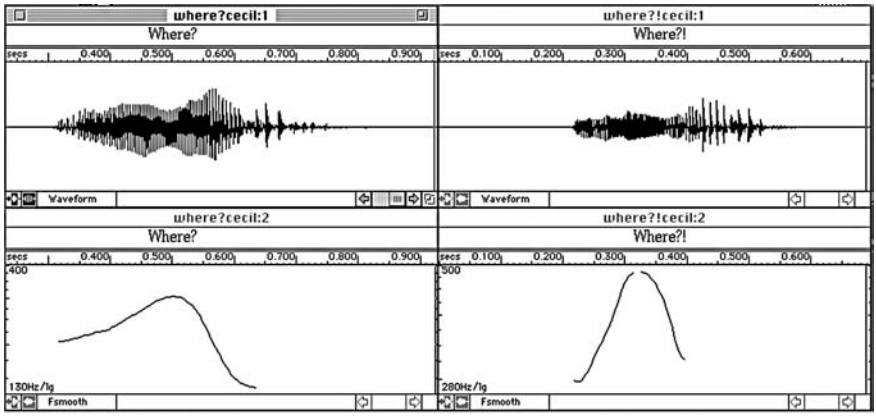
In the examples in Figure 17 below, the first instance of *Monday* is said with simple falling intonation in (a), as a "neutral" reply to a question. The second utterance, shown in (b), is an exclamation, emphatic and potentially angry, spoken in high key — as shown by the maximum pitch being close to 500 Hz — and falling to a low pitch of almost 130 Hz. The third utterance in (c) is a surprised echo question which rises to high key, approximately 420 Hz.

In Figure 18 below, the two examples of *Where* both show the rising-falling intonation typical of *wh*-questions; the first, 18(a), is a more "neutral" question while the second, 18(b), is incredulous or exasperated; the latter shows a much steeper rise and fall in pitch, with maximum fundamental frequencies of 305 Hz and 487 Hz, respectively.



(a) *Monday.*\ (b) [H] *Monday!*/\ (c) *Monday?*/ [H]

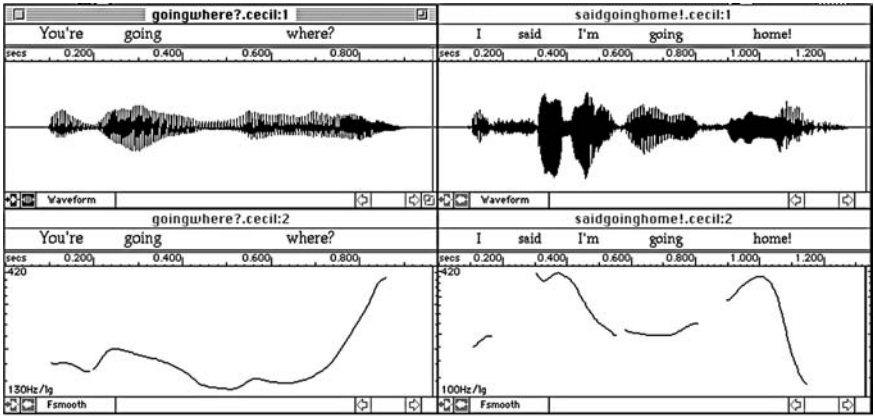
Figure 17. Three renditions of *Monday*



(a) *Where?*/\ (305 Hz) (b) *Where?!*/\ [H] (487 Hz)

Figure 18. Two renditions of *Where?*

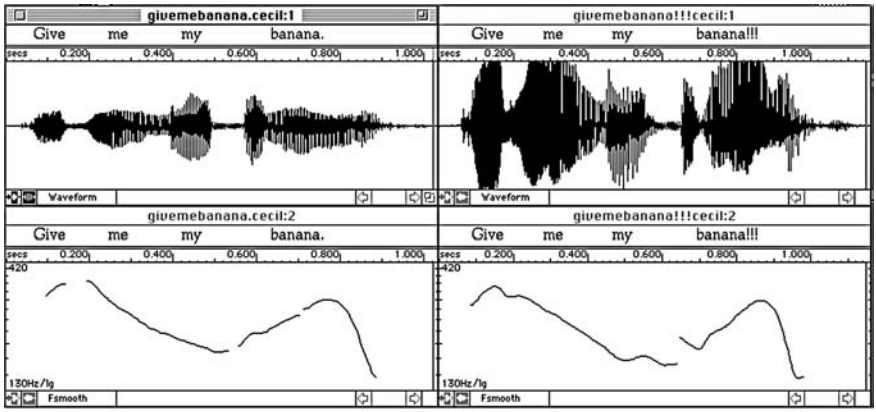
In the figure below, the follow-up question, 19(a), *You're going WHERE?* is asked with great surprise and the pitch rises to high key on the word *where*. In the reply, 19(b), the repetition by the speaker shows high peaks on *said* and *home*, suggesting emotion such as exasperation or impatience.



(a) *You're going WHERE?* (b) *I SAID I'm going HOME.*

Figure 19. Follow-up utterances

In the examples in Figure 20 below, the left display, (a), shows a fairly neutral imperative while the right display, (b), shows an insistent imperative. Of note is the fact that the intonation curves and pitch ranges are very similar for this particular speaker, but the waveform shows much greater intensity (loudness) in the insistent than in the neutral utterance. Features of voice quality also play a role here, but these types of features are not easily measured by the current software (see endnote 1 to this chapter).[7]



(a) *GIVE me my baNAna.* (b) *GIVE me my baNAna!!*

Figure 20. Two renditions of an imperative

Discourse functions of intonation

Stress (accent) for information focus

A typical exercise for practicing sentence stress (or accent) in non-neutral utterances is to vary the placement of stress in various sentences. Using the same sequence of words, different elements of a sentence can be stressed to illustrate how accent is realized in different positions as well as how the placement of stress signifies new information, signals emphasis, or answers different questions.

As a listening exercise, Dauer (1993) suggests having the instructor or a student read a sentence several times with a different stress placement each time and asking students to identify the word which was stressed, either orally or on a written transcript (p. 230):

> *When did you finish your HOMEwork?* (neutral; not your dinner)
> *When did you FINish your homework?* (not begin it)
> *When did YOU finish your homework?* (not someone else)
> *WHEN did you finish your homework?* (what time exactly)

Students could also be asked to give an appropriate answer to the question they thought they had heard — either given multiple-choice answers provided in written form or in an open-ended oral activity.

Conversely, learners could be given a statement (e.g., *That's my new black leather jacket*) and asked to read it with stress appropriate for given questions. The following are examples in Hagen and Grogan (1992, p. 136); cf. also Prator and Robinett (1985) for similar suggestions:[8]

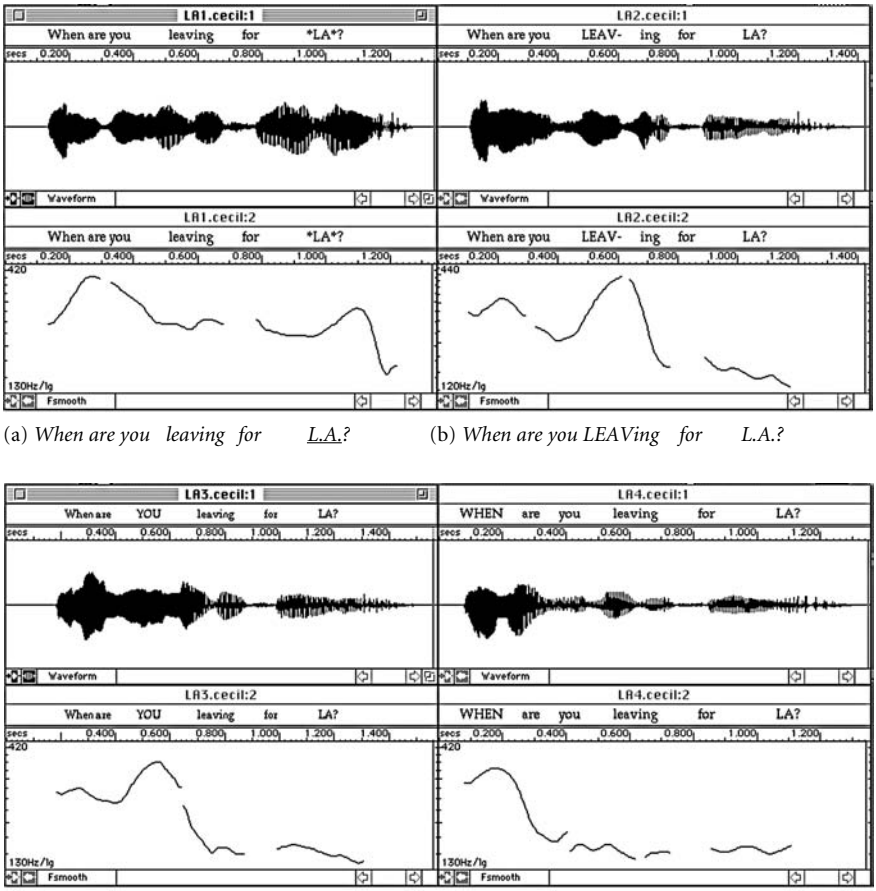
What's that?	*That's my new black leather JACKET.*
Whose new jacket is that?	*That's MY new black leather jacket.*
Is that your new jacket, or your old one?	*That's my NEW black leather jacket.*
Which is your new black leather jacket?	*THAT'S my new black leather jacket.*

Of course, all of the above answers are not very lifelike because they are unusually elaborate. In actual conversation, the answers would likely be syntactically less complete in that the shared information (the entire noun phrase) would generally not be repeated. But as an exercise, it demonstrates the moveable placement of stress in English to learners.

The examples (a)–(d) in Figure 21 below show intonation curves for the four questions:

(a) *When are you leaving for L.A.?*
(b) *When are you LEAVing for L.A.?*
(c) *When are YOU leaving for L.A.?*
(d) *WHEN are you leaving for L.A.?*

The most striking feature is that the stressed word in each question has a definite falling-rising pitch contour, no matter where it occurs in the utterance. All of these *wh*-questions have overall falling intonation.



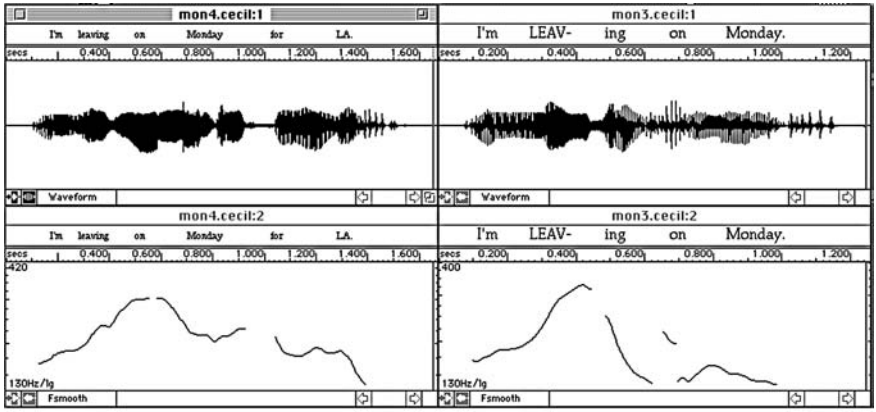
(a) *When are you leaving for L.A.?* (b) *When are you LEAVing for L.A.?*

(c) *When are YOU leaving for L.A.?* (d) *WHEN are you leaving for L.A.?*

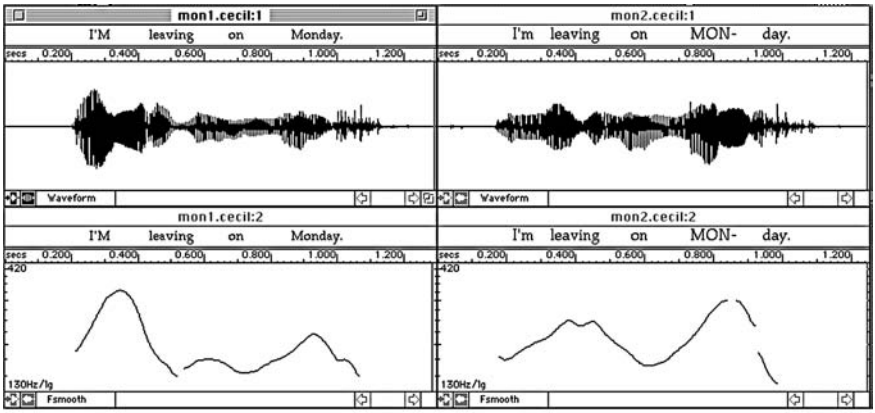
Figure 21. Differing placement of sentence accent in questions

Replies to the questions above are shown in Figure 22:

(a) *I'm leaving on MONday for L.A.*
(b) *I'm LEAVing on Monday.*
(c) *I'M leaving on Monday.*
(d) *I'm leaving on MONday.*



(a) *I'm leaving on MONday for LA.* (b) *I'm LEAVing on Monday.*



(c) *I'M leaving on Monday.* (d) *I'm leaving on MONday.*

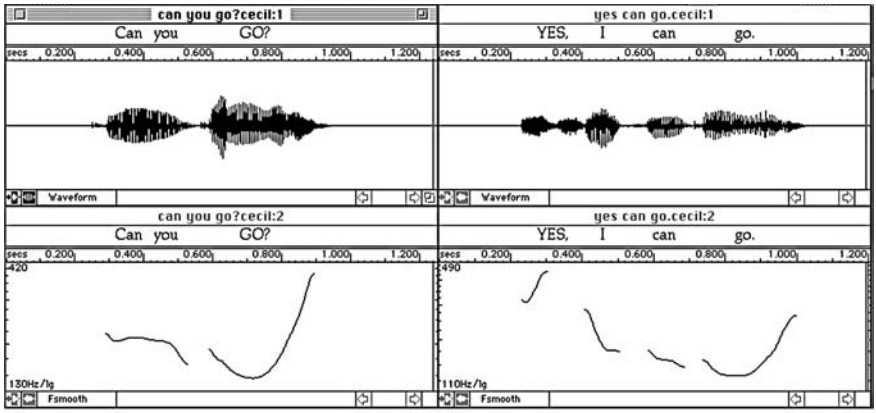
Figure 22. Sentence accent placement in corresponding replies

Although a number of handbooks on pronunciation do provide examples similar to those presented here, few include actual fundamental frequency (pitch) curves. Inspection of the pitch contours, however, could be enlightening for learners as a method of visualization, whether such patterns are displayed on an overhead projector or if the learners use computer software to display the pitch contours they themselves produce and use the displays as feedback.

Stress (accent) for contrast and emphasis

Another type of stress that can be practiced is contrastive stress. Of note is that contrastive stress (accent) is sometimes moved from its normal/neutral position at sentence end closer to the beginning of a sentence. Intonational features commonly found in English for signaling emphasis, contrast, and contradiction of expectations, respectively, are very high or very low pitch, steeper rising or falling curves and greater intensity and longer duration of stressed syllables.[9] In the example in Figure 23 below, the yes-no question *Can you GO?* is answered emphatically as *YES, I can go* with stress on the word *YES,* as if the answer might be surprising to the hearer or contradictory to the hearer's expectation, implying "Of course I can go, why do you ask? Did you have any doubt?" The word *YES* is spoken in a very high key.

