INTONATIONAL PRIMITIVES

Amalia Arvaniti

Paper submitted to Marc van Oostendorp, Colin Ewen, Beth Hume and Keren Rice (Eds.), *Companion to Phonology*, Wiley-Blackwell.

1. Introduction

It is a well-known truism that no utterance is ever produced in a strict monotone; all utterances, in all languages, show some pitch modulation. Such changes in pitch – impressionistically described as rises and falls – are due to changes in fundamental frequency or F0, the physical property of the speech signal that is determined by the basic rate of vibration of the vocal folds and gives rise to the percept of pitch.

Although pitch modulations exist in all languages, their origin and function differs, in that pitch patterns may be specified at both the lexical and phrasal level or only at the phrasal level, resulting in more or less dense tonal specifications respectively (Gooden, Drayden & Beckman, 2009). The term *intonation* is used to refer to phrasal tonal patterns, while the terms *pitch accent* and *tone* are traditionally used to refer to lexical tonal specifications. Simplifying somewhat, in languages like English, Italian or Greek and many other European languages the entire F0 contour is specified at the phrasal level by means of a complex interplay between metrical structure, prosodic phrasing, syntax and pragmatics; these factors determine where pitch movements will occur and of what type they will be. In languages referred to as *tone languages* – such as Mandarin, Thai or Igbo – most syllables are lexically specified for tone and tonal changes affect lexical meaning; in languages often referred to as *pitch accent languages* – such as Japanese, Swedish, or Serbian – tone operates in a similar fashion, except that at most one syllable in each word is lexically specified for tone. In both tone and pitch accent languages additional tonal patterns are specified at the phrasal level. Here the focus is on languages without lexical tonal specifications, since it is the intonation of these languages that has been mostly examined.

Determining the structure of pitch modulation and the primitives that make up pitch contours in languages without lexical tone is challenging, since F0 changes are not as discrete and easily identifiable as in "tonal" languages, their connections to segmental material are less easy to determine, and associated meanings are harder to pinpoint since they deal with information structure and pragmatic interpretation rather than lexical semantics. The following examples illustrate these points. In Figure 1, two utterances are shown, *Me?!* and *A ballgown designer?!,* both using the rise-fall-rise melody that implies incredulity (Ward & Hirschberg, 1985; Hirschberg & Ward, 1992). They are plausible responses to a career advisor's pronouncement that according to test results designing ballgowns is the recommended career choice for the speaker, who has all along dreamed of becoming an aerospace engineer (for similar examples, see Ladd, 2008:45-46). Although the short contour can be informally described as rise-fall-rise, the longer contour cannot be described in a similar fashion, as it shows a long low-level stretch between a rise-fall and a final rise. In Figure 2, Greek contours very similar to the English ones in Figure 1 are shown, though in the case of Greek these contours are used for wh-

questions (Arvaniti & Baltazani, 2005; Arvaniti & Ladd, 2009). As can be seen, the same issue with overall shape arises here as well. Further, as Arvaniti & Baltazani (2005) note, the Greek melody in Figure 2 can also be used for polite requests employing an imperative; e.g. [ˈðose sti maˈria ˈliɣo neˈraci] "give Maria some water" (*lit.* give to Maria a-little water-DIM.). Finally Figure 3 illustrates two instances of another English melody: unlike the contours in Figures 1 and 2 which look different from each other but convey the same meaning in each case, the contours of Figure 3 are realizations of the same melody but convey different meaning, depending on the utterance: the melody used is the default for *That's twenty dollars*, but sounds blasé or sarcastic when used with *That's really awesome*. Note that this is not so because the melody is wrong for *That's really awesome*: the use is legitimate and meaningful (if the speaker wishes to be sarcastic or convey her indifference).

