ELSEVIER

Speech Communication 33 (2001) 263–296

SPEECH COMMUNICATION

www.elsevier.nl/locate/specom

# Developments and paradigms in intonation research

Antonis Botinis [a,*], Bjorn Granström [b], Bernd Möbius [c]

[a] *Department of Languages, University of Skövde, P.O. Box 408, S-541 28, Skövde, Sweden*
[b] *Department of Speech, Music and Hearing, Royal Institute of Technology (KTH), Stockholm, Sweden*
[c] *Institute of Natural Language Processing, University of Stuttgart, Stuttgart, Germany*

## Abstract

The present tutorial paper is addressed to a wide audience with different discipline backgrounds as well as variable expertise on intonation. The paper is structured into five sections. In Section 1, "*Introduction*", basic concepts of intonation and prosody are summarised and cornerstones of intonation research are highlighted. In Section 2, "*Functions and forms of intonation*", a wide range of functions from morpholexical and phrase levels to discourse and dialogue levels are discussed and forms of intonation with examples from different languages are presented. In Section 3, "*Modelling and labelling of intonation*", established models of intonation as well as labelling systems are presented. In Section 4, "*Applications of intonation*", the most widespread applications of intonation and especially technological ones are presented and methodological issues are discussed. In Section 5, "*Research perspective*" research avenues and ultimate goals as well as the significance and benefits of intonation research in the upcoming years are outlined. © 2001 Elsevier Science B.V. All rights reserved.

## Zusammenfassung

Dieser Überblicksartikel richtet sich an eine breite Leserschaft aus unterschiedlichen Disziplinen und mit unterschiedlichem Kenntnisstand hinsichtlich der Intonationsforschung. Der Beitrag ist in fünf Abschnitte gegliedert. In Abschnitt 1, "*Introduction*", werden die grundlegenden Konzepte der Prosodie und Intonation und die wichtigsten Richtungen der Intonationsforschung skizziert. Eine Bandbreite von Funktionen der Intonation, von der morpholexikalischen und Phrasenebene bis zur Ebene des Diskurses und Dialogs, werden in Abschnitt 2, "*Functions and forms of intonation*", diskutiert, und Formen der Intonation werden anhand von Beispielen aus verschiedenen Sprachen illustriert. In Abschnitt 3, "*Modelling and labelling of intonation*", werden etablierte Intonationsmodelle sowie Systeme der Annotation und Etikettierung intonatorischer Merkmale präsentiert. Die am weitesten verbreiteten Anwendungen der Intonation, mit einem Schwerpunkt auf der Sprachtechnologie, werden in Abschnitt 4, "*Applications of intonation*", vorgestellt, und es werden einige methodische Probleme diskutiert. Schließlich werden in Abschnitt 5, "*Research perspectives*", mögliche Richtungen und Ziele für zukünftige Forschungsarbeiten auf dem Gebiet der Intonation aufgezeigt. © 2001 Elsevier Science B.V. All rights reserved.

## Résumé

Cet article de synthèse s'adresse à un large public provenant de différentes disciplines mais également aux spécialistes de l'intonation. Il comprend cinq parties. Dans la première partie, ou *Introduction*, sont brièvement rappelés les concepts de base de l'intonation et de la prosodie et mises en lumière les périodes charnières de la recherche intonative.

[*] Corresponding author. Tel.: +46-500-448-915; fax: +46-500-448-949.
*E-mail address:* antonis.botinis@isp.his.se (A. Botinis).

0167-6393/01/$ - see front matter © 2001 Elsevier Science B.V. All rights reserved.
PII: S 0 1 6 7 - 6 3 9 3 ( 0 0 ) 0 0 0 6 0 - 1

Dans la deuxième partie, *Fonctions et formes de l'intonation*, un large éventail des fonctions des niveaux morpholexical, phrastique, discursif ou dialogal est examiné; des formes intonatives provenant de langues différentes sont présentées et comparées. Dans la troisième partie, *Modélisation et transcription de l'intonation*, on fait référence aux modèles courants de l'intonation et aux systèmes d'étiquetage opérationnels. Dans la quatrième partie, *Les applications de l'intonation*, sont présentées les applications les plus courantes de l'intonation, plus particulièrement les applications technologiques; une discussion est engagée sur les problèmes méthodologiques. Dans la cinquième partie, *Perspectives de recherche*, des directions de recherche sont tracées; on précise les buts à atteindre et on souligne le sens et l'intérêt de la recherche intonative pour les années à venir.

*Keywords:* Pitch; Tone; Stress; Accent; Intonation models; Intonation applications

## 1. Introduction

The present tutorial paper aims to outline developments and paradigms of intonation research and discuss related contributions to the field. Intonation structures and intonation analyses in languages with different prosodic typology, Chinese, Greek and Swedish in the first place, will be presented with actual examples in relation to the relevant discussion. After a general *Introduction*, the paper develops with *Functions and forms of intonation*, *Modelling and labelling of intonation*, *Applications of intonation* and is concluded with *Research perspectives*. Topics of intonation research which are, to our mind, fairly representative and have produced substantial knowledge are considered but no comprehensive review, either contemporary or historical, is intended whatsoever. General overviews of intonation are found in, among others, Cruttenden (1997), Hirst and Di Cristo (1998b), Ladd (1996) and Rossi (1999, 2000). Intonation description and analysis of different languages, on the other hand, are reported in, e.g., American English (Pierrehumbert, 1980), Danish (Grønnum, 1998), Dutch (t'Hart et al., 1990), French (Di Cristo, 1998), German (Möbius, 1993), Greek (Botinis, 1989), Japanese (Pierrehumbert and Beckman, 1988) and Swedish (Bruce, 1977).

### 1.1. Definition of intonation

Intonation is defined as the combination of tonal features into larger structural units associated with the acoustic parameter of *voice fundamental frequency* or $F_0$ and its distinctive variations in the speech process. $F_0$ is defined by the quasi-periodic number of cycles per second of the speech signal and is measured in Hz. The production of intonation is defined by the number of times per second that the vocal folds complete a cycle of vibration and is controlled by muscular forces of the larynx, which determine the tension of the vocal folds, as well as aerodynamic forces of the sublaryngeal (respiratory) system. The perception of intonation is defined by the perceived *pitch*, which roughly corresponds to $F_0$ realisations.

Despite the acoustic and perceptual definition of $F_0$ and pitch, respectively, these two terms are used rather interchangeably in much of the international literature. On the other hand, *intonation* and *prosody* may also be found in an interchangeable use, although, most usually, the term intonation is confined to tonal ($F_0$) features specifically, whereas the term prosody, in addition to tonal features, involves temporal (duration) and dynamic (sound pressure level) features. Furthermore, in a broad sense, both local and global tonal distribution may be referred to as intonation whereas, in a narrow sense, only global tonal distribution is referred to as intonation proper and inherently lexical tonal features are attributed to the area of prosody.

In related as well as unrelated languages, intonation and tonal features in general may have very similar or completely different *linguistic* functions, which may vary from morphological and lexical levels to phrase and sentence as well as discourse and dialogue levels. Intonation may also have a *paralinguistic* function and be most characteristic with reference to various expressive functions, such as surprise, anxiety and threat, as well as an *extralinguistic* (or *non-linguistic*) function with reference to personal characteristics and indexing, such as sex, age and socio-economic status.

### 1.2. Background of intonation

Aspects of intonation and tonal distinctions have been studied throughout man's literate history. In classical Greece e.g., Plato and Aristotle discuss the prosodic system and basic questions about accentual distinctions are raised. The term itself is derived from the Greek "tónos" (tension) through the Latin "intonatio" and old French "intonation". In modern times, intonation has been studied extensively from both a theoretical and experimental point of view.

In the framework of structural linguistic theory, particularly in the first half of the 20th century, formal descriptions of phonological systems and distinctions among different languages are established and the role of prosody in linguistic analysis and theory is discussed (Bloomfield, 1933; Trubetzkoy, 1939; Martinet, 1954; Hockett, 1955; Malmberg, 1967). Stress, pitch, and juncture variations are classified and up to four pitch levels are distinguished (cf. Pike, 1945; Trager and Smith, 1951). The role of intonation in linguistic theory is also emphasised in the framework of generative (transformational) grammar, especially in the second half of the 20th century, with mainstream work on the relation between intonation and syntax as well as semantics (e.g., Bresnan, 1971; Chomsky, 1972; Jackendoff, 1972; Stockwell, 1972). Tonal analysis also has a central place in autosegmental phonology, i.e., post-generative phonology (see Goldsmith, 1976a,b, 1990). Intonation and information structure relations are also investigated and basic thematic notions with intonation correlates denoting the most important part of the utterance are brought to light (e.g., Bolinger, 1958; Daneš, 1960; Halliday, 1967; Lambrecht, 1996). The ultimate interpretation of an utterance into a given context, i.e., the pragmatics of intonation, as well as the relation of intonation to the intended meaning, in the broad framework of the speech act theory (see Searle, 1969, 1976, 1979), has also drawn considerable attention in the study of intonation. Intonation is further widely acknowledged with reference to the organisation of text, discourse and dialogue as well as various interactive functions and considerable research is being carried out with reference to these areas (Brazil et al., 1980; Brown et al., 1980; Brown and Yule, 1983; Brazil et al., 1997).

Experimental phonetics, although dating from earlier times in Europe and the USA (e.g. Wheatstone, 1837; Helmholtz, 1877; Bell, 1879), has had a crucial turning point with the historic invention of the spectrograph in the 1940s (reported in Potter, 1945; Koenig et al., 1946; Potter et al., 1947). Ever since, the steady development of sophisticated laboratory devices and the increased interest on prosodic phenomena have contributed to an upstepping of research on the main aspects of intonation. Thus, significant research has been conducted on the physiology of intonation (Ladefoged and McKinney, 1963; Ladefoged, 1967; Lieberman, 1967; Collier, 1975; Atkinson, 1978; Ohala, 1978), the acoustics of intonation (Hadding-Koch, 1961; Lehiste, 1970; Cooper and Sorensen, 1981), as well as the perception of intonation (Fry, 1958; Hadding-Koch and Studdert-Kennedy, 1964; Rossi, 1978; t'Hart et al., 1990). Although the main bulk of research has been conducted on functional aspects of intonation, microprosodic effects, i.e., the effects of different segments on the realisation of $F_0$, have also been studied extensively (see Lehiste, 1970; Di Cristo and Hirst, 1986; Fischer-Jørgensen, 1990; Whalen and Levitt, 1995; Fourakis et al., 1999). Reports and publications with intonation data have grown immensely during the past two or three decades and the repertoire of knowledge on a considerable number of different languages is getting steadily bigger. Contrastive and dialectal studies are also accelerating research areas and language-dependent as well as language-independent intonation features are continuously brought into light (Gårding, 1977a; Bruce and Gårding, 1978; Gårding et al., 1982; Beckman and Pierrehumbert, 1986; Vaissière, 1995).

Accumulated knowledge on the nature of intonation, mostly on forms and functions, resulted to the development of intonation models with predictive power which, formalised mainly in the 1970s (e.g. Bruce, 1977; Thorsen, 1978;

Pierrehumbert, 1980), have been tested and applied to a considerable number of languages (Gårding et al., 1982; Cutler and Ladd, 1983; Hirst and Di Cristo, 1998a). The modelling of intonation, in addition to forms and functions, is closely related to labelling and transcription of intonation for which several systems have been proposed, among them Tone and Break Indices (ToBI), applied initially in American English (see Silverman et al., 1992; also Beckman and Ayers, 1997) and INternational Transcription System for INTonation (INTSINT), applied in several languages (see Hirst and Di Cristo, 1998a).

The significance of intonation is also widely acknowledged with reference to speech and language technology, speech pathology and phoniatrics, as well as applied linguistics and language education. In speech synthesis, e.g., the contribution of intonation is not only confined to the intelligibility of tonal and prosodic distinctions such as stress, focus and phrase boundaries but has a decisive effect to the naturalness of the system as a whole (for evaluation of speech synthesis see, among others, Monaghan and Ladd, 1990; Van Santen, 1993; Véronis et al., 1998; Tatham and Morton, 2000; Di Cristo et al., 2000; for an overview of speech synthesis systems in Europe see Monaghan, 1998).

### *1.3. Study and research of intonation*

Intonation is mainly treated in university course books in phonetics and linguistics and is taught in all its major dimensions, from abstract to concrete representations and relations, with reference to phonology as well as speech physiology, speech acoustics and speech perception. Traditionally, intonation and prosody have been taught as part of larger courses, but nowadays there are autonomous courses in these areas in many Universities.

There are even established teaching and research centres, i.e., "schools of prosody" in a traditional sense, where a respectful tradition has been developed and the state-of-the-art in the study of prosody is continuously being pushed forward. Department of Phonetics at Provence University (Aix-en-Provence, France), Institute for Perceptual Research (IPO, Einhoven, Netherlands), Department of Linguistics and Phonetics at Lund University (Lund, Sweden), Department of Linguistics at the Ohio State University (Columbus, USA), and AT&T Research Labs and Lucent Technologies Bell Labs (New Jersey, USA) ought to be mentioned. Significant research is however being carried out in phonetics laboratories and linguistics departments in most of Europe and the USA as well as by individual researchers and research groups worldwide.

Intonation is a central concern for many established as well as emerging disciplines ranging from theoretical linguistics and experimental phonetics to computer sciences and signal processing. As any phonetics area, intonation has interdisciplinary dimensions with reference to speech physiology, speech acoustics, and speech perception, which are related to human anatomy, physics and auditory sciences, respectively. On the other hand, intonation applications are related to language technology, language pathology and language education. In short, intonation studies and research may be mostly found in phonetics and linguistics departments and, to a considerable degree, in various language, technology, and medicine university departments. In addition, research and applications of intonation are carried out in research institutes and industrial companies with reference to high technology speech and language products.

## 2. Functions and forms of intonation

Intonation may have a continuous form with a complex structure, as a result of different contributions related to linguistic as well as paralinguistic and extralinguistic functions. The relation of function and form, much like other aspects of phonetics, is one-to-many and the question of invariance and variability is an actual theme in intonation studies too. The speech communication contexts in which intonation may have a distinctive function are limitless in principle, although the forms are assumed to be of a finite nature with variable recurrent combinations, which is an object of current intensive research. The analysis of intonation is based on the decomposition of complex structures with fairly continuous forms into local and global features, which may have discrete linguistic functions. Although much remains to be done, there is considerable knowledge on basic aspects of functions and forms of intonation at lexical, phrase, and sentence levels as well as steadily accelerating knowledge at discourse and dialogue levels.

### *2.1. Functions of intonation*

The main functions of intonation are centred round the notions of *prominence*, *grouping* and *discourse*, which are related to various grammatical components as well as linguistic levels. Prominence is related to weight structuring of linguistic units such as syllables and words. Grouping is related to coherence and segmentation structuring of speech units into prosodic units. Discourse is related to structuring of prosodic units with reference to topics of discussion and turn regulations between speakers involved in a conversation. Intonation is thus involved in linguistic structuring with variable distinctive functions in accordance with the level of application.

#### *2.1.1. Lexical functions*

Intonation may have a distinctive function at lexical level with reference to the prosodic categories of *tone*, *stress* and *accent* and thus languages with corresponding distinctions are often referred to and classified as *tone languages* (such as Chinese, Thai and Vietnamese), (*dynamic*) *stress languages* (such as Greek, Italian, Russian and Spanish) and (*pitch*) *accent languages* (such as Japanese and Swedish). Although the overwhelming majority of the European languages are stress languages, tone languages are widespread in many parts of the world including Asia, Australia, Africa and America whereas accent languages are found in many language families with either stress or tone dominant distinctions. It should be noted, however, that a prosodic taxonomy is mainly indicative rather than restrictive. In Chinese, e.g., in addition to tones there are stress distinctions whereas, in Swedish, there are accent as well as stress distinctions. On the other hand, Japanese has only accent distinctions.

*2.1.1.1. Tone distinctions.* Tone is associated with different tonal patterns within a syllable, which may have a distinctive function with reference to *static* and *dynamic* tones (e.g., low versus high or rising versus falling) in lexicon/morphology. Chinese, e.g., regardless of the neutral tone, has four distinctive tones, H, R, L, and F, also referred to as Tones 1, 2, 3, 4, respectively, and determined by the *height* and *shape* of the tonal patterns. These tones are described as *high-level* (H), *mid-rising* (R), *low-dipping* (L), and *high-falling* (F). Polysyllabic words may have different combinations of contrastive tones in a row whereas monosyllabic words may have all four contrastive tones with the corresponding transcription such as *mā* 'mother', *má* 'hemp', *mǎ* 'horse', and *mà* 'scold'. Thus, tone, much like segmental but not prosodic distinctions in general, is in a *paradigmatic* contrast. Tones in languages where the contrast is based on the height of the tones are known as *register tones* whereas tones involving contrastive shapes are known as *contour tones*.