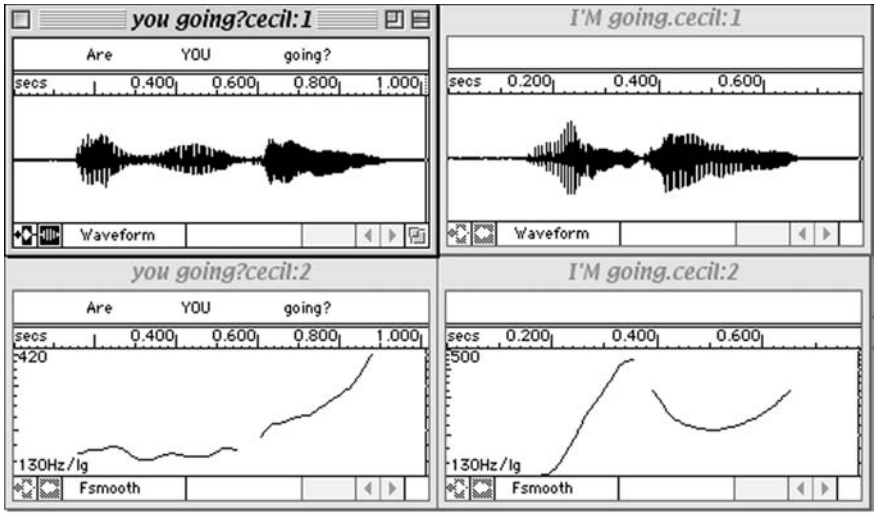
(a) *Can you GO?*
(b) *YES, I can go.*



(a) *Can you GO?/* [H] (b) [H] *YES, I can go./* [M]

Figure 23. Contrastive stress (accent)

Another example of discourse emphasis is the pair of utterances in Figure 24 below, with stress on the pronouns *you* in (a) and *I* (in (b). Generally, pronouns are not stressed or accented in discourse — at least not for the reason of signaling new information, since by nature pronouns refer to entities that have been introduced previously in the discourse or whose noun referents are known. But depending on what has gone before in the discourse, the pronouns could be stressed; in this example, the pronoun *I* is spoken in very high key in (b), and the utterance ends with rising intonation as if to imply "Why do you ask?" or "Why did you doubt it?" Rising intonation in general might also signal non-finality in the sense that the responder expects a rejoinder.



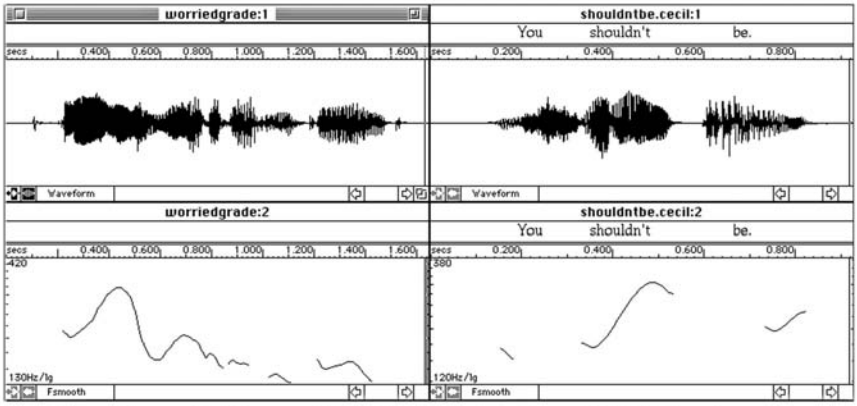
(a) *Are YOU going?*/ [H] (b) [H] *I'M going.*/ [M]

Figure 24. Emphatic accent

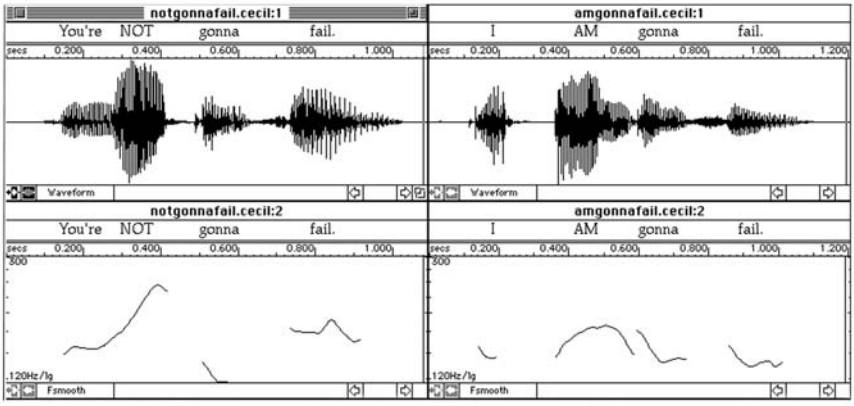
Stress (accent) to insist or contradict

In the sample dialogue in Figure 25 below, sentence stress occurs in *final* position in the first utterance, but — due to the pragmatic stress on words such as pronouns and auxiliary verbs — at or near the *beginning* of the sentence in the remaining utterances. Unlike so-called content words, such function words are not generally stressed, but in this context they are.

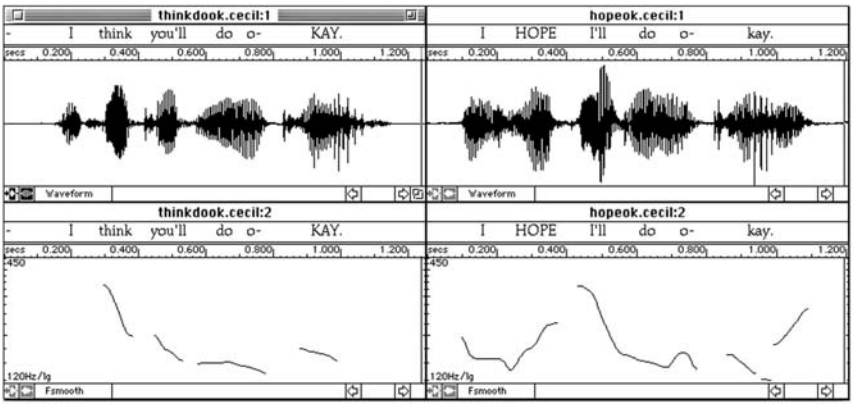
A: (a) *I'm really worried about my GRADE.* (A thinks s/he might fail)
B: (b) *You SHOULDn't be.* (c) *You're NOT gonna fail.* (B insists that A isn't going to fail)
A: (d) *I AM gonna fail.* (A insists s/he will fail)
B: (e) *I think you'll do okay.* (B expresses a contrasting belief)
A: (f) *I HOPE I'll do okay.*



(a) *I'm really worried about my GRADE.*\ [L] (b) *You SHOULDn't be.*/ [M]



(c) *You're NOT gonna fail._* [M] (d) *I AM gonna fail.*\ [L]



(e) [H] *I think you'll do o kay.*\ (f) *I HOPE I'll do o kay.*\/ [M]

Figure 25. Sample dialogue with insistence and contradiction

With the above example serving as a model of students' interaction with the teacher, similar exchanges between peers (student + student) can subsequently be practiced. Working in pairs or in small groups, the students would be asked to construct their own dialogues based on this model.

Intonation at transition points

In addition to the use of stress or accent to indicate information focus, contrast, and emphasis, pitch contours at points of transition in discourse are critical. Brazil (1975) maintains that high key functions in utterance-final position to signal contrast and set up expectations whereas low key implies finality and may discourage the interlocutor from responding. DuBois et al. (1992) posit three main types of "transitional continuity" at the end of intonation units: final, continuing, and appeal, with intonation contours that fall to a low pitch generally understood as "final," those that rise slightly or remain level throughout to be "continuing," and those rising to a high pitch considered to be an "appeal." In the examples given earlier in Figure 23, the question in (a) ends in high rising pitch and is an appeal for information. The reply in (b) starts in high key similar to that at the end of the question and indicates a sort of contrast: "Contrary to what you might think, yes, I can go." The reply ends in a rising contour, which may invite or encourage the interlocutor to continue the conversation (a sort of appeal, in the sense of DuBois et al.). Similarly, in

Figure 24, the question in (a) also ends in high key, and the reply in (b), though a statement, ends in mid key with a slightly rising contour, as if to imply, "Why do you ask?" — again, a sort of appeal.

In examining the intonation contour at transition points in the examples in Figure 25, we see that the statement in (a) ends in a low falling contour, which might signal the end of the speaker's turn. The reply in (b), ending in a slight rise and mid key, indicates that the interlocutor might want to continue. The (generally) level contour at the end of (c) might signal a willingness to yield the floor. The low-falling contour in (e) might reflect a desire on the part of the speaker to reassure the other.

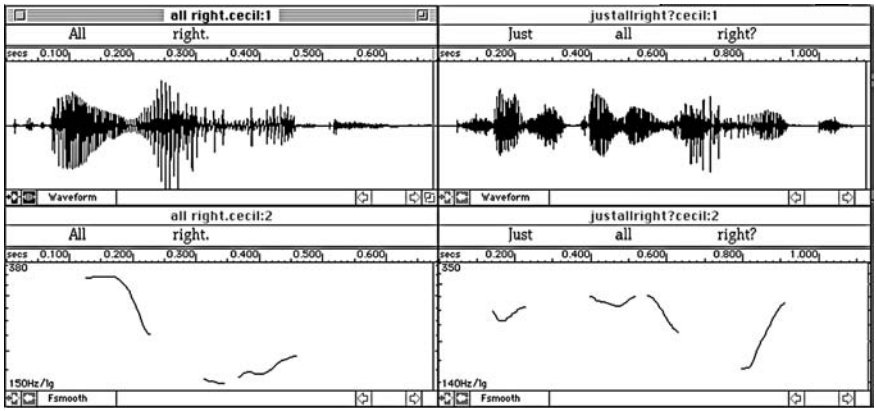
Interactive functions of intonation

In the handbook *Intonation in Context*, Bradford (1988) provides a course on intonation for upper-intermediate and advanced learners of (British) English as supplementary practice from a communicative approach. While earlier approaches to the teaching of intonation related intonation to either grammatical features or attitudes, the activities suggested by Bradford are based on Brazil (1975, 1978, 1985), who viewed intonation primarily as a feature of discourse and the developing interaction between speakers. The following is an example of a very short conversation which students first listen to, then mark on a transcript for tones and finally repeat (cf. Bradford, p. 22). They then practice the response *All right, I think* using different intonations and note the difference in meanings; the initial response is a statement or assessment, whereas the "echo" or "quoting" of the interlocutor is a request for clarification or expansion. This is a very common conversational strategy which demonstrates that the listener is attentive and wants to continue the topic; \/ represents the fall-rise tone, \ the fall and // represents a tone unit boundary (cf. also Couper-Kuhlen, 1996 on "The prosody of repetition: on quoting and mimicry"):

L: HELLO TONY // \/ DID you go for your INterview yesterday//
T: // \ HI Lisa// \ YES // \ I DID //
L: // \ HOW did it GO //
T: // \ All RIGHT // \/ I THINK //
L: // \/ All RIGHT // \/ You DON'T sound very SURE //

The displays in Figure 26 below show falling intonation followed by a level pitch at the end of *ALL right* in (a) and a falling-rising intonation on the echo-

question *Just all RIGHT?* in (b), as these were rendered by a native speaker of American English.



(a) *ALL right._* (b) *Just all RIGHT?\/*

Figure 26. Two renditions of *All right*

The above example is intended to illustrate that students can "role-play" using existing dialogues in whichever textbook is being used. A first step is to have them read the dialogues in a neutral, unemotional manner to practice the "unmarked" forms of intonation. They can subsequently be asked to assume various roles or attitudes and in so doing can be sensitized to the use of intonation to convey attitude or emotion (e.g., interest, annoyance, indifference, enthusiasm, impatience, surprise) while also practicing interactional strategies like asking for confirmation, quoting, or indicating expectations. Lastly, they can be asked to generate their own dialogues on related topics with slightly modified purposes, intentions, or outcomes.

Other discourse and turn-taking strategies that speakers use in conversation include interrupting, changing the subject, repairing following misspeaking or misunderstanding, and encouraging and discouraging responses from their interlocutor. Research in the area of discourse strategies, with particular emphasis on analyzing databases of natural conversations, is currently being conducted (cf. Couper-Kuhlen & Selting, 1996b; Yang, 1995). There are few studies that have addressed this issue in L2 learning. As mentioned in Chapter 5, Hurley (1992), for instance, found that drops in loudness and pitch are turn-relinquishing signals in English, and if Arabic speakers of English use non-nativelike volume, this could be construed as an effort to hold the floor (pp. 272–273). As

more becomes known about the components of prosody that help to shape conversation and interaction, language instructors will be better able to teach learners about the intonational features necessary for successful communication. At the moment, further examples of pitch concord and other prosodic features that are present in naturally occurring, cooperative interactions are not available, but there is relevant ongoing research, e.g., with the extensive Corpus of Spoken American English (CSAE), [http://www.linguistics.ucsb.edu/research/sbcorpus/default.htm]. A study by Wennerstrom (1998) reports on lectures in English by native speakers of Chinese, whose intonation was analyzed in terms of effectiveness for signaling the distinction between content and function words, linking related constituents, contrasting new items and given items, and signaling topic shift. These four measures had been chosen for their contribution to the cohesion of the lectures, and Wennerstrom found that good macro-level intonation helped speakers score higher on a global language test and made them more comprehensible. She advocated that future studies of intonation focus on discourse contexts rather than on individual utterances.

Sociolinguistic functions of intonation

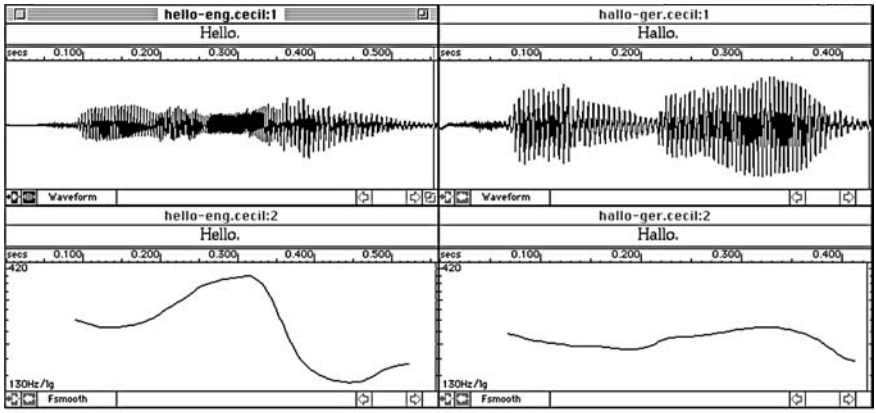
As outlined earlier in this chapter, some of the sociolinguistic functions of intonation include the following:

- indicating roles and status of speakers
- characterizing formal vs. informal speech
- signaling politeness, deference, etc.
- revealing gender differences and age differences
- indicating regional origins

Cross-linguistic studies of intonation have begun to shed light on these functions, which can vary subtly among languages. As discussed in Chapter 5, Loveday (1981) studied politeness among Japanese and English speakers and found more sharply-defined differences in both absolute pitch and within-utterance pitch variation between Japanese males and females in uttering Japanese politeness formulas than he did between English males and females in uttering English formulas. Low intonation contours are judged by native speakers of English to indicate boredom and detachment, and if male Japanese speakers transfer their low contours from Japanese to English when trying to be polite, this might result in misunderstandings by native speakers of English.

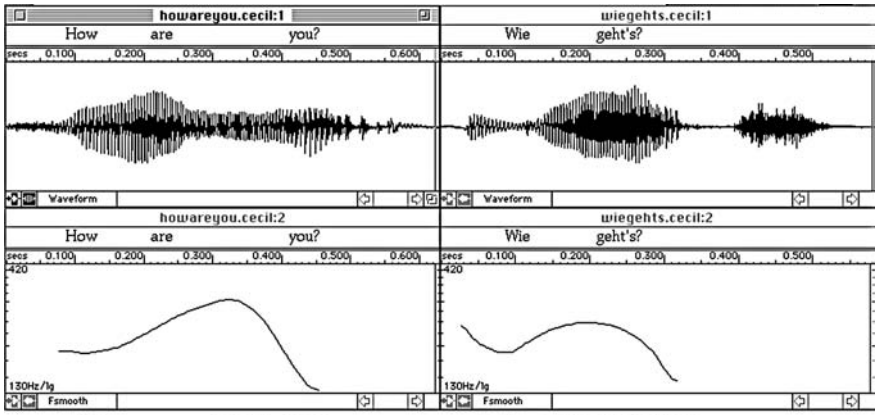
Van Bezooijen (1995) corroborated earlier studies that found Japanese women to have higher pitches than Dutch women. Gimson (1980) posited that in English the greeting *Good morning*, when said within a high pitch range and with a rising contour, might express that the speaker is cheerful and friendly, whereas the same phrase said within a low pitch range would tend to make the greeting more routine or perfunctory (p. 112). In extrapolating to other languages such as German, *Guten Morgen* said within a low pitch range would not necessarily imply cheerlessness, unfriendliness, or impoliteness, but nor would it necessarily be routine or perfunctory (see Moulton, 1962, p. 138). Anecdotal evidence suggests that German women generally use lower pitch ranges and less dynamic intonation patterns than American women, and that when American English intonation patterns are transferred to German, speakers are often thought to be silly or frivolous.

