Werner, S., & Keller, E. (1994). Prosodic aspects of speech. In E. Keller (ed.), *Fundamentals of Speech Synthesis and Speech Recognition: Basic Concepts, State of the Art, and Future Challenges* (pp. 23-40). Chichester: John Wiley.

Prosodic Aspects of Speech 2

Stefan Werner[1] and *Eric Keller*[2]

[1] Linguistics and Phonetics, Joensuu University, FIN-80101 JOENSUU, Finland and Laboratoire d'analyse informatique de la parole (LAIP) Université de Lausanne, CH-1015 LAUSANNE, Switzerland
[2] Laboratoire d'analyse informatique de la parole (LAIP) Université de Lausanne, CH-1015 LAUSANNE, Switzerland

This chapter contains a general introduction to the concepts of prosody, described primarily from a phenomenological point of view. Definitions of prosody are discussed, descriptions of its articulatory, acoustic and perceptual manifestations are provided, and the status of prosody within the larger domains of linguistics and phonetics is examined. Also included is a short overview of recent prosodic research in speech synthesis and recognition.

According to the classic definition, prosody has to do with speech features whose domain is not a single phonetic segment, but larger units of more than one segment, possibly whole sentences or even longer utterances. Consequently, prosodic phenomena are often called *supra-segmentals* (cf. the title of a classical study on prosody, Lehiste 1970). They appear to be used to structure the speech flow and are perceived as stress or accentuation, or as other modifications of intonation, rhythm and loudness.

Four principal manifestation levels of prosodic phenomena can be distinguished. Although there is considerable divergence between authors with respect to the use of terminology in prosody (see below), a first differentiation of prosodic phenomena according to these four manifestation levels is probably useful.

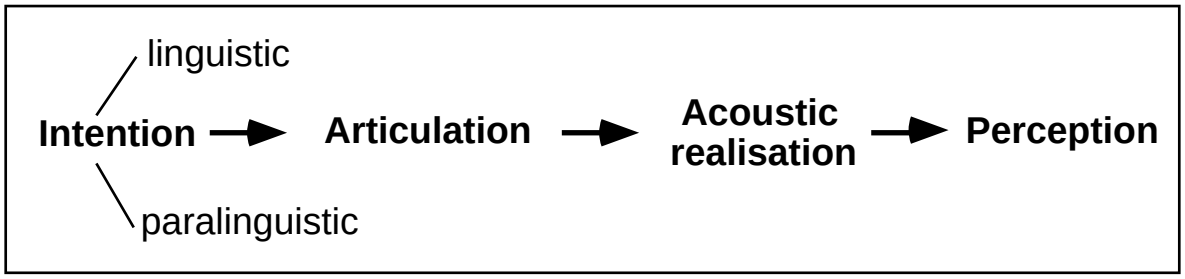

Figure 1: Stages of oral communication

The Linguistic Intention Level

In any language with oral expression, the speaker can be assumed to employ prosodic coding, as other elements of speech, with a certain *intention*. This intention can influence both linguistic and paralinguistic expression. By *linguistic expression* is meant any oral expression using language signs. *Paralinguistic phenomena* include non-verbal vocalisations like onomatopoeia and certain interjection-like expressions as well as speaking styles that make an utterance sound, e.g. angry, urgent or ironic.

Prosody clearly plays a major role in both types of phenomenon[1]. However, since the present volume is concerned with the more formal forms of speech used in automatic speech processing, this text will concentrate exclusively on the manifestations and the semiotic structure of the *linguistic* use of prosody. In this context, it is important to establish the exact *physical nature* of prosodic phenomena, as well as the use of prosodic phenomena for communicating linguistic *distinctions*. Examples of linguistic distinctions that tend to be communicated by prosodic means are the question-statement distinction, or the semantic emphasis of an element with respect to previously enunciated material. Systematic knowledge of how these phenomena are used in human speech can be expected to play a significant role in improving the naturalness of synthetic speech. Such knowledge may also some day contribute to the improved performance of speech recognition systems.

From the linguistic point of view, prosody is generally thought of as relating different linguistic elements to each other, above all by accentuating certain elements of a text, by marking boundaries and by

[1] Interestingly, evidence from speech pathology shows that linguistic and paralinguistic prosody seem to be processed in different hemispheres of the brain; see e.g., Fromkin, 1987. At the same time, this is not the only right-left hemisphere difference affecting prosody, since right-hemispheric anterior lesions have been reported to produce a monotonous and unemphatic, but otherwise unaffected speech — thus interfering selectively with stress and fundamental frequency and leaving other aspects (e.g. timing) unaffected.

defining transitions between words, phrases or sentences. Linguistically, differentiated prosodic phenomena are usually categorised as relating either to *tone, intonation,* or *stress (accent).* These terms and their use will be described below in more detail.

The Articulatory Manifestation Level

At the articulatory level, prosodic phenomena are physically manifested as a series of modifications of articulatory movement which can be observed with sophisticated machinery (magnetography, ultrasound, X-ray etc.). Since prosodic phenomena are fundamentally to be understood as a distinctive layer of phenomena superimposed on the normal articulatory speech train, prosodic phenomena do not result in separate, identifiable articulations. Rather, they are manifested as systematic modifications of assumed underlying "neutral" articulatory behaviours.

So for example, the stressed syllable /ej/ in "**á**tony" /ˈejtənɪ/ does not involve an articulatory movement distinctive of a more neutral, destressed articulation of the same syllable in "**a**tónic" /əˈtɑnɪk/. Rather, articulatory movements for the production of the stressed diphthong /ej/ would *tend* to be larger, longer in duration and more distinctive from other diphthongal movements than articulations of the unstressed variant. Pertinent physical observations of prosodic manifestations thus typically include variations in the *amplitude of articulatory movements*, variations in *air pressure*, or specific patterns of *electric impulses* in nerves leading to the articulatory musculature, especially those innervating the larynx.

The Acoustic Realisation Level

Muscle activity in the respiration system and along the vocal tract leads to the emission of sound waves. This acoustic realisation of prosodic phenomena can be observed and quantified using acoustic signal analysis. The main acoustic parameters bearing on prosody are *fundamental frequency, intensity* and *duration* (see chapter 1, this volume). Stressed syllables, for example, tend to be higher in fundamental frequency, of greater amplitude and longer in duration than comparable unstressed syllables[2].