*2.1.1.2. Stress distinctions.* Stress is associated with syllabic prominence which, in combination with other prosodic features (i.e., duration, intensity, and vowel quality), may have a *distinctive function* in lexicon/morphology. The word *ˈnomos* 'law' in Greek e.g., with stress in the first syllable, has a different lexical meaning from *noˈmos* 'county', with stress in the final syllable, and stress distribution is thus more or less free, i.e., has a distinctive function. Greek has a relatively simple stress pattern, i.e., one stress per word at lexical level in principle, whereas other languages such as Germanic may have a complex stress pattern reflecting the lexical and morphological constituency of the words. Traditionally, the more prominent stress is referred to as *main* (or *primary*) *stress* and the less prominent stress as *secondary stress*. In some languages, however, stress distribution is bound on a syllable and has thus a *demarcative function*, i.e. the lexical boundaries may be predicted from the position of stress. Finish and Polish e.g. have stress distribution on the first and second last (penultimate) syllable, respectively. Stress represents a typical *syntagmatic* contrast, i.e., a syllable stands out and is more prominent in relation to other syllables at word domain.

*2.1.1.3. Accent distinctions.* Accent is also associated with syllabic tonal prominence in lexicon/morphology. The tonal pattern is however the distinctive factor, in comparison to stress the tonal pattern of which may have a large variability. In Japanese, accent may have a three-way distinction in series of words like *'kaki* 'oyster', *ka'ki* 'hedge' and *kaki* with first syllable versus last syllable versus accentless distribution. In Swedish, on the other hand, the words *´tanken* 'tank' and *`tanken* 'thought' do have accent distinction on the first syllable, i.e., acute (accent 1) versus grave (accent 2). Furthermore, the first syllable is more prominent than the second one and thus, in Swedish, stress distribution is a basic condition for the application of accent. Accent has a distinct classification in relation to tone and stress with regards to prosodic typology, although a syntagmatic contrast, much like stress, is the most usual description (cf. Garde, 1968).

### *2.1.2. Phrase and sentence functions*

At phrase and sentence levels intonation may be associated with prominence relations and phrasing as well as sentence type variations and distinctions.

*2.1.2.1. Tonal prominence.* Tonal prominence, apart from prosodic distribution at lexical level (see above), is mainly associated with *focus* (and more or less synonymous terms such as *nucleus*, *sentence stress* and *focal accent*) distribution at phrase and/or sentence levels. Focus has been studied from a wide range of perspectives with reference to linguistic theory and thus syntactic, semantic, pragmatic and information structures and representations (see, among others, Gussenhoven, 1984; Rossi, 1985; Ladd, 1996; Lambrecht, 1996; Cruttenden, 1997; Hirst and Di Cristo, 1998a). Focus has mostly been related to presupposition, according to which *focus* designates *new information* and *presupposition shared* or *old* information in a speaker–listener relation (cf. Jackendoff, 1972). The focus/presupposition concept may be found in a similar or closely related use with *rheme/theme*, *comment/topic*, *new/given* and *foreground/background* terminology and thus focus, somewhat simplified, has a highlighting function associated with the most important information in a speech unit. On the other hand, although focus traditionally refers to information structuring, in much of the current international literature as well as here focus simply refers to the prosodic distribution. Emphasis and contrast, with the corresponding prosodic terminology *emphatic stress* and *contrastive stress*, have also a highlighting function with more or less similar use to the focus one (for relevant discussion see Hirst and Di Cristo, 1998a).

*2.1.2.2. Tonal phrasing.* Phrasing (also referred to as *grouping*) is associated with the segmentation of utterances into variable prosodic units and prosodic theory and phonological studies refer to several prosodic categories and units ranging from syllable to utterance (Silkerk, 1984; Nespor and Vogel, 1986). For intonation description and analysis however the *syllable*, *stress group* and *intonation phrase* are mostly acknowledged (cf. Thorsen, 1978; Botinis, 1989; Hirst, 1993; Hirst and Di Cristo, 1998b).

The syllable may be an anchoring point of tonal structures associated with prosodic distinctions at lexical level, i.e., tone, stress or accent. The basic distinction between stressed and unstressed syllables, e.g., may determine the distribution of tonal "turning points", i.e., an abrupt tonal change associated with the stressed syllable, and may thus have a substantial effect on the structure of intonation.

The stress group (and fairly synonymous terms such as *foot*, *tonal unit*, *prosodic word*, etc.) is the immediate prosodic unit above the syllable and refers to a stressed syllable and any unstressed syllable(s) up to, but not including, the next stressed syllable. A prosodic sequence of stress and unstressed syllables is thus structured into stress groups, which may correlate with distinct tonal gestures the onset and offset of which are aligned with the beginning and end of the corresponding stress groups. The distribution and relation among stress groups within an utterance define the rhythm, which is a basic characteristic of languages with different prosodic structures.

The intonation phrase (also referred to as *intonation unit*, *prosodic sentence*, *breath group*, etc.) refers to coherent intonation structures with no major prosodic break. An intonation phrase may consist of a single syllable up to syntactic phrases, clauses and sentences and a larger utterance may thus be pronounced with one or several intonation phrases, which may not have any predictive one-to-one correspondence with syntactic or semantic units. Accordingly, the correlation of intonation phrase boundaries and syntactic boundaries is casual rather than causal. On the other hand, intonation phrases may be associated with different sentence types, which may define as statements, questions, commands, etc. Intonation phrasing may also be related to information units, with reference to which speech units should be marked as more or less autonomous information units on behalf of the speaker.

In between the stress group and the intonation phrase, acknowledged for the description and analysis of different languages, an intermediate category has been suggested, i.e. *intermediate phrase*, according to which an intonation phrase may be decomposed into several intermediate phrases (see Beckman and Pierrehumbert, 1986; Pierrehumbert and Beckman, 1988).

### 2.1.3. Discourse and dialogue functions

At discourse and dialogue levels intonation may structure larger speech units above sentence level in different ways, in accordance with intraspeaker as well as interspeaker interactive functions between/among speakers. Discourse intonation and dialogue intonation are more or less overlapping terms and may be found rather interchangeably in the international literature. Somewhat simplified, discourse and dialogue intonation may structure thematic units such as *topics* and *sub-topics*, i.e., what the discussion is about and aspects of the discussion respectively, as well as turn units between speakers such as turn-taking, turn-keeping and turn-leaving interplay, i.e., the contribution of each speaker to the development of spoken discourse. In phonetic studies, both terms usually refer to the study of *spontaneous speech* in contrast to controlled read speech in a laboratory condition, i.e. *lab speech*. The study of intonation is also related to the analysis of read texts, within the subject-area of *text linguistics*, referred to as *text intonation* in current literature.

## 2.2. Forms of intonation

Intonation is based on the vibration of the vocal folds, which is an inherent characteristic of the speech production process and thus, in other words, once there is speech there is normally intonation too. Monotonous intonation would be laborious to maintain from a physiological point of view, as there are variations of subglottal pressure due to biological reasons such as breathing. On the other hand, from a perceptual point of view, monotonous intonation would be tiresome and uninteresting, which is not compatible with the fundamental function of speech to open and maintain a channel of information exchange. Once intonation variations are inherently related to speech production, an attribution of distinctive functions fulfils a basic principle of speech communication economy, i.e., to produce the maximum linguistic information with the least effort.

The forms of intonation are the merger of various physiological, linguistic, paralinguistic and extralinguistic contributions into any speech unit in principle. The physiology of voiced sounds is associated with measurable tonal production, as a result of vocal folds vibrations, whereas voiceless sounds are missing tonal production. There is however perceptual concatenation and thus intonation is perceived in a continuous rather than a gap-like way. Furthermore, microprosodic variability is considerable and thus high vowels may have higher tonal realisation than low vowels, voiceless stops may trigger higher tonal onset of the succeeding vowel than voiced stops, etc. Linguistic categories such as stress and focus may be associated with higher tonal patterns and/or tonal changes whereas finality of variable speech units such as phrase, sentence and discourse may be associated with a lowered intonation and/or tonal changes. Paralinguistic factors such as excitement, involvement and aggressiveness may increase the tonal range whereas sadness, boredom and indifference may decrease tonal variations (although there is large variability among different speakers and languages). Extralinguistic features such as age and sex have a physiologically determined effect on the height of intonation, which mainly depends on the size and form of the vocal folds, i.e., smaller vocal folds produce higher intonation, and hence the higher intonation of children versus women versus men. Furthermore, hierarchical relations, cultural attitudes and socioeconomic status, among other extralinguistic factors, may have considerable effects on intonation.

In prosodic studies, including intonation, the "isolation method", i.e. the analysis of a phenomenon at a time is a standard method and the choice of the speech materials is in accordance with the objectives of the analysis. Even "nonsense" materials, i.e. speech productions with no meaning are fairly common, much like other aspects of experimental phonetics. In spontaneous speech and dialogues, however, which are the most authentic types of speech production, the speech material is more or less unrestrictive and the decomposition of intonation and the factoring out of different contributions on a speech unit are much more complicated. No matter what type of speech material or what particular aspects of intonation are analysed, the decomposition of intonation is usually based on a few dimensions and parameters. Two dimensions are most relevant, i.e. a local and a global one, whereas the parameters are mostly related to tonal change events and tonal range magnitudes. The alignment of intonation with the segmental realisation of the speech material is also an important aspect of intonation analysis (see Bruce, 1977; Botinis, 1989; House, 1990).

### *2.2.1. Lexical forms*

There is a complex relation between lexical words and the corresponding intonation form. First, with reference to prominence, intonation variations may be associated with prosodic categories like tone, stress and accent although in variable ways, which mainly depend on the local and global prosodic contexts. Second, with reference to phrasing, tonal boundaries may not correlate with word boundaries, unless a word boundary coincides with an intonation phrase boundary, and thus intonation may have a continuous form irrespective of word boundaries. Third, tonal distribution at lexical level is in a trade-off relation with higher level intonation such as phrase, sentence and discourse and thus a decomposition of intonation associated with each level is a standard procedure for intonation description and analysis. Tonal analysis of lexical words in citation forms may be found in the literature whereas key words in simple, declarative sentence carriers, pronounced with no special focus or emphasis, is a standard context for intonation analysis at lexical level, as there is limited interference from higher level intonation. Intonation forms associated with the latter context are usually referred to as a "neutral" intonation (for factors and principles of tonal distribution see Monaghan, 1993).

*2.2.1.1. Tone patterns.* Fig. 1 shows tonal distribution of four contrastive tones in (Mandarin) Chinese in one-word declarative utterances consisting of a segmentally homophonous syllable, i.e., *mā*, *má*, *mǎ* and *mà*.

The tonal pattern of the four contrastive tones is in fairly good accordance with traditional classifications. The high, rising, low and falling tones have relatively high-level, rising, convex falling-rising and falling tonal shapes respectively. Thus, the high tone is typically static whereas the rising and falling tones, and even the low tone, are more or less dynamic.

Fig. 2 shows tonal distribution of four contrastive tones in (Mandarin) Chinese in a simple declarative utterance context (Xu, 1999). The

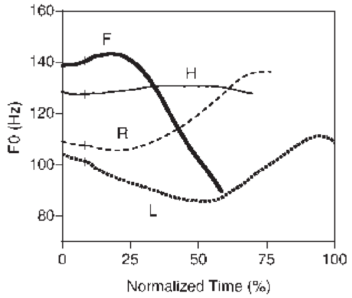

Fig. 1. Tonal shapes of four Mandarin (Chinese) tones (F, H, R, L) produced in one-word (ma) utterances (adopted from Xu, 1997).

speech material consists of three words, two disyllabic and one intervening monosyllabic, i.e., five syllables, pronounced in a neutral way. All syllables have a high tone except for the first word's second syllable, which has all four contrastive tones, i.e., HH*māomī*{HR*māomí* ∼ HL*māomǐ* ∼ HF*māomì*} H*mō* H*māomī* 'Kitty' {'cat-fan' ∼ 'cat-rice' ∼ cat-honey'} 'touches' 'Kitty'.

The basic tonal patterns of the four contrastive tones are also reasonably maintained in the utterance context, with reference to the high versus low tonal levels in combination with the falling versus rising tonal movements. However, whereas the sequence of the high tones is associated with a rather even tonal pattern throughout the utterance, the alternation of consecutive tones triggers two main tonal coarticulation effects: anticipatory and carryover. These effects are related to the falling and low tones, which trigger a tonal raising as well as a tonal lowering of the preceding and following high tones respectively. The rising tone has also a similar effect with regard to the preceding syllable but hardly the following one (see Xu and Wang, 2001, this issue). The raising of the preceding tone and lowering of the following tone are related to the notion of *downstep*, i.e., the tendency of consecutive tones to form a rightward lowering pattern (see also Section 2.2).

The later part of the syllable, which exhibits the maximum contrast in tone production, is also assumed to be the most relevant part in tone perception as the tonal pattern of the earlier part of the syllable is subject to both perturbation by the initial consonants (cf. Hombert, 1978) and carryover influence by the preceding tone (Xu, 1999; Xu and Wang, 2001, this issue).

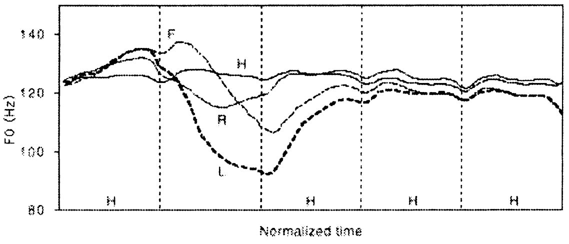

Fig. 2. Tonal shapes of four Mandarin (Chinese) tones (F, H, R, L) produced in a H tone sentence context (adopted from Xu, 1999).

*2.2.1.2. Stress patterns.* Fig. 3 shows tonal distribution of stress distinctions of Greek in a simple declarative utterance context. The speech material consists of the test words *'nomos* 'law' and *no'mos* 'county' in the context *i ma'ria 'iksere* to _ ka'la 'Maria knew the _ well' pronounced in a neutral way (Botinis, 1989, 1998).

The stressed syllables of the test words are included in a tonal gesture, i.e. a tonal hop, which consists of three phases: a rise, a plateau, and a fall. The onset of the tonal rise is aligned with the very beginning of the stressed syllable, i.e. the consonant, and the offset is completed at the end of the stressed syllable. The tonal plateau spans the poststressed syllable whereas the tonal fall is completed by the beginning of the next stressed syllable, although, in this context, the tonal fall is fairly suppressed. This suppression depends mainly on tonal coarticulation effects related to the next and final stressed syllable, which is subjected to the final juncture interference, triggering a tonal lowering. Thus, the tonal gesture spans the whole stress group, irrespective word boundaries. The division of the speech material into stress groups is evident across the test sentences, which consist of four lexical words each and thus four stressed syllables and the respective stress groups.

The tonal rise, whenever associated with a stressed syllable, is aligned with the very beginning of the stressed syllable as a rule in Greek whereas the tonal plateau and the tonal fall may show considerable variability, which depends on the immediate prosodic context. The tonal plateau mainly depends on the size of the interstress interval whereas the tonal fall may be truncated, e.g.

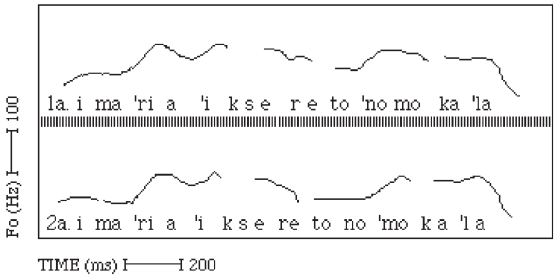

Fig. 3. Greek tonal patterns of stress productions in a sentence context (adopted from Botinis, 1989, 1998).

in tonal upsteping patterns where the tonal rise and the tonal plateau, but not the tonal fall, are the tonal correlates for consecutive stress groups. In Greek, the entire syllable rather than the vocalic part is the relevant unit of stress as, apart from tonal evidence, there is duration and intensity evidence, according to which there is considerable augmentation of both consonantal and vocalic parts (see Botinis, 1989; Fourakis et al., 1999; Botinis et al., 1999).

The tonal rise of the stressed syllable as well as the tonal pattern of the stress group in Greek is also fairly regular in different languages such as Danish and German (see Thorsen, 1982; Bannert, 1985; Bannert and Thorsen, 1988; Möbius, 1993). However, instead of, or in addition to a tonal rise, languages may have a tonal fall, which may also depend on the prosodic context, and thus a tonal change, rather than a tonal rise or tonal fall, associated with stress distinctions (see Gårding, 1977b; Hyman, 1977; Gårding et al., 1982; Thorsen, 1982). The "hat pattern", introduced for the analysis of Dutch intonation, is a typical example, according to which consecutive stressed syllables may be correlated with an alternation of tonal rises and falls whereas the corresponding interstress interval may form a tonal plateau (t'Hart and Collier, 1975; t'Hart et al., 1990; t'Hart, 1998).