In Figures 27 and 28 below, the American English greetings *Hello* 26(a) and *How are you?* 27(a) are compared with the German greetings *Hallo* 26(b) and *Wie geht's?* 27(b). Note that the English pitch curves are more "dynamic" in the sense that they exhibit steeper rises and falls as well as wider and higher pitch ranges. Learners can be made aware that German intonation is characterized by somewhat "flatter" pitch curves but that this does not necessary imply less enthusiasm or friendliness.



(a) *H e l l o.* /\ (b) *H a l l o.* /\

Figure 27. English vs. German greetings



(a) *How are you?* /\ (b) *Wie geht's?* /\

Figure 28. English vs. German greetings

Gibbon (1998) states that the indexical function of pitch height to indicate degree of emotion is widespread and that it certainly applies to German. "However, the range of pitch modulation in German is in general much less than in English […], which may lead to misjudgments of intention or attitude. British female voices, in general relatively high-pitched, tend to sound aggressive and over-excited to the German hearer, and conversely, German males may sound 'bored' or 'unfriendly' to the British hearer" (p. 89).

Examples of less formal speech were presented in Chapter 6 with regard to rhythm and how certain sounds and syllables are elided or omitted in fast, informal speech (cf. also Johns-Lewis, 1986b, "Prosodic differentiation of discourse modes" for studies on read vs. spoken language). Although the pitch patterns may not be too different in formal vs. informal speech, rhythm may be a more important factor in differentiating the two, and learners should accordingly be given opportunities to practice speaking in different situations and contexts.

In teaching students how intonation is used to express politeness or deference, instructors would want to combine intonation training with a discussion of the syntactic formulas and expressions that are used in a given language. For example, in English, one could compare the command "Give me that!" (falling intonation) with the polite request using the subjunctive "Would you please give me that?" (falling intonation). This in turn could be contrasted with the very polite request that is an appeal "Would you please give me that?" (rising

intonation). Instructors should then describe situations in which these variants might occur.[10]

With regard to different regional varieties of English, (Standard) American English and (Standard) British English are the two most frequently researched. They are also considered distinct standards for the teaching of English as a foreign language. Hirst (1998) points out that both national standards of English are spoken with a number of different accents, and that for British English, one particular accent, so-called *Received Pronunciation* (RP), is generally presented as a model for foreign learners despite the fact that the proportion of the population of England that actually speaks RP has been estimated to be as small as 3%. Bolinger (1998) maintains that "With minor differences, American English shares a single intonation system with English in general, particularly Southern British. The differences are not in the configurations [...] but in frequency and pragmatic choice" (p. 45).

Bolinger's (1998) views on intonation are also interesting from a sociolinguistic standpoint in that he describes the intonation of American English as highly iconic. Accordingly, he believes that it must be studied in relation to the speaker's overall gestural code, especially facial expression and expressive body language. "A higher pitch is typically associated with higher positions of the eyebrows, shoulders, and often hands and arms..." (p. 45)

Hirst (1998) recommends that considerably more research into the intonation of other accents be conducted, and I believe this is particularly true regarding the sociolinguistic functions of intonation, which have not been studied as extensively as many of the other functions.

Adapting existing materials

This section will discuss how to adapt existing materials and situations so that they can be used as sources of intonation and discourse management practice. Examples will be given of some of the social functions of discourse intonation that can be practiced in a typical L2 classroom, particularly in signaling points of speaker change. The principles guiding use of intonation in discourse management and the fulfillment of social functions can and should be integrated into the earliest stages of language instruction. Their integration does not require new or extensively revised texts; even the most basic materials can be modified. If more traditional, grammar-based textbooks are used, an overwhelming percentage of exercises will consist of statements unrelated in con-

tent which are supposed to be manipulated toward specific grammatical ends. Such statements can be adapted for use in different discourse situations featuring various speech acts for interactive purposes. The instructor's mission in this situation is to point out not how the sentences are constructed (not "intensive grammar practice"), but what they do, i.e., what *pragmatic* functions they serve. If more recent, communication-based textbooks are used, they will likely contain contextualized exercises, and in this case, instructors must make intonation practice an integral component of instruction.

Below are simple examples of utterances in a discourse situation that reflect the four functions of intonation that comprise communicative competence. With dialogues incorporated into pair work or small group discussions, a simple question-answer sequence can be expanded to include various speech acts as well as different intonational patterns. In the dialogue below, none of the four persons involved needs to be a teacher, so it is ideally suited for small group work. The purpose of this exercise is to heighten students' sense of options about stress placement and sentence-final intonation (cf. Chun, 1988b, pp. 83–86).

Speaker A: (1) [H] *What did you think of the eXAM?*
Speaker B: (2) [H] *I thought it was HARD.*
Speaker C: (3) *ExCUSE me?*
(4) *You found it HARD?*
Speaker B: (5) *Yeah, TOTally.*
Speaker D: (6) *Well, I didn't.*
(7) *I thought it was pretty EASY.* [L]

The dialogue below is from Hagen and Grogan (1992), with stressed words capitalized (p. 134):

Speaker A: (8) *I have to buy a BOOK.*
Speaker B: (9) *What KIND of book?*
Speaker A: (10) *A TEXTbook.*
Speaker B: (11) *For GRAMmar or READing?*
Speaker A: (12) *I have THOSE books already. This one's for pronunciAtion.*

The dialogues above are typical of what might be found in textbooks, but they can be modified and used in group work with four different types of focus or suggested variations, as shown below. One goal is to get the students to correlate the pitch, for example of request (3), with an appropriate register (polite or not). If (3) is uttered within a mid key rather than a high key, it is a polite request for clarification rather than an expression of surprise or disbelief.

Level 1 variations: Grammatical competence

Although the above exchanges contain a minimum of lexical and syntactic difficulty, they do entail indirect practice of certain grammatical and lexical elements (e.g., verb tenses, use of *did* + verb vs. preterit, antonyms). Some of these grammatical features correlate more readily with specific intonations. The intonational features of greatest importance at this level are those of stress or accent, i.e., features that make a word in a sentence prominent. In the dialogues above, the capitalized syllables and words are those which might under "neutral" circumstances be accented. For some of the utterances, however, other possibilities exist — in sentence (2), for instance, both *I* and *hard* could be accented. In other utterances, on the other hand, the placement of accent may not be variable, e.g., in utterance (7), the word *easy* is being contrasted with *hard* and thus must be stressed (one would not, for example, stress *it* (the exam) since it is already an established topic or "old information"). The pronoun *I* in (7) would not normally be stressed since it is stressed in (6).

Level 2 variations: Attitudinal competence

Using the first dialogue, students could practice role plays and express different attitudes. In the first instance, all of the speakers in both dialogues could be matter-of-fact in tone, but could then be assigned specific characteristics. For example, Speaker B could be "whiny" while Speakers C and D could be arrogant and boastful. In the second dialogue, Speaker B could be nosy and inquisitive while Speaker A could be annoyed and impatient.

Level 3 variations: Discourse competence

Discourse competence involves the ability to string together utterances to form a coherent discourse. Despite the limited number of utterances in the sample dialogues, they still contain a variety of sentence types and discourse management strategies. Questions are posed and information is sought in (1), (4), (9), (11); responses are given in (2), (5), (10) and (12); confirmation is requested in (3) and (4); surprise is expressed in (4); opinions are stated in (2) and (6) and challenged in (4); and a differing opinion is offered in (7). Kramsch (1986) describes the discourse management skills used here to include the following: "turn-taking" [four speakers each take the floor and relinquish it]; "linking and

expanding" [utterance (4) restates (is linked to) utterance (2); utterance (7) expands on (6); utterances (9) and (11) expand on utterances (8) and (10) by requesting additional information]; "negotiation" [questions (3) and (4) request clarification and confirmation of what the previous speaker said]; and "repair" [although there was no misunderstanding, (5) confirms what the speaker had originally said and adds emphasis]. These constellations can readily be adjusted to derive additional role plays that the students might play out as alternate scenarios, e.g., Speaker A should not be polite; B should interrupt; C should not yield, etc.

In sentences (6) and (7), a contradictory opinion is expressed and particular word(s) contrasting with a previous word or idea are stressed: *Well,* **I** *didn't* and **I** *thought it was pretty* **easy**. This added stress on *easy* is greater than the stress used in a neutral statement. When students practice marking this emphatic or contrastive stress, knowledge of lexical antonyms is also reinforced. In utterances (10) and (12), stress on *text* and *those* emphasizes the new pieces of information that are being introduced.

In terms of DuBois et al.'s (1992) notion of transitional continuity, possible intonation contours to be used at the ends of the utterances might include: a rising "appeal" in (1), (3), and (4), perhaps a level contour for (6) indicating that the speaker wants to continue with another utterance, and a low falling contour in (7) indicating finality or "topic closed."

In practicing language used in social interaction, students can also incorporate the variable *key*, or the pitch relative to that used in the immediately preceding utterance, which functions as a control factor in conversation. As noted in the examples throughout this chapter, utterances can be described as being in a particular key: H (high), M (mid), or L (low). Since utterance (1) in the first dialogue begins a new topic, the beginning of the question might be uttered in high key (H). Since Speaker B then takes the floor for the first time and continues with the topic begun by Speaker A, high key might also be used at the beginning of utterance (2). By contrast, if Speaker D wishes to conclude the discussion of that topic, low key (L) might be used on utterance (6), along with low falling pitch at the end of the utterance to signal finality at this potential transition point. The instructor can introduce the concept of interactive pitch by pointing out keys that would not be used.

To structure such exercises, students form groups of four. Each person receives an index card assigning the recipient a specific role, e. g., Speaker A is the "discussion leader," responsible for introducing new topics; Speaker B is responsible for responding to questions and stating an opinion; Speaker C is

supposed to interrupt and ask for clarification of any piece of information (see example below) and is then to express surprise or disbelief; and Speaker D is to contradict what one of the others has said. Students are to play the roles indicated on the card they receive for one or more assigned topics. Then the cards can be redistributed and the roles rotated. In this way, students will learn to listen to the prevailing key of a conversation and manage the discourse so that they are perceived as cooperating group members. Taping these dialogues would make it possible for learners to listen to and analyze their use of key, or — if signal analysis software were used — learners could obtain measurements of their fundamental frequency and be given a visual indication of whether their use of key corresponded to that of their interlocutor.

Level 4 variations: Sociolinguistic competence

In languages which distinguish grammatically between formal and informal registers, conversations can be practiced using various combinations of speakers. In the examples above, the speakers are all socially equal; if presumed to be students, informal registers might be natural, but a modified situation could involve equals in a formal situation (perhaps professional people taking an adult evening course) where a more formal register might likely be found. Depending on the scenario suggested, students would need to use appropriate verb forms and perhaps more polite intonation as well — making it not only a grammar exercise, but also a role play. Students must learn, for example, to disagree politely: utterance (6), *Well, I didn't,* would probably be too abrupt in a formal situation, where it could easily be perceived as arrogant, especially depending on the tone of voice. A more polite rejoinder might be *I actually didn't find it so difficult,* particularly if said in a relatively mild tone of voice.

Interactive competence

"Interactive competence" overlaps discourse competence and sociolinguistic competence and is integral to both. It involves, for instance, the ability for interlocutors to negotiate in case of actual or potential misunderstanding. In the sample dialogues, utterances (3) and (4) request confirmation. Utterance (4) *You found it hard?* is a so-called "echo question" because it paraphrases what the previous speaker said. Although syntactically a statement, it becomes a question if rising intonation is used.

There are other ways of requesting clarification: e.g., as a rejoinder to utterance (2) in the first sample dialogue, *I thought it was hard*, either of the following would also be possible:

WHAT did you say was hard? / [H]
[H] *WHO thought it was hard?* \

If the first utterance, *WHAT did you say was hard?*, started with stress on the *wh*-word and at low pitch and continued at a rising pitch over the rest of the question and given high key at the end, it could represent a request for clarification while expressing surprise. On the other hand, if the second question, *WHO thought it was hard?*, were said with high pitch on the *wh*-word and falling intonation at the end, it might seem more like a matter-of-fact *wh*-question with emphasis on *who*. It might also be somewhat less polite, seeming like a demand for clarification: "[Tell me,] who thought it was hard?". In that case, it is possible that the speaker would also be surprised, but the implication might be "How could anyone find it hard?" Students should be made aware that English uses falling intonation to indicate assertiveness, while rising intonation can indicate doubt, surprise, or politeness.[11]

In order to be able to have students produce these different intonations naturally, teachers should set up different interactive scenarios. That is, the students need classroom directives beyond being told or encouraged merely to "have a conversation." Specificity of the scenarios will itself help to simulate greater "naturalness." For example, students can role play and differentiate between how "an impolite kid" vs. "a fearful employee" might request clarification. After practicing scenarios, students should find it easier to recognize the nuances of social interaction, which are often conveyed by intonation rather than grammar. They may also begin to understand native speakers more readily and accurately.

As an example, a boss/uncertain employee transaction could be used to teach students how to ask for clarification. One way of requesting clarification is by restating the question. In the examples below, the "responses" (2) and (4) both paraphrase the original questions, and the shifts into first person grammatically entail using different verb forms as well as syntactically different constructions (with or without auxiliary verbs):

Question: (1) *What did you think of the restructuring?*\
Response: (2) *What I thought of the restructuring?*/
["Are you asking what I thought of the restructuring?"]

Question: (3) *You found it unfair?/* or \
Response: (4) *Did I find it unfair?/*

For these echo questions, rising intonation would probably be used. As with other echo-questions and requests for confirmation, the overall pitch contour and key of the question would probably be higher than usual.

Final pitch at the transition points can help signal various social nuances which can be played out and discussed. For example, final rising pitch in (2) and (4) could signal that the speaker is unsure of the original question and is genuinely asking or appealing for confirmation — or it could signal heightened politeness. By contrast, if terminal falling pitch is used with the questions in (1) and (3), falling intonation would indicate greater certainty or assertiveness, i.e., a speaker who uses falling intonation is more expectant of an affirmative answer than if rising, questioning intonation were used. In the boss/employee role play, the employee would therefore use the rising version to indicate apprehension or fear; the boss would use the falling pattern indicative of greater certainty or self-assurance and even finality in the sense of "case closed."

As shown above with the simple examples of beginning-level dialogues, new materials will not necessarily have to be to be created for classroom use — old dialogues can be varied with various interactional twists. In keeping with the *ACTFL Oral Proficiency Guidelines* (see Chapter 5), the majority of recent elementary foreign language textbooks in the U. S. deal with topics similar to those in the examples above, and even the simplest dialogues (or partial dialogues) can be elaborated upon to allow for inclusion of different speech acts as well as for intonation practice.

This would also be an opportune time to practice the words and phrases used for the so-called back-channeling by which listeners provide feedback to their interlocutors, e.g., *yeah, I see, really, mhm, oh, is that true.* The social skills that adult learners possess are not automatically transferred to conversation in a foreign language — especially with regards to intonation — but even if they are, these are the verbal cues (or "gambits") that learners lack instinctively. If asked as genuine questions, e.g., *Really? Is that true?,* these English utterances would probably be said with high rising pitch indicating interest and urging the speaker to continue. If great intensity were used and the pitch were to drop radically, the effect could be that of cutting the other person off (*Yeah, I see*). If low intensity is used with sustained or level intonation as in *mhm*, it could signal that the listener is paying attention but does not have anything immediate to add, with the sense of "go on…," which might, indeed, be the same connotation of a native speaker's use of *mhm.*

In another gambit that is common in natural conversations, a listener may echo the speaker's utterance with partial or total repetition of the utterance or a paraphrase of it. This is often done in low key and perhaps a softer voice (murmuring). Though in a sense not discourse, it still encourages the other speaker to continue.