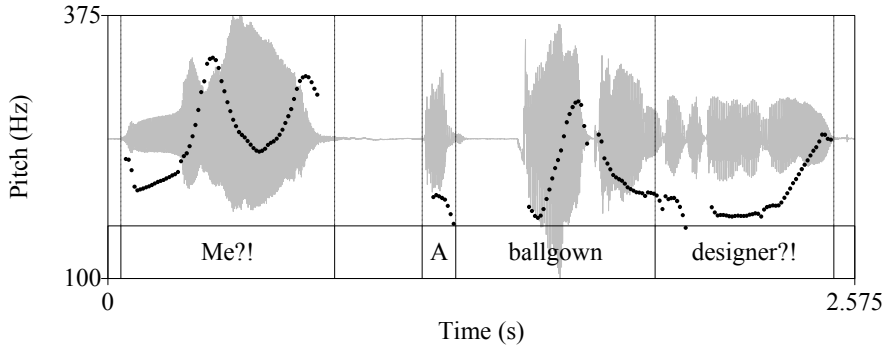

Figure 1: Waveforms, and F0 contours of two English utterances illustrating the incredulity contour (Hirschberg & Ward, 1992); on the left, *Me?!*; on the right, *A ballgown designer?!*. Vertical lines indicate word boundaries.

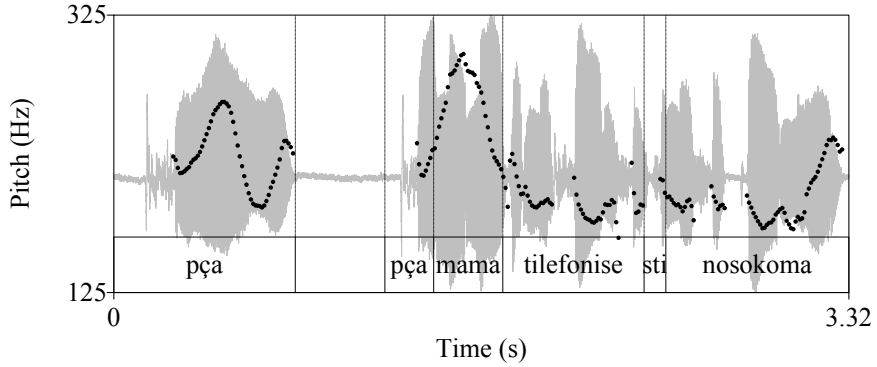

Figure 2: Waveforms, transcriptions and F0 contours of two Greek wh-questions, on the left, [ˈpça] "which, fem.", on the right, [ˈpça maˈma tileˈfonise sti nosoˈkoma] "which mom called the nurse?". Vertical lines indicate word boundaries.

These examples illustrate three main points about intonation. First, they show that the shape of intonational contours with a given pragmatic interpretation can vary substantially depending on the segmental material with which they are uttered. Such differences are not random, but related to the overall prosodic structure of the utterance

with which the contours are associated, including the number of syllables and the position of stresses (where applicable). Second, the examples show that contours do not have a constant meaning either within or across languages; within a language, their interpretation may well depend on lexical and other choices that accompany the use of the melody; across languages differences can be arbitrary. A successful theory of intonation should be able to capture these properties: it should be able to explain the connection between intonation and meaning and make generalizations from surface F0 data with sufficient predictive power to generate new contours of the same basic melody to "fit" new utterances of varying lengths and structures.

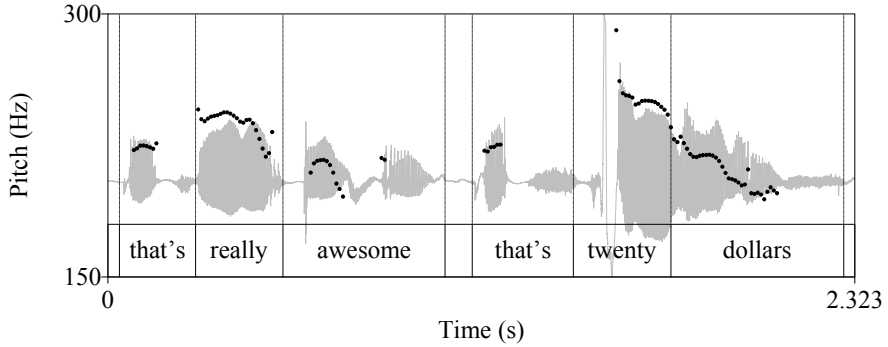

Figure 3: Waveforms, and F0 contours of two English utterances; on the left, *That's twenty dollars*; on the right, *That's really awsome.* Vertical lines indicate word boundaries.