In several prosodic contexts, such as the postfocal one, stress distinctions may not correlate with any tonal change but a low tonal flattening which is a regular tonal pattern in Greek as well as many languages (see Section 2.1.4). On the other hand, unstressed syllables may correlate with a tonal change, either a rise or a fall, which are not associated with stress distinctions at lexical level but with prosodic distinctions at higher levels such as sentence and discourse (see Sections 2.2 and 2.3).

In summary, a stressed syllable may correlate with a tonal change whereas an unstressed syllable is usually the carrier of a tonal change already started on the stressed syllable (see the stress group notion above). An unstressed syllable may also correlate with a tonal change but this is for intonation boundary distinctions associated with higher level intonation structuring and not stress distinctions at lexical level.

Tonal distribution has traditionally been assumed as the main perceptual factor for stress distinctions. Even a hierarchy has been proposed according to which: (1) a change in $F_0$, (2) increased duration, (3) increased intensity, and (4) a change in vowel quality (or timbre), in this ranking order, constitute the unmarked prosodic universality (see Fry, 1958; Bolinger, 1958; Lehiste, 1970; Hyman, 1977; Berinstein, 1979; Hirst and Di Cristo, 1998a). In a series of perceptual experiments in Greek (Botinis, 1989, 1998) tonal alignment as well as tonal height of stressed syllables were manipulated and resynthesised speech stimuli were subjected to perceptual analysis. Only $F_0$ manipulations were carried out whereas duration and intensity were constant (in an LPC environment) and the conclusions are based on 10 listeners' responses. Fig. 4(a) shows tonal displacement in 8 equal steps, from *'nomo* to *no'mo* and vice versa, respectively, and thus 16 stimuli in total for the tonal alignment experiment. Fig. 4(b), on the other hand, shows displacement of tonal height in 7 equal steps in *'nomo* and *no'mo*, respectively, and thus 14 stimuli in total for the tonal height experiment.

Tonal alignment displacements comprising totally 16 stimuli divided the listeners' responses into fairly categorical groups. Rightward displacements (Fig. 4(a), over) divided the corresponding eight stimuli into two main groups: a group comprising stimuli 1–5, identified as *'nomo* (over 90%), and another group comprising stimuli 7–8, identified as *no'mo* (over 90%) whereas stimulus 6 was ambiguous. Leftward displacements (Fig. 4(a), under) also divided the corresponding eight stimuli into two groups: a group comprising stimuli 9–12, identified as *no'mo* (over 90%), and another group comprising stimuli 14–16, identified as *'nomo* (over 90%) whereas stimulus 13 was ambiguous. Thus, displacements of tonal alignment may cause a complete identification change provided that a critical point is crossed over. It should be noted however that this critical point is not the same for the two reference words as duration and intensity are presumably in a trade-off relation with tonal alignment.

Tonal height displacements comprising totally 14 stimuli (Fig. 4(b)) had no or negligible percep-

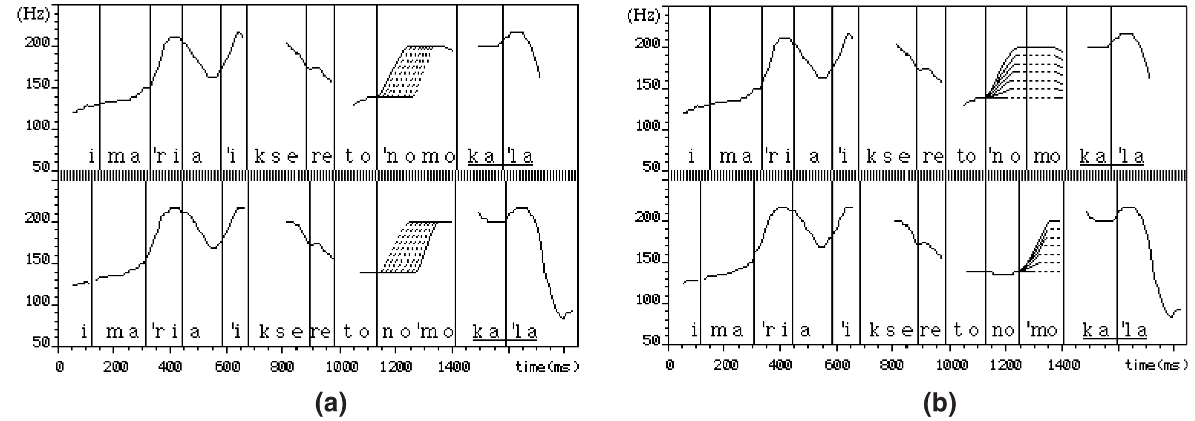

Fig. 4. Greek reference tonal patterns (solid lines) and manipulated synthetic stimuli (broken lines) with regards to stress perception in tonal alignment (a) and tonal range (b) dimensions (adopted from Botinis, 1989, 1998).

tual effect on the test words. The listeners' responses were divided into two groups: a group comprising stimuli 1–7, identified as *'nomo*, and a group comprising stimuli 8–14, identified as *no'mo*, in accordance with the respective reference productions. Thus, the two contrastive words retain their original identity even with a neutralised tonal pattern and this presumably depends on duration and intensity perceptual interference. This is in accordance with variability of stress productions, in accordance with prosodic contexts, which may not correlate with tonal changes (see Section 2.1.4).

The effects of tonal alignment versus tonal height displacements is an open question as, in the former case, an all-or-none effect is produced whereas, in the latter case, there is hardly any effect. However, we may assume that duration and intensity are acoustic and perceptual correlates of stress distinctions and tonal distribution is related with relative semantic weighting of linguistic units. A tonal change neutralisation does not thus have any considerable effect whereas a tonal change displacement may be interpreted as an experimental artefact as a Greek native speaker "knows" that a tonal change is aligned with a stressed syllable and thus alignment displacements may cause identification changes.

*2.2.1.3. Accent patterns.* Fig. 5 shows tonal distribution of the Swedish word accents in one-word declarative utterances consisting of the words "flyger" (fly, present), with acute accent, and "flyga" (fly, infinitive), with grave accent.

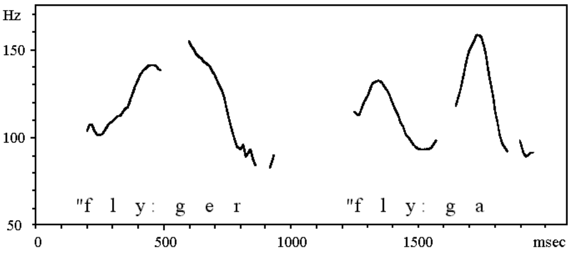

Fig. 5. Tonal patterns of the acute accent (left) and the grave accent (right) in Swedish produced in one-word utterances (adopted from Bruce, 1998).

The acute accent (left) is characterised by one tonal top whereas the grave accent (right) is characterised by two tonal tops. This is a traditional description of the Swedish word accents according to which the acute accent triggers a rising–falling tonal pattern whereas the grave accent triggers a falling–rising–falling tonal pattern in one-word utterances (see Bruce, 1998). In larger utterances, however, these patterns are subject to modifications with respect to the degree of prominence and superimposed tonal structures (see below).

*2.2.1.4. Lexical and focus interplay.* Fig. 6 shows tonal distribution of focus as well as focus and

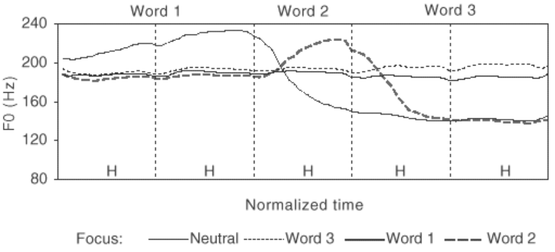

Fig. 6. Global effects of Mandarin (Chinese) focus in a H tone sentence context (adopted from Xu, 1999; also Xu and Wang, 2001, this issue).

tone interplay in Chinese. The speech material consists of the utterance H*māomī*H*mō*H*māomī* 'Kitty' 'cat-fan' 'touches' 'Kitty' with the same (H) tone sequence. Focus was produced by question-answer pairing, according to which the question was leading the placing and realisation of focus in the answer (examples of this methodology are reported in, among others, Bruce, 1977; Botinis, 1989; Xu and Wang, 2001, this issue).

Two tonal patterns are evident in Fig. 6. The first tonal pattern is related to the neutral as well as focus-final (word 3) productions, which have a fairly similar tonal level in accordance with the high lexical tone sequence. The second tonal pattern, on the other hand, is related to the focus-initial (word 1) as well as focus-medial (word 2), which is divided into two parts: a focal part and a postfocal part which have an expanded and a compressed tonal range, respectively.

Beyond the focus and tone interplay shown in Fig. 6, the effects of focus are evident in all tone contrasts in Chinese. The high points of the H, R, and F tones become higher and the low points of the R, F, and L tones become lower as a result of the tonal range expansion triggered by focus application at a local level. On the other hand, the postfocal compression is applied for all tonal contrasts at a global level (see Xu, 1999; Xu and Wang, 2001, this issue). Thus, the focus and tone interplay is a typical example in Chinese as the tone contrast may boost the highest and lowest tonal points of the focus locally whereas focus may compress the tonal variation of the tone contrast postfocally. Both local and global effects of focus are evident in many languages with different prosodic systems (see Hirst and Di Cristo, 1998a,b; also further examples below).

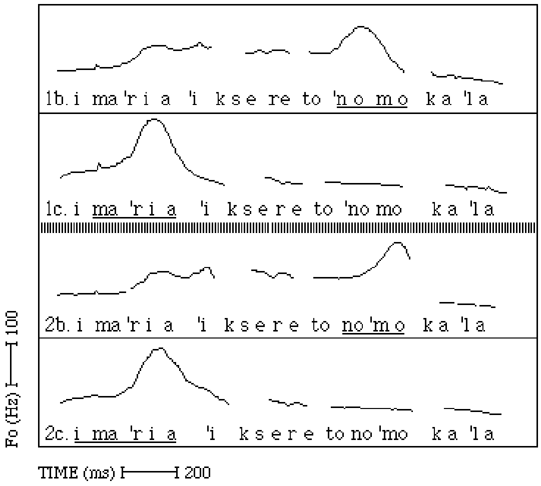

Fig. 7. Greek global effects of focus in different positions (adopted from Botinis, 1989, 1998).

Fig. 7 shows tonal distribution of focus as well as focus and stress interplay in Greek. The speech material includes the words *'nomos* 'law' and *no'mos* 'county' in the carrier sentence *i ma'ria 'iksere to* ___ *ka'la* 'Maria knew the ___ well' pronounced in different question–answer pairings and thus with different focus distribution (Botinis, 1989, 1998).

The effects of focus are not confined to the words in focus but are rather evident throughout the sentences. The fairly regular tonal pattern associated with the corresponding stress groups is heavily distorted in three main ways: first, the postfocus tonal pattern is flattened and lowered; second, the tonal range of the stressed syllable of the words *'nomo* and *no'mo* in focus is expanded; third, the prefocus tonal pattern is considerably compressed.

Fig. 8 shows tonal distribution of focus as well as focus and accent interplay in Swedish. The speech material includes the words *flyger* 'fly, presens' with acute accent and *flyga* 'fly, infinitive' with grave accent in the carrier sentence _ *med mattan* '_ with carpet' pronounced in a question–answer pairing with "flyger" and "flyga" in focus (Bruce, 1998).

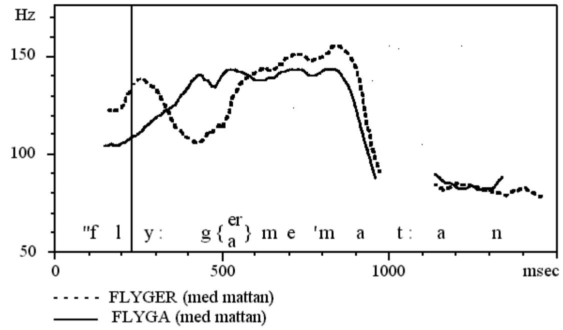

Fig. 8. Swedish global effects of focus in relation to acute and grave accents (adopted from Bruce, 1998).

The basic tonal pattern of both acute and grave accents is also kept in a focus context. Thus the acute accent has one tonal top whereas the grave accent has two tonal tops. It should be noted that the focus contribution is a tonal rise after the accented syllable, which is realised earlier for the acute accent than the grave accent. The focus tonal rise is spread on the right including the next accented syllable, which is aligned with a tonal fall. On the other hand, the postfocus material has a compressed tonal pattern at a low level. Thus, in Swedish, focus distribution has both a local and a global tonal effect (cf. Bruce, 1977; Fant et al., 2000), much like in Chinese and Greek.

Focus and stress/accent interplay with fairly similar material for Greek and Swedish is shown in Fig. 9. The sentence *Mona saw Molly in London* was pronounced in a neutral way (0), as well as with *Mona* (1), *Molly* (2) and *London* (3) in focus in Greek and Swedish.

Greek and Swedish show some basic similarities. First, the neutral productions show a fairly even tonal alternation in accordance with the stress/accent distributions (0). Second, the most prominent tonal patterns are associated with focus applications at a local level (1, 2, 3). Third, there is a tonal compression at global level, which is most evident for the postfocus material in both Greek and Swedish but more for the prefocus material in Greek than Swedish. A basic dissimilarity between the two languages, on the other hand, is related to the local effects of focus: in Greek, there is an expansion of the tonal range associated with the stressed syllable of the word in focus whereas, in Swedish, there may be a distinct tonal gesture of focus. The distribution and timing of the focus tonal gesture is dependent on the accent distinction of the word in focus, earlier for the acute accent and later for the grave accent (see Bruce, 1977; Fant et al., 2000).

Apart from different tonal distributions of focus in Greek and Swedish, there are basic differences in perception too. In a series of perceptual experiments with resynthesised stimuli, tonal manipulations had partly similar and partly different effects in Greek and Swedish (Botinis and Bannert, 1997).

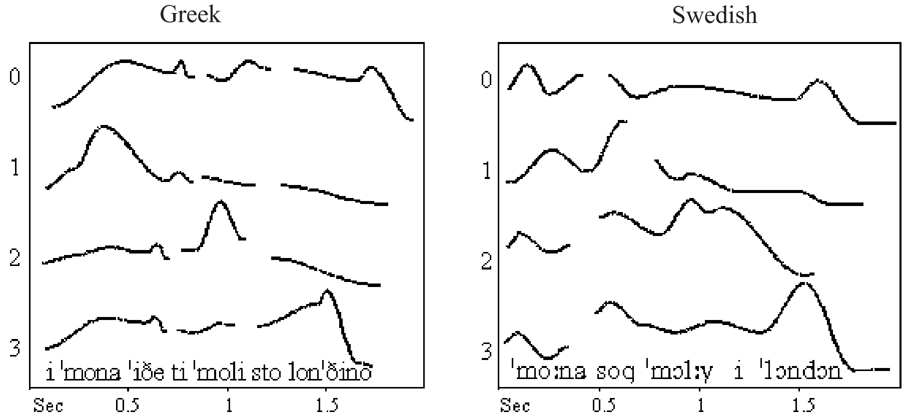

Fig. 9. Greek (left) and Swedish (right) tonal contours of test sentence productions with focus-neutral (0) as well as focus-initial (1), focus-medial (2) and focus-final distribution (3). In Swedish, Mona has grave accent whereas Molly and London have acute accents (adopted from Botinis and Bannert, 1997).

First, tonal range manipulations had a substantial effect but did not cause a complete perceptual change in accordance with the focus-target manipulation in either Greek or Swedish. Second, poststressed tonal flattening manipulations had a major perceptual effect on both Greek and Swedish and caused a complete perceptual change in Greek but not in Swedish. Third, tonal shift manipulations had a major perceptual effect on both Greek and Swedish and caused a complete perceptual change in Greek but not in Swedish. Fourth, tonal neutralisation manipulations had a major perceptual effect and caused a defocalisation in both Greek and Swedish. In summary, Greek is more sensitive to tonal manipulations than Swedish with reference to focus perception. This is in accordance with the acoustics of focus where duration is a constant correlate of focus in Swedish but not in Greek (Botinis et al., 1999). Thus, it seems that tonal and duration patterns are in a trade-off relation for focus perception in Swedish but less in Greek, the tonal pattern of which is by far the critical perceptual correlate of focus.

### 2.2.2. *Phrase and sentence forms*

Apart from, and in combination with, morphological and/or syntactic markers, intonation may define phrasing structures as well as different types of sentences. Aspects of intonation phrasing as well as sentence types and intonation interplay will be discussed below.