---

2 It is clearly understood that these are tendencies, not rules. For example, it is often found that stressed syllables are characterized by only one or two of the classic three parameters (fundamental frequency, amplitude, duration), and that sometimes, a stressed syllable can actually be lower on any of these parameters than its unstressed variant.

The Perceptual Level

Finally, speech sound waves usually enter the ear of a hearer who derives linguistic and paralinguistic information from prosodic phenomena via perceptual processing. At this point, psycholinguistic tests can provide evidence of hearer reactions to prosodic phenomena. This sort of test can verify the salience of different prosodic markers in speech, as well as the acoustic differentiations necessary to provoke minimal perceptual distinctions between different speaker intentions. At the level of perception, it is common to classify prosodic phenomena in terms of the hearer's subjective experience, such as *pauses, length, pitch/melody* and *loudness* .

Prosodic Phenomena in Detail

The acoustic domain, intermediary between the speaker's production and the hearer's perception, traditionally constitutes the main area of automatic speech processing for synthesis and recognition. The reproduction of acoustic signals uses generally available technology, the recording of acoustic manifestations of speech is inexpensive and non-invasive, and acoustic signals are easily evaluated by means of well-established techniques. By contrast, articulatory and perceptual analyses remain much more labour-intensive and often involve the use of complex, expensive and sometimes invasive technology. In the following description of prosodic phenomena, priority is thus given to relating linguistic distinctions to acoustic aspects of prosody.

Stressing or "Accentuation"

A given syllable can be pronounced with more or less perceived loudness, as in the examples "**á**tony" *vs.* "**a**tónic" given above. This phenomenon is generally called "stress" in English and is often called "accent" with respect to other languages. According to their different domains, three types of stress can be distinguished:

— word stress
— phrasal stress
— sentence stress

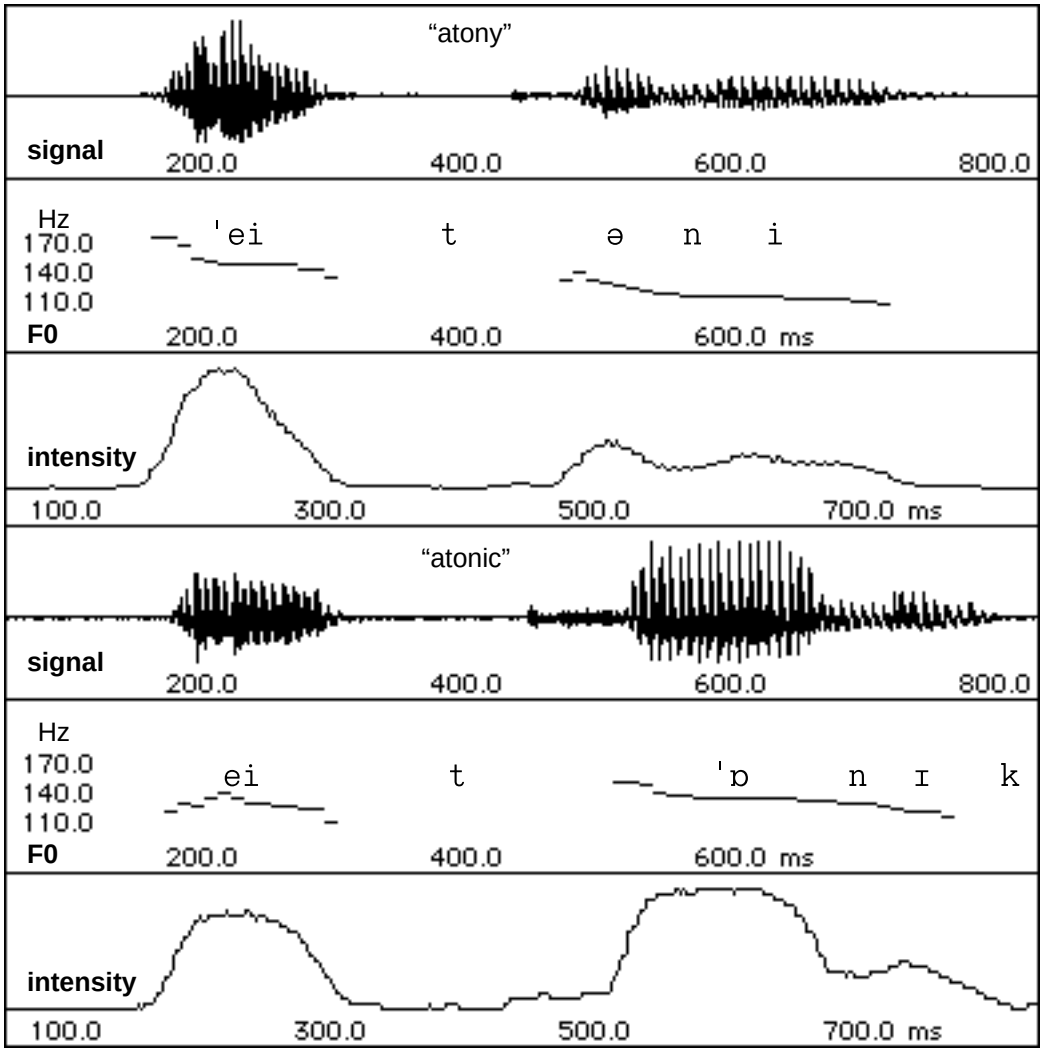

Figure 2. "átony" *vs.* "atónic", fundamental frequency and amplitude envelope analysis. The stressed syllable /ej/ shows a higher average F_0 and a greater amplitude than its unstressed variant.

Several rules concerning the distribution of stress and their systematic relations are widely agreed upon among linguists:

1. Although the *domain* of a stress may include whole words, phrases and sentences, it is always a single syllable that actually bears the stress.

2. In the procedure to identify the stressed syllable in a word, phrase or sentence that contains several stressed syllables, the main stress is identified before all other stresses.