Also to be recommended is teaching students how to fill pauses with common hesitation phenomena in order to buy time when they have the floor or to encourage some response from the listener. In English, "you know?," "you know what I mean?," and "I mean" are commonly used in various sociolinguistic versions, as when younger speakers and speakers especially in informal contexts use phrases such as "y'know," "right?," "'kay?," and "like…". The intonation used with these fillers can vary from low key for parenthetical remarks to rising intonation with high key among certain groups of speakers.

All of these examples show possibilities for using natural conversations and interactions in teaching and practicing how intonation functions in negotiating meaning. Simply knowing the phrases and conversational structures is in itself not enough to insure a speaker's higher-level proficiency. In a role play for highlighting social nuances, however, use of intonation will serve as a learnable skill and can be something for the teacher to evaluate in terms of appropriateness — in much the same way a native speaker would evaluate and comprehend utterances in a real conversation.

As seen in many of the above examples, the general principle to be applied in adapting existing materials for intonation practice is that by altering the situation or surrounding context, the role, status, or attitudes of the speakers, and the intended or expected outcome of a conversation or interaction, occasions for various possible intonations can be made available. It is important to note that by using such role plays or situations, it is not only emotions that are being elicited; students are also being asked to use various types of discourse strategies (persuading, hedging, appeasing, etc.) in order to negotiate interactional meaning with another person (cf. Chun, 1994).

Looking toward the future: Basic principles for conceptualization and design of computer software to train intonation

As discussed briefly in Chapter 5 and then detailed in Chapters 6 and 7, if computers are used to train students in the production and perception of intonation, intonation contours and key can be displayed for them. Of particu-

lar importance is that the exercises students are asked to do contain, whenever possible, meaningful discourse-level phenomena. Using computer-generated visual displays of pitch curves, the same sentence could be spoken several times in various contexts, with different intonations expressing different pragmatic meanings. Learners could first compare native speakers' renditions with each other and then practice producing the different curves themselves. Similarly, visual images on computers of such phenomena as rhythm, linking, contraction, and other forms of reduction associated with fluent speech can also be used, as discussed in Chapter 6. An overarching principle is that the overall discourse context steer the (relative) "correctness" of the intonation, not an arbitrary normed standard of absolute correctness or appropriateness.

Four specific ways in which technology can be integrated into intonation instruction and research were suggested and discussed in Chapter 5: (1) using computers to provide learners with visualizations of their intonational patterns as specific feedback so that they can compare them to those of native speakers and go on to improve their own speech production; (2) using computers to provide learners with authentic speech and cultural input to represent the diversity of speech sounds and the great variation that exists within a language and in turn enable them to hone their perceptual abilities; (3) using computer software to facilitate, record, and analyze interactions between and among speakers; and (4) using computers for research purposes, i.e., to record students' performance, progress, and steps toward self-correction.

Below are desiderata to be borne in mind for the conceptualization and design of multimedia software for teaching discourse intonation to language learners (cf. Chun, 1998):

- perception as well as production must be trained; awareness must be heightened; attention must be focused on the holistic, overall shape and characteristics of intonation as well as on particular features, first on the part of instructors, and then by learners
- the features to be recognized and produced must include *stress* and *rhythm*, on the one hand, and *intonation* and *pitch patterns,* on the other, along with *key*
- programs must incorporate and promote practice and feedback — software must provide tutoring components, not just serve as tools for signal analysis
- programs must go beyond the sentence level and address the multiple levels of competence — *grammatical, attitudinal, discourse,* and *sociolinguistic*
- initial units may contain read or elicited speech (both for perception and production exercises and activities) to allow for practice of basic intonation patterns, but in later units, authentic speech must be used whenever possible and be used for:

- close analysis (repeated listenings and viewing of intonation curves) and
- activities requiring genuine interaction that can be recorded, processed with a signal analysis program and subsequently viewed in display
- the ultimate goal must be for learners (1) to be able to comprehend native speakers' intonation patterns, and (2) to be able to produce intonational patterns that allow them to communicate effectively and to interact successfully with others.

Conclusion

The discussion and examples in this chapter suggest how intonation can be taught as an integral part of higher-level discourse. The use of signal analysis software was proposed as a supplementary aid for teaching intonation. Computer software that digitizes speech, extracts pitch, and displays the melodic curves as well intensity levels of utterances can help students visualize the intonation patterns used in the language they are learning. Using appropriate software is also more efficient for learners than conventional audiotapes because computer programs allow for immediate retrieval and playback of audio segments without the need to rewind and locate a specific point in data as with an audiotape. The visualization and audio playback capabilities should heighten learners' ability to analyze the intonation contours of the target language in native speakers' speech and then to analyze and internalize these contours in their own speech.

As a conclusion to this book, I reiterate that the shift in emphasis in both linguistic and applied linguistic research to focus on discourse-level phenomena in language use has led to the proposition that intonation has important functions in conversation and interaction and that these functions correlate directly with the components of communicative competence we seek to develop in language learners. After defining and describing the components of intonation in Chapter 1 and the major linguistic theories of intonation in Chapter 2, I reviewed in Chapter 3 the functions of intonation that have been suggested in the literature. Four main categories of functions were then proposed, *grammatical, attitudinal, discourse, and sociolinguistic,* which correspond to the elements of communicative competence being advocated by the language acquisition profession.

Chapters 4 and 5 described the history of pronunciation teaching and research, highlighting an earlier lack of attention to intonation but a current revival of interest in the importance of suprasegmentals, particularly in the

domain of ESL (and EFL).[12] For the teaching of languages other than English, although the profession has clearly moved beyond a focus on instilling *grammatical* competence to include achievement of *discourse, interactive,* and *sociocultural* competence, the *ACTFL Guidelines* (adhered to by many in the U. S.) have not explicitly delineated the role of intonation in these types of competence. Instructional materials and approaches currently assure that some types of classroom interaction provide opportunities for learners to manage discourse, e.g., to initiate conversations, pose questions, interrupt others, ask for clarification, change the subject, and give feedback to their interlocutors. However, the role of intonation in negotiating the features of discourse management has not been specified, and I suggest that intonational components be incorporated into discourse activities in the classroom as well as in evaluations of oral proficiency. Intonation is a primary tool for both "doing things with speech" (illocutionary function) and "having things done" (perlocutionary function). By not teaching intonation, we may be keeping students from reaching complete self-expression by being able to use a wide range of speech acts, which is crucial to upper-level proficiency ratings.

In the final two chapters, Chapters 6 and 7, suggestions were presented for teaching rhythm and stress at the word, sentence, and discourse levels and for teaching intonation patterns (pitch contours and key) crucial for attaining the four components of communicative competence: *grammatical, attitudinal, discourse, and sociolinguistic.* Signal analysis software, which is becoming increasingly more accessible and is being developed by linguists and applied linguists, shows promising pedagogical applications, particularly when combined with linguistic research into the prosodic components of real interaction and conversation and with second language acquisition research into how prosody is acquired by L2 learners.

Notes

1. In addition to rhythm and pitch, a third important suprasegmental component is that of *voice quality settings*, which is said to characterize a speaker's accent. Esling and Wong (1983) describe voice quality settings as "long-term postures of the larynx, pharynx, tongue, velopharyngeal system and lips, as well as long-term laryngeal configurations reflected in the diverse phonation types [...] Voice quality settings may function linguistically, to characterize the particular language or dialect or social group to which a speaker belongs; or they may function paralinguistically, to signal mood or emotion in conversational contexts; or they may also function extralinguistically to characterize or identify the speaker" (p. 89).

2. Hirst and Di Cristo (1998) provide basic descriptions of the intonation systems of twenty languages, with characterizations of basic non-emphatic patterns as well as patterns that express modality and expressivity; focalization and contextual patterns; and stylized, formulaic patterns. The characterizations reveal that non-emphatic statements and assertions as well as the equivalents of *wh*-questions in German and Spanish generally have falling intonation patterns. However, it is interesting to note that intonation in Spanish at the sentence level is chiefly characterized by the various patterns and distribution of stressed syllables of the sentence (p. 156). In French, simple declarative sentences and neutral *wh*-questions have globally rising-falling patterns.

3. Rising terminal intonation is also used in German, French, and Spanish with yes-no questions. In French, the final rise on the last stressed syllable of a yes-no question has a greater range and a steeper slope than "continuatives" (Hirst & DiCristo, 1998, p. 202, and see below).

4. Cf. Griffiths, 1991 for a review of pausological research in L2; O'Connell and Kowal, 1983.

5. In French, yes-no tag questions end with terminators such as *oui?*, *si?*, *non?*, *hein?*, *n'est-ce pas?*. The rising-falling pattern associated with statements is used, but a subsequent rise is associated with the terminator (Di Cristo, 1998, p. 204).

6. In German, the tags *nicht* (*wahr*) 'not (true)' and *oder* 'or' are generally spoken with rising intonation (marked with ´), but in rare instances are spoken with falling intonation (`) if the expected reply is in the affirmative and the tag used is *oder doch* ('or indeed'); the symbol ˆ indicates rising-falling intonation (cf. Fox, 1984, p. 104).

Das Konzert war ˆhimmlisch, ´nicht wahr?	'The concert was heavenly, wasn't it?'
Sie kommen ´mit, ´oder?	'You're coming along, aren't you?'
Sie kommen nicht `mit, `oder doch?	'You're not coming along, or are you (after all)?'

7. Hirst and Di Cristo (1998) describe the use of intonation to signal attitude, modality and expressivity in twenty languages. Rising-falling intonation is characterized in German, for example, as indicating "certainty" and "obviousness," whereas rising or falling-rising intonation is said to connote "deference," "politeness," or "uncertainty." In French, the attitudinal pitch pattern most frequently used, according to Hirst and Di Cristo, can be termed "implicative" and generally indicates that the speaker is insisting on a fact which seems particularly evident to him or her (p. 208). In Spanish, general intonation patterns of questions are said to potentially undergo alterations to show different secondary meanings such as "courtesy," "repetition," "confirmation," or "order."

8. In German, the functions of prosody are characterized by complex interactions between word order and sentence stress. Sentence stress is influenced by the predominant SOV word order, which reduces the tendency to sentence-final stress in contrast to English (see Gibbon, 1998, p. 94). German is a highly inflected language with relatively free word order. Word order is a major focalization device that is used for topicalization and for marking new information in utterances (see Gibbon, 1998, p. 81). Therefore, an instructor of German would have to incorporate these considerations into exercises that show the use of

word order to mark focus either instead of, or in addition to, using intonation. It has been suggested by Schubiger (1958, 1980) and Ehlich (1979) that the functional load on prosody may be lower in German than in English. To my knowledge, however, no follow-up studies are available.

9. Gibbon (1998) states that in German, "emphatic or emotive accents are not necessarily different in kind from other accents, but basically just have 'more of everything'" (p. 91). They have broader frequency modulation and more extreme syllable lengthening than non-emphatic accents. In French, Di Cristo (1998) asserts that focus accents are used either for intensification or for contrast: a syllable or word can be highlighted by extra pitch prominence, and contrast is characterized by a global rising-falling pitch pattern (p. 209).

10. Here again, of course, different languages work differently. In German, rising intonation is used for a range of functions, from signaling questions to indicating connotations of deference, politeness or uncertainty (see Gibbon, 1998, p. 88).

11. The stylized pitch contour, often termed "chanted," "flat" or "call" contours, is used in German to signal the opening, sustaining, and closing of a channel of communication (Gibbon, 1998, p. 91). It is used not only specifically in calls, but in a range of other functions, such as greeting, leave-taking, thanking, and (unlike in other languages) in signaling the need for discourse repairs caused by mishearing.

12. In fact, Murphy (1997) described phonology courses offered by TESOL programs offering an M.A. degree in the U.S. and reported that the top three suggestions for improvement were (1) availability of more recorded samples of ESL/L2 learners' speech for assessment and analysis, (2) access to improved computer software, and (3) more emphasis on suprasegmentals.

Bibliography

Abberton, E., & Fourcin, A. J. (1975). Visual feedback and the acquisition of intonation. In E. H. Lenneberg & E. Lenneberg (Eds.), *Foundations of language development*, (pp. 157–165). New York: Academic Press.

ACTFL Proficiency Guidelines. (1986). In A. Omaggio, *Teaching language in context: Proficiency-oriented instruction* (pp. 501–511). Boston: Heinle.

Acton, W. (1984). Changing fossilized pronunciation. *TESOL Quarterly*, *18*, 71–85.

Adams, C. (1979). *English speech rhythm and the foreign learner.* The Hague: Mouton.

Alcoba, S. & Murillo, J. (1998). Intonation in Spanish. In D. Hirst & A. Di Cristo (Eds.), *Intonation systems: A survey of twenty languages* (pp. 152–166). Cambridge: Cambridge University Press.

Altmann, H. (Ed.). (1988). *Intonationsforschungen.* Tübingen: Niemeyer.

Anderson, K. O. (1979). Applied prosodic analysis: A pedagogical model for English and German intonation. In H. Hollien & P. Hollien (Eds.), *Current issues in the phonetic sciences* (Part I) (pp. 227–234). Amsterdam: John Benjamins.

Anderson-Hsieh, J. (1990). Teaching suprasegmentals to international teaching assistants using field-specific materials. *English for Specific Purposes*, *9*, 195–214.

Anderson-Hsieh, J. (1992). Using electronic visual feedback to teach suprasegmentals. *System*, *20*, 51–62.

Anderson-Hsieh, J. (1994). Interpreting visual feedback on suprasegmentals in computer assisted pronunciation instruction. *CALICO Journal*, *11*(4), 5–22.

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of non native pronunciation and deviance in segmentals, prosody and syllable structure. *Language Learning*, *42*, 529–555.

Anderson-Hsieh, J., & Koehler, K. (1988). The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning*, *38*, 561–605.

Anderson-Hsieh, J., & Venkatagiri, H. (1994). Syllable duration and pausing in the speech of Chinese ESL speakers. *TESOL Quarterly*, *28*(4), 807–812.

Armstrong, L. E., & Ward, I. C. (1926). *Handbook of English intonation.* Leipzig and Berlin: Teubner.

Arnold, G. F., & Gimson, A. C. (1973). *English pronunciation practice* (2nd ed.). London: University of London Press.

Arteaga, D. L., (2000). Articulatory phonetics in the first-year Spanish classroom. *Modern Language Journal*, *84*, 339–354.

Austin, J. L. (1962). *How to do things with words.* Cambridge, MA: Harvard University Press.

Avery, P., & Ehrlich, S. (Eds.). (1987). *The teaching of pronunciation: An introduction for Teachers of English as a Second Language. TESL Talk, 17(1)*, Ontario, Canada: Ministry of Citizenship and Culture.

Avery, P., & Ehrlich, S. (1992). *Teaching American English pronunciation.* Oxford: Oxford University Press.

Bachman, L., & Palmer, A. S. (1984). Some comments on the terminology of language testing. In C. Rivera (Ed.), *Communicative competence approaches to language proficiency assessment: Research and applications*. Clevedon: Multilingual Matters.

Bachman, L., & Savignon, S. J. (1986). The evaluation of communicative language proficiency: A critique of the ACTFL Oral Interview. *Modern Language Journal*, *70*, 380–90.

Bailey, C. (1978). Native accent and learning English as a foreign language. *IRAL*, *16*, 229–40.

Ballmer, T. T. (1980). The role of pauses and suprasegmentals in a grammar. In H. Dechert & M. Raupach (Eds.), *Temporal variables in speech* (pp. 211–220). Mouton: The Hague.

Barrutia, R., & Terrell, T. D. (1982). *Fonética y fonología españolas*. New York: Wiley.

Beebe, L. M. (1980). Sociolinguistic variation and style shifting in second language acquisition. *Language Learning*, *30*, 433–47.

Bell, A. (1978). Syllabic consonants. In J. H. Greenberg (Ed.), *Universals of human language* (pp. 153–201). Stanford: Stanford University Press.

Benware, W. A. (1986). *Phonetics and phonology of Modern German*. Washington, D. C.: Georgetown University Press.

Berkovits, R. (1984a). Duration and fundamental frequency in sentence-final intonation. *Journal of Phonetics*, *12*, 255–265.