*2.2.2.1. Intonation phrasing*. Intonation phrasing is related to two complementary notions: boundary signalling and coherence structuring. Boundary signalling is associated with the segmentation of a wide-range of prosodic categories and structures, including stress groups, intonation phrases, and variable discourse units. At the same time, coherence structuring may also be defined by boundary signalling which determines the degree of attachment of successive speech units. Usually, intonation phrasing is defined by global as well as local tonal correlates. The global correlates are related to declination structures, which follow a "resetting" tonal pattern in accordance with the intonation phrasing of the corresponding speech material. The local correlates, on the other hand, are related to intonation phrase boundaries which may be associated with a tonal change, either a fall or a rise. Duration patterns, most usually lengthening of the boundary material (as well as silent pauses), may also be correlated with intonation phrasing.

Fig. 10 shows one aspect of intonation phrasing in Swedish, i.e., the tonal structuring of two consecutive intonation phrases with a distinctive function. The speech material consists of the sentences *fast man offrade bonden*, *och löparen hälsade kungen* 'but we sacrificed the pawn, and the bishop greeted the king' and *fast man offrade bonden och löparen*, *hälsade kungen* 'though we sacrificed the pawn and the bishop, the king greeted us'.

The main difference between the two sentences is confined to a deep versus a shallow tonal valley at the boundary of the word *bonden*, which defines intonation phrasing in accordance with the alternative syntactic structures.

Fig. 11 shows another aspect of intonation phrasing in Swedish, which have been extracted from a larger speech context. This extract consists of the speech material "... om det så är bönor, eller malet kaffe ..." (... whether (coffee) beans, or ground coffee ...).

The main tonal correlate of phrasing is a tonal rise at the final boundary of the word *bönor* which reaches the top of the tonal variations within the phrase. The distribution of this tonal rise is not

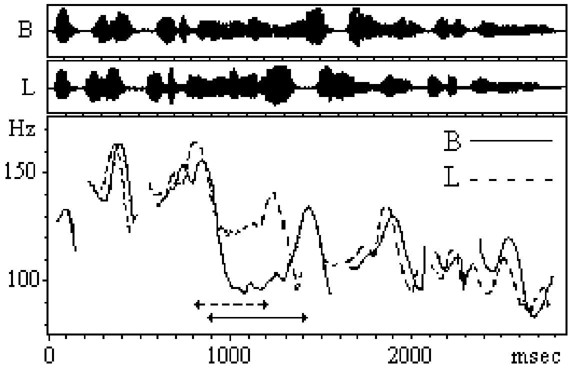

Fig. 10. Tonal patterns of prosodic phrasing in Swedish. Arrows indicate the main difference between the "bonden" (solid lines) and the "löparen" (dashed lines) productions (adopted from Bruce et al., 1993).

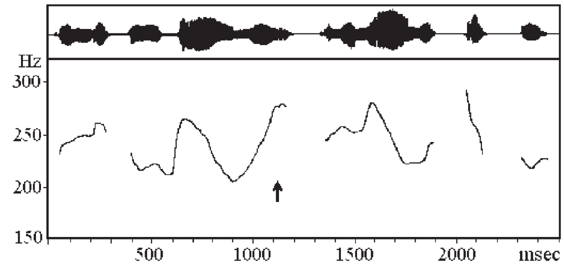

Fig. 11. Intonation phrasing in Swedish. The arrow indicates the tonal boundary (adopted from Bruce, 1998).

directly associated with accentual prominence but, in combination to a phrasing function, has a continuation function in relation to the following material (see Sections 2.2.3).

Thus, phrasing in Swedish may be defined by low/falling but also by high/rising tonal patterns in accordance with the linguistic context and the type of material in the first place.

*2.2.2.2. Sentence types and intonation forms.* There is a basic distinction between sentence typology, at least with regards to traditional morphological and syntactic criteria, and sentence functioning. A sentence may have a distinct typological form such as declarative, interrogative and imperative with the corresponding function, i.e. statement, question and command. However, the relation of typology and function is not one-to-one and thus a declarative sentence may e.g. function as a statement but also as a question. In Greek, e.g., the sentence *'efiγe i ma'ria* may function as a statement (Maria (has) left) as well as a total (yes–no) question (did Maria leave?). In the absence of any morphological or syntactic markers, this distinction is assumed to be based on the prosodic and in particular on the intonation form. In other languages, however, total questions may be associated with another question marker or a multiple combination of morphological, syntactic and intonation ones. On the other hand, in different languages, distinct sentence functions may be associated with local or global tonal features, or a combination of both (see Grønnum, 1998; Van Heuven and Haan, 2000; Jun and Fougeron, 2000).

Statements are generally characterised by a global rightward intonation lowering which is mainly related to junctures (*alias* boundary tones) and tonal range variations. The initial juncture, i.e., the tonal onset at the very beginning of the sentence, is usually higher than the final juncture, i.e., the tonal offset at the very end of the sentence. The tonal tops mostly but also the tonal valleys of the entire sentence are in descending form from left to right. The general rightward lowering is referred to as "declination" whereas related notions include "downdrift" and "downstep". In addition, contextual rules such as "catathesis" and "final lowering" may further modify the global lowering pattern. On the other hand, although declination phenomena have been described for many languages for simple sentences, in relatively complex sentences even mirror image patterns of declination may be found and, thus, a sentence may have a semiglobal rising part on the left and a falling one on the right. Even upstepping patterns, the mirror image of catathesis (i.e. anathesis), may be found. Thus, declarative sentences with a statement function may usually be characterised by a falling or even a rising–falling global tonal pattern.

Fig. 12 shows stylised intonation forms of different sentence types in standard (Copenhagen) Danish. The sentence types represent (1) syntactically unmarked questions, (2) morphologically and/or syntactically marked questions, non-final declarative and interrogative clauses and (3) final declarative statements.

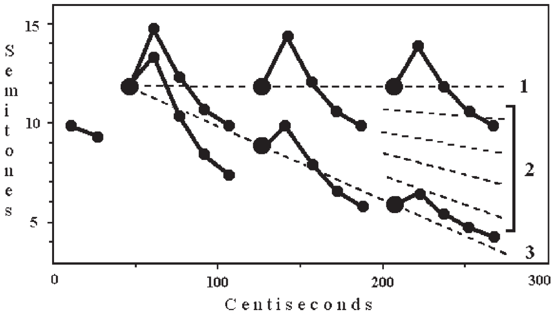

Fig. 12. Sentence types and intonation forms in standard Danish. Large dots indicate stressed syllable and small dots unstressed syllables. Full lines represent local stress groups and broken lines global intonation patterns (adopted from Grønnum, 1998).

The stress groups in Danish have a regular bell-like pattern with (1) a tonal change (rise) aligned with the stress syllable, (2) a tonal top at the post-stressed syllable and (3) a tonal fall to the end of the stress group much like other languages such as German and Greek (see Section 2.2.1.2). The different sentence types are correlated with different declination slopes with reference to the stressed syllables (but also tonal range variations). Thus, the syntactically unmarked questions do not show any declination pattern and final declarative statements show maximum declination whereas other types of sentences are in between. This taxonomy is related to the development of an intonation model in Danish (an elaborated version of the model is reported in (Grønnum, 1995); see also Section 3.2).

Fig. 13 shows intonation forms and sentence functions in French. The sentence functions correspond to (1) question, (2) continuation, (3) command and (4) vocative of the one-word utterance “Anne Marie”.

The distinct sentence functions in Fig. 13 are related to both global and local tonal features. Question (a) and continuation (b) functions have a final tonal rise whereas command (c) and vocative (d) have a final tonal fall. On the other hand, command has a left tonal dominance whereas vocative has a right one (see Di Cristo, 1998).

### 2.2.3. Discourse and dialogue intonation forms

Intonation forms encountered in isolated sentence productions may be heavily modified in the context of larger discourse and dialogue units. Aspects of topic and dialogue boundaries associated with intonation forms will be presented below.

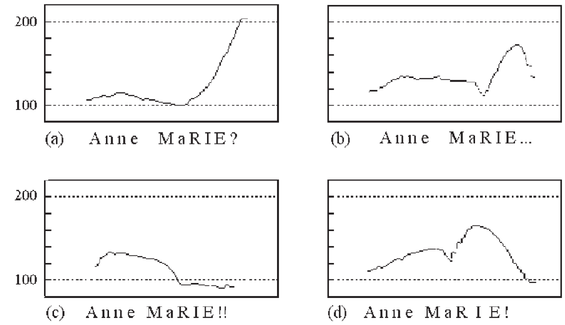

Fig. 13. Question (a), continuation (b), command (c) and vocative (d) intonation forms in French produced in one-word utterance (adopted from Di Cristo, 1998).

*2.2.3.1. Topic boundaries and intonation forms.* Topic boundaries, apart from syntax and morphology, may be marked by a variety of local and global tonal patterns. Most usually, higher tonal patterns are associated with topic (or aspects of topic, i.e., subtopics) initiality and lower tonal patterns with topic finality. Fig. 14 shows a topic initiality in Greek consisting of the spontaneous speech material *li′pon kseki′name a′mesos* “well, we start right away”.

The word *li′pon* is the topic initiality of the speech material in Fig. 14. There is a tonal rise aligned with the first (unstressed) syllable as if that syllable were stressed. However, this tonal rise is most likely a discourse marker related to the onset of the topic rather than lexical prominence and stress. Presumably, lexical prominence at the second (stressed) syllable is correlated with an augmentation of duration and intensity. It should be noted that there is no word $*$*′lipon* (with stress on the first syllable) in the Greek lexicon.

Fig. 15 shows intonation forms of topic marking in a spontaneous dialogue. The speech material, apart from the reorganisation repetition *′pezis ja to* … ‘you play for’, consists of four intonation phrases: (1) *′pezis ja to pade′loni ′dzin* ‘you play for a trouser jean, (2) *a′ksias ′ðeka xi′ljadon ðrax′mon tis ka′rera* ‘worth ten thousand drachmas of Karera, (3) *′pezis ja to ′futer tis*

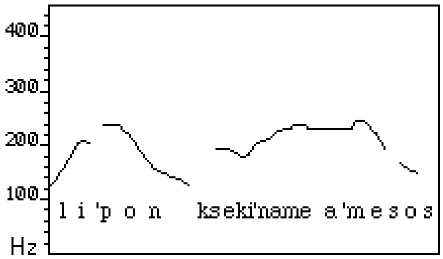

Fig. 14. Tonal marking of topic initiality in Greek realised in the first (unstressed) syllable of the word li′pon produced in a spontaneous dialogue context (adopted from Botinis, 1992).

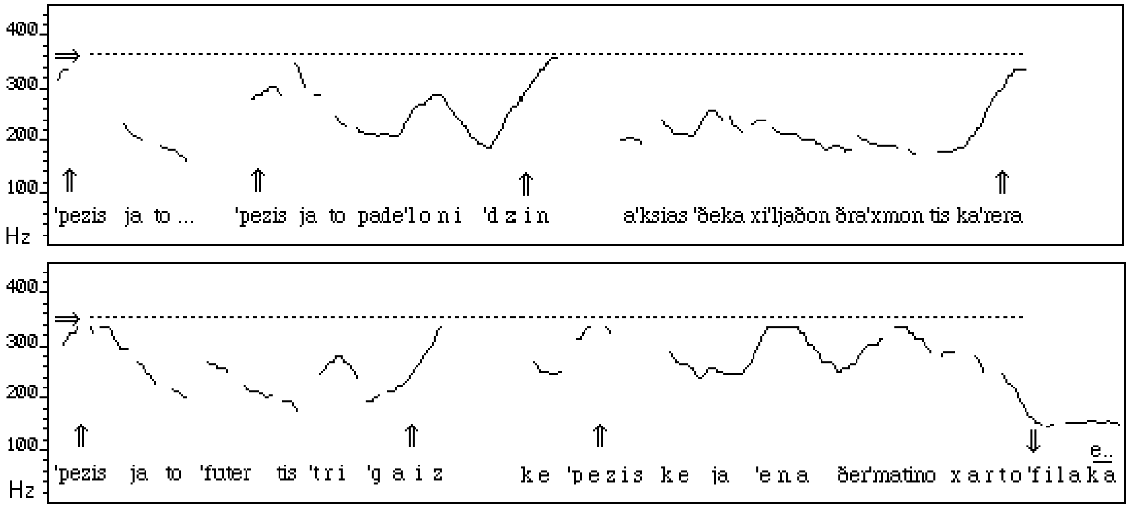

Fig. 15. Intonation phrases with continuation rises as well as turn finality tonal marking in Greek produced in a spontaneous dialogue context. Arrows indicate main points of discourse and dialogue intonation forms (adopted from Botinis, 1992).

*'tri 'gaiz* ‘you play for the blouse of Three Gaiz and (4) *ke 'pezis ke ja 'ena ðer'matino xarto'filaka* ‘and you play for a leather briefcase’.

The speech material is structured in intonation phrases with a final boundary tonal rise (often referred to as “continuation rises” in the international literature), regardless of the stressed/unstressed distribution. The tonal rises are presumably “turn-keeping” markers, in the sense that the speaker wants to keep his turn within a topic, whereas the tonal fall is a “turn-leaving” and/or topic finality marker, in the sense that the speaker has concluded his turn and/or his topic. There are hardly any declination tendencies, either at the onset or the offset of the intonation phrases, which is an indication that declination may not be evident in certain types of speech material and contexts. On the other hand, there is a tonal fall aligned with the last stressed syllable of the final word *xarto'filaka* ‘briefcase’, which is most probably a (sub)topic/turn finality rather than a prominence marker. This tonal pattern has an interspeaker dialogue effect, according to which the second conversationalist resumes his turn with *e* . . .

Thus, discourse-triggered tonal rises may be aligned with unstressed syllables and tonal falls with stressed syllables. These patterns are not usually found in isolated sentence productions and stress group realisations. This is an indication that concrete tonal distribution may be related to different levels of abstraction and have different functions in accordance to the level of application.

*2.2.3.2. Dialogue boundaries and intonation forms.* Among other factors, such as type of speech material and linguistic as well as situation context, intonation realisation is related to tonal patterns across different speakers. Thus, there is a tonal interplay between speakers which is evident, among other domains, at turn-unit boundaries. Fig. 16 shows tonal patterns of turn-unit boundaries in a spontaneous dialogue between two speakers. One speaker is leading the development of the dialogue and another speaker follows in a cooperative way. Fig. 16(a) consists of three turn-units: (1) *sefxari'sto* ‘thanks’ (first speaker), (2) *'ela ja ke xa'ra su* ‘well, by-by’ (second speaker) and (3) *ja ja* ‘by-by’ (first speaker). Fig. 16(b) consists of two turn-units: (1) *ðila'ði 'nane kli'sto* ‘which means off’ (first speaker) and (2) *ne* ‘yes’ (second speaker).

In Fig. 16(a) and (b) the turn-units between the two speakers form a coherent intonation structure. These patterns are rather regular between speakers involved in a cooperative dialogue even in cases of

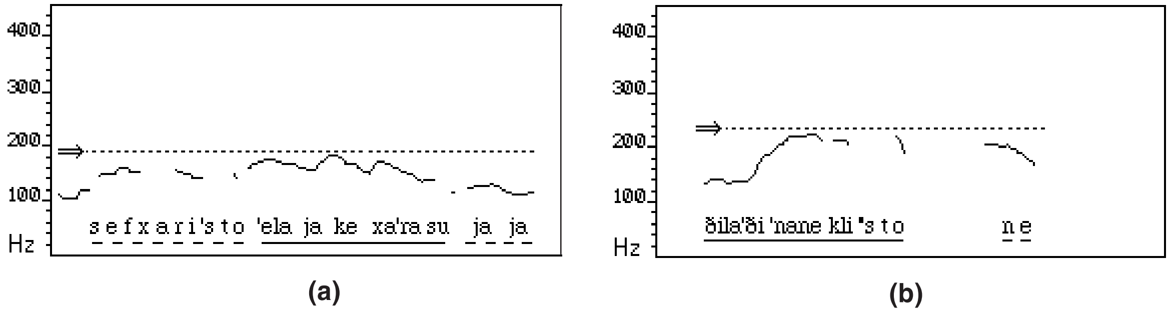

Fig. 16. Interspeaker pitch-concord tonal patterns. Solid and broken lines correspond to two different speakers (adopted from Botinis, 1992).

interspeaker personal differences such as age and sex (see Botinis, 1992). The tonal choices of a speaker may thus have interspeaker effects and, accordingly, the ultimate intonation form in dialogues and spontaneous discourse is a constant interplay between/among speakers, much like vocabulary and other grammatical components.