3. Phrasal stress and sentence stress generally coincide with word stress. The different levels of stress supersede and reinforce each other. For example, in the sentence

The manifest**á**tion of stress is different from its signi**f**icance.

there are two main (primary) phrasal stresses, one on "manifestation", and the other on "significance". Each of these two words have their own,

relatively complex stressing pattern[3]. It is found that the same word stress pattern is maintained when the phrasal stress is changed. For example, in the two sentences

The manifest**á**tion of its sign**í**ficance is not too **é**vident.
The manifest**á**tion of its sign**í**ficance is not too **é**vident.

only one of the two words carries primary phrasal stress and the other carries a reduced phrasal stress. (Exactly which word carries which stress evidently depends on the semantic intention). Despite the reduction of phrasal stress, each word maintains its "essential" lexical stress pattern.

The acoustic realisation of stress generally makes use of at least two, and often all three acoustic parameters of prosody (fundamental frequency, intensity and duration). Figure 2 illustrates these differences with respect to the syllable /ej/ in "**á**tony" *vs.* "**a**tónic". It can be seen that the stressed syllable /ej/ shows a higher average F_0 and a greater amplitude than its unstressed variant.

Intonation

The second important prosodic phenomenon is what is perceived as speech melody. This is generally called "intonation". Speakers of most Indo-European languages can communicate sentence mode, such as the declarative *vs.* the interrogative mode, by means of variations of the melody with which the sentence is conveyed. For example, the two sentences

He did it!
He did it?

are generally distinguished by a difference in intonation patterns, where the first sentence carries a falling and the second carries a rising intonation pattern (Figure 3).

Also, intonation can serve to communicate paralinguistic comments on the contents of an utterance. For example, in the following utterances

Speaker 1: That's a good solution.
Speaker 2: [doubtful] That's a good solution?

the intonational pattern expresses a complementary intonational layer of doubt which is not captured by the segmental information alone.

---

[3] In a system where stress levels are indicated by numbers preceding each syllable, the patterns would be something like ^{2}man-^{3}i-^{4}fest-^{1}a-^{3}tion and ^{4}sig-^{1}ni-^{3}fi-^{2}cance.

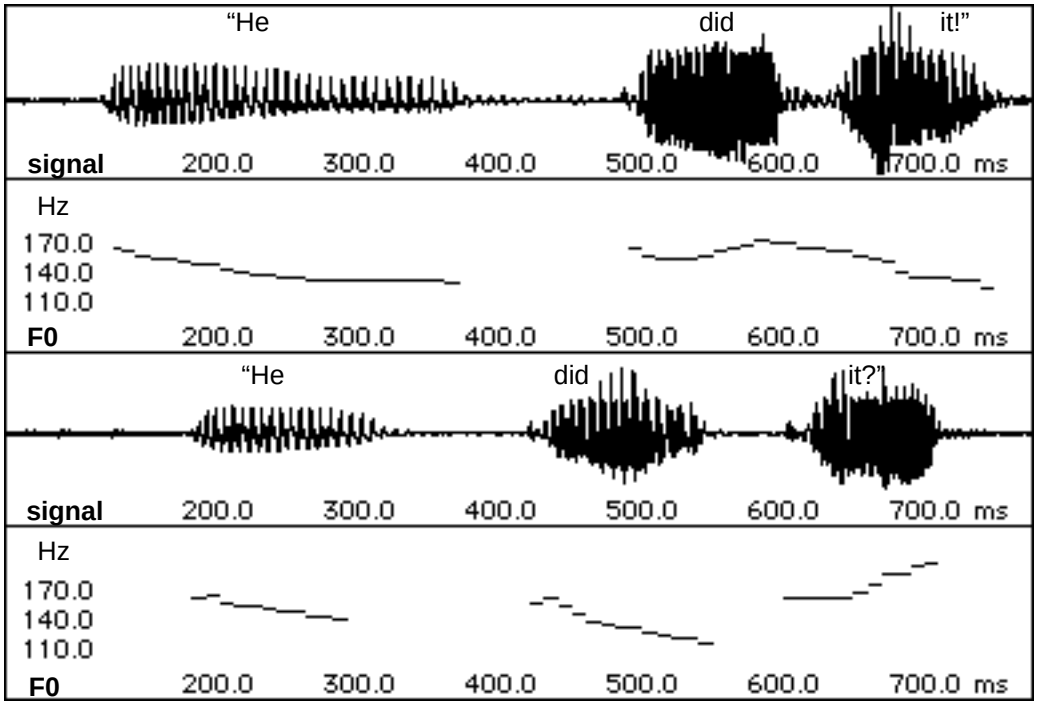

Figure 3. Fundamental frequency traces for the exclamatory and the interrogative forms of the sentence "He did it". In tune with the perceptual impression, F_0 falls during the final word of the exclamatory form and rises for the interrogative form of the sentence.

In its acoustic manifestation, intonation is primarily related to fundamental frequency (F_0). So for example, the utterance-final rises and falls of perceived intonation in "he did it!" *vs.* "he did it?" is well captured by the F_0 inflections shown in Figure 3.

In the context of speech synthesis, an interesting question concerns the interaction between variations in F_0 induced by stress and those induced by intonation. A simple hypothesis would suggest that the two effects are cumulative. So for example, in the two sentences

The manifest**á**tion of stress is different from its sign**í**ficance.
Is the manifest**á**tion of stress different from its sign**í**ficance?

the frequency difference between the unstressed and the stressed /i/s in "s**i**gn**i**ficance" of the declarative sentence should simply cumulate with the frequency rise in the intonational contour of the interrogative sentence.

However, actual frequency measurement shows that the story is quite a bit more complex. For Figure 4, the word "significance" was recorded in the declarative and interrogative contexts given above. Sound reproduction showed a fully acceptable American pronunciation of the two sentences.

The frequency measurements were in complete disagreement with the hypothesis. First, the frequency difference between unstressed and

stressed /i/ of the declarative "significance" was the inverse of what had been expected, since the first unstressed /i/ had a slightly higher average mid-vowel frequency (126 Hz) than the stressed /i/ (120 Hz). Second, the interrogative form showed the same tendency, just more so (140 Hz for the unstressed, 114 Hz for the stressed /i/). It turns out in this example that interrogative intonation is not only characterised by a rising intonation, but also by a greater 'dip' in the fundamental frequency prior to the word-final rise in intonation. This 'dip' is often found in multi-syllabic words with non-initial stress.