Berkovits, R. (1984b). A perceptual study of sentence-final intonation. *Language and Speech*, *27*, 291–308.

Bieritz, W. D. (1974). Phonologische Interferenzen: französische Aussprachefehler deutscher Studenten. *IRAL*, *9*, 93–130.

Blakemore, D. (1988). The organization of discourse. In F. J. Newmeyer (Ed.), *Linguistics: The Cambridge survey* Vol. IV (pp. 229–250). Cambridge: Cambridge University Press.

Bleiching, D. (1992). Prosodisches Wissen im Lexikon. In G. Gürz (Ed.), *KONVENS 92*, Nürnberg (pp. 59–68). Berlin: Springer.

Bloomfield, L. (1933). *Language*. Chicago: University of Chicago Press.

Bluhme, H., & Burr, R. (1971). An audio-visual display of pitch for teaching Chinese tones. *Studies in Linguistics*, *22,* 51–57.

Bolinger, D. (1951). Intonation: Levels versus configurations. *Word*, *7*, 199–210.

Bolinger, D. (1957a). On certain functions of accents A and B. *Litera*, *4*, 80–89.

Bolinger, D. (1957b). Disjuncture as a cue to constructs. *Word*, *13*, 246–255.

Bolinger, D. (1958). A theory of pitch accent in English. *Word*, *14*, 109–149.

Bolinger, D. (1978). Intonation across languages. In J. Greenberg (Ed.), *Universals of Human Language* Vol. 2 (pp. 473–519). Stanford: Stanford University Press.

Bolinger, D. (1981). *Two kinds of vowels, two kinds of rhythm*. Bloomington: Indiana University Linguistics Club.

Bolinger, D. (1985). The inherent iconism of intonation. In J. Haiman (Ed.), *Iconicity in syntax* (pp. 97–108). Amsterdam: John Benjamins.

Bolinger, D. (1986). *Intonation and its parts: Melody in spoken English*. Stanford: Stanford University Press.

Bolinger, D. (1989). *Intonation and its uses: Melody in grammar and discourse*. Stanford: Stanford University Press.

Bolinger, D. (1998). Intonation in American English. In D. Hirst & A. Di Cristo (Eds.), *Intonation systems: A survey of twenty languages* (pp. 45–55). Cambridge: Cambridge

University Press.
Bongaerts, T., van Summeren, C., Planken, B., & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition, 19*(4), 447–465.
Borzone de Manrique, A. M., & Signorini, A. (1983). Segmental duration and rhythm in Spanish. *Journal of Phonetics, 11*(2), 117–28.
Botinis, A., Granström, B., & Möbius, B. (2001). Developments and paradigms in intonation research. *Speech Communication, 33,* 263–296.
Bowen, J. D. (1975). *Patterns of English pronunciation.* New York: Newbury House.
Boyce, S., & Menn, L. (1979). Peaks vary, endpoints don't: Implications for intonation theory. *Proceedings of the Fifth Annual Meeting of the Berkeley Linguistics Society,* (pp. 373–84). Berkeley: Berkeley Linguistics Society.
Boyle, J. P. (1987). Perspectives on stress and intonation in language learning. *System, 15,* 189–195.
Bradford, B. (1988). *Intonation in context.* Cambridge: Cambridge University Press.
Brazil, D. (1975). *Discourse intonation I.* Birmingham: English Language Research Monographs.
Brazil, D. (1978). *Discourse intonation II.* Birmingham: English Language Research Monographs.
Brazil, D. (1985). *The communicative value of intonation in English.* Birmingham: Bleak House Books and English Language Research.
Brazil, D. (1994) *Pronunciation for advanced learners of English.* Cambridge: Cambridge University Press.
Brazil, D., Coulthard, M., & Johns, C. (1980). *Discourse intonation and language teaching.* London: Longman.
Breiner-Sanders, K. E., Lowe, P. Jr., Miles, J., & Swender, E. (1999). ACTFL Proficiency Guidelines — Speaking, Revised. *Foreign Language Annals, 33*(1), 13–17.
Breitung, H. (Ed.) (1994). *Phonetik — Intonation — Kommunikation.* Munich: Goethe-Institut.
Broselow, E., Hurtig, R. R., & Ringen, C. (1987). The perception of second language prosody. In G. Ioup & S. H. Weinberger (Ed.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 350–364). Cambridge, MA: Newbury House.
Brown, A. (1988). Functional load and the teaching of pronunciation. *TESOL Quarterly, 22,* 593–606.
Brown, A. (1991). *Teaching English pronunciation: A book of readings.* London and New York: Routledge.
Brown, G. (1977). *Listening to spoken English.* London: Longman.
Brown, G. (1983). Prosodic structure and the given/new distinction. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and measurements* (pp. 67–77). Berlin: Springer.
Brown, G., Currie, K. L., & Kenworthy, J. (1980). *Questions of intonation.* London: Helm.
Brown, G., & Yule, G. (1983). *Teaching the spoken language: An approach based on the analysis of conversational English.* Cambridge: Cambridge University Press.
Brown, R. (1990). A maverick vowel: The notorious schwa. *Hispania, 73,* 1158–1161.
Bygate, M. (1987). *Speaking.* Oxford: Oxford University Press.

Byrnes, H. (1987). Speech as process. *Foreign Language Annals*, *20*, 301–310.
Byrnes, H., & Canale, M. (Eds.). (1986). *Defining and developing proficiency. Guidelines, implementations and concepts.* Lincolnwood, IL: National Textbook Co.
Byrnes, H., Thompson, I., & Buck, K. (Eds.). (1989). *ACTFL Oral Proficiency Interview. Tester training manual.* New York: ACTFL.
Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. Richards & R. Schmidt (Eds.), *Language and Communication* (pp. 2–27). London: Longman.
Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, *1*, 1–47.
Caplan, D. (1972). Clause boundaries and recognition latencies for words in sentences. *Perception and Psychophysics*, *12*, 73–76.
Capretz, P. (1987). *French in Action.* New Haven: Yale University Press.
Catford, J. C. (1987). Phonetics and the teaching of pronunciation. In J. Morley (Ed.), *Current perspectives on pronunciation* (pp. 87–100). Washington, D. C.: TESOL (Teachers of English to Speakers of Other Languages).
Celce-Murcia, M. (Ed.). (1991). *Teaching English as a Second or Foreign Language* (2nd ed.). New York: Newbury House.
Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages.* Cambridge: Cambridge University Press.
Celce-Murcia, M. (1987). Teaching pronunciation as communication. In J. Morley (Ed.), *Current perspectives on pronunciation* (pp. 1–12). Washington, D. C.: TESOL (Teachers of English to Speakers of Other Languages).
Celce-Murcia, M., & Goodwin, J. M. (1991). Teaching pronunciation. In M. Celce-Murcia (Ed.), *Teaching English as a Second or Foreign Language* (2nd ed.) (pp. 136–153). New York: Newbury House.
Chafe, W. (1972). Discourse structure and human knowledge. In J. B. Carroll & R. O. Freedle (Eds.), *Language comprehension and the acquisition of knowledge* (pp. 41–69). New York: John Wiley.
Chafe, W. (1974). Language and consciousness. *Language*, *50*, 111–133.
Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics and points of view. In C. N. Li (Ed.), *Subject and topic* (pp. 25–55). New York: Academic Press.
Chafe, W. (1979). The flow of thought and the flow of language. In T. Givón (Ed.), *Discourse and syntax* (pp. 159–181). New York: Academic Press.
Chafe, W. (1980). The deployment of consciousness in the production of a narrative. In W. Chafe (Ed.), *The Pear Stories: Cognitive, cultural, and linguistic aspects of narrative production* (pp. 9–50). Norwood, NJ: Ablex.
Chafe, W. (1987). Cognitive constraints on information flow. In R. Tomlin (Ed.), *Coherence and grounding in discourse* (pp. 21–51). Amsterdam: Benjamins.
Chafe, W. (1993). Prosodic and functional units of language. In J. A. Edwards & M. D. Lampert (Eds.), *Talking data: Transcription and coding methods for language research* (pp. 33–42). Hillsdale, NJ: Lawrence Erlbaum.
Chan, M. (1987). *Phrase by phrase.* Englewood Cliffs, NJ: Prentice-Hall.

Chaudron, C. (1985). Intake: On models and methods for discovering learners' processing of input. *Studies in Second Language Acquisition*, *7*, 1–14.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper.

Chun, D. M. (1988a). The neglected role of intonation in communicative competence and proficiency. *Modern Language Journal*, *72*, 295–303.

Chun, D. M. (1988b). Teaching intonation as part of communicative competence: Suggestions for the classroom. *Die Unterrichtspraxis*, *21*, 81–88.

Chun, D. M. (1989). Teaching tone and intonation with microcomputers. *CALICO Journal*, *7*, 21–46.

Chun, D. M. (1991). The state of the art in teaching pronunciation. In J. E. Alatis (Ed.), *Georgetown University Round Table on Languages and Linguistics 1991–Linguistics and language pedagogy: The state of the art* (pp. 179–193). Washington, D. C.: Georgetown University Press.

Chun, D. M. (1994). Using computer networking to facilitate the acquisition of interactional competence. *System*, *22*(1), 17–31.

Chun, D. M. (1998). Signal analysis software for teaching discourse intonation. *Language Learning and Technology*, *2*(1), 61–77, http://llt.msu.edu/vol2num1/article4/index.html.

Clark, J. L. D. (1986). Development of a tape-mediated, ACTFL/ILR scale-based test of Chinese speaking proficiency. In C. W. Stansfield (Ed.), *Technology and language testing* (pp. 129–146). Washington. D. C.: TESOL (Teachers of English to Speakers of Other Languages).

Clark, J., & Yallop, C. (1990). *An introduction to phonetics and phonology*. Oxford: Basil Blackwell.

Clennell, C. (1997). Raising the pedagogic status of discourse intonation teaching. *ELT Journal, 51*(2), 117–125.

Cohen, A., & 't Hart, J. (1967). On the anatomy of intonation. *Lingua*, *19*, 177–192.

Conrad, B. K. (1991). *The relationship between empathy and pronunciation ability: A study of elementary-level college students of German*. Unpublished doctoral dissertation, University of Texas at Austin, Austin, TX.

Cook, V. J. (1968). *Active intonation*. London: Longman.

Cook, V. J. (1979). *Using intonation*. London: Longman.

Cooper, W. E., & Sorensen, J. M. (1981). *Fundamental frequency in sentence production*. Berlin: Springer.

Corder, S. P. (1967). The significance of learners' errors. *IRAL*, *5*, 161–169.

Coulthard, M. (1977). *An introduction to discourse analysis* (2nd ed., 1985). London: Longman.

Coulthard, M., & Brazil, D. (1981). The place of intonation in the description of interaction. In D. Tannen (Ed.), *Analyzing discourse: Text and talk* (pp. 94–112). Washington, D. C.: Georgetown University Press.

Coulthard, M., & Montgomery, M. (Eds.). (1981). *Studies in discourse analysis*, London: Routledge & Kegan Paul.

Couper-Kuhlen, E. (1983). Intonatorische Kohäsion. Eine makroprosodische Untersuchung. *Zeitschrift für Literaturwissenschaft und Linguistik*, *49*, 74–100.

Couper-Kuhlen, E. (1986). *An introduction to English prosody*. Tübingen: Niemeyer.
Couper-Kuhlen, E. (1996). The prosody of repetition: On quoting and mimicry. In E. Couper-Kuhlen & M. Selting (Eds.), *Prosody in conversation* (pp. 366–405). Cambridge: Cambridge University Press.
Couper-Kuhlen, E. (1997). "Intonation" and discourse: Views from within. Paper presented at the UCSB Linguistics Colloquium, Santa Barbara, CA.
Couper-Kuhlen, E., & Selting, M. (Eds.). (1996a). *Prosody in conversation*. Cambridge: Cambridge University Press.
Couper-Kuhlen, E., & Selting, M. (1996b). Towards an interactional perspective on prosody and a prosodic perspective on interaction. In E. Couper-Kuhlen & M. Selting (Eds.), *Prosody in conversation* (pp. 11–56). Cambridge: Cambridge University Press.
Crombie, W. (1985). *Process and relation in discourse and language learning*. Oxford: Oxford University Press.
Crothers, J. (1978). Typology and universals of vowel systems. In J. H. Greenberg (Ed.), *Universals of human language* (pp. 93–152). Stanford: Stanford University Press.
Cruttenden, A. (1981). Falls and rises: Meanings and universals. *Journal of Linguistics*, *17*, 77–91.
Cruttenden, A. (1986). *Intonation*. Cambridge: Cambridge University Press.
Cruttenden, A. (1997). *Intonation* (2nd ed.). Cambridge: Cambridge University Press.
Cruz-Ferreira, M. (1987). Non-native interpretive strategies for intonational meaning: An experimental study. In A. James & J. Leather (Eds.), *Sound patterns in second language acquisition* (pp. 103–120). Dordrecht: Foris.
Crystal, D. (1969). *Prosodic systems and intonation in English*. Cambridge: Cambridge University Press.
Crystal, D. (1975). *The English tone of voice*. London: Edward Arnold.
Crystal, D. (1985). *A dictionary of linguistics and phonetics*. Oxford: Basil Blackwell.
Crystal, D., & Quirk, R. (1964). *Systems of prosodic and paralinguistic features in English*. The Hague: Mouton.
Currie, K., & Yule, G. (1982). A return to fundamentals in the teaching of intonation. *IRAL*, *20*, 228–32.
Cutler, A. (1983). Speakers' conceptions of the function of prosody. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and measurements* (pp. 79–91). Berlin: Springer.
Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, *40*(2), 141–201.
Cutler, A., & Ladd, D. R. (Eds.). (1983). *Prosody: Models and measurements*, Berlin: Springer.
Cutler, A., & Pearson, M. (1986). On the analysis of prosodic turn-taking cues. In C. Johns-Lewis (Ed.), *Intonation in discourse* (pp. 139–55). San Diego: College Hill.
Dalbor , J. B. (1980). *Spanish pronunciation: Theory and practice* (2nd ed.). New York: Holt, Rinehart, and Winston.
Dalbor , J. B. (1997). *Spanish pronunciation: Theory and practice* (3rd ed.). New York: Holt, Rinehart, and Winston.
Dalton, C., & Seidlhofer, B. (1994). *Pronunciation*. Oxford: Oxford University Press.
Dansereau, D. (1995). Phonetics in the beginning and intermediate oral-proficiency-oriented French classroom. *French Review*, *68*(4), 638–651.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, *11*(1), 51–62.

Dauer, R. M. (1993). *Accurate English: A complete course in pronunciation.* Englewood Cliffs, NJ: Regents/Prentice Hall.

de Bot, K. (1980). The role of feedback and feedforward in the teaching of pronunciation: An overview. *System*, *8*, 35–45.

de Bot, K. (1981a). Visual feedback of intonation, an experimental approach. In B. Sigurd & J. Svartvik (Eds.), *Proceedings of the Sixth AILA Congress* (pp. 243–245). Lund: University of Lund.

de Bot, K. (1981b). Intonation teaching and pitch control. *ITL Review of Applied Linguistics*, *52*, 31–42.

de Bot, K. (1983). Visual feedback of intonation I: Effectiveness and induced practice behavior. *Language and Speech*, *26*(4), 331–350.

de Bot, K., & Mailfert, K. (1982). The teaching of intonation: Fundamental research and classroom applications. *TESOL Quarterly*, *16*, 71–77.

Dechert, H. W., & Raupach, M. (1987). Prosodic patterns of proceduralized speech in second and first language narratives. In A. James & J. Leather (Eds.), *Sound patterns in second language acquisition* (pp. 81–102). Dordrecht: Foris.

Delattre, P. (1963). Comparing the prosodic features of English, German, Spanish, and French. *International Review of Applied Linguistics*, *1*, 193–210.

Delattre, P. (1965). Some characteristics of German intonation for the expression of continuation and finality. *Phonetica*, *13*, 134–161.

Delattre, P. (1966). A comparison of syllable length conditioning among languages. *International Review of Applied Linguistics*, *4*, 183–198.