## 3. Modelling and labelling of intonation

In the past 20 years or so two major classes of intonation models have been developed. There are, on the one hand, *phonological* models that represent the prosody of an utterance as a sequence of abstract units. $F_0$ contours are generated from a sequence of phonologically distinctive tones, or categorically different pitch accents, that are locally determined and do not interact with each other (tone sequence models). On the other hand, there are *acoustic–phonetic* models that interpret $F_0$ contours as complex patterns that result from the superposition of several components (superposition or overlay models).

Apart from these two prevalent types of intonation models there are several important approaches that defy a categorisation as being either of the superpositional or the tone sequence type. For instance, *perception-based* models exploit the observation that some acoustic–prosodic properties of speech, while measurable, cannot always be perceived by the listener. These models tend to downplay the linguistic functions of intonational events. The Kiel intonation model, by contrast, emphasises the *functional* aspect and shows how phonetic details, such as the location of an $F_0$ peak relative to the segmental structure, can change the meaning of the utterance. Finally, *acoustic stylisation* approaches aim at a robust computational analysis and synthesis of $F_0$ contours.

### 3.1. Phonological models of intonation

Janet Pierrehumbert's Ph.D. dissertation (Pierrehumbert, 1980) is widely regarded as the single most influential work in the field of intonational phonology. Pierrehumbert's intonation model builds upon metrical (Liberman and Prince, 1977) and autosegmental phonology (Leben, 1976; Goldsmith, 1976a) as well as Bruce's analysis of Swedish word accents (Bruce, 1977).

In Pierrehumbert's model an intonational phrase, the largest prosodic constituent, is represented as a sequence of high (H) or low (L) *tones*. H and L are members of a primary phonological opposition. The tones do not interact with each other but merely follow each other sequentially in the course of an utterance. There are three types of tones:

*Pitch accents*, either as single tones ($H^*, L^*$) or bitonal ($H^* + L, H + L^*, L^* + H, L + H^*$). Pitch accents are assigned to prosodic words; the "*" symbol indicates the association and alignment between the tone and the accented syllable of the prosodic word.

*Phrase accents*, marked by the "−" symbol (H−, L−). Phrase accents indicate the offset pitch of intermediate phrases and thus control the pitch movement between a pitch accent and a boundary tone.

*Boundary tones*, denoted by the "%" symbol (H%, L%). Boundary tones are aligned with the edges of an intonational phrase. The initial and final boundary tones control the onset and offset pitch, respectively, of the intonational phrase.

The model thus introduces a three-level hierarchy of intonational domains which obey the *strict layer hypothesis*: An intonational phrase consists of one or more intermediate phrases; each intermediate phrase is composed of one or more prosodic words. The intonation contour of an utterance is described as a sequence of relative (H and L) tones. Well-formed sequences are predicted by a finite-state grammar (Fig. 17).

This abstract tonal representation is converted into $F_0$ contours by applying a set of *phonetic realisation rules*. The phonetic rules determine the $F_0$ values of the H and L tones, based on the metric prominence of the syllables they are associated with, and on the $F_0$ values of the preceding tones. Calculation of the $F_0$ values of tones is performed strictly from left to right, depending exclusively upon the already processed tone sequence and not taking into account the subsequent tones. The phonetic rules also compute the temporal alignment of the tones with the accented syllables.

Pierrehumbert's intonation model is predominantly sequential; what is treated in other frameworks as the correlates of the phrase structure of a sentence or as global trends, such as question or declarative intonation patterns, is conceptualised as elements of the tonal sequence and their (local) interaction. In this model, the English question intonation is embodied in the tonal sequence $L^*H - H\%$, and there is no separate phrase-level "question intonation contour" that these tones are superimposed on. Similarly, the downward trend observed in some types of sentences, particularly in list intonation, is accounted for by the *downstep* effect triggered by certain accents, such as $H^* + L$, rather than being attributed to a phrase-level intonation that affects all pitch accents.

There are a few aspects of the model that are hierarchical or non-local. The model is situated at the interface between intonation and metrical phonology and inherits the hierarchical organisation of the metrical stress rules. Another element whose effect is global is declination, onto which the linear sequence of tones is overlaid. Given these properties, Ladd (1988) characterised Pierrehumbert's model as a hybrid between the *superposition*

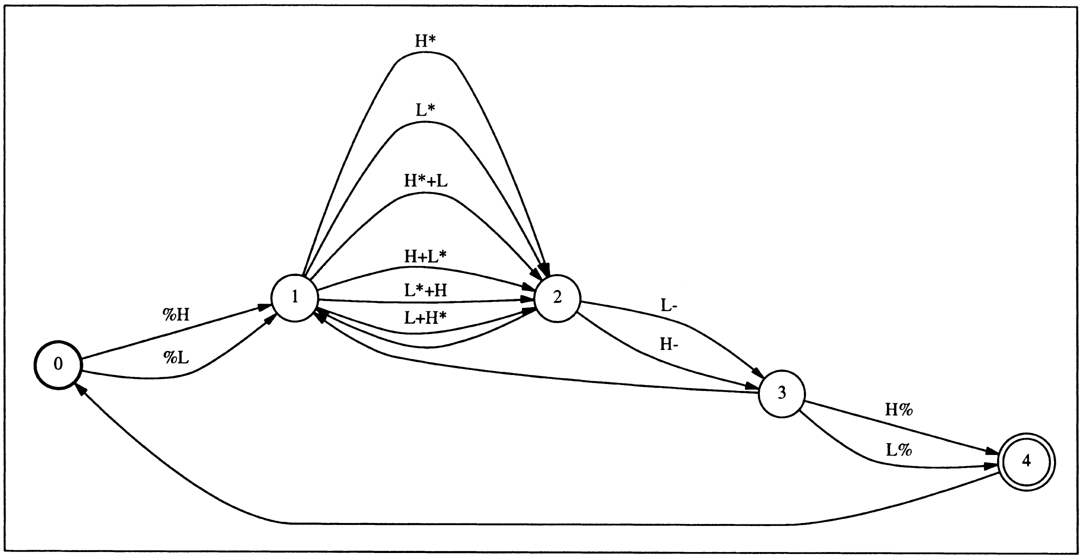

Fig. 17. Phonological model with pitch accent, phrase accent and boundary tone labelling (adopted from Pierrehumbert, 1980).

and the *tone sequence* approach. Furthermore, discourse structure is hierarchically organised, and the information is used to control $F_0$ scaling so that the pitch height of discourse segments reflects the discourse hierarchy (Hirschberg and Pierrehumbert, 1986; Silverman, 1987). The strongest position with respect to the local nature of tone scaling was taken by Liberman and Pierrehumbert (1984) who concluded that most of the observed downward trend is attributed to downstep and that there is no evidence of declination in English.

Ladd's phonological intonation model (Ladd, 1983) is based on Pierrehumbert's work but integrates some aspects of the IPO approach (t'Hart et al., 1990) and the Lund intonation model (Bruce, 1977; Gårding, 1983; see also Section 3.2). Like Pierrehumbert, Ladd applies the framework of autosegmental and metrical phonology. He attempts to extend the principles of feature classification from segmental to suprasegmental phonology, which would also facilitate cross-linguistic comparisons. In Ladd's model, $F_0$ contours are analysed as a sequence of structurally relevant points, viz., accent peaks and valleys, and boundary end points, each of which is characterised by a bundle of features. Acoustically, each tone is described in terms of its height and its position relative to the segmental chain. Tones are connected by straight-line or smoothed $F_0$ transitions. Key elements of this model are also presented in Ladd's more recent "Intonational Phonology" (Ladd, 1996).

The tone sequence theory of intonation has been formalised into the tone and break indices (ToBI) transcription system (Silverman et al., 1992). ToBI was originally designed for transcribing the intonation of three varieties of spoken English, viz., general American, standard Australian and southern British English, and the authors were sceptical about the possibility of using it to describe the intonation systems of other dialects of English, let alone other languages. After all, the tone sequence theory provides an inventory of phonological entities; the ToBI system may thus be characterised as a broad phonemic system. The phonetic details of $F_0$ contours in a given language have to be established independently. These considerations notwithstanding, the basic principles of ToBI have by now been adopted to develop transcription systems for a large number of languages, including Japanese, German, Italian, and Bulgarian. ToBI labels, in conjunction with $F_0$ generation rules, are also frequently used in the intonation components of text-to-speech (TTS) synthesis systems.

### *3.2. Acoustic–phonetic models of intonation*

Grønnum (Thorsen) developed a model of Danish intonation (sumarised in Grønnum, 1992, 1995) that is conceptually quite different from the tone sequence approach. Her intonation model is hierarchically organised and includes several simultaneous, non-categorical components of different temporal scopes. The components are *layered*, i.e., a component of short temporal scope is superimposed on a component of longer scope.

Grønnum's model integrates the following components. The highest level of description is the text or paragraph, which requires a discourse-dependent intonational structuring (*text contour*). Beneath the text there are influences of the sentence or the utterance (*sentence intonation contour*) and of the prosodic phrase (*phrase contour*). The lowest linguistically relevant level is represented by *stress group patterns*. The four components are language-dependent and actively controlled by the speaker. The model also includes a component that describes *microprosodic* effects, such as vowel intrinsic and coarticulatory $F_0$ variations, which are generally assumed not to be under the conscious control of the speaker. Finally, a Danish-specific component models the *stød*, a creaky voice phenomenon at the end of phonologically long vowels, or on the post-vocalic consonant in the case of short vowels.

All components of the model are highly interactive and jointly determine the $F_0$ contour of an utterance. Therefore, for the interpretation of observed natural $F_0$ curves, a hierarchical concept is needed that allows the analytical separation of the effects of a multitude of factors on the intonation contour.

Similar to Grønnum's work, the Lund intonation model (Bruce, 1977; Gårding, 1983) analyses the intonation contour of an utterance as the complex result of the effects of several factors. A *tonal grid*, whose shape is determined by the sentence mode and by pivots at major syntactic boundaries, serves as a reference frame for local $F_0$ movements. It is thus implicitly assumed that the speaker pre-plans the global intonation contour. At first glance, the Lund model also includes elements of the tone sequence approach in that it represents accents by sequences of high and low tones. But in the Lund model position and height of the tones are determined by the tonal grid, which is, by definition, a non-local component. Yet, the Lund model suggests that it is possible to integrate aspects of the superpositional and the tone sequence approaches.

The classical *superpositional* intonation model has been presented by Fujisaki (Fujisaki, 1983, 1988). It can be characterised as a functional model of the production of $F_0$ contours by the human speech production apparatus, more specifically by the laryngeal structures; the approach is based on work by Öhman and Lindqvist (1966). The model represents each partial glottal mechanism of fundamental frequency control by a separate component. Although it does not include a component that models intrinsic or coarticulatory $F_0$ variations, such a mechanism could easily be added in case it is considered essential for, e.g., natural-sounding speech synthesis.

Fujisaki's model additively superimposes a basic $F_0$ value (*Fmin*), a *phrase component*, and an *accent component*, on a logarithmic scale (Fig. 18). The control mechanisms of the two components are realised as critically damped second-order systems responding to impulse commands in the case of the phrase component, and rectangular commands in the case of the accent component. These functions are generated by two different sets of parameters: (1) amplitudes and timing of phrase commands, and damping factors of the phrase control mechanism; (2) amplitudes and timing of the onsets and offsets of accent commands, and the damping factors of the accent control mechanism.

The values of these parameters are constant for a defined time interval: the parameters of the phrase component within one prosodic phrase; the parameters of the accent component within one accent group; and the basic value Fmin within the whole utterance.

The $F_0$ contour of a given utterance can be decomposed into the components of the model by applying an *analysis-by-synthesis* procedure. This is achieved by successively optimising the parameter values, eventually yielding a close approximation of the original $F_0$ curve. Thus, the model provides a parametric representation of intonation contours.

The model has been applied to a number of languages, including Japanese, Swedish, Mandarin Chinese, French, Greek, German, and English. With the exception of English, where the model failed to produce certain low or low-rising accentual contours (Liberman and Pierrehumbert, 1984; Taylor, 1994), very good approximations of natural $F_0$ contours were generally obtained. The compatibility of several key assumptions of the tone sequence approach with a Fujisaki-style model has been discussed and, to some extent, experimentally shown in Möbius's work on German intonation (Möbius, 1993, 1995), in which a linguistic motivation and interpretation of the phrase and accent commands has been attempted.

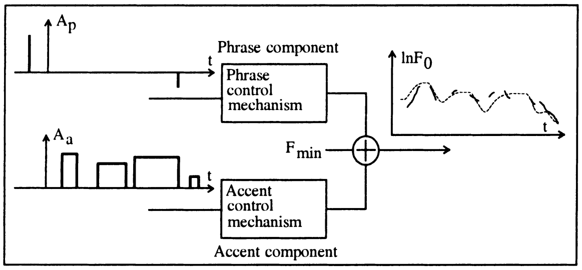

Fig. 18. Superpositional model with phrase and accent components (adopted from Fujisaki, 1988).

The concept of superposition is also exploited in the Bell Labs intonation model (Van Santen and Möbius, 1997, 2000), which focuses on the temporal aspects of intonation. In this model an $F_0$ contour is computed by adding up three types of time-dependent curves: a *phrase curve*, which depends on the type of phrase, e.g., declarative versus interrogative; *accent curves*, one for each accent group; and *segmental perturbation curves*. The model incorporates results from (Van Santen and Hirschberg, 1994) who had shown that there is a relationship between accent group duration and $F_0$ peak location. Another important factor is the segmental structure of onsets and codas of stressed syllables.

In the Bell Labs model the phonological unit of a pitch accent is the *accent group*, not the accented syllable. An accent group is defined as an entity that consists of an accented syllable followed by zero or more unaccented syllables. The time course of an accent curve depends on the segmental and temporal structure of the entire accent group, not only on the properties of the accented syllable. This dependency is complicated but regular: all else being equal, the pitch peak, as measured from the start of the accented syllable, is shifted to the right as any part of the accent group is lengthened. It is shown that this regularity can be captured by a simple linear *alignment* model.

Based on these findings, the model predicts $F_0$ peak location in a given accent group by computing a weighted sum of the onset and rhyme durations of the stressed syllable, and the duration of the remainder of the accent group. It is assumed that the three factors exert different degrees of influence on peak location. For any given segmental structure, the ensemble of regression weights is called an *alignment parameter matrix*, and for each given pitch accent type the alignment parameter matrix characterises how accent curves are aligned with accent groups. The accent curves also have to be scaled to some appropriate pitch range, reflecting the relative prominence of the pitch accent.

$F_0$ curves of pitch accents that belong to the same perceptual or phonological class are generated from a common basic shape or *template* by applying a common set of alignment parameters. Predicted accent curve shapes can thus be considered as time-warped versions of a common template. It is stipulated that pitch accents of the same class sound the same because they are aligned in the same way with the segmental structure of the accent group they are associated with. Conversely, two accent curves are phonologically distinct if they cannot be generated from the same template using the same alignment parameter matrix.

This model is used in the Bell Labs TTS system for English, French, German, Italian, Spanish, Russian, Romanian, and Japanese (Van Santen et al., 1998; Venditti et al., 1998).

### 3.3. Perceptual models of intonation

The starting point of the best-known perceptual model of intonation, the model developed at IPO (Institute of Perception Research, Eindhoven), was the observation that certain $F_0$ movements are perceptually relevant whereas others are not. Intonation analysis according to the IPO method (t'Hart et al., 1990) consists of three steps. First, the perceptually relevant movements are *stylised* by straight lines. The procedure results in a sequence of straight lines, a *close copy contour*, that is perceptually indistinguishable from the original intonation contour: the two contours are *perceptually equivalent*. The motivation for stylising the original intonation is that the enormous variability of raw $F_0$ curves presents a serious obstacle for finding regularities.

In a second step, common features of the close copy contours, expressed in terms of duration and range of the $F_0$ movements, are standardised and collected as an inventory of discrete, phonetically defined types of $F_0$ rises and falls. These movements are categorised according to whether or not they are *accent lending*; for example, in both Dutch and German, $F_0$ rises occurring early in a syllable cause this syllable to be perceived as stressed, while rises of the same duration and range, but late in the syllable, are not accent lending. Similarly, late falls produce perceived syllabic stress but early ones do not. The notion of accent lending adds a functional aspect to the otherwise purely melodic character of the model.

In the third and final step, a grammar of possible and permissible combinations of $F_0$ movements is written. The grammar describes both the grouping of pitch movements into longer-range contours and the sequencing of contours across prosodic phrase boundaries. The contours must comply with two criteria: they are required to be perceptually similar to, and as acceptable as, naturally produced contours. Thus, the complete model describes the melodic possibilities of a language.

The IPO model was originally developed for Dutch, but it was later also applied to English (de Pijper, 1983), German (Adriaens, 1991), and Russian (Odé, 1989). It has been implemented in speech synthesis systems for Dutch (Terken, 1993; Van Heuven and Pols, 1993), English (Willems et al., 1988), and German (Van Hemert et al., 1987).