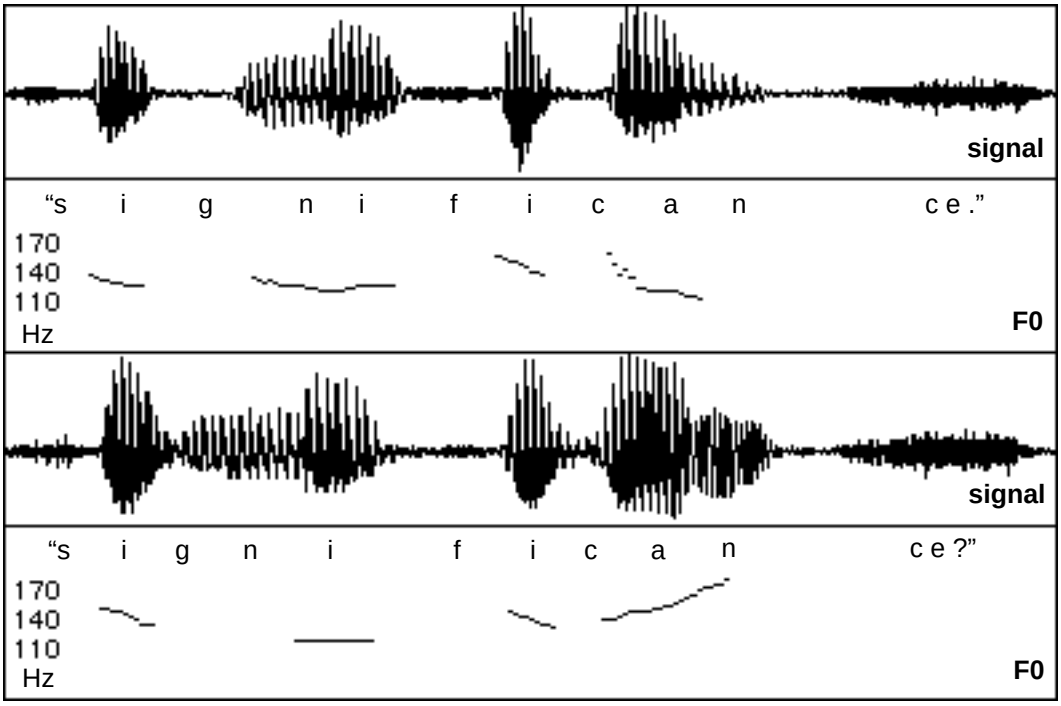

Figure 4. The word "significance" at the end of a declarative (above) and interrogative (below) sentence. The stressed /i/ shows a lower average F_0 in the interrogative than in the declarative context. The example shows that under certain circumstances such as pre-sentence-final 'F_0-dip', stress can be characterised by particularly low F_0.

The consequences of such intricate F_0 interactions are manifold. First of all, intonation is not at all simple to predict. The research of the last thirty years has shown that F_0 structure is highly complex and shows great inter- and intra-speaker variability. It is subject to variations due to the number of syllables, placement of main and secondary stress, placement within the sentence and interactions with intonational variables. To top it off, the analyst can bank on few procedural certainties: it is often unclear which speech segments should be measured (should one include or exclude nasal and voiced fricative segments?), and which are "accepted" or even "acceptable" categorisations of intonational primitives.

Second, intonation as used in most speech synthesis is still quite rudimentary. Few systems implement intonational modulations that go

much beyond the obligatory question-final F_0 rises. Fewer still implement believable word-level modulations. There is still much to be done to formalise resistant algorithms that predict intonational modulations with any degree of believability.

Rhythm, Speech Rate and Other Durational Effects

A third prosodic phenomenon relates to variations in speed of speech production. This gives rise to different perceptual impressions, depending on the length of the stretch of speech that is modified.

If an entire utterance or discourse is spoken either at a fast, comfortable or slow speed, this corresponds to a modification of *rhythm* or *speech rate*.

If time variations are of local nature, the durational effects are likely related to *stressing (accentuation)*. Local slowing, resulting in an increase in the duration of an entire word or at least some of its syllables, is representative of stressing. Even when pronounced without particular loudness, a slowed word tends to signal the particular importance of the content conveyed by the word. On the other hand, local acceleration signifies lessened semantic importance.

In the context of automatic speech processing, it is important to note that neither local nor global rate modifications are entirely linear. That is, it is not possible to produce entirely natural-sounding speech by simply accelerating or decelerating normal speech by a fixed rate. Speech rate modifications affect vocalic segments a great deal more than consonantal segments of speech, and within consonantal segments, VOTs of stop consonants and the duration of transitional portions of V-C or C-V boundaries are less variable than the durations of fricatives or nasals.

Tone

In tone languages (a well-known example is Mandarin Chinese), certain words are distinguished from others only by the direction and the contour of F_0 change. In these cases, melodic oppositions have phonemic value: word meaning is established by intonation. This distinctive use of a prosodic marker is exceptional, because generally, prosodic phenomena tend to modify words, phrases or sentences that have a meaning independently of their concrete prosodic realisation (Figure 5).