Although the above observations are by and large shared by most intonational models, the ways in which they treat these basic properties shows substantial variation. As discussed in more detail in section 2, many researchers have treated F0 contours as gestalts or *configurations* that is, as holistic pitch movements that encompass entire utterances and have a uniform meaning. In other models, melodies are seen as being composed of primitives of some sort. These primitives are considered to be either local configurations (or *dynamic* tones), such as local rises and falls, or level tones, such as high, mid and low. Here I review both the controversy between advocates of gestalt approaches to intonation and those who proposed analyses based on the decomposition of melodies into smaller elements, and the disagreement between researchers who use dynamic tones (that is, local configurations) as the primitives of intonational structure and those who advocate the use of level tones instead.

As I show, however, focusing only on the form of intonational primitives avoids an even more fundamental question, namely which parts of a contour should be represented at all. This question is addressed in more detail in section 3.2, where the main argument is advanced that an inordinate attention to phonetic form and a reluctance to accept that intonation is part of a language's phonology have hampered research and have lead to many analyses that by and large aim at reproducing entire melodies, but have little predictive power and cannot successfully generalize beyond the F0 contours they reproduce. As argued in section 3.2, sparse representations that aim at capturing only the

linguistically significant aspects of each contour can better handle both intonational form and intonational meaning.

2. Configurational models

2.1. Melodies as gestalts

As mentioned, many researchers have treated intonation contours as gestalts, such as Bolinger (1951), Jones (1972), Cooper & Sorensen (1981), Hirst & Di Cristo (1998), Grabe, Kochanski & Coleman (2003), Xu (2005). In these works, it is most often the case that intonational contours are seen as holistically and directly reflecting certain functional or structural aspects of speech, such as the distinction between questions and statements or that between levels of phrasing.

According to Jones (1972: 279) – who in the last edition of *An Outline of English Phonetics* followed several earlier intonational analyses, notably those of Armstrong & Ward (1926) and Kingdon (1958) – English has two basic tunes, Tune 1 and Tune 2. These are a fall and rise respectively, that "may be spread over a large number of syllables, or […] be compressed into smaller spaces." Bolinger (1951: 208) also concludes his critique of level-based analyses (see section 3.1) by saying that "intonation could not be a more appropriate illustration of the Gestalt." More recently, Cooper & Sorensen (1981) presented a series of experiments in which contours are treated as undivided wholes and peak height is taken to directly reflect levels of phrasing.

Modern versions of the gestalt approach include INTSINT (INternational Transcription System for INTonation; e.g. Hirst & Di Cristo, 1998, chap. 1; Hirst, Di Cristo & Espesser, 2000), OXIGEN (OXford Intonation GENerator; Grabe et al., 2003), and PENTA (Parallel Encoding and Pitch Target Approximation; e.g. Xu, 2005). In INTSINT entire intonation contours are transcribed using symbols that represent pitch movements. The movements, however, are not seen as primitives but rather as a means to transcribing the course of F0 over an entire utterance (hence their descriptive labels – *Higher*, *Lower*, *Upstepping*, *Downstepping*, *Same*, *Top* and *Bottom* –which express the course of F0 relative to preceding points and the overall range of the speaker). In OXIGEN polynomials are used to model overall contour shape differences between statements and questions in British English. Finally, in PENTA each syllable in a contour has its own pitch specification, while global aspects of the overall melody are directly linked to functional effects (a feature shared with OXIGEN); e.g., the use of an utterance as statement or question is said to lead to changes in overall pitch shape from fall to rise respectively (Xu, 2005).

Configurational approaches have been quite popular for several reasons. First, their appeal is intuitive: F0 contours are more or less continuous so, as Bolinger (1951: 206) put it, intonation can be seen as "a pattern […] in the fundamental, down-to-earth sense of a continuous line that can be traced on a piece of paper" (though, as noted in Arvaniti, 2007, the fact that F0 looks continuous on paper or a computer monitor does not necessarily mean that it is perceived in this fashion). Second, the relationship between

shape and function seems sufficiently natural in many cases that it has been argued to derive from the biological underpinnings of pitch production (e.g. Ohala, 1983; for a thorough discussion of the *biological code*, see Gussenhoven, 2004, chap. 5, who, however, does not adopt a configurationalist approach to intonation). This view is supported by certain general trends, such as the use of rising F0 for questions and falling F0 for statements for which it is possible to find empirical evidence in several languages (e.g. Grabe et al., 2003).