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, *19*(1), 1–16.

Derwing, T. M., Munro, M. J., & Wiebe, G. (1997). Pronunciation instruction for 'fossilized' learners — Can it help? *Applied Language Learning*, *8*, 217–235.

Dickerson, W. B. (1989). *Stress in the speech stream.* Urbana, IL: University of Illinois Press.

Di Cristo, A. (1998). Intonation in French. In D. Hirst & A. Di Cristo (Eds.), *Intonation systems: A survey of twenty languages* (pp. 195–218). Cambridge: Cambridge University Press.

DuBois, J. W., Schuetze-Coburn, S., Paolino, D., & Cumming, S. (1992). *Discourse transcription.* Santa Barbara: Santa Barbara Papers in Linguistics.

DuBois, J. W., Schuetze-Coburn, S., Cumming, S., & Paolino, D. (1993). Outline of discourse transcription. In J. A. Edwards & M. D. Lampert (Eds.), *Talking data: Transcription and coding in discourse research* (pp. 45–89). Hillsdale, NJ: Erlbaum.

Duncan, S. J. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, *23*, 283–292.

Duncan, S. J. (1973). Towards a grammar for dyadic conversation. *Semiotica*, *9*(1), 29–46.

Duncan, S. J. (1974). On the structure of speaker-auditor interaction during speaking turns. *Language in Society*, *3*(2), 161–180.

Duncan, S. J. (1975). Interaction units during speaking turns in dyadic, face-to-face conversations. In A. Kendon, R. M. Harris, & M. R. Key (Eds.), *Organization of behavior in*

face-to-face interaction (pp. 199–213). The Hague: Mouton.
Duncan, S. J., & Fiske, D. (1977). *Face-to-face interaction: Research, methods, and theory.* Hillsdale, NJ: Erlbaum.
Ecklund, C. L., & Wiese, P. (1981). French accent through video analysis. *Foreign Language Annals*, *14*, 17–23.
Edmondson, W., House, J., Kasper, G., & Stemmer, B. (1984). Learning the pragmatics of discourse: A project report. *Applied Linguistics*, *5*, 113–125.
Ehlich, K. (1979). Formen und Funktionen von 'HM' eine phonologisch-pragmatische Analyse. In H. Weydt (Ed.), *Die Partikeln der deutschen Sprache* (pp. 503–517). Berlin: de Gruyter.
Elliott, A. R. (1995a). Field independence/dependence, hemispheric specialization, and attitude in relation to pronunciation accuracy in Spanish as a foreign language. *The Modern Language Journal*, *79*(3), 357–371.
Elliott, A. R. (1995b). Foreign language phonology: Field independence, attitude, and the success of formal instruction in Spanish pronunciation. *Modern Language Journal*, *79*(4), 530–542.
Elliott, A. R. (1997). On the teaching and acquisition of pronunciation within a communicative approach. *Hispania*, *80*(1), 95–108.
English, S. L. (1988). *Say it clearly: Exercises and activities for pronunciation and oral communication.* New York: Collier Macmillan.
Enkvist, N. E., & Nordström, H. (1977). On textual aspects of intonation in Finland-Swedish newscasts. Unpublished manuscript, Text Linguistics Research Group, Abo Akademi.
Esling, J. H. (1991). Researching the effects of networking: Evaluating the spoken and written discourse generated by working with CALL. In P. Dunkel (Ed.), *Computer-assisted language learning and testing: Research issues and practice* (pp. 111–131). New York: Newbury House/Harper Collins.
Esling, J. H. (1992). Speech technology systems in applied linguistics instruction. In M. C. Pennington & V. Stevens (Eds.), *Computers in applied linguistics: An international perspective* (pp. 244–272). Clevedon: Multilingual Matters.
Esling, J. H., & Wong, R. F. (1983). Voice quality settings and the teaching of pronunciation. *TESOL Quarterly*, *17*, 89–95.
Essen, O. v. (1964). *Grundzüge der hochdeutschen Satzintonation.* Ratingen: A. Henn.
Esser, J. (1978). Contrastive intonation of German and English: Problems and some results. *Phonetica*, *35*, 41–55.
Esser, J. (1988). *Comparing reading and speaking intonation.* Amsterdam: Rodopi.
Feld, E. S., Schumann, W., & von Nardroff, E. (1973). *Anfang und Fortschritt: An introduction to German* (2nd ed.). New York: Macmillan.
Feldman, D. M. (1973). Measuring auditory discrimination of suprasegmental features in Spanish. *IRAL*, *11*, 195–209.
Feldman, D. M. (1977). Instrumental methods for the teaching of English pronunciation. In H. P. Kelz (Ed.), *Phonetische Grundlagen der Ausspracheschulung.* Hamburg: Buske.
Felix, S. W. (1978). *Linguistische Untersuchungen zum natürlichen Zweitsprachenerwerb.* München: Wilhelm Fink.

Felix, S. W. (Ed.). (1980). *Second language development: Trends and issues.* Tübingen: Narr.
Felix, S. W., & Wode, H. (Eds.). (1983). *Language development at the crossroads.* Tübingen: Narr.
Féry, C. (1993). *German intonational patterns.* Tübingen: Niemeyer.
Fidelman, C. (1992). In the German/French Body [Computer Program]. Boston: Northeastern University.
Fidelman, C., & Keller, E. (1992). SpeechLab [Computer Program]. Boston: Northeastern University.
Fischer, L. B. (1986a). The use of Visi-Pitch in the analysis of Chinese language suprasegmentals. Pine Brook, NJ: Kay Elemetrics Corporation.
Fischer, L. B. (1986b). The use of audio/visual aids in the teaching and learning of French. Pine Brook, NJ: Kay Elemetrics Corporation.
Fischer, W. B., & Richardson, P. N. (1989). *Wie, bitte?* New York: John Wiley and Sons.
Flege, J. E. (1981). The phonological basis of foreign accent: A hypothesis. *TESOL Quarterly*, *15*, 443–455.
Flege, J. E., & Port, R. (1981). Cross-language phonetic interference from Arabic to English. *Language and Speech*, *24*, 125–138.
Ford, C. E., & Thompson, S. A. (1996). Interactional units in conversation: Syntactic, intonational, and pragmatic resources for the management of turns. In E. Ochs, E. A. Schegloff, & S. A. Thompson (Eds.), *Interaction and grammar* (pp. 134–184). New York: Cambridge University Press.
Fourcin, A. J., & Abberton, E. (1971). First applications of a new laryngograph. *Medical and Biological Illustration*, *21*, 172–182.
Fox, A. (1984). *German intonation: An outline.* Oxford: Clarendon Press.
Fries, C. C. (1953). *Patterns of English sentences: Nineteen lessons from an intensive course in English.* Ann Arbor: University of Michigan, English Language Institute.
Fries, C. C. (1954). *English pronunciation: Exercises in sound segments, intonation, and rhythm.* Ann Arbor: University of Michigan, English Language Institute.
Frommer, J. G., & Weitz, M. C. (1981). 'Sound' learning: Using taped interviews in foreign language courses. *The French Review*, *55*, 233–240.
Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, *1*, 126–152.
Fudge, E. (1984). *English word-stress.* London: George Allen and Unwin.
Gårding, E., Botinis, A., & Touati, P. (1982). A comparative study of Swedish, Greek and French intonation. *Working Papers, Lund University Department of Linguistics*, *22*, 137–52.
Gårding, E., & Abramson, A. S. (1965). A study of the perception of some American English intonation contours. *Studia Linguistica*, *19*, 61–79.
Geluykens, R. (1987). Intonation and speech act type: An experimental approach to rising intonation in queclaratives. *Journal of Pragmatics*, *11*, 483–94.
Gibbon, D. (1976). *Perspectives of intonation analysis.* Bern: Herbert Lang.
Gibbon, D. (1998). Intonation in German. In D. Hirst & A. Di Cristo (Eds.), *Intonation systems* (pp. 78–95). Cambridge: Cambridge University Press.
Gibbon, D., & Richter, H. (Eds.). (1984). *Intonation, accent and rhythm: Studies in discourse phonology*, Berlin: W. de Gruyter.

Gibbon, D., & Selting, M. (1983). Intonation und die Strukturierung eines Diskurses. *Zeitschrift für Literaturwissenschaft und Linguistik*, *49*, 53–73.

Gierut, J. A., & Dinnsen, A. (1987). On predicting ease of phonological learning. *Applied Linguistics*, *8*(4), 35–57.

Gilbert, J. B. (1978). Gadgets: Nonverbal tools for teaching pronunciation. *CATESOL Occasional Papers*, *4*, 68–78.

Gilbert, J. B. (1984). *Clear speech: Pronunciation and listening comprehension in American English* (2nd ed., 1993). Cambridge: Cambridge University Press.

Gimson, A. C. (1980). *An introduction to the pronunciation of English* (3rd ed.). London: Edward Arnold.

Goldbeck, T., & Sendlmeier, W. (1988). Wechselbeziehungen zwischen Satzmodalität und Akzentuierung in satzfinaler Position bei der Realisierung von Intonationskonturen. In H. Altmann (Ed.), *Intonationsforschungen* (pp. 305–321). Tübingen: Niemeyer.

Goldsmith, J. (1990). *Autosegmental and metrical phonology*. Oxford: B. Blackwell.

Griffiths, R. (1990). Speech rate of NNS comprehension: A preliminary study in time-benefit analysis. *Language Learning*, *40*(3), 311–336.

Griffiths, R. (1991). Pausological research in an L2 context: A rationale and review of selected studies. *Applied Linguistics*, *12*(4), 345–364.

Grosser, W. (1997). On the acquisition of tonal and accentual features of English by Austrian learners. In J. Leather & A. James (Eds.), *Second-language speech (pp.* 211–228). Berlin: Mouton de Gruyter.

Grover, C., Jamieson, D., & Dobrovolsky, M. (1987). Intonation in English, French, and German: Perception and production. *Language and Speech*, *30*(3), 277–295.

Gumperz, J. (1982). *Discourse strategies*. Cambridge: Cambridge University Press.

Gussenhoven, C. (1987). Focus, mode and the nucleus. In C. Gussenhoven, D. Bolinger, & C. Keijsper (Eds.), *On accent* (pp. 1–50). Bloomington: Indiana University Linguistics Club.

Gussenhoven, C., Bolinger, D., & Keijsper, C. (1987). *On accent*. Bloomington: Indiana University Press.

Gutknecht, C. (1975). The function of intonation in the teaching of English. In C. Gutknecht (Ed.), *Contributions to applied linguistics* (pp. 117–131). Frankfurt: Peter Lang.

Gutknecht, C. (1978). Intonation and language learning: The necessity for an integrative approach. *Studies in Second Language Acquisition*, *1*(2), 25–36.

Guzmán, A. (1992). Entonación y ritmo en la communicación oral. *Hispania*, *75*, 209–213.

Hadding-Koch, K., & Studdert-Kennedy, M. (1965). Intonation contours evaluated by American and Swedish test subjects. In *Proceedings of the 5th International Congress of Phonetic Sciences* (pp. 326–331). Basel: S. Karger.

Hagen, S. A. (1988). *Sound advice*. Englewood Cliffs, NJ: Prentice-Hall.

Hagen, S. A., & Grogan, P. E. (1992). *Sound advantage: A pronunciation book*. Englewood Cliffs, NJ: Regents/Prentice Hall.

Hahn, L., & Dickerson, W. (1999). *Speechcraft: Discourse pronunciation for advanced learners*. Ann Arbor: The University of Michigan Press.

Hall, C. (1992). *Modern German pronunciation*. Manchester and New York: Manchester University Press.

Halliday, M. A. K. (1963). The tones of English. *Archivum Linguisticum*, *15*, 1–28.
Halliday, M. A. K. (1964). Intonation in English grammar. *Transactions of the Philological Society*, pp. 143–169.
Halliday, M. A. K. (1967a). *Intonation and grammar in British English*. The Hague: Mouton.
Halliday, M. A. K. (1967b). Notes on transitivity and theme in English, Part 2. *Journal of Linguistics*, *3*, 199–244.
Halliday, M. A. K. (1970). *A course in spoken English: Intonation*. London: Oxford University Press.
Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
Hatch, E. (1992). *Discourse and language education*. Cambridge: Cambridge University Press.
Helmke, B., & Wu, Y.-S. (1980). Individual differences and foreign language pronunciation achievement. *Revue de Phonétique Appliquée*, *53*, 25–34.
Hengstenberg, P. (1980). *Suprasegmentalia und Aspekte ihrer Vermittlung in sprachlichen Lehr- und Lernprozessen*. Tübingen: Narr.
Henning, W. A. (1966). Discrimination training and self-evaluation in the teaching of pronunciation. *IRAL*, *4*, 7–17.
Henrichsen, L., Green, B., Nishitani, A., & Bagley, C. L. (1999). *Pronunciation matters*. Ann Arbor: The University of Michigan Press.
Hewings, M. (1987). Intonation and feedback in the EFL classroom. In M. Coulthard (Ed.), *Discussing discourse* (pp. 221–235). Birmingham: English Language Research, University of Birmingham.
Hewings, M. (Ed.) (1990). *Papers in discourse intonation*. Birmingham: English Language Research, University of Birmingham.
Hewings, M., & Goldstein, S. (1999). Pronunciation plus — Practice through interaction. Cambridge: Cambridge University Press.
Hieke, A. E. (1981). Audio-lectal practice and fluency acquisition. *Foreign Language Annals*, *14*, 189–194.
Hieke, A. E. (1984). Linking as a marker of fluent speech. *Language and Speech*, *17*, 343–354.
Hieke, A. E. (1985). A componential approach to oral fluency evaluation. *Modern Language Journal*, *69*, 135–142.
Higgs, T. V. (Ed.). (1984). Teaching for proficiency: The organizing principle. Lincolnwood, IL: National Textbook Co.
Hill, C., & Beebe, L. M. (1980). Contraction and blending: The use of orthographic clues in teaching pronunciation. *TESOL Quarterly*, *14*(3), 299–323.
Hirschberg, J., & Litman, D. (1987). Now let's talk about 'now': Identifying cue phrases intonationally. In Association for Computational Linguistics, *Proceedings of the Twenty-fifth Annual Meeting* (pp. 163–171). Stanford: ACL.
Hirschberg, J., & Pierrehumbert, J. (1986). The intonational structuring of discourse. In Association for Computational Linguistics, *Proceedings of the Twenty-fourth Annual Meeting* (pp. 136–144). New York: ACL.
Hirschfeld, U., & Stock, E. (Eds.) (2000). *Phonothek interaktiv* [CD-ROM]. Berlin/Munich: Langenscheidt.
Hirst, D. (1998). Intonation in British English. In D. Hirst & A. Di Cristo (Eds.), *Intonation systems: A survey of twenty languages* (pp. 56–77), Cambridge: Cambridge University Press.

Hirst, D., & Di Cristo, A. (1998). *Intonation systems: A survey of twenty languages*. Cambridge: Cambridge University Press.

Hurley, D. S. (1992). Issues in teaching pragmatics, prosody, and non-verbal communication. *Applied Linguistics*, *13*(3), 259–281.

Hymes, D. (1967). On communicative competence [Unpublished manuscript].

Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics* (pp. 269–293). Harmondsworth, UK: Penguin Books.

Ioup, G., & Tansomboon, A. (1987). The acquisition of tone: a maturational perspective. In G. Ioup & S. H. Weinberger, (Eds.) *Interlanguage phonology: The acquisition of a second language sound system* (pp. 333–349). Cambridge, MA: Newbury House.

Ioup, G., & Weinberger, S. H. (Eds.). (1987). *Interlanguage phonology: The acquisition of a second language sound system*. Cambridge, MA: Newbury House.

Isačenko, A. V., & Schädlich, H. J. (1966). Untersuchungen über die deutsche Satzintonation. *Studia Grammatica*, *7*, 7–67.

James, A., & Ketteman, B. (Eds.). (1983). *Dialektphonologie und Fremdsprachenerwerb*, Tübingen: Gunter Narr.

James, A., & Leather, J. (Eds.). (1987). *Sound patterns in second language acquisition*. Dordrecht: Foris.