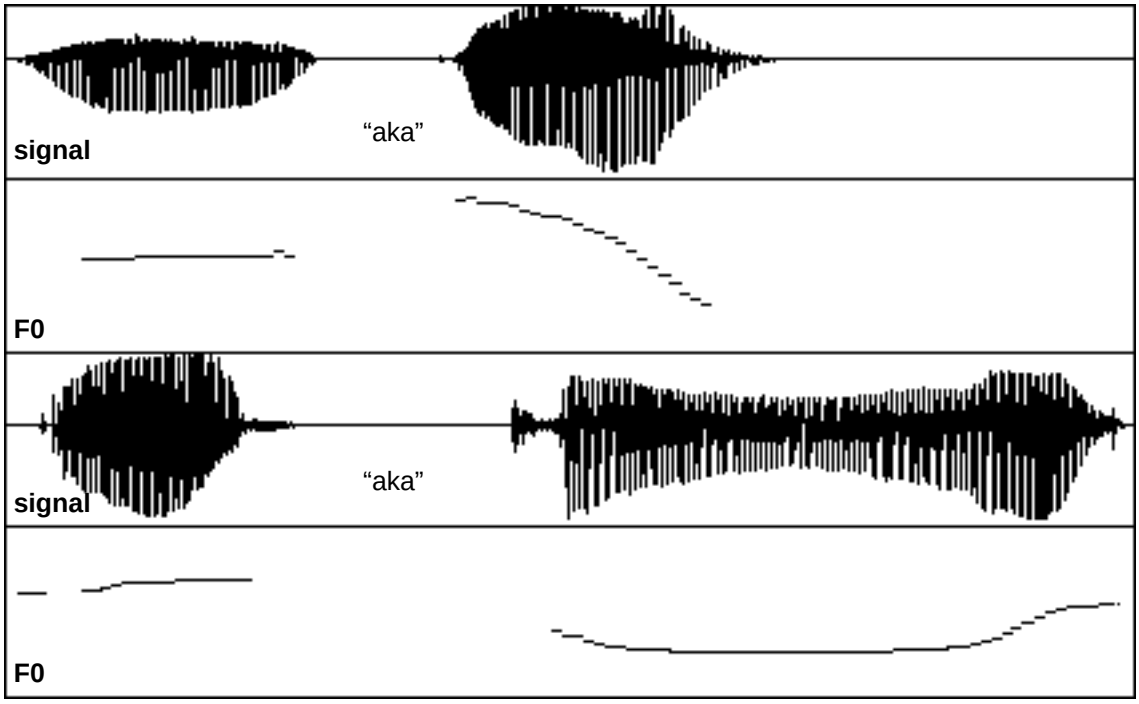

Figure 5. Two words with different significance from a tone language (Mandarin Chinese). The two words are distinguished solely by their intonational pattern.

Junctures

At transitions between words, there are often specific rules at work for stress placement and pausing, as well as for local modification of the segment structure.

A common finding is that pauses between words (if they do occur) tend to get longer as the phrase proceeds. The longest pauses are generally found between phrases (see Chapter 3, this volume). However, junctures are not always handled in the same manner. For instance, many languages suppress pauses between a word-final consonant and a word-initial vowel: French “il a” -> /ila/, not /il#a/, in accordance with the so-called “liaison rule”. Such *liaisons* are obligatory in certain contexts (e.g. between a pronoun and the succeeding verb), but are prohibited in other contexts. For instance between two vowels enclosing a word boundary, the liaison is prohibited and a glottal stop is inserted: “des haricots” -> /de?aʀiko/.

Microprosody

On a lower level, prosody is also influenced by segmental structure. It can be shown, at least for laboratory speech, that F_0 (and sometimes other prosodic parameters as well) are influenced by the surrounding sounds

(e.g. voiced *vs.* voiceless consonant, stops *vs.* fricatives) and by the sound type (e.g. low *vs.* high vowel). For example, the initial F_0 of a vowel tends to be lower in post-plosive than in post-fricative position, and vowel duration are generally shorter in voiceless than in voiced contexts. These effects, however, are very difficult to trace in continuous speech, due to the more dominant manifestations of higher-level prosody.

Prosodic Universals and Language-Specific Differences

A common question in automatic speech processing concerns the universality of prosodic phenomena: Can a prosodic speech synthesis component for language x handle prosodic phenomena in language y? Indeed, many prosodic phenomena operate in a nearly indistinguishable manner in the various languages. Some of these probably have their roots in neurological and physiological aspects of human speech processing. Other phenomena, however, pattern quite differently from language to language (for a review of this question, see Hirst and Di Cristo, in press). Here is a short overview of the most common observations.

Physiologically-Based Universals: Declination and Range Reduction

Among the phenomena with apparent physiological roots, one may count "declination" and "range reduction". Within a breath-group, F_0 shows a general tendency for a decline, and at the beginning of each breath-group, the fundamental frequency starts again at a higher frequency (the "baseline reset") (Figure 6). Since this declination runs parallel with a gradual loss in subglottal pressure, it is quite likely that it is physiologically based. Indeed, declination lines have been documented for a large number of languages.

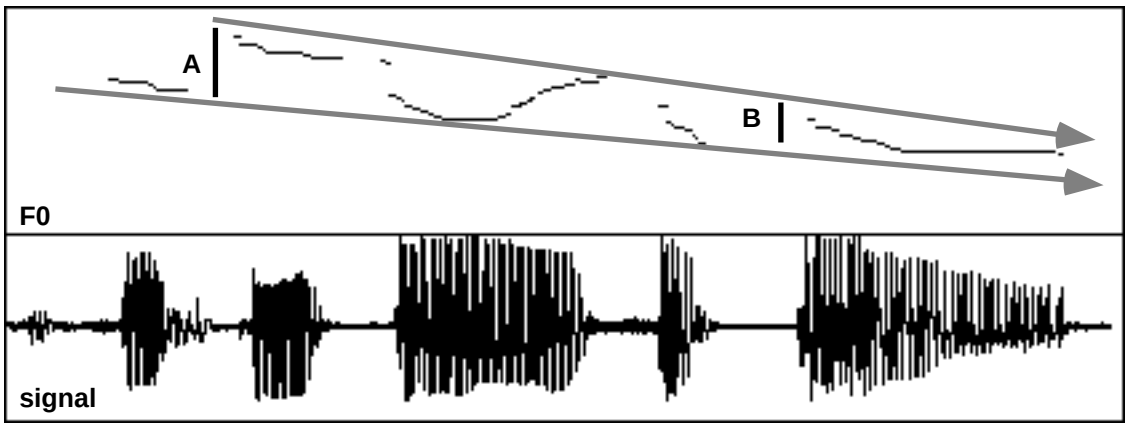

Figure 6. F_0 declination trend and reduction in F_0 range in the French sentence "certains pays se plaignent". The range at A is larger than the range at B, later on in the sentence.

Within a breath-group, the *variation* of the fundamental frequency also decreases. Specifically, it is found that the upper limit of the F_0 range decreases faster than the lower bound (which, in fact, often does not decrease at all). It can be assumed that at the reduced subglottal pressures found at the end of a breath group, the glottal tension necessary to achieve high fundamental frequencies may be more difficult to obtain.