Stylisation in the IPO model is performed by a human experimenter. The method can therefore produce inconsistent results when the same original contour is stylised more than once, either by the same or by different researchers, which may yield different parameter values. It is claimed, however, that in practice any inconsistencies are below perceptual thresholds (Adriaens, 1991, p. 38) and thus negligible.

Methods for automatic stylisation of intonation contours on perceptual grounds have been proposed by Mertens and d'Alessandro (Mertens, 1987; D'Alessandro and Mertens, 1995). The authors base their approach on two assumptions. First, perception studies have provided evidence that $F_0$ contours should always be interpreted along with co-occurring segmental and prosodic properties of the speech signal, not in isolation – a point also emphasised by Kohler (1991). Second, it is hypothesised that the syllable may be the appropriate domain for the perception of intonation, and that perceived pitch contours within a syllable can be further decomposed into elementary contours, viz. *tonal segments*. During the stylisation process the tonal segments that make up a syllabic pitch contour are determined by applying thresholds on the slopes of rising and falling $F_0$ curves. Finally, a pitch target is assigned to each tonal segment. This approach has been applied to Dutch and French, in both automatic speech recognition and speech synthesis tasks (Mertens, 1989; Malfrère et al., 1998).

### 3.4. A functional model of intonation

The Kiel intonation model (KIM; Kohler, 1991) can be characterised as a generative, functional model of German intonation, based on research on $F_0$ production and perception. Unlike many other intonation models, first, KIM does not ignore microprosodic $F_0$ variations and, second, it integrates syntactic, semantic, pragmatic, and expressive functions (*meaning functions*). The model applies two types of rules. Input information to the symbolic feature rules is a sequence of segmental symbols annotated for stress as well as pragmatic and semantic features. The rules convert this input into sequences of binary features, such as [±late] or [±terminal]. Finally, parametric rules generate duration and $F_0$ values and control the alignment of the $F_0$ contour elements with the segmental structure of the target utterance. The parametric rules include rules for the downstepping of accent peaks during the course of the utterance as well as microprosodic rules.

One of the starting points in the development of KIM was the study of accent *peak shifts* (Kohler, 1987, 1990), which discovered three distinct locations of the $F_0$ peak relative to the segmental structure of the stressed syllable: early peaks signal established facts that leave no room for discussion; medial peaks convey new facts or start a new argument; late peaks put emphasis on a new fact and contrast it to what may already exist in the speaker's (or listener's) mind. Thus, shifting a peak backwards from the early location causes a category switch from given to new information. KIM has been implemented in the German version of the INFOVOX speech synthesis system (Carlson et al., 1990).

### 3.5. Acoustic stylization models of intonation

The Tilt intonation model (Taylor, 2000) was designed to provide a robust computational analysis and synthesis of intonation contours. The model analyses intonation as a sequence of phonetic *intonation events*, such as pitch accents and boundary tones. Whereas in customary terminology pitch accents and boundary tones are assumed to be phonological entities, in the Tilt model they are events that are characterised by continuous acoustic–phonetic parameters – an approach that has been criticised by some authors (e.g., Ladd, 1996) as being paralinguistic. Each of these events consists of a rising and a falling component of varying size; rise or fall may also be absent. The mid point of an event is defined as the end of the rise and the start of the fall.

The events are described by the so-called Tilt parameters: (a) amplitude or $F_0$ excursion of the event; (b) its duration; (c) a dimension-less shape parameter that is computed as the ratio of rise and fall and ranges between $+1$ (rise only) and $-1$ (fall only), a value of 0 indicating that the rising and the falling component are of the same size; (d) $F_0$ position, expressing the distance (in Hz) between a baseline and the mid point of the event; (e) position of the event in the utterance.

The model proposed by Möhler (1998a) implements an $F_0$ parameterisation procedure that is similar to the Tilt model. It uses parameters that express the shape and steepness of intonation events. Further parameters control the alignment of the $F_0$ curve with the syllable and the scaling of the event within the speaker's local pitch range. The perceptual adequacy of the parameterisation was tested and confirmed in a series of perception experiments. The model has been implemented in the German version of the Festival TTS system (Möhler, 1998b). For synthesis or prediction of intonation contours, the model offers an interface to syntactic and semantic analysis by way of interpreting prosodic labels and pitch range in its input.

Both Möhler's and the Tilt parameters are appropriate for the synthesis of $F_0$ contours because, other than in ToBI-based approaches, no rules for the realisation of $F_0$ from abstract units are required. Both models may be characterised as $F_0$ coding or $F_0$ data reduction, with potential problems concerning their linguistic interpretability. The authors argue, however, that the parameters of their models are meaningful; for instance, the amplitude parameter is related to perceived prominence, $F_0$ position may be used to model downstep or declination, and the timing of the event may have linguistic function (cf. Kohler's early and late peaks).

Automatic analysis and generation of intonation contours is also performed by a collection of tools that were developed in Aix-en-Provence (see Hirst et al., 1994; Véronis and Campione, 1998; Véronis et al., 1998; Di Cristo et al., 2000) in the context of the MULTEXT project (Multilingual Text Tools and Corpora). The toolkit allows the automatic modelling of the $F_0$ curve from the speech signal following the method described in (Hirst et al., 1991). The output of this approach is a sequence of target points, specified in time and frequency, that represents a stylisation of the $F_0$ curve. For $F_0$ synthesis the target points are interpolated by a quadratic spline function. The target points also serve as input to the symbolic coding of intonation according to the INTSINT system (International Transcription System for Intonation; Hirst and Di Cristo, 1998b). Finally, the symbolic coding of intonation is automatically aligned with the segmental annotation of the utterance (Fig. 19).

The INTSINT system provides a narrow phonetic transcription of intonation, which has been shown to be applicable to a number of languages with quite diverse intonation systems. In fact, as Hirst and Di Cristo (1998b) point out, the attempt to design a transcription system that would be equally suitable for both English and French intonation was one of the original motivations for the development of INTSINT. Unlike ToBI, which presupposes that the inventory of $F_0$ patterns of the language in question is already known, INTSINT can be used to explore and analyse the intonation of languages whose tonal systems have not already been described in acoustic detail.

## 4. Applications of intonation

Since intonation forms such a central part of human speech communication, not only conveying diverse linguistic information, but also information about the speaker, the speaker's mood and attitude, it certainly ought to be useful in many

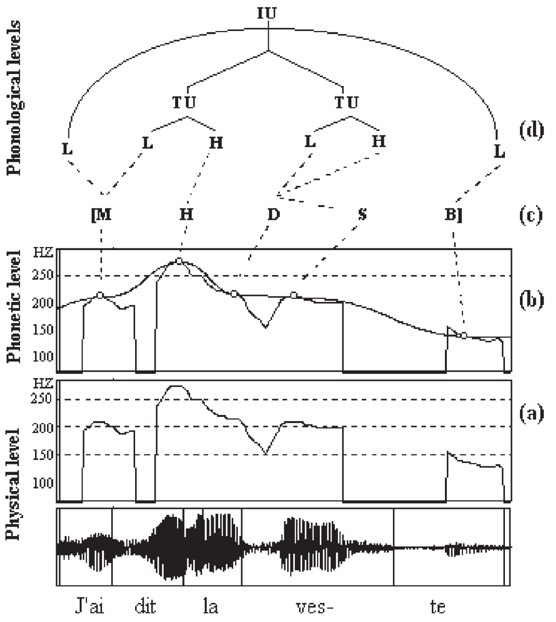

Fig. 19. Physical, phonetic and phonological levels for the French utterance "J'ai dit la veste" (I said the jacket). From bottom to top: (a) oscillogram of raw $F_0$ curve; (b) modelled curve; (c) representation of surface intonation structure; (d) representation of deep intonation structure (adopted from Di Cristo et al., 2000).

applications. One attraction of intonational features in many technical applications is that they are quite immune to transmission distortions and noise. The confounded nature of intonation might however be one of the obstacles. It is quite hard to separate the different kinds of information, to be reliably used. The phonology and phonetics of intonation are often difficult to separate. It is well known that intonational realisations vary widely with speaker and dialect. One example is Öhman's, so-called Scandinavian accent orbit (Öhman, 1967) based on data from Meyer (1937, 1954) describing how word accents vary within Scandinavia. In Swedish, it is possible to find tonal patterns that are similar for a pair of dialects, but signal opposite distinctions and a wrong word accent assignment in e.g. a speech synthesis system may not be perceived as an error, but as the system talking with a regional accent.

Apart from language technology and speech synthesis, where intonation is an established application, diverse areas of medical as well as educational applications where intonation is less commonplace are being developed.

### 4.1. Technological applications

The most wide-spread technological application of intonation is related to the development of speech synthesis systems. Other areas are however also being developed within the speech and language technology context, the mainstream of which are presented below.

#### 4.1.1. Speech synthesis

One motivation for the development of computational models for intonation is the use in speech synthesis systems. In recent times we have seen considerable improvements in the rendering of intonation and prosody in general, with emphasis being placed on improved naturalness. Traditionally speech synthesis systems are capable of talking only in one voice, in one speaking style. This by-passes the difficulty with the inherent variability in the phonetics of intonation. Lately more emphasis has been placed on different speaking styles, different voices and also speech in different emotions and other pragmatic effects which increase the need for a more flexible intonation model (see Morton and Tatham, 1995; Tams and Tatham, 1995; Morlec et al., 2001). An EU project, voices attitudes and emotions in speech synthesis (VAESS) with that ambition has just finished. Examples from recent work can be found in (Bertenstam et al., 1997). There is also currently some emphasis on the problems of rendering synthetic intonation used as part of a dialogue system (e.g., Tatham and Morton, 1995).

#### 4.1.2. Speech recognition

Studies to apply prosodic knowledge to speech recognition have a long tradition. One early example is Lea's Ph.D. dissertation from 1972, followed by a series of studies. Much of the early work is summarised in (Lea, 1980). However, the use of intonation and other prosodic factors in speech understanding systems has largely been neglected in commercial systems. The most obvious reason is the difficulty of integrating suprasegmental information in the mainstream speech recognition algorithms like hidden Markov models (HMM), which are basically segmental in nature. An efficient and statistically sound combination of the two sources of knowledge is hence hard to obtain. Another reason is the lack of large prosodically annotated speech databases, so important for training data-driven speech recognisers. This is changing with the increased use of notational systems like ToBI.

On the research level, we have seen a considerable activity in prosody driven speech understanding in the last few years. Most of the studies are restricted in scope, and could be regarded more as feasibility studies, which have shown the effectiveness of prosodic information in speech understanding. In a recent paper from the Verbmobil project (Niemann et al., 1997; Lieske et al., 1997), it is proudly claimed that they are the first to put prosody in efficient operation in a complete speech understanding system. Analysis of phrasal breaks and accentuation is claimed to bring down the processing time in the searches and to boost the performance considerably.

#### *4.1.3. Speaker verification*

Since intonation realisations vary so much with speakers, intonation should be an interesting candidate for use in speaker verification systems. However, scanning the literature, it is obviously not included in mainstream systems. Already in 1972, Atal had published a paper, based on his Ph.D. dissertation where he reports on a small experiment on speaker recognition based on intonation. 10 female speakers were used, reading the same, all voiced sentence six times (Atal, 1972). 97% correct identification was obtained. This is certainly encouraging, but the practice of using a single sentence is largely abandoned due to the risk of impostors using recordings. Other ways of representing the intonation, applicable to arbitrary utterances, must be sought. Matsui and Furui (1990) demonstrated that introducing local pitch and delta pitch in a vector quantisation based method decreased the error rate in a nine speaker identification experiment from 4% to 1%, in a text-independent mode. In this case the sentence used for identification was 30 s long, impractical in a real application.

A related study, with the restricted scope of speaker sex identification, is worth mentioning. Parris and Carey (1996) added pitch estimations in vowels to an acoustic classifier and reports a performance increase from a 4.2% to a 0.7% error rate (pitch alone yielded an error rate of 2.8%). In a system-prompted verification system, we could speculate whether local intonation measures related to an intonation model could be efficient. Deviations from the model could be used to parametrise idiosyncratic speaker behaviour.

#### *4.1.4. Language identification*

Automatic language identification could be important especially in different telecom applications, when the spectral content of the speech could be expected to be distorted. Intonation cues are in this case especially interesting. The varied prosodic structure of languages could be exploited in this application. In a study by Thymé-Gobbel and Hutchins (1996) a variety of prosodic features were used in a language identification task, including Chinese, Japanese, Spanish and English. Syllable based pitch measures (pitch contour and differential pitch) were shown to be quite efficient for discrimination. As is the case for the other recognition tasks described in previous sections, the intonation cues need to be combined with other types of information.

### *4.2. Medical applications*

Since intonation has such a profound importance for human communication and also mirrors personality and emotional state it is natural to find the use of intonation research in medical applications. Expertise on this area is rather limited, but we will mention a couple of cases from two different areas. One in psychiatry and one in rehabilitation of voice disorders.

The hypothesis that a person's emotional state could be inferred from an analysis of his/her intonational patterns was tested in a study by Nilsonne (1987). Several different $F_0$ measures were used, like $F_0$ mean, rate of change, $F_0$ histograms etc. from a story read by the patients. A high correlation to psychological variables was documented. Intonation tests, using both production and perception, could potentially be used in other areas, e.g. diagnosis of brain damage.

In voice pathology leading to laryngectomy the normal voice laryngeal source needs to be replaced. There are essentially two ways of doing this. One is to rely on oesophageal speech and to rely on a new voice source that is formed by a constriction in the upper oesophagus. This voice source could be powered by swallowed air or air from the lungs routed through a new passage created by surgery, connected to a mechanical valve on the neck. Surprisingly good functional intonation is obtained by many patients (Nord et al., 1995).

However, some patients have to resort to an "artificial larynx", essentially a vibrator placed on the neck. Normally the frequency and on/off is controlled manually. As is obvious from the study of intonation, complex and fast variation of $F_0$ is needed. This puts a high demand on both dexterity and training, where intonation functions are made conscious. This is at least true for the training phase, while eventually the control should be more unconscious and automatic. In this respect the training of laryngectomised patients has much in common with training of intonation for second language learners.

One device, Servox Inton, has a built-in feature that should ease its operation. The device is developed in the Dutch intonation tradition and relies on a study by Van Geel (1983). It is built on the observation that stress groups often have a declination. The device could manually be switched between two pitch levels and automatically adds a declination. To the present audience it should come as no surprise that this strategy created problems in many cases. Final rising intonation is commonplace, and in e.g. Swedish the quite important word accent distinction was not possible to produce. The problem must be even more acute in languages with syllable tone distinctions. However, the underlying idea that an intonation model could be used is interesting. However, if it is to be at all useful, it needs to be programmable to adapt to the different needs of languages, dialects and personalities.

### 4.3. Educational applications

Intonation much like phonetics in general has traditionally had a significant place in language teaching and especially in foreign language teaching (Jones, 1956; Gimson, 1962; Brazil et al., 1980; Bannert, 1990). Language teaching and pronunciation training is also associated with speech and language technology, some examples of which are presented below.

#### 4.3.1. Prosodic training on your own voice

The idea of minimising the influence of a teacher's voice, by using the student's segmental realisations with superimposed teacher prosody was tried in a different study of the Spell project, using LPC techniques (Wang et al., 1993). A somewhat similar training environment has been reported by Meron and Hirose (1996) and a demonstrator on speech technology in language training has been produced (Sundström, 1997).

The architecture of the demonstrator can be seen in Fig. 20. The resynthesis is based on a combination of PSOLA analysis (Hamon et al., 1989) and prosodic information from a teacher model. Normally this is an utterance of a teacher scaled to the appropriate pitch level and range of the student. Provisions are also made for using the

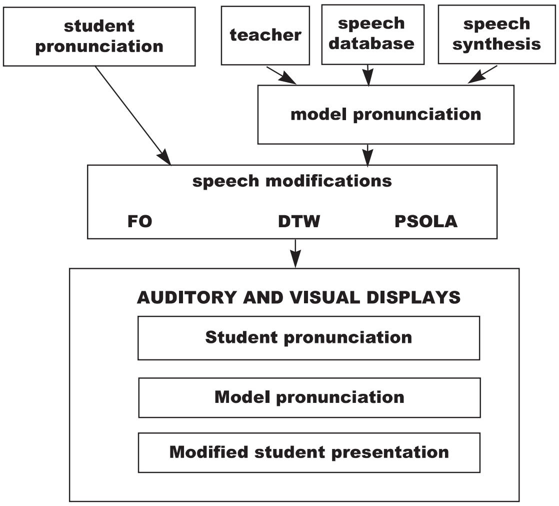

Fig. 20. Structure of the experimental training module (adopted from Sundström, 1997).

prosody model within the KTH text-to-speech system. This makes it possible to use the system without a teacher on-line or to use training material that is not in the teacher database. This material could be supplied as text by the teacher, by the student or by a program that is monitoring the student's progress. In the training, model governed variation in intonation could be introduced through different commands to the synthesiser, such as emphasis, speaking style or even dialect. The visual auditory display could show both the production of the "teacher", the student and the combined student production with mapped prosody. No evaluation with real students has been performed yet.