James, A., & Leather, J. (Eds.). (1997). *Second-language speech*. Berlin: Mouton de Gruyter.

James, A., & Westney, P. (Eds.). (1981). *New linguistic impulses in foreign language teaching*. Tübingen: Gunter Narr.

James, E. (1976). The acquisition of prosodic features of speech using a speech visualizer. *IRAL*, *14*(3), 227–243.

James, E. (1977). The acquisition of second language intonation using a visualizer. *Canadian Modern Language Review*, *33*(4), 503–506.

James, E. (1979). Intonation through visualization. In H. Hollien & P. Hollien (Eds.), *Current issues in the phonetic sciences* Vol. IV, Amsterdam Studies in the Theory and History of Linguistic Science (pp. 295–301). Amsterdam: John Benjamins.

Jassem, W., & Demenko, G. (1986). Extracting linguistic information from F_0 traces. In C. Johns-Lewis (Ed.), *Intonation in discourse* (pp. 1–17). London: Croom-Helm.

Johansson, I. (1978). Functional aspects of sentence intonation in Swedish. *Umea Studies in the Humanities*, *16*, 1–209.

Johns-Lewis, C. (Ed.). (1986a). *Intonation in discourse*. San Diego: College Hill.

Johns-Lewis, C. (1986b). Prosodic differentiation of discourse modes. In C. Johns-Lewis (Ed.), *Intonation in discourse* (pp. 199–219). London: Croom Helm.

Johnston, J., & Milne, L. (1995). Scaffolding second language communicative discourse with teacher-controlled multimedia. *Foreign Language Annals*, *28*(3), 315–329.

Jones, D. (1909). *The pronunciation of English*. Cambridge: Cambridge University Press.

Jones, D. (1918). *An outline of English phonetics*. Leipzig: Teubner.

Jourdain, S., & Schuler, M. V. (1999). Morte la Bête Noire: Two techniques for teaching the French /R/. *French Review*, *72*(1), 216–228.

Keller, E. (1988). Signalyze [Computer Program] . Lausanne: InfoSignal.

Keller, E. (1989). Two-stage acquisition of Chinese tones [Unpublished manuscript].

Kelm, O. R. (1987). An acoustic study on the differences of contrastive emphasis between native and non-native Spanish speakers. *Hispania*, *70*, 627–633.

Kelz, H. P. (Ed.). (1977). Phonetische Grundlagen der Ausspracheschulung. Hamburg: Buske.

Kenworthy, J. (1987). *Teaching English pronunciation*. London: Longman.

Ketteman, B. (1977). Frequency manipulation as perceptual phonetic correction procedure. In H. P. Kelz (Ed.), *Phonetische Grundlagen der Ausspracheschulung* (pp. 285–311). Hamburg: Buske.

Kim, H.-Y. (1995). Intake from the speech stream: Speech elements that L2 learners attend to. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 65–83). Honolulu: Second Language Teaching and Curriculum Center.

Kingdon, R. (1958). *The groundwork of English intonation*. London: Longman.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustical and perceptual evidence. *Journal of the Acoustical Society of America*, *59*, 1208–1221.

Koster, C. J., & Koet, T. (1993). The evaluation of accent in the English of Dutchmen. *Language Learning*, *43*(1), 69–92.

Kramsch, C. (1981). *Discourse analysis and second language teaching*. Washington, D. C.: Center for Applied Linguistics.

Kramsch, C. (1983). Interaction in the classroom: Learning to negotiate roles and meanings. *Die Unterrichtspraxis*, *16*, 175–90.

Kramsch, C. (1986). From language proficiency to interactional competence. *Modern Language Journal*, *70*, 366–372.

Krashen, S. D. (1985). *The input hypothesis*. London: Longman.

Kreiman, J. (1982). Perception of sentence and paragraph boundaries in natural conversation. *Journal of Phonetics*, *10*, 163–175.

Ladd, D. R. (1980). *The structure of intonational meaning*. Bloomington: Indiana University Press.

Ladd, D. R. (1987). Review of Bolinger, Intonation and its parts. *Language*, *63*(3), 637–643.

Ladd, D. R. (1990). Intonation: Emotion vs. grammar. *Language*, *66*(4), 806–816.

Lado, R., Fries, C. C., & Robinett, B. W. (1954). *English pronunciation*. Ann Arbor: University of Michigan Press.

Lambrecht, K. (1994). *Information structure and sentence form*. Cambridge: Cambridge University Press.

Landick, M. (1995). The mid-vowels in figures: Hard facts. *French Review*, *69*(1), 88–102.

Lane, H. L., & Buiten, R. (1966). A self-instructional device for conditioning accurate prosody. In A. Valdman (Ed.), *Trends in language teaching* (pp. 159–174). New York: McGraw Hill.

Lane, H. L., & Schneider, B. (1963). Methods for self-shaping echoic behavior. *Modern Language Journal*, *47*, 154–60.

Lantolf, J. P. (1976). On teaching intonation. *Modern Language Journal*, *60*, 267–274.

Lantolf, J. P., & Frawley, W. (1985). Oral proficiency testing: A critical analysis. *Modern Language Journal*, *69*, 337–345.

Leahy, R. M. (1980). A practical approach for teaching ESL pronunciation based on distinctive feature analysis. *TESOL Quarterly*, *14*(2), 209–19.

Leather, J. (1983). Second-language pronunciation learning and teaching. *Language Teaching*, *16*, 198–219.

Leather, J. (1987). F_0 pattern inferences in the perceptual acquisition of second language

tone. In A. James & J. Leather (Eds.). Sound patterns in second language acquisition (pp. 59–80). Dordrecht: Foris.
Leather, J., & James, A. (1991). The acquisition of second language speech. *Studies in Second Language Acquisition, 13*, 305–341.
Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.
Lehiste, I. (1976). Suprasegmentals. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 225–239). New York: Academic.
Lehiste, I. (1979). Sentence boundaries and paragraph boundaries — perceptual evidence. In *The elements: A parasession on linguistic units and levels* (pp. 99–109). Chicago: Chicago Linguistic Society.
Lehiste, I., & Peterson, G. E. (1961). Some basic considerations in the analysis of intonation. *Journal of the Acoustical Society of America, 33*, 419–425.
Lenneberg, E. H. (1967). *Biological foundations of language*. New York: Wiley.
Lenneberg, E. H., & Lenneberg, E. (Eds.) (1975). *Foundations of language development*. New York: Academic Press.
Leon, P., & Martin, P. (1972). Applied linguistics and the teaching of intonation. *Modern Language Journal, 56*, 139–44.
Lepetit, D. (1985). Interlangue intonative et fossilisation. *Studies in Second Language Acquisition, 7*, 308–322.
Lepetit, D. (1989). Cross-linguistic influence in intonation: French/Japanese and French/English. *Language Learning, 39*(3), 397–413.
Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
Levis, J. M. (1999). Intonation in theory and practice, revisited. *TESOL Quarterly, 33*(1), 37–63.
Liberman, M. (1975). *The intonational system of English*. Unpublished doctoral dissertation, MIT, Cambridge, MA.
Liberman, M. (1979). *The intonational system of English*. New York: Garland.
Liberman, M., & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff & R. Oehrle (Eds.), *Language sound structure* (pp. 157–233). Cambridge: MIT Press.
Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry, 8*, 249–336.
Liberman, M., & Sag, I. A. (1974). Prosodic form and discourse function. In Chicago Linguistic Society, *Papers from the Tenth Regional Meeting* (pp. 416–427). Chicago.
Lieb, H. (1980). Intonation als Mittel verbaler Kommunikation. *Linguistische Berichte, 68*, 34–48.
Lieberman, P. (1965). On the acoustic basis of perception of intonation by linguists. *Word, 21*, 40–54.
Lieberman, P. (1967). *Intonation, perception, and language*. Cambridge, MA: MIT Press.
Lieberman, P. (1986). The acquisition of intonation by infants: Physiology and neural control. In C. Johns-Lewis (Ed.), *Intonation in discourse* (pp. 239–257). London: Croom Helm.
Lieberman, P., & Blumstein, S. E. (1988). *Speech physiology, speech perception, and acoustic phonetics*. Cambridge: Cambridge University Press.

Lifschitz, S. (1933). Two integral laws of sound perception relating loudness and apparent duration of sound impulses. *Journal of the Acoustical Society of America*, *7*, 213–219.

Lindblom, B., & Sundberg, J. (1971). A quantitative theory of cardinal vowels and the teaching of pronunciation. In G. E. Perren & J. L. M. Trim (Ed.), *Applications of linguistics* (pp. 319–328). London: Cambridge University Press.

Lindsey, G. (1981). Intonation and pragmatics. *Journal of the International Phonetic Association*, *11*(1), 2–21.

Linell, P. (1982). The concept of phonological form and the activities of speech production and perception. *Journal of Phonetics*, *10*, 37–72.

Liskin-Gasparro, J. E. (1982). *ETS oral proficiency testing manual*. Princeton: Educational Testing Service.

Lohnes, W., & Strothmann, F. W. (1980). *German: A structural approach* (3rd ed.). New York: Norton.

Loveday, L. (1981). Pitch, politeness and sexual role: An exploratory investigation into the pitch correlates of English and Japanese politeness formulae. *Language and Speech*, *24*(1), 71–89.

Lowe, P., & Hughes, H. (1990). In-Tube-View: TOPS (Testing Oral Proficiency System). Paper presented at the 1990 ACTFL Annual Meeting, Nashville, TN.

Lowenthal, K., & Bull, D. (1984). Imitation of foreign sounds: What is the effect of age? *Language and Speech*, *27*, 95–98.

Luthy, M. J. (1983). Nonnative speakers' perceptions of English "nonlexical" intonation signals. *Language Learning*, *33*(1), 19–36.

MacCarthy, P. (1975). *The pronunciation of German*. London: Oxford University Press.

Mallard, A. R., Ringel, R. L., & Horii, Y. (1978). Sensory contributions to control of fundamental frequency of phonation. *Folia Phoniatrica*, *30*, 199–213.

Mantini, L. C. (1980). The use of a structured sequence of recorded material for language stress-rhythm and intonation study. *Foreign Language Annals*, *13*(5), 373–383.

Matthews, P. H. (1997). *The Concise Oxford Dictionary of Linguistics*. Oxford: Oxford University Press.

McCandless, P., & Winitz, H. (1986). Test of pronunciation following one year of comprehension instruction in college German. *Modern Language Journal*, *70*, 355–362.

McCarthy, M. (1991). *Discourse analysis for language teachers*. Cambridge: Cambridge University Press.

McGinnis, S. (1997). Tonal spelling versus diacritics for teaching pronunciation of Mandarin Chinese. *Modern Language Journal*, *81*(2), 228–236.

Meier, R. (1984). *Bibliographie zur Intonation*. Tübingen: Max Niemeyer.

Menn, L., & Boyce, S. (1982). Fundamental frequency and discourse structure. *Language and Speech*, *25*, 341–83.

Moeller, J., Liedloff, H., & Kent, C. (1986). *German Today, One* (3rd ed.). Boston: Houghton.

Molholt, G. (1988). Computer-assisted instruction in pronunciation for Chinese speakers of American English. *TESOL Quarterly*, *22*(1), 91–111.

Molholt, G., & Presler, A. M. (1986). Correlation between human and machine ratings of Test of Spoken English reading passages. In C. W. Stansfield (Ed.), *Technology and language testing* (pp. 111–128). Washington, D. C.: TESOL (Teachers of English to Speakers of Other Languages).

Morgan, B. (1997). Identity and intonation: Linking dynamic processes in an ESL classroom. *TESOL Quarterly, 31*(3), 431–450.

Morley, J. (1979). *Improving spoken English.* Ann Arbor: The University of Michigan Press.

Morley, J. (1984). *Listening and language learning in ESL: Developing self-study activities for listening comprehension*, Orlando: Harcourt Brace Jovanovich.

Morley, J. (Ed.). (1987). *Current perspectives on pronunciation: Practices anchored in theory*, Washington, D. C.: TESOL (Teachers of English to Speakers of Other Languages).

Morley, J. (1991). The pronunciation component in Teaching English to Speakers of Other Languages. *TESOL Quarterly*, *25*(3), 481–520.

Morley, J. (1992). *Rapid review of vowel and prosodic contexts.* Ann Arbor: University of Michigan Press.

Morley, J. (Ed.). (1994). Pronunciation pedagogy and theory: New views, new directions. Alexandria, VA: TESOL (Teachers of English to Speakers of Other Languages).

Moulton, W. G. (1962). *The sounds of English and German.* Chicago: University of Chicago Press.

Moyer, A. (1999). Ultimate attainment in L2 phonology: The critical factors of age, motivation and instruction. *Studies in Second Language Acquisition*, *21*, 81–108.

Munro, M. J. (1995). Nonsegmental factors in foreign accent: Ratings of filtered speech. *Studies in Second Language Acquisition*, *17*(1), 17–34.

Munro, M. J., & Derwing, T. M. (1995a). Foreign accent, comprehensibility and intelligibility in the speech of second language learners. *Language Learning*, *45*, 73–97.

Munro, M. J., & Derwing, T. M. (1995b). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, *38*, 289–306.

Munro, M. J., & Derwing, T. M. (1998). The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning*, *48*, 159–182.

Murphy, J. M. (1991). Oral communication in TESOL: Integrating speaking, listening, and pronunciation. *TESOL Quarterly*, *25*(1), 51–75.

Murphy, J. M. (1997). Phonology courses offered by MATESOL programs in the U. S. *TESOL Quarterly, 31*(4), 741–764.

Nash, R. (1972). Phonemic and prosodic interference and their effects on intelligibility. In A. Rigault & R. Charbonneau (Ed.), *Proceedings of the Seventh International Congress of Phonetic Sciences* (pp. 570–573). The Hague: Mouton.

National Standards in Foreign Language Education (a collaborative project of ACTFL, AATF, AATG, AATI, AATSP, ACL/APA, AATR, CLASS/CLTA, & NCSTJ/ATJ). (1996). Standards for Foreign Language Learning: Preparing for the 21st Century. Yonkers, NY: ACTFL (American Council on the Teaching of Foreign Languages).

Neufeld, G. G. (1978). On the acquisition of prosodic and articulatory features in adult language learning. *Canadian Modern Language Review*, *34*, 168–94.

Neufeld, G. G. (1980). On the adult's ability to acquire phonology. *TESOL Quarterly*, *14*, 285–298.

Neuner, G., Schmidt, R., Wilms, H., & Zirkel, M. (1979). *Deutsch aktiv.* Berlin: Langenscheidt.

O'Connell, D. C., & Kowal, S. (1983). Pausology. In W. Sedelow & S. Sedelow (Eds.), *Computers in Language Research* (2nd ed.) (pp. 221–301). The Hague: Mouton.

O'Connor, J. D., & Arnold, G. F. (1961). *Intonation of colloquial English* (2nd ed., 1973). London: Longman.

Olsen, L., & Samuels, J. (1973). The relationship between age and accuracy of foreign language pronunciation. *Journal of Educational Research*, *66*, 263–268.

Omaggio, A. C. (1986). *Teaching language in context: Proficiency-oriented instruction*. Boston: Heinle.

Omaggio Hadley, A. C. (1993). *Teaching language in context: Proficiency-oriented instruction* (2nd ed.). Boston: Heinle.

Oreström, B. (1983). *Turn-taking in English conversation*. Lund: CWK Gleerup.

Orion, G. F. (1988). *Pronouncing American English: Sounds, stress, and intonation*. New York: Newbury House.

Palmer, H. E. (1922). *English intonation, with systematic exercises*. Cambridge: Heffer.

Pavlik, C. (1986). *Speak up!* Rowley, MA: Newbury House.

Pegolo, C. (1985). The role of rhythm and intonation in the silent reading of French as a foreign language. *Reading in a Foreign Language*, *3*, 313–327.

Pennington, M. C. (1989a). *Teaching languages with computers: The state of the art*. La Jolla: Athelstan.

Pennington, M. C. (1989b). Teaching pronunciation from the top down. *RELC Journal*, *20*, 20–38.

Pennington, M. C. (Ed.). (1996). *The Power of CALL*. Houston: Athelstan.