Other Generalised Observations

Some other observations of prosodic phenomena are generalised over a large number of languages without apparent or evident connection with physiological origins. Among these are *word grouping* , *syllabification* and the *content word vs. function word opposition*.

Word grouping. In spontaneous speech, words can generally be observed to be grouped according to prosodic principles (Figure 7). For example, words connected (syntactically or otherwise) to a certain stress-bearing word tend to show strong prosodic affinity. In French read speech, for example, particularly long pauses, strong rises in F_0 and drops in amplitude typically coincide at the end of phrase groups (Keller *et al.*, 1993). Such groups are variously called “tone groups”, “stress groups”, “prosodic phrases” or “prosodic words”. It remains that these tendencies can be broken, because a speaker may choose to change the grouping deliberately.

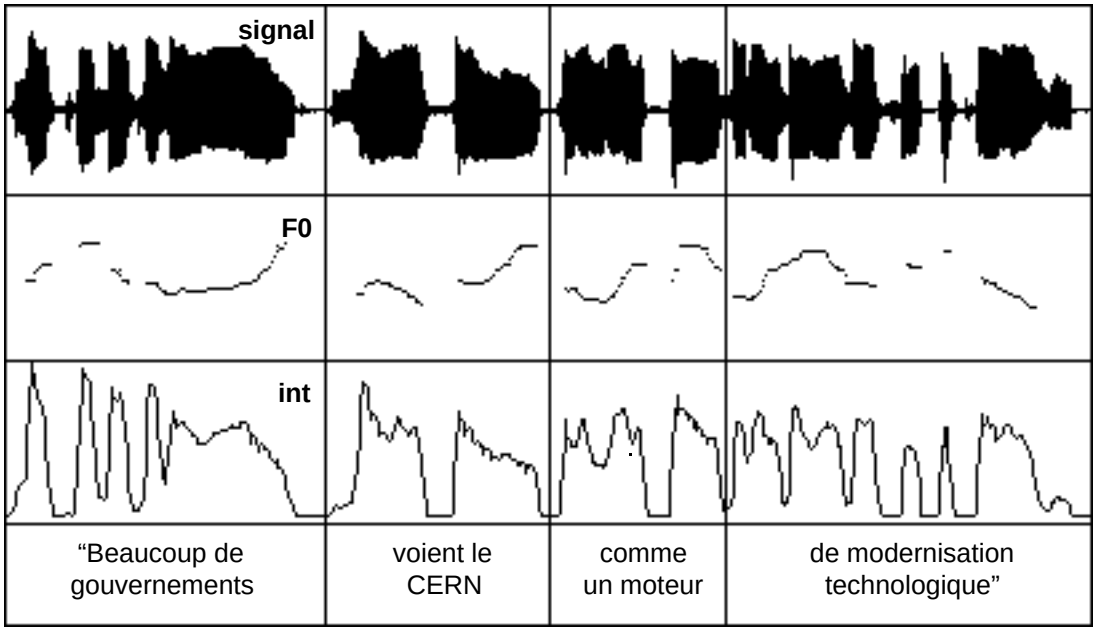

Figure 7. Word grouping in French. English word grouping is nearly indistinguishable.

Syllabification: Individual syllables are often clearly discernible from the acoustic signal (Figure 8). Most easily they are identified in a display of intensity over time where every major peak corresponds to one syllable. This phenomenon has been observed in every language analysed so far.

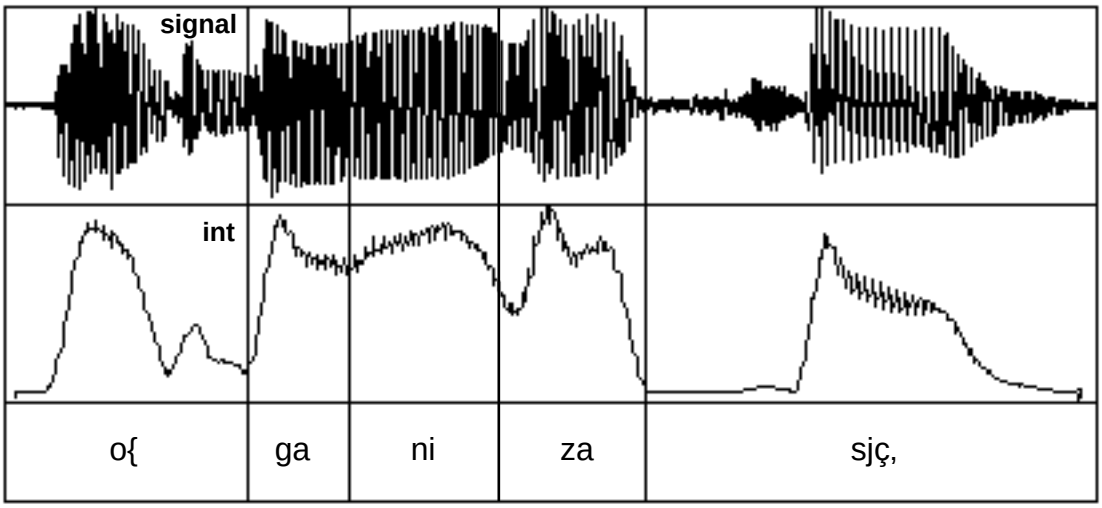

Figure 8. Syllable boundaries in the French word "organisation", as illustrated by features of the amplitude envelope (RMS) curve.

Content word vs. function word opposition: Content words, i.e. lexemes carrying an independent semantic meaning (like nouns, verbs, adjectives) are more likely to contain stressed syllables than function words, or, more generally speaking, grammatical lexemes and morphemes (like articles, conjunctions, suffixes, inflectional morphemes etc.). This dichotomy, however, is not categorical – exceptions occur and various morpheme classes can be ordered according to their gradual probability of receiving stress rather than according to a rigorous classification into one of two groups (Grosjean and Gee 1987).

Language-Specific Prosodic Differences

At the same time as there are many widespread prosodic phenomena, prosody structure is not the exactly same for every language. Differences that one should be on the lookout for in the context of automatic speech processing concern the exact manifestation of stress, the placement of stress and pauses, the communicative functions of intonation patterns, and the interaction between the parameters F_0, intensity and duration.