Despite their popularity, configurational approaches face a major problem when it comes to accounting for intonational form. Specifically, if melodies were undivided wholes they should simply shrink and stretch to "fit" the segmental duration of the utterance they co-occur with. There is plenty of evidence, however, that when tunes are matched with utterances of varying lengths and different metrical structures they are not realized in this simple manner. This was observed by 't Hart, Collier & Cohen (1990, chap. 4), among the first researchers to use instrumental rather than impressionistic data for intonation research (e.g. Cohen & 't Hart, 1968; 't Hart & Cohen, 1973; 't Hart & Collier, 1975 inter alia). They noticed that in their Dutch corpora sequences of pitch movements would appear on a single syllable in some instances but separated by several syllables in others (cf. Figures 1 and 2). Importantly, 't Hart et al. found that this *elasticity*, as they named it, did not affect the melodic identity of the contour (determined by means of perceptual experiments; see section 2.3), even though it radically altered the overall contour shape (thus, IPO's concept of elasticity can be juxtaposed to the *compression* envisaged by Jones, 1972, which implies greater uniformity in the squeezing and stretching of contours).

Results from several later studies support the original observations of the IPO researchers that certain aspects of the contour are important for listeners, while overall contour shape is not. Pierrehumbert & Steele (1989) varied in steps the position of the pitch peak in English melodies that can be holistically described as rise-fall-rise and found that listeners imitating these stimuli produced not a continuum but two different contours, one with an early and one with a late peak. Similar results have also been presented by Redi (2003), following Pierrehumbert & Steele's imitation protocol (argued by Gussenhoven, 1999, to be the best way to examine categoriality in intonation). Similarly categorical responses to intonational continua that would be holistically seen as instances of the same contour have been obtained by Rietveld & Gussenhoven (1995) for Dutch, D'Imperio & House (1997) for Neapolitan Italian, and Botinis (1989) for Greek, inter alia. The contours in Figures 1 and 2 also illustrate this general point. In the monosyllabic utterances, the rise-fall-rise stretches over the entire syllable. In the longer utterances, however, the rise co-occurs with the first stressed syllable (with some peak delay) and the final rise is realized on the last syllable, while the fall and subsequent low-level stretch vary depending on the language and length of the utterance. As a result, the contours of the longer utterances are not stretched out copies of the shorter contours, nor are the shorter contours compressed versions of the longer contours; these differences, amply discussed in Arvaniti & Ladd (2009), are illustrated in Figure 4, using the Greek contours of Figure 2.

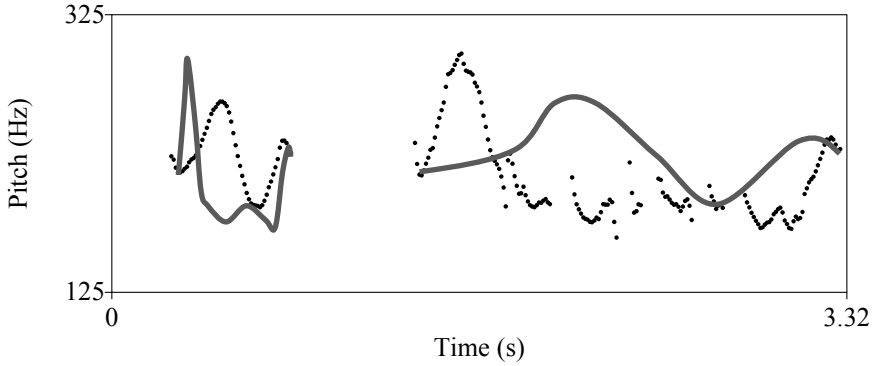

Figure 4: F0 contours of the Greek wh-questions shown in Figure 2; the thick gray lines are compressed and stretched out copies of the long and short contour respectively.

Differences in contour shape may also relate to the number and position of stressed syllables in an utterance and the location of the word in focus. This is experimentally shown in Arvaniti, Ladd & Mennen (2006a) who studied the default melody of Greek polar questions in which the position of focus can vary. They show that the shape of the contour is strongly affected by the position of the stressed syllables and of the word in focus (see also Arvaniti & Baltazani, 2005; Arvaniti, 2007). The focus effect in particular is illustrated in Figure 5, which shows the two melodies that can be used with the sentence [ˈpinun lemoˈnaða] "they-drink lemonade" when it is uttered as a question with focus on the verb (dotted line) or the noun (solid line). As can be seen, no description in terms of overall shape can possibly cover both contours; at best, the late focus question would be characterized as rise-fall-rise-fall and the early focus question as rise-fall, but this ignores the location of the rise-fall part that the two contours share and the significance of this location both for understanding the pragmatics of the two questions and for their naturalness (for details, see Arvaniti et al., 2006a).