#### *4.3.2. Foreign language training*

For persons with normal hearing, auditory feedback is obviously efficient in acquiring the mother tongue. However, we know from experience that correct prosody is notoriously difficult to master in a foreign language. Could visual feedback improve this situation?

Even if the literature is not in full agreement on this point, there is a strong indication that visual displays together with the normal auditory feedback results in more efficient intonation training.

Weltens and de Bot (1984) describe a series of experiments that show clear results of improvement in the audio-visual condition compared to auditory feedback only. Some indication is given that the advantage is stronger for inexperienced learners. A surprising result is that delaying the feedback by 250 ms or until the sentence is completed did not diminish the effect compared to immediate feedback.

In a recent study by Öster (1997), the Speech Viewer has been used in an experimental teaching class for grown-up immigrants learning Swedish. The addition of the computer supported pronunciation training contained both prosodic and segmental training. The attitude to this method was very positive and many of the students show quite striking improvements. Long term retention of the improvement still needs to be studied.

In the EU Spell project several alternative intonation training strategies were investigated. The problem with acceptable variability was solved by heavily smoothing the student's $F_0$ curve, normalising it to the model utterance and allowing some variation within a so called "pitch tunnel" between the "pitch anchor points" of the $F_0$ curve (Rooney et al., 1992). The segmentation that is needed to correctly identify these points in the student utterances is performed with HMM techniques.

#### *4.3.3. Providing feedback for deaf students*

For many years, intonation displays have been developed for use in speech and language training. Simple analogue instruments with lights and dials were used to provide feedback for deaf persons in an effort to place $F_0$ within an appropriate range (Risberg, 1976).

An early, classical example of using computers is reported by Nickerson et al. (1976). In this case the training was designed as a play, a precursor of computer games. The $F_0$ of a subject was controlling the vertical position of a dot (ball) that travelled from left to right on the screen. A stylised obstacle should be avoided and a target point (the basket) should be reached by varying the $F_0$. Many of these early devices never found their way into real life applications and their pedagogical impact has been questioned.

Around 1985 a PC-based speech training aid was developed at IBM in Paris. It contained a variety of programs aimed at teaching deaf students different aspects of pronunciation. Both game-like programs and more analytic modules were contained in the package, many aimed at intonation. This device has been further developed and commercialised as the IBM Speech Viewer. For deaf and severely hard-of-hearing persons the visual feedback of their intonation is essential in establishing acceptable pronunciation habits. The expectation is that associated tactile and proprioceptive feedback could help in maintaining the skills.

Clearly, better theories, descriptions and models of intonation, and prosody in general, could be profitably exploited in many areas, with reference to interdisciplinary technological, medical, and educational applications. Significant work is being carried on not only in established areas of intonation applications but also in areas where segmental phonetics has been the main concern.

## 5. Research perspectives

Before concluding this tutorial, some general considerations are to be made, namely, the meaning as well as the benefits of intonation studies is in a wider perspective. Intonation, as well as phonetics and linguistics in general, is *par excellence* a humanities area and the relevant knowledge concerns the human as a social being. Apart from knowledge as such, linguistic expertise may however have multi-dimensional effects with regards to social relations and quality of life. Speech and language technology, e.g., require integrated linguistic knowledge of both theoretical and application aspects which have given a boost to linguistic research during the last years.

There is a circular relation between linguistic knowledge and linguistic applications: there is by far much more knowledge than diverse applications may adequately respond to and, at the same time, applications require more and more thorough linguistic knowledge. Furthermore, the repertoire of linguistic applications grows steadily bigger, which puts new demands on linguistic knowledge. In addition, applications may have empirical implications for existing hypotheses and theoretical aspects as well as lead to new questions, methodological procedures and theory development.

As outlined in previous sections, the main applications of intonation are related to technology, medicine and education areas. Technological applications and especially speech technology constitute the most wide-spread among the applications of intonation. However, in recent years, speech technology does have a direct application on language education as well as language pathology within the broad areas of educational technology and medical technology respectively. Thus the closer link between intonation research and technological applications opens new perspectives to technological, educational and medical aspects of interdisciplinary research in the future years.

Intonation research is steadily widening and new issues as well as investigation areas are being developed. Well-established research paradigms in controlled speech environments are standard sources of new knowledge production and spontaneous speech analyses as well as discourse studies are well on the way with promising expectations in the immediate future. Apart from basic knowledge about aspects of discourse intonation such as accentuation, boundary signalling and topic changes, much is to be learned on the nature and meaning of intonation in spontaneous discourse. A basic taxonomy of tonal categories and tonal units as well as distinctive functions with regards to the organisation of spoken discourse are eminent issues.

More fundamental questions about the nature of intonation such as the contribution of intonation to the speech communication process as well as the relation of intonation to information units ought also to be considered. Intonation is part of the linguistic system and thus a natural question is its degree of contribution. In other words, what would be the effects of intonation neutralisation and lack of tonal distinctions in speech communication? And, subsequently, what would the compensation from other linguistic components? On the other hand, language, and thus intonation, is the means to a communicative end which is the indented message in, presumably, an organised structure of information units. What is the relation between abstract messages and concrete language forms? How can we study the way messages and thus higher order thoughts are organised and realised through language? This type of question may be related to invariance and variability issues, in the sense that linguistic forms may have several degrees of variability, in accordance with the levels of abstraction up to and including message and information structure units in speech communication.

In summary, the way the study of intonation is carried out encompasses a wide-range of methodological approaches and theoretical backgrounds. New horizons are continuously opening and more and more disciplines are involved which are expected to give further boosting of intonation studies in the upcoming years.

## Acknowledgements

For comments and much useful feedback our thanks go to Robert Bannert, Gösta Bruce, Albert Di Cristo, Gunnar Fant, Nina Grønnum, Carlos Gussenhoven, Alex Monaghan, Mark Tatham, Jean Véronis and Yi Xu.

## References

Adriaens, L.M.H., 1991. Ein Modell Deutscher Intonation. Ph.D. Dissertation, Technical University Eindhoven.

Atal, B.S., 1972. Automatic speaker recognition based on pitch contours. J. Acoust. Soc. Amer. 52, 1687–1697.

Atkinson, J.A., 1978. Correlation analysis of the physiological factors controlling fundamental voice frequency. J. Acoust. Soc. Amer. 63, 211–222.

Bannert, R., 1985. Towards a model for German prosody. Folia Linguistica XIX, 321–341.

Bannert, R., 1990. På Väg mot Svenskt Uttal. Studentlitteratur, Lund.

Bannert, R., Thorsen, N., 1988. Empirische Studien zur Intonation des Deutschen und Dänischen: Ähnlichkeiten und Unterschiede. Kopenhagener Beiträge zur Germanistischen Linguistik 24, 26–50.

Beckman, M.E., Ayers, M., 1997. (3rd version). Guidelines for ToBI labelling. Department of Linguistics, The Ohio State University (URL:http://www.ling.ohio-state.edu/phonetics/E_ToBI).

Beckman, M.E., Pierrehumbert, J.B., 1986. Intonation structure in Japanese and English. In: Phonology Yearbook, Vol. 3, pp. 255–309.

Bell, A.G., 1879. Vowel theories. American Journal of Otology 1, 163–180. Reprinted in: Bell, A.G. (Ed.), 1916. The Mechanisms of Speech. 8th ed. Funk and Wagnalls, New York, pp. 117–129.

Berinstein, A.E., 1979. A cross linguistic study of perception and production of stress. UCLA Working Papers in Phonetics, pp. 1–59.

Bertenstam, J., Granström, B., Gustafson, K., Hunnicutt, S., Karlsson, I., Meurlinger, C., Nord, L., Rosengren, E., 1997. The VAESS communicator: a portable communication aid with new voice types and emotions. Phonum 4, Department of Phonetics, Umeå University, Sweden, pp. 57–60.

Bloomfield, L., 1933. Language. Holt, Rinehart and Winston, New York.

Bolinger, D.L., 1958. A theory of pitch accent in English. Word 14, 109–149.

Botinis, A., 1989. Stress and Prosodic Structure in Greek. Lund University Press, Lund.

Botinis, A., 1992. Accentual distribution in Greek discourse. Travaux de l'Institut de Phonétique d'Aix 14, 13–52.

Botinis, A., 1998. Intonation in Greek. In: Hirst, D., Di Cristo, A. (Eds.), Intonation Systems: A Survey of Twenty Languages. Cambridge University Press, Cambridge, pp. 288–310.

Botinis, A., Bannert, R., 1997. Tonal perception of focus in Greek and Swedish. In: Proceedings of the ESCA Workshop on Intonation. Athens, Greece, pp. 47–50.

Botinis, A., Erkenborn, S., Isacsson, C., Westin, P., 1999. Prosodic variability and segmental durations in Greek and Swedish. In: Proceedings of the Swedish Phonetics Conference Fonetik 99. Gothenburg, Sweden, pp. 41–44.

Brazil, D., Coulthard, M., Johns, C., 1980. Discourse Intonation and Language Teaching. Longman, London.

Brazil, D., Hewings, M., Cauldwell, R., 1997. The Communicative Value of Intonation in English. Cambridge University Press, Cambridge.

Bresnan, J., 1971. Sentence stress and syntactic transformations. Language 47, 257–280.

Brown, G., Yule, G., 1983. Discourse Analysis. Cambridge University Press, Cambridge.

Brown, G., Currie, K., Kenworthy, J., 1980. Questions of Intonation. Croom Helm, London.

Bruce, G., 1977. Swedish Word Accents in Sentence Perspective. Gleerup, Lund.

Bruce, G., 1998. Allmän och Svensk Prosodi. In: Practical Linguistics, Vol. 16. Department of Linguistics, Lund University.

Bruce, G., Gårding, E., 1978. A prosodic typology for Swedish dialects. In: Gårding, E., Bruce, G., Bannert, R. (Eds.), Nordic Prosody. Gleerup, Lund, pp. 219–228.

Bruce, G., Granström, B., Gustafson, K., House, D., 1993. Prosodic phrasing in Swedish. In: Proceedings of the ESCA Workshop on Prosody, Lund, Sweden, pp. 180–183.

Carlson, R., Granström, B., Hunnicutt, S., 1990. Multi-language text-to-speech development and applications. In: Ainsworth (Ed.), Advances in Speech, Hearing, and Language Processing. JAI Press, London, pp. 269–296.

Chomsky, N., 1972. Deep structure, surface structure and semantic interpretation. In: Chomsky, N. (Ed.), Studies on Semantics on Generative Grammar. Mouton, The Hague, pp. 62–119.

Collier, R., 1975. Physiological correlates of intonation patterns. J. Acoust. Soc. Amer. 58, 249–255.

Cooper, W.E., Sorensen, J.M., 1981. Fundamental Frequency in Sentence Production. Springer, New York.

Cruttenden, A., 1997. Intonation, 2nd ed. Cambridge University Press, Cambridge.

Cutler, A., Ladd, D.R. (Eds.), 1983. Prosody: Models and Measurements. Springer, Heidelberg.

D'Alessandro, C., Mertens, P., 1995. Automatic pitch contour stylisation using a model of tonal perception. Comput. Speech Language 9, 257–288.

Daneš, F., 1960. Sentence intonation from a functional point of view. Word 16, 34–54.

de Pijper, J.R., 1983. Modelling British English Intonation. Foris, Dordrecht.

Di Cristo, A., 1998. Intonation in French. In: Hirst, D., Di Cristo, A. (Eds.), Intonation Systems: A Survey of Twenty Languages. Cambridge University Press, Cambridge, pp. 195–218.

Di Cristo, A., Hirst, D., 1986. Modelling French micromelody: analysis and synthesis. Phonetica 43, 11–30.

Di Cristo, A., Di Cristo, Ph., Campione, E., Véronis, J., 2000. A prosodic model for text-to-speech synthesis in French. In: Botinis, A. (Ed.), Intonation: Analysis, Modelling and Technology. Kluwer Academic Publishers, Dordrecht (in press).

Fant, G., Kruckenberg, A., Liljencrants, J., 2000. acoustic–phonetic prominence in Swedish. In: Botinis, A. (Ed.), Intonation: Analysis, Modelling and Technology. Kluwer Academic Publishers, Dordrecht (in press).

Fischer-Jørgensen, E., 1990. Intrinsic $F_0$ in tense and lax vowels with specific reference to German. Phonetica 47, 99–140.

Fourakis, M., Botinis, A., Katsaiti, M., 1999. Acoustic characteristics of Greek vowels. Phonetica 56, 28–43.

Fry, D.B., 1958. Experiments in the perception of stress. Language and Speech 1, 126–152.

Fujisaki, H., 1983. Dynamic characteristics of voice fundamental frequency in speech and singing. In: MacNeilage, P.F. (Ed.), The Production of Speech. Springer, New York, pp. 39–55.

Fujisaki, H., 1988. A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour. In: Fujimura, O. (Ed.), Vocal Physiology: Voice Production, Mechanisms and Functions. Raven, New York, pp. 347–355.

Garde, P., 1968. L'Accent. Press Universitaires, Paris.

Gårding, E., 1977a. The Scandinavian Word Accents. Gleerup, Lund.

Gårding, E., 1977b. The importance of turning points for the pitch patterns of Swedish accents. In: Hyman, L.M. (Ed.), Studies in Stress and Accent. Southern California Occasional Papers in Linguistics 4, Los Angeles, pp. 27–35.

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

Gårding, E., Botinis, A., Touati, P., 1982. A comparative study of Swedish, Greek and French intonation. Working Papers 22, Department of Linguistics & Phonetics, Lund University, pp. 137–153.

Gimson, A.C., 1962. An Introduction to the Pronunciation of English. Arlond, London.

Goldsmith, J.A., 1976a. Autosegmental Phonology. Ph.D. Dissertation, MIT (distributed by IULC and published 1979 by Garland Press, New York).

Goldsmith, J.A., 1976b. An overview of autosegmental phonology. Linguistic Analysis 2, 23–68.

Goldsmith, J.A., 1990. Autosegmental and Metrical Phonology. Basil Blackwell, Oxford.

Grønnum (Thorsen), N., 1992. The Groundworks of Danish Intonation: An Introduction. Museum Tusculanum Press, Copenhagen.

Grønnum (Thorsen), N., 1995. Superposition and subordination in intonation – a non-linear approach. In: Proceedings of the 13th International Congress – Phon. Sc. Stockholm, pp. 124–131.

Grønnum (Thorsen), N., 1998. Intonation in Danish. In: Hirst, D., Di Cristo, A. (Eds.), Intonation Systems: A Survey of Twenty Languages. Cambridge University Press, Cambridge, pp. 131–151.

Gussenhoven, C., 1984. On the Grammar and Semantics of Sentence Accents. Foris, Dordrecht.

Hadding-Koch, K., 1961. Acoustico-phonetic Studies in the Intonation of Southern Swedish. Gleerup, Lund.

Hadding-Koch, K., Studdert-Kennedy, M., 1964. An experimental study of some intonation contours. Phonetica 1, 175–185.

Halliday, M.A.K, 1967. Notes on transitivity and theme in English. J. Linguist. 3, 199–244.

Hamon, C., Moulines, E., Charpentier, F., 1989. A diphone system based on time-domain prosodic modifications of speech. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing 89, pp. 238–241.

Helmholtz, H., 1877. On the Sensations of Tone (translated by Ellis, A.J., 1885, Dover, New York).

Hirschberg, J., Pierrehumbert, J., 1986. The intonational structuring of discourse. In: Proceedings of the 24th Annual Meeting of the Association Computational Linguistics. New York, pp. 136–144.

Hirst, D., 1993. Detaching intonation phrases from syntactic structure. Linguist. Inquiry 24, 781–788.

Hirst, D., Di Cristo, A. (Eds.), 1998a. Intonation Systems: A Survey of Twenty Languages. Cambridge University Press, Cambridge.

Hirst, D., Di Cristo, A., 1998b. A survey of intonation systems. In: Hirst, D., Di Cristo, A. (Eds.), Intonation Systems: A Survey of Twenty Languages. Cambridge University Press, Cambridge, pp. 1–44.

Hirst, D.J., Ide, N., Véronis, J., 1994. Coding fundamental frequency patterns for multi-lingual synthesis with INTSINT in the MULTEXT project. In: Proceedings of the Second ESCA/IEEE Workshop on Speech Synthesis. New Paltz, NY, pp. 77–80.