Pennington, M. C. (1998). The teachability of phonology in adulthood: A re-examination. *International Review of Applied Linguistics in Language Teaching 36*(4), 323–341.

Pennington, M. C. (1999). Computer-aided pronunciation pedagogy: Promise, limitations, directions. *Computer Assisted Language Learning 12*(5), 427–440.

Pennington, M. C. (2000) Focusing learners' attention on prosody. Revised version of a plenary address delivered on 20 May 2000 at the University of Durham Seminar on Exploring Research-based Innovations for Teaching L2 Pronunciation.

Pennington, M. C., & Ellis, N. C. (2000). Cantonese speakers' memory for English sentences with prosodic cues. *Modern Language Journal, 84*(3), 372–389.

Pennington, M. C., & Esling, J. H. (1996). Computer-assisted development of spoken language skills. In M. C. Pennington (Ed.), *The Power of CALL (pp.* 153–189). Houston: Athelstan.

Pennington, M. C., & Richards, J. C. (1986). Pronunciation revisited. *TESOL Quarterly*, *20*(2), 207–225.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. Unpublished doctoral dissertation, MIT, Cambridge, MA.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollock (Eds.), *Intentions in communication* (pp. 271–312). Cambridge, MA: MIT Press.

Pike, K. L. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Press.

Piper, A. (1986). Conversation and the computer: A study of conversational spin-off generated among learners of English as a foreign language working in groups. *System*, *14*(2), 187–98.

Politzer, R. L., & Weiss, L. (1969). Developmental aspects of auditory discrimination, echo response and recall. *Modern Language Journal*, *53*, 75–85.

Prator, C. H., & Robinett, B. W. (1985). *Manual of American English pronunciation* (4th ed.). New York: Holt, Rinehart and Winston.

Preu, O. (1973). Ein methodischer Ansatz zur Einführung der deutschen Satzintonation im Phonetikunterricht für Ausländer. *Deutsch als Fremdsprache*, *10*(1), 45–49.

Pritchard, R. M. O. (1985). The teaching of French intonation to native speakers of English. *IRAL*, *23*, 117–147.

Purcell, E., & Suter, R. (1980). Predictors of pronunciation accuracy: A reexamination. *Language Learning*, *30*, 271–287.

Pürschel, H. (1975). *Pause und Kadenz: Interferenzerscheinungen bei der englischen Intonation deutscher Sprecher.* Tübingen: Niemeyer.

Quilis, A., & Fernández, J. A. (1992). *Curso de fonética y fonología españolas: para estudiantes angloamericanos* (14th ed.). Madrid: Consejo Superior de Investigaciones Cientificas.

Rando, E. (1980). Intonation in discourse. In L. R. Waugh & C. H. van Schooneveld (Eds.), *The melody of language* (pp. 243–278). Baltimore: University Park Press.

Rausch, R., & Rausch, I. (2000). *Deutsche Phonetik für Ausländer* (6th ed.). Berlin/Munich/Leipzig: Langenscheidt/Verlag Enzyklopädie.

Richards, J. C. (1986). Focus on the learner. *University of Hawaii Working Papers in ESL*, *5*, 61–86.

Riney, T. J., Takada, M., & Ota, M. (2000). Segmentals and global foreign accent: The Japanese flap in EFL. *TESOL Quarterly, 34*(4).

Rogerson, P., & Gilbert, J. B. (1990). *Speaking clearly* (Teacher's book and Student's book). Cambridge: Cambridge University Press.

Rooney, E., Hiller, S., Laver, J., & Jack, M. (1992) Prosodic features for automated pronunciation improvement in the SPELL system. In: *Proceedings of the International Conference on Spoken Language Processing 92* (pp. 413–416). Banff: ICSLP.

Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language*, *50*, 696–735.

Sag, I. A., & Liberman, M. (1975). The intonational disambiguation of indirect speech acts. In Chicago Linguistic Society, *Papers from the Eleventh Regional Meeting* (pp. 487–497). Chicago.

Sanders, D. A. (1977). *Auditory perception of speech.* Englewood Cliffs, NJ: Prentice Hall.

Savignon, S. J. (1972). *Communicative competence: An experiment in foreign language teaching.* Philadelphia: The Center for Curriculum Development.

Savignon, S. J. (1983). *Communicative competence: Theory and classroom practice.* Reading, MA: Addison-Wesley.

Savignon, S. J. (1985). Evaluation of communicative competence: The ACTFL Provisional Proficiency Guidelines. *Modern Language Journal*, *69*, 129–34.

Scanlan, T. (1987). Improving fluency in spoken French through a study of native pause behavior. *Foreign Language Annals*, *20*, 345–352.

Schaffer, D. (1983). The role of intonation as a cue to turn taking in conversation. *Journal of Phonetics*, *11*, 243–257.

Schaffer, D. (1984). The role of intonation as a cue to topic management in conversation.

Journal of Phonetics, 12, 327–344.
Schairer, K. E. (1992). Native speaker reaction to non-native speech. *Modern Language Journal, 76*(3), 309–319.
Schegloff, E. A., & Sacks, H. (1973). Opening up closings. *Semiotica, 8*(4), 289–327.
Scherer, K. R. (1979). Acoustic concomitants of emotional dimensions: Judging affect from synthesized tone sequences. In S. Weitz (Ed.), *Non-verbal communication: Readings with commentary* (2nd ed.) (pp. 249–253). New York: Oxford University Press.
Schiffrin, D. (1988). Conversation analysis. In F. J. Newmeyer (Ed.), *Language: The Socio-cultural Context* (pp. 251–276). Cambridge: Cambridge University Press.
Schmid, B. (1986). A comparative study of children's and adults' acquisition of tone accents in Swedish. *Language Learning, 36*(2), 185–210.
Schmidt, H. (1983). Mhm. Intonation und kommunikative Funktion von Rezipienten-signalen im Französischen und Deutschen. *Zeitschrift für Literaturwissenschaft und Linguistik, 49*, 101–123.
Schubiger, M. (1958). *English intonation: Its form and function.* Tübingen: Max Niemeyer.
Schubiger, M. (1980). English intonation and German modal particles II: A comparative study. In L. R. Waugh and C. H. van Schooneveld (Eds.), *The melody of language: Intonation and prosody (pp.* 279–298). Baltimore: University Park Press.
Schuetze-Coburn, S. (1993). Discourse prosody. [Unpublished manuscript].
Schuetze-Coburn, S., Shapley, M., & Weber, E. G. (1991). Units of intonation in discourse: A comparison of acoustic and auditory analyses. *Language and Speech, 34*(3), 207–234.
Scovel, T. (1988). *A time to speak: A psycholinguistic inquiry into the critical period for human speech.* Cambridge, MA: Newbury House.
Scuffil, M. (1982). *Experiments in comparative intonation: A case-study of English and German.* Tübingen: Niemeyer.
Searle, J. (1969). *Speech acts.* Cambridge: Cambridge University Press.
Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics, 10*, 209–232.
Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure.* Cambridge, MA: MIT Press.
Selting, M. (1987). Descriptive categories for the auditive analysis of intonation in conversation. *Journal of Pragmatics, 22*, 777–791.
Selting, M. (1988). The role of intonation in the organization of repair and problem handling sequences in conversation. *Journal of Pragmatics, 12*, 293–322.
Selting, M. (1992). Prosody in conversational questions. *Journal of Pragmatics, 17*, 315–345.
Selting, M. (1995). *Prosodie im Gespräch: Aspekte einer interaktionalen Phonologie der Konversation.* Tübingen: Niemeyer.
Shapley, M. (1989). *Fundamental frequency variation in conversational discourse.* Unpublished doctoral dissertation, UCLA, Los Angeles, CA.
Sheeler, W. D., & Markley, R. W. (1986). *Sounds and rhythm: Focus on vowels.* New York: Regents.
Shelly, S. L. (1998). Teaching the sound system(s): The case of mid vowels. *French Review 71*(4), 598–607.
Simoes, A. R. M. (1996). Phonetics in second language acquisition: An acoustic study of fluency in adult learners of Spanish. *Hispania, 79*(1), 87–95.

Sinclair, J. M., & Coulthard, R. M. (1975). *Towards an analysis of discourse*. Oxford: Oxford University Press.

Singleton, D. (1989). *Language acquisition: The age factor*. Clevedon, UK: Multilingual Matters.

Snow, C. E., & Hoefnagel-Höhle, M. (1978). The critical period for second language acquisition: Evidence from second language learning. *Child Development*, *49*, 1114–1128.

Snow, C. E., & Hoefnagel-Höhle, M. (1977). Age differences in the pronunciation of foreign sounds. *Language and Speech*, *20*, 357–65.

Spaai, G. W. G., & Hermes, D. J. (1993). A visual display for the teaching of intonation. *CALICO Journal*, *10*(3), 19–30.

Spolsky, B. (1984). A note on the dangers of terminological innovation. In C. Rivera (Ed.), *Language proficiency and academic achievement* (pp. 41–43). Clevedon: Multilingual Matters.

Stevens, K. N. (1972). Quantal nature of speech. In E. E. David & P. B. Denes (Eds.), *Human communication: A unified view* (pp. 51–66). New York: McGraw-Hill.

Stevens, K. N., & Perkell, J. S. (1977). Speech physiology and phonetic features. In M. Sawashima & F. S. Cooper (Eds.), *Dynamic aspects of speech production* (pp. 323–341). Tokyo: University of Tokyo Press.

Studdert-Kennedy, M., & Hadding, K. (1973). Auditory and linguistic processes in the perception of intonation contours. *Language and Speech*, *16*, 293–313.

Swan, M., & Smith, B. (Eds.). (1987). *Learner English: A teacher's guide to interference and other problems*. Cambridge: Cambridge University Press.

Sweet, H. (1890). *A primer of spoken English*. Oxford: Clarendon Press.

't Hart, J., & Collier, R. (1975). Integrating different levels of intonation analysis. *Journal of Phonetics*, *3*, 235–55.

't Hart, J., Collier, R., & Cohen, A. (1990). *A perceptual study of intonation: An experimental-phonetic approach to speech melody*. Cambridge: Cambridge University Press.

Tarone, E. (1980). Communication strategies, foreigner talk, and repair in interlanguage. *Language Learning*, *30*, 417–431.

Tarone, E. (1987). The phonology of interlanguage. In G. Ioup & S. H. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 70–85). Cambridge: Newbury House.

Temperly, M. S. (1983). The articulatory target for final -s clusters. *TESOL Quarterly*, *17*, 421–436.

Test of Spoken English (TSE): Manual for Score Users. (1992). Princeton: Educational Testing Service.

Thompson, I. (1991). Foreign accents revisited: The English pronunciation of Russian immigrants. *Language Learning*, *41*(2), 177–204.

Thorsen, N. G. (1980). A study of the perception of sentence intonation: Evidence from Danish. *Journal of the Acoustical Society of America*, *67*(3), 1014–1030.

Touati, P. (1987). *Structures prosodiques du suédois et du français: Profils temporels et configurations tonales*. Lund: Lund University Press.

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

Tranel, B. (1987). *The sounds of French: An introduction*. Cambridge: Cambridge University Press.

Trim, J. L. M. (1975). *English pronunciation illustrated* (2nd ed.). Cambridge: Cambridge University Press.

Tumposky, N. R. (1982). Activities for practicing stress and intonation. *English Teaching Forum*, *20*, 15–18.

Uhmann, S. (1996). On rhythm in everyday German conversation: beat clashes in assessment utterances. In E. Couper-Kuhlen & M. Selting (Eds.), *Prosody in conversation* (pp. 303–365). Cambridge: Cambridge University Press.

Uldall, E. (1961). Ambiguity: Question or statement? or "Are you asking me or telling me?". *Proceedings of the Fourth International Congress of Phonetic Sciences* (pp. 779–783). The Hague: Mouton.

Uldall, E. (1964). Dimensions of meaning in intonation. In D. Abercrombie (Ed.), *In honour of Daniel Jones: Papers contributed on the occasion of his 80th birthday* (pp. 271–279). London: Longman.

Umeda, S. N., & Quinn, A. M. S. (1981). Word duration as an acoustic measure of boundary perception. *Journal of Phonetics*, *9*, 19–28.

van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech*, *38*(3), 253–265.

van Els, T. (1984). *Applied linguistics and the learning and teaching of foreign languages*. Baltimore: Edward Arnold.

van Els, T., & de Bot, K. (1987). The role of intonation in foreign accent. *Modern Language Journal*, *71*, 147–155.

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

Vardanian, R. M. (1964). Teaching English through oscilloscope displays. *Language Learning*, *3/4*, 109–118.

Wales, R., & Taylor, S. (1987). Intonation cues to questions and statements: How are they perceived? *Language and Speech*, *30*, 199–211.

Walz, J. (1980). An empirical study of pronunciation errors in French. *The French Review*, *53*, 424–432.

Ward, G., & Hirschberg, J. (1985). Implicating uncertainty: The pragmatics of fall-rise intonation. *Language*, *61*, 747–776.

Weiss, E. (1980). *Deutsch: Entdecken wir es!* New York: Webster-McGraw.

Weiss, R. (1970). The role of perception in teaching the German vowels to American students. In G. Nickel (Ed.), *Proceedings of the Fourth International Congress of Applied Linguistics* (pp. 513–524). Stuttgart: Hochschul.

Wells, R. S. (1945). The pitch phonemes of English. *Language*, *21*, 27–39.

Weltens, B., & de Bot, K. (1984). Visual feedback of intonation II: Feedback delay and quality of feedback. *Language and Speech*, *27*(1), 79–88.

Wennerstrom, A. (1994). Intonational meaning in English discourse: A study of non-native speakers. *Applied Linguistics*, *15*(4), 399–420.

Wennerstrom, A. (1998). Intonation as cohesion in academic discourse: A study of Chinese speakers of English. *Studies in Second Language Acquisition*, *20*(1), 1–25.
White, C. M. (1981). Tonal pronunciation errors and interference from English. *Journal of the Chinese Language Teachers' Association*. *16*, 27–56.
Wichern, P. (1979). Visuele terugmelding als hulpmiddel bij het aanleren van onbekende intonatieverlopen. Nijmegen: Intern Rapport Instituut Fonetiek.
Wichern, P. U. M., & Boves, L. (1980). Visual feedback of F_0 curves as an aid in learning intonation-contours. *Proceedings Institute of Phonetics Nijmegen*, *4*, 53–63.
Winkler, C. (1979). *Untersuchungen zur Kadenzbildung in deutscher Rede*. München: Hueber.
Winkler, P. (1980). Phonetische und konversationsanalytische Interpretationen: Eine Vor-orientierung. *Linguistische Berichte*, *68*, 67–84.
Wipf, J. (1985). Toward improving second language pronunciation. *Die Unterrichtspraxis*, *18*(1), 55–63.
Wong, R. (1987). *Teaching pronunciation: Focus on English rhythm and intonation*. Englewood Cliffs, NJ: Prentice Hall.
Wunderli, P. (1979). Satz, Paragraph, Text — und die Intonation. In J. Petöfi (Ed.), *Text vs. sentence: Basic questions of text linguistics* (pp. 319–341). Hamburg: Buske.
Wylie, L. (1985). Language learning and communication. *The French Review*, *58*(6), 777–785.
Yang, L. (1995). *Intonational structures of Mandarin discourse*. Unpublished doctoral dissertation, Georgetown University, Washington, D. C.
Young, R. (1988). Computer-assisted language learning conversations: Negotiating an outcome. *CALICO Journal*, *5*(3), 65–83.
Yule, G. (1980a). Intonation and givenness in spoken discourse. *Studies in Language*, *4*, 271–286.
Yule, G. (1980b). Speakers' topics and major paratones. *Lingua*, *52*, 33–47.
Yule, G. (1989). The spoken language. *Annual Review of Applied Linguistics*, *10*, 163–172.
Yule, G. (1990). Review of Kenworthy, Wong, and Morley. *System*, *18*, 107–111.
Yule, G., Hoffman, P., & Damico, J. (1987). Paying attention to pronunciation: The role of self-monitoring in perception. *TESOL Quarterly*, *21*(4), 765–768.

Subject index

Author index