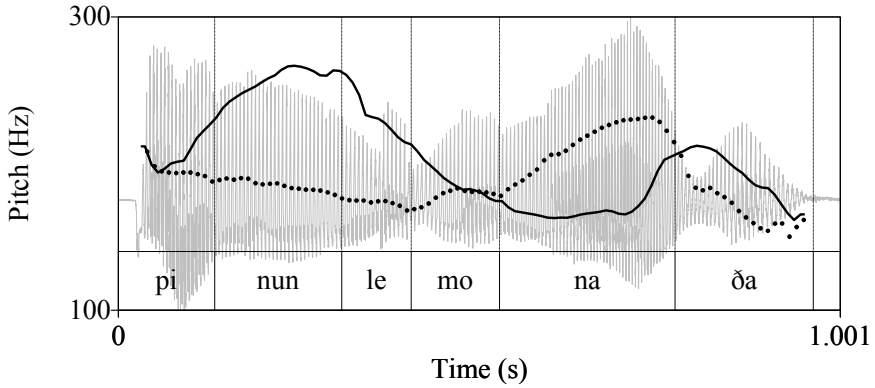

Figure 5: Waveform, transcription and F0 contours for the phrase [ˈpinun lemoˈnaða] "they-drink lemonade" uttered as a question with focus on [ˈpinun] "they-drink" (dotted contour) or on [lemoˈnaða] "lemonade" (solid contour); the former could be glossed as "do they like lemonade?" and the latter as "is it lemonade that they are drinking?". Vertical lines indicate syllable boundaries.

Holistic approaches, then, suffer from two problems. First, they cannot represent in the same manner contours that superficially look different, like the contours in Figure 4 and Figure 5. At the same time, holistic approaches cannot account for the systematic differences between the two types of realizations: e.g. they cannot account in a principled manner for the fact that the syllable [nun] is Figure 5 is low in the dotted, early focus contour but rising in the solid, late focus contour. In short, configurational approaches cannot represent different instantiations of the same melody in a way that can either capture their similarity or predict their differences, thereby failing one of the main criteria for an adequate theory of intonational phonology mentioned in section 1.

2.2. Intonational gestalts and meaning

In addition to issues with intonational form, gestalt approaches encounter problems with intonational meaning. In gestalt models overall contour shape is said to associate with differences in meaning. Yet, as illustrated in section 1, the relationship between melody and meaning is not one-to-one: the same melody can lend different pragmatic nuances to different utterances, while the same meaning can be expressed by superficially different-looking contours. This lack of one-to-one correspondence has been repeatedly noted over the years (among many, Lehiste, 1970: 95 ff., and references therein; Bolinger, 1964; Ladd, 1978; Pierrehumbert, 1980; 't Hart, et al., 1990, chap. 4), and prompted Pike (1945: 23 ff.) to strongly caution against the practice of investigating contour meaning on the basis of grammatical structure (such as pitting statements against questions). Overall then melodies do not appear to have specific functions, and indeed attempts to describe the melodies of specific pragmatic nuances, such as irony, have proved unsuccessful (e.g. Bryant & Fox Tree, 2005).

In addition, cross-linguistic research has shown that functional effects of the sort favored by gestalt approaches are expressed in language specific ways. Such findings abound and strongly argue against a natural or direct relationship between intonational form and function. For instance, it has been argued that focus is universally expressed as local pitch expansion with a concomitant reduction in pitch range post-focally (e,g. Xu, 2005). Yet a review of the literature clearly shows that this is far from a universal mechanism for the prosodic marking of focus. For example, in Greek polar questions the word in focus has the lowest F0 in the entire utterance and pitch expansion is associated with the post-focal region (Arvaniti et al., 2006a; see Figure 5 for an illustration). Taiwanese relies on duration to mark focus rather than changes in pitch (Pan, 2007), while in many other languages, pitch expansion is but one, optional, manifestation of an overall prosodic reorganization under focus (e.g. Chen, 2006, to appear, for Mandarin; de Jong, 1995, and Harrington, Fletcher & Beckman, 2000, for English; Baltazani, 2006, and Arvaniti, Ladd & Mennen, 2006b, for Greek; Jun, 2005, for Korean; Venditti, Maekawa & Beckman, 2008, for Japanese). Perhaps the strongest counter-argument against the view that the relationship between focus and pitch range is natural and direct is the fact that not all languages can use intonation to mark focus (e.g. Swerts, Kramer & Avensani, 2002, on Italian; see Ladd 2008, chap. 6 for a discussion). If the relationship between focus and