The Relevance of Prosody for Speech Synthesis and Speech Recognition

Prosody plays an important role in speech synthesis, both in terms of intelligibility and naturalness, which rely heavily on the assignment of correct prosodic structures. (The two concepts in fact seem to be positively correlated, cf. Carlson *et al.*, 1979). F_0, intensity and duration have to be modelled in accordance with information from segment structure, lexicon, syntax and semantics. Inversely, speech recognition needs to extract prosodic information in order to produce correct semantic and syntactic interpretations and successfully identify words. How, then, can one provide and retrieve prosodic knowledge for synthesis and recognition? What are the rules or constraints, and where can they be derived from?

These questions can only be answered by an examination of the way a particular language implements the essential features of prosody, which are *grouping* and *emphasis*. It has to be verified whether boundaries between groups of words are marked by temporal variation, by local modification of stress (e.g. so-called de-accentuation), by global intonation patterns or by a combination of all these — to name but one typical example. This knowledge, combined with rules covering universal prosodic features, then has to be translated into systems of control or measurement parameters.

An example of a comparatively simple set of control parameters for intonation contours in speech synthesis is the following list, taken from the description of a multi-language text-to-speech system (Olaszy, 1991[4]), with sample values for a male voice:

Table 1. Sample control parameters for F_0 control in speech synthesis (Olaszy 1991).

Parameter	**min**	**max**	**unit**
F_0 starting point	95	125	Hz
F_0 end point	95	125	Hz
degree of F_0 rise	5	25	%
degree of F_0 fall	5	25	%
steepness	0.25	2	Hz / ms
jump up	0	+30	Hz
jump down	0	-30	Hz
vowel lengthening	1.5	3	times

Prosodic parameters like these are influenced by a complex variety of factors. To illustrate this complexity, figure 9 provides an overview of

4 Many more interesting examples of models and parameters can be found in this and the other four volumes of the ICPhS XII proceedings.

linguistic and phonetic factors involved in determining syllable-level prosody.

Which of these factors are more relevant than others and how they translate into combined measures of F_O, intensity and duration, still remains largely to be explored. To complicate things even further, many influences have to be understood as two-way relationships; not only is prosody influenced by segment structure, but segmental features — like reduction and coarticulation — are often co-determined by prosodic features. In fact, this is probably more frequent than is traditionally acknowledged (cf. Fant, 1991).

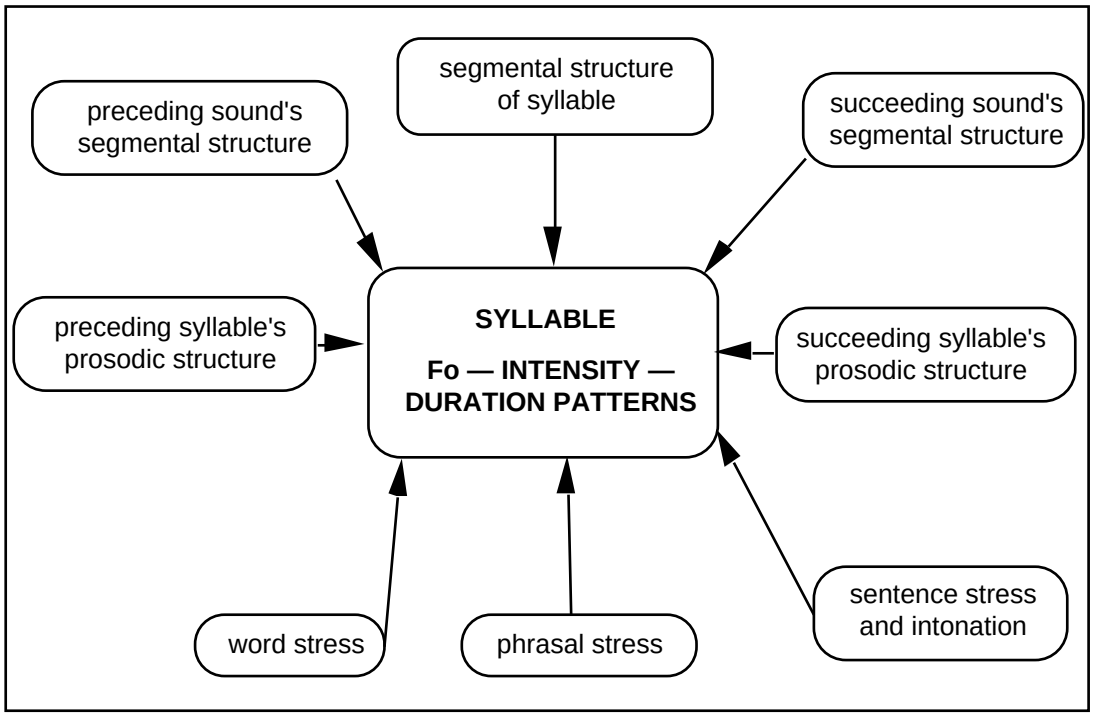

Figure 9. Linguistic and phonetic factors influencing syllable prosody.

Conclusion: Some Terminological Remarks and Relationships to Other Linguistic Domains

Prosodic phenomena are both wide-ranging and pervasive in speech. Consequently, speech synthesis without good prosodic component lacks a "certain human quality" and is quite tedious to listen to. Also, the speech signal is considerably modified by prosodic markers, and thus poses difficulties to speech recognition devices that are not programmed to look for phenomena extending over long stretches of speech. It is therefore imperative that practitioners of automatic speech processing become familiar with this domain.

This is more easily said than done. Writings in this field are full of terminological fuzziness, and work authored by persons with linguistic

training is often totally esoteric to outsiders. We therefore end this chapter with a few remarks concerning terminology and links with various current forms of linguistic theory.

With respect to the terminology used in prosody, it must be admitted that there is still no such thing as a "standardised technical terminology" for prosody. To help in diffusing some of the confusion, here is an incomplete list of terms used in the literature with a particular diversity of meanings, terms that are particularly prone to misinterpretation.