Hirst, D.J., Nicolas, P., Espesser, R., 1991. Coding the $F_0$ of a continuous text in French: an experimental approach. In: Proceedings of the 12th International Congress – Phon. Sc. Aix-en-Provence. France, pp. 234–237.

Hockett, C.F., 1955. A Manual of Phonology. Waverley Press, Baltimore.

Hombert, J.M., 1978. Consonant types, vowel quality, and tone. In: Fromkin, V.A. (Ed.), Tone: A Linguistic Survey. Academic Press, New York, pp. 77–111.

House, D., 1990. Tonal Perception in Speech. Lund University Press, Lund.

Hyman, L.M., 1977. On the nature of linguistic stress. In: Hyman, L.M. (Ed.), Studies in Stress and Accent. Southern California Occasional Papers in Linguistics 4, Los Angeles, pp. 37–82.

Jackendoff, R., 1972. Semantic Interpretation in Generative Grammar. MIT Press, Cambridge, MA.

Jones, D., 1956. Outline of English Phonetics, 8th ed. Heffer, Cambridge.

Jun, S.-A., Fougeron, C., 2000. A phonological model of French intonation. In: Botinis, A. (Ed.), Intonation: Analysis, Modelling and Technology. Kluwer Academic Publishers, Dordrecht (in press).

Koenig, W., Dunn, H.K., Lacy, L.Y., 1946. The sound spectrograph. J. Acoust. Soc. Amer. 18, 19–49.

Kohler, K.J., 1987. The linguistic functions of $F_0$ peaks. In: Proceedings of the 11th International Congress – Phon. Sc. Tallinn, pp. 149–152.

Kohler, K.J., 1990. Macro and micro $F_0$ in the synthesis of intonation. In: Kingston, J., Beckman, M.E. (Eds.), Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech. Cambridge University Press, Cambridge, pp. 115–138.

Kohler, K.J., 1991. Prosody in speech synthesis: the interplay between basic research and TTS application. J. Phonetics 19, 121–138.

Ladd, D.R., 1983. Phonological features of intonational peaks. Language 59, 721–759.

Ladd, D.R., 1988. Declination 'reset' and the hierarchical organization of utterances. J. Acoust. Soc. Amer. 84, 530–544.

Ladd, D.R., 1996. Intonational Phonology. Cambridge University Press, Cambridge.

Ladefoged, P., 1967. Three Areas of Experimental Phonetics. Oxford University Press, London.

Ladefoged, P., McKinney, N.P., 1963. Loudness, sound pressure and sub-glottal pressure in speech. J. Acoust. Soc. Amer. 35, 454–460.

Lambrecht, K., 1996. Information Structure and Sentence Form. Cambridge University Press, Cambridge.

Lea, W.A., 1980. Prosodic aids to speech recognition. In: Lea, W.A. (Ed.), Trends in Speech Recognition. Prentice-Hall, Englewood Cliffs, NJ, pp. 166–205.

Leben, W., 1976. The tones in English intonation. Linguist. Anal. 2, 69–107.

Lehiste, I., 1970. Suprasegmentals. MIT Press, Cambridge, MA.

Liberman, M.Y., Pierrehumbert, J., 1984. Intonational invariants under changes in pitch range and length. In: Aronoff, M., Oehrle, R.T. (Eds.), Language Sound Structure. MIT Press, Cambridge, MA, pp. 157–233.

Liberman, M.Y., Prince, A., 1977. On stress and linguistic rhythm. Linguistic Inquiry 8, 249–336.

Lieberman, Ph., 1967. Intonation, Perception, and Language. MIT Press, Cambridge, MA.

Lieske, C., Bos, J., Gambäck, B., Emele, M., Rupp, C.J., 1997. Giving prosody a meaning. In: Proceedings of the European Conference on Speech Communication and Technology Eurospeech 97. Rhodes, Greece, pp. 1431–1434.

Malfrère, F., Dutoit, T., Mertens, P., 1998. Automatic prosody generation using suprasegmental unit selection. In: Proceedings of the Third ESCA Workshop on Speech Synthesis. Jenolan Caves, Australia, pp. 323–328.

Malmberg, B., 1967. Structural Linguistics and Human Communication. Springer, New York.

Martinet, A., 1954. Accent et tons. Miscellanea Phonetica 2, 13–24.

Matsui, T., Furui, S., 1990. Text-independent speaker recognition using vocal tract and pitch information. In: Proceedings of the International Conference on Spoken Language Processing 90. Kobe, Japan, pp. 137–140.

Meron, Y., Hirose, K., 1996. Language training system utilising speech modification. In: Proceedings of the International Conference on Spoken Language Processing 96. Philadelphia, PA, pp. 1449–1452.

Mertens, P., 1987. L'intonation du Franąis: de la description linguistique à la reconnaissance automatique. PhD Dissertation, Katholieke Universiteit Leuven, Leuven.

Mertens, P., 1989. Automatic recognition of intonation in French and Dutch. In: Proceedings of the European Conference on Speech Comminication and Technology Eurospeech Eurospeech 89. Paris, pp. 46–49.

Meyer, E.A., 1937. Die Intonation im Schwedischen I: Die Sveamundarten. Studies Scand. Philol. 10, University of Stockholm.

Meyer, E.A., 1954. Die Intonation im Schwedischen II: Die norrländischen Mundarten. Studies Scand. Philol. 11, University of Stockholm.

Möbius, B., 1993. Ein quantitatives Modell der deutschen Intonation: Analyse und Synthese von Grundfrequenzverläufen. Niemeyer: Tübingen.

Möbius, B., 1995. Components of a quantitative model of German intonation. In: Proceedings of the 13th International Congress – Phon. Sc. Stockholm. Sweden, pp. 108–115.

Möhler, G., 1998a. Theoriebasierte Modellierung der deutschen Intonation für die Sprachsynthese. University of Stuttgart, Stuttgart.

Möhler, G., 1998b. IMS Festival (http://www.ims.uni-stuttgart.de/phonetik/synthesis/index.html).

Monaghan, A.I.C., 1993. What determines accentuation?. J. Pragmat. 19, 559–584.

Monaghan, A.I.C., 1998. State-of-the-art summary of European synthetic prosody R&D (http://www.compapp.dcu.ie/alex/PUB/soap.html).

Monaghan, A.I.C., Ladd, D.R., 1990. Symbolic output as the basis for evaluating intonation in text-to-speech systems. Speech Communication 9, 305–314.

Morlec, Y., Bailly, G., Aubergé, V., 2001. Generating prosodic attitudes in French: Data, model and evaluation. Speech Communication 33 (4), 357–371.

Morton, K., Tatham, M., 1995. Pragmatic effects in speech synthesis. In: Proceedings of the European Conference on Speech Communication and Technology Eurospeech 95. Madrid, Spain, pp. 1819–1822.

Nespor, M., Vogel, I., 1986. Prosodic Phonology. Foris, Dordrecht.

Nickerson, R.S., Kalikow, D.N., Stevens, K.N., 1976. Computer-aided speech training for the deaf. J. Speech Hearing Disorders 41, 120–132.

Niemann, H., Nöth, E., Kieβling, A., Kompe, R., Batliner, A., 1997. Prosodic Processing and its use in Verbmobil. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing 97. München, Germany, pp. 75–789.

Nilsonne, Å., 1987. Speech in depression: a methodological study of prosody. Ph.D. Dissertation, Karolinska Institute, Stockholm.

Nord, N., Hammarberg, B., Lundström, E., 1995. Laryngectomee speech in noise – voice effort, speech rate and intelligibility. Scand. J. Logopedics Phoniatrics 20, 107–112.

Odé, C., 1989. Russian Intonation: A Perceptual Description. Rodopi, Amsterdam.

Ohala, J.J., 1978. Production of tone. In: Fromkin, V.A. (Ed.), Tone: A Linguistic Survey. Academic Press, New York, pp. 5–39.

Öhman, S.E.G., 1967. Word and sentence intonation: a quantitative model. STL-QPSR 2–3, 20–54.

Öhman, S.E.G., Lindqvist, J., 1966. Analysis-by-synthesis of prosodic pitch contours. STL-QPSR 4, 1–6.

Öster, A.-M., 1997. Auditory and visual feedback in spoken L2 teaching. Phonum 4, Department of Phonetics, Umeå University, Sweden, pp. 145–148.

Parris, E., Carey, M. 1996. Language independent gender identification. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing 96. Atlanta GE, pp. 685–688.

Pierrehumbert, J.B., 1980. The Phonology and Phonetics of English Intonation. PhD Dissertation, MIT, MA (Published in 1988 by IULC).

Pierrehumbert, J.B., Beckman, M.E., 1988. Japanese Tone Structure. MIT Press, Cambridge, MA.

Pike, K.L., 1945. The Intonation of American English. University of Michigan Press, Ann Arbor.

Potter, R.K., 1945. Visible patterns of science. Science 102, 463–470.

Potter, R.K., Kopp, G.A., Green, H.C., 1947. Visible Speech. Van Nostrand, New York.

Risberg, A., 1976. Visual aids for speech correction. Amer. Ann. Deaf, 178–194.

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. Banff, Alberta, pp. 413–416.

Rossi, M., 1978. Interaction of intensity glides and frequency glissandos. Language and Speech 21, 284–396.

Rossi, M., 1985. L'intonation et l'organisation de l' énoncé. Phonetica 42, 135–153.

Rossi, M., 1999. L'Intonation, le Système du Francais: Description et Modélisation. Editions Ophrys, Paris.

Rossi, M., 2000. Intonation: past, present, future. In: Botinis, A. (Ed.), Intonation: Analysis, Modelling and Technology. Kluwer Academic Publishers, Dordrecht (in press).

Searle, J.R., 1969. Speech Acts. Cambridge University Press, Cambridge.

Searle, J.R., 1976. A classification of illocutionary acts. Language in Society 5, 1–23.

Searle, J.R., 1979. Expression and Meaning. Cambridge University Press, Cambridge.

Silkerk, E.O., 1984. Phonology and Syntax: The Relation between Sound and Structure. MIT Press, Cambridge, MA.

Silverman, K., 1987. The Structure and Processing of Fundamental Frequency Contours. Ph.D. Dissertation, University of Cambridge, Cambridge.

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., Hirschberg, J., 1992. ToBI: A standard for labelling English prosody. In: Proceedings of the International Conference on Spoken Language Processing 92. Banff, Alberta, pp. 867–870.

Stockwell, R.P., 1972. The role of intonation: reconsiderations and other considerations. In: Bolinger, D.L. (Ed.), Intonation. Penguin Books, Harmondsworth, pp. 87–109.

Sundström, A., 1997. Speech technology in language learning. M.Sc. thesis, KTH, Stockholm (in Swedish).

Tams, A., Tatham, M., 1995. Describing speech styles using prosody. In: Proceedings of the European Conference on Speech Communication and Technology Eurospeech 95. Madrid, Spain, pp. 2081–2084.

Tatham, M., Morton, K., 1995. Speech synthesis in dialogue systems. In: Dalsgaard, P. (Ed.), Spoken Dialogue Systems. (Visgo, Denmark, ESCA, 1995), pp. 221–225.

Tatham, M., Morton, K., 2000. Speech prosodics for synthesis - perspectives. In: Proceedings of the Swedish Phonetics Conference Fonetik 2000. Skövde, Sweden, pp. 133–136.

Taylor, P.A., 1994. A Phonetic Model of Intonation in English. Indiana University Linguistics Club, Bloomington.

Taylor, P.A., 2000. Analysis and synthesis of intonation using the Tilt model. J. Acoust. Soc. Amer. 107, 1697–1714.

Terken, J., 1993. Synthesizing natural – sounding intonation for Dutch: rules and perceptual evaluation. Comput Speech Language 7, 27–48.

t'Hart, J., 1998. Intonation in Dutch. In: Hirst, D., Di Cristo, A. (Eds.), Intonation Systems: A Survey of Twenty Languages. Cambridge University Press, Cambridge, pp. 96–111.

t'Hart, J., Collier, R., 1975. Integrating different levels of intonation analysis. J. Phonetics 3, 235–255.

t'Hart, J., Collier, R., Cohen, A., 1990. A perceptual Study of Intonation. Cambridge University Press, Cambridge.

Thorsen (Grønnum), N., 1978. An acoustical analysis of Danish intonation. J. Phonetics 6, 151–175.

Thorsen (Grønnum), N., 1982. On the variability of $F_0$ patterning and the function of $F_0$ timing in languages where pitch cues stress. Phonetica 39, 302–316.

Thymé-Gobbel, A., Hutchins, S., 1996. On using prosodic cues in automatic language identification. In: Proceedings of the International Conference on Spoken Language Processing 96, Philadelphia, PA, pp. 1768–1771.

Trager, G.L., Smith, H.L., 1951. An Outline of English Structure. Battenburg Press, Norman, Oklahoma.

Trubetzkoy, 1939. Grundzüge der Phonologie (translated 1969 by Baltaxe, C.A.M. Principles of Phonology, University of California Press, Berkley).

Vaissière, J., 1995. Phonetic explanations for cross-linguisitic similarities. Phonetica 52, 123–130.

Van Geel, R., 1983. Pitch inflection in electrolaryngeal speech. Ph.D. Dissertation, University of Utrecht.

Van Hemert, J.P., Adriaens-Porzig, U., Adriaens, L.M., 1987. Speech synthesis in the SPICOS project. In: Tillmann, H.G., Willée, G. (Eds.), Analyse und Synthese Gesprochener Sprache. Olms, Hildesheim, pp. 34–39.

Van Heuven, V.J., Haan, J., 2000. Phonetic correlates of statement versus question intonation in Dutch. In: Botinis, A. (Ed.), Intonation: Analysis, Modelling and Technology. Kluwer Academic Publishers, Dordrecht (in press).

Van Heuven, V.J., Pols, L.C.W. (Eds.), 1993. Analysis and Synthesis of Speech. Mouton de Gruyter, Berlin and New York.

Van Santen, J.P.H., 1993. Perceptual experiments for diagnostic testing of text-to-speech systems. Comput. Speech Language 7, 49–100.

Van Santen, J.P.H., Hirschberg, J., 1994. Segmental effects on timing and height of pitch contours. In: Proceedings of the International Conference on Spoken Language Processing 94, Yokohama, pp. 719–722.

Van Santen, J.P.H., Möbius, B., 1997. Modelling pitch accent curves. In: Proceedings of the ESCA Workshop on Intonation. Athens, Greece, pp. 321–324.

Van Santen, J.P.H., Möbius, B., 2000. A qualitative model of $F_0$ generation and alignment. In: Botinis, A. (Ed.), Intonation: Analysis, Modelling and Technology. Kluwer Academic Publishers, Dordrecht (in press).

Van Santen, J.P.H., Möbius, B., Venditti, J., Shih, C., 1998. Description of the Bell Labs intonation system. In: Proceedings of the Third ESCA Workshop on Speech Synthesis. Jenolan Caves, Australia, pp. 293–298.

Venditti, J.J., Maeda, K., van Santen, J.P.H., 1998. Modelling Japanese boundary pitch movements for speech synthesis. In: Proceedings of the Third ESCA Workshop on Speech Synthesis. Jenolan Caves, Australia, pp. 317–322.

Véronis, J., Campione, E., 1998. Towards a reversible symbolic coding of intonation. In: Proceedings of the International Conference on Spoken Language Processing 98. Sidney, pp. 2899–2902.

Véronis, J., Di Cristo, Ph., Courtois, F., Chaumette, C., 1998. A stochastic model of intonation for text-to-speech synthesis. Speech Communication 26, 233–244.

Wang, H.D., Degryse, D., Carraro, F., 1993. A prosody modification approach for auditory user feedback in the SPELL pronunciation teaching system. In: Proceedings of the European Conference on Speech Communication and Technology Eurospeech 93. Berlin, pp. 991–994.

Weltens, B., de Bot, K., 1984. The visualisation of pitch contours: some aspects of its effectiveness in teaching foreign languages. Speech Communication 3, 157–163.

Whalen, D.H., Levitt, A.G., 1995. The universality of intrinsic $F_0$ of vowels. J. Phonetics 23, 349–366.

Wheatstone, C., 1837. Reed organ pipes, speaking machines, etc. Westminster Review 27, 30–37 (reprinted in Scientific Papers of Sir Charles Wheatstone, 1879. Taylor and Francis, London, pp. 48–367).

Willems, N., Collier, R., t'Hart, J., 1988. A synthesis scheme for British English intonation. J. Acoust. Soc. Amer. 84, 1250–1261.

Xu, Y., 1997. Contextual tonal variations in Mandarin. J. Phonetics 25, 61–83.

Xu, Y., 1999. Effects of tone and focus on the formation and alignment of $F_0$ contours. J. Phonetics 27, 55–105.

Xu, Y., Wang, E., 2001. Pitch targets and their realization: Evidence from Mandarin Chinese. Speech Communication 33 (4), 319–337.