intonation is natural and direct, there is no good explanation of why some languages do not avail themselves of this option.

The presence of an arbitrary relationship between intonation and meaning (as in all aspects of linguistic structure) is also evident in cross-linguistic data from questions, which, as mentioned, are often assumed to end in rises while statements end in falls (e.g. Ohala, 1983; Gussenhoven, 2004: chap. 4; Xu, 2005). But this idea is not supported cross-linguistically. Low or falling intonation is used with questions in typologically diverse languages, such as Bengali (Hayes & Lahiri, 1991), Chickasaw (Gordon, 2005), Bininj Gun-wok (Bishop & Fletcher, 2005), many Niger-Congo languages (Rialland, 2007), Greek (Arvaniti et al. 2006a, b; Arvaniti & Ladd, 2009) and Romani (Arvaniti & Adamou, in press). It is conversely the case that statements end in rises in many languages, including Bengali (Hayes & Lahiri, 1991), Chickasaw (Gordon, 2005) and many varieties of English (e.g. Grabe et al., 2000; Fletcher, Grabe & Warren, 2005).

Overall then, cross-linguistic research has confirmed the observation made by many scholars over the last century that the relationship between meaning and melody (as in global contour shape) is arbitrary and many-to-many. In turn, these results support the view that no useful generalizations about either intonational meaning or intonational form can be made on the basis of contour shape and its relationship to broad functional effects.

2.3. Intonation as a composite of rises and falls

If it is accepted that melodies are composed of independent elements, the question that arises is the nature of these elements. In this respect, most researchers have favored approaches in which the primitives are dynamic tones or movements, such as rises and falls, though the possibility of level tones (monotones) within such systems is also acknowledged (Bolinger, 1964; Crystal,1972; Ladd, 1978; Bolinger, 1986; 't Hart et al., 1990).

Perhaps the most thoroughly tested of these models is that of the IPO ('t Hart et al., 1990, and references therein). The IPO model was developed on the basis of Dutch, but it has also been used for the description of intonation in other languages, such as English, German and Russian (see 't Hart et al., 1990, chap. 4, and references therein). In this system, the main elements are rises and falls, a choice justified on perceptual grounds: the IPO researchers noticed that pitch changes were realized more slowly than possible by laryngeal control (as determined by the studies of Ohala & Ewen, 1973, and Sundberg, 1979), and concluded that the purpose for this slow execution must be to give listeners the perception of "pitch movement" rather than of a jump in pitch ('t Hart et al., 1990:71). Rises and falls are composed of four "perceptual features": pitch direction, timing relative to syllable boundaries, rate of pitch change and excursion size. In addition, rises and falls combine into larger configurations or *contours*; e.g. a rise-plateau-rise creates a "hat-pattern' while a rise-fall creates a "pointed-hat." This architecture makes the system configurational in two ways, since both the primitives and their combinations are configurations.

An important feature of the IPO approach is that meaning does not play a part in establishing either the primitives or the contours on the grounds that "intonation features have no intrinsic meaning" ('t Hart et al., 1990: 110). Instead, decisions as to the number and nature of melodies are based on experimental evidence derived from the *close-copy technique* in which listeners are asked whether stylized versions of various contours sound the same or different. By using experiments like these the IPO researchers determined the limits within which contours may vary phonetically; in turn, variants that are considered different by listeners are used to establish the elements of the intonational system.

The IPO stance towards the role of meaning in the analysis of intonation is the exact opposite of that taken by the British school (among many, Crystal, 1972; O'Connor & Arnold, 1973; Halliday, 1967, 1970), where differences in meaning are