In addition to the *syllable* (whose own definition is not universally agreed on), there are a number of other linguistic entities suggested as basic units of prosody: foot, stress foot, mora, tone group, intonational group, stress group, syntagm, rhythmic unit, prosodic word, prosodic phrase. In general, the most frequently used terms, like foot or tone group, are also the ones with the most problematic or even contradictory definitions. One way of injecting a greater degree of objectivity into these definitions is to base them on easily measurable criteria which are found in a large number of languages. For example, the "prosodic phrase" might usefully be defined by the presence of exceptionally long pauses, since in several languages, pauses of twice to four times the usual duration have been observed to occur at places that also corresponded to the end of a major syntactic structure. Generally speaking, a more empirically-oriented approach to these definitional problems would be all for the better.

Accent and stress both refer to basically the same phenomenon, i.e. emphasis or, more generally, prominence. But in the literature they are seldom used as synonyms. Very often, word-level prominence is referred to as word stress, whereas accentuation on the sentence level is called "sentence accent" in this framework[5]. Another distinction has to do with the actual physical realisation, e.g. pitch accent *vs.* stress (produced by means of duration and intensity).

Also the more detailed classification of sentence and phrasal accent (or stress) is not straightforward. There are no generally agreed-upon distinctions between focal, emphatic, contrastive, highlighting, etc., accents.

Tone: It is important not to confuse the tones of a tone language with intonational phenomena in non-tone languages that are sometimes also called "tones" (or tone levels or patterns).

Lines: In addition to the above-mentioned baseline and declination line, researchers also postulate a number of other abstract lines: top line, level line, reference line, grid line, focus line, and so on. Definitions can vary according to the theoretical framework and / or the experimental set-up employed.

[5]Again, prominent exceptions exist, see for example Gårding (1983).

With respect to other domains of linguistics, prosody is intricately connected with all other structural levels of language. The following are several examples of areas where phonetic research in prosody and linguistic theory interact.

– The correct identification of sentence stress is crucial for semantic and pragmatic analysis. On the other hand, information from semantics and pragmatics is needed for successful interpretation of sentence stress (e.g. emphasis *vs.* contrast).

– Prosodic identification of different phrase and sentence types as well as the marking of phrase boundaries and grouping of words interact with the syntactic organisation of the utterance. There have been many endeavours to predict prosodic phenomena either solely on the basis of syntactic structure or integrating additional information from lexicology and morphology. But more recently, evidence in apparent contradiction with syntactic structures has been taken into account in new theories of the structuring of speech (e.g. Gee and Grosjean, 1983; Beckman and Edwards, 1990).

– New approaches in phonological theory, auto-segmental (first conceived in Goldsmith, 1976) and metrical (both grid and tree versions first proposed in Liberman and Prince, 1977) phonology, owe much to inspiration from prosody: Classic segmental phonological analysis could not easily account for such phenomena as tone or stress distributions. Phonological frameworks and rule systems in turn motivate the investigation of specific prosodic facts whose relevance otherwise might still remain unobserved and/or facilitate the systematic aquisition and storage of prosodic data (like the Delta system, cf. Hertz, 1990).

References

Beckman, M.E., & Edwards, J. (1990). Lengthenings and shortenings and the nature of prosody. In J. Kingston, & M.E. Beckman (Eds.), *Between the Grammar and Physics of Speech (Papers in Laboratory Phonology I)* (pp. 152-178). Cambridge: Cambridge University Press.

Carlson, R., Granström, B., Klatt, D.H. (1979). Some notes on the perception of temporal patterns in speech. In B. Lindblom, & S. Öhman (Eds.), *Frontiers of speech communication research* (pp. 233-244). London: Academic Press.

Fant, G. (1991). What can basic research contribute to speech synthesis? *Journal of Phonetics, 19,* 75-90.

Fromkin, V.A. (1987). A note on the suprasegmental representation of prosody. In R. Shannon & L. Shockey (Eds.), *In Honor of Ilse Lehiste* (pp. 99-107). Foris: Dordrecht.

Gårding, E. (1983). A generative model of intonation. In A. Cutler, & D.R. Ladd (Eds.), *Prosody: Models and measurements* (pp. 11-25). Berlin: Springer.

Gee, J.P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology, 15,* 411-458.

Grosjean, F., & Gee, J.P. (1987). Prosodic structure and spoken word recognition. In U.H. Frauenfelder, & L.K. Tyler (Eds.), *Spoken word recognition* (pp. 135-155). Cambridge, MA: MIT Press.

Goldsmith, J. (1976). An overview of autosegmental phonology. *Linguistic Analysis, 2*, 23-68.

Hertz, S. (1990). The Delta programming language: An integrated aproach to nonlinear phonology, phonetics, and speech synthesis. In J. Kingston, & M.E. Beckman (Eds.), *Between the grammar and physics of speech (Papers in Laboratory Phonology I)* (pp. 215-257).

Hirst, D., & di Cristo, A. (in press). *Intonation systems: A survey of twenty languages.*

Keller, E., Zellner, B., Werner, S., and Blanchoud, N. (1993). The prediction of prosodic timing: Rules for final syllable lengthening in French. In D. House, & P. Touati (Eds.). *Proceedings of an ESCA workshop on Prosody (Working Papers 41).* (pp. 212-215) Lund: Dept. of Linguistics and Phonetics, Lund University.

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

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry, 8*, 249-336.

Olaszy, G. (1991). A crosslinguistic description of intonation contours of a multilanguage text-to-speech system. In *Actes du XIIème Congrès International des Sciences Phonétiques. Vol. 4* (pp. 210-213). Aix-en-Provence: Université de Provence.

Further reading

Docherty, G., & Ladd, D.R., (Eds.). (1992). *Gesture, segment, prosody (Papers in Laboratory Phonology II).* Cambridge: Cambridge University Press.

Cutler, A., & Ladd, D.R., (Eds.). (1983). *Prosody: Models and measurements. (Springer Series in Language and Communication 14).* Berlin: Springer.

Halle, M., & Vergnaud, J.-R. (1987). *An essay on stress*. Cambridge, MA: MIT Press.

House, D., & Touati, P. (Eds.). (1993). *Proceedings of an ESCA workshop on prosody (Working Papers 41).* Lund: Dept. of Linguistics and Phonetics, Lund University.

t'Hart, J., Collier, R., & Cohen, A. (1990). *A perceptual study of intonation: An experimental-phonetic approach to speech melody (Cambridge studies in speech science and communication).* Cambridge: Cambridge University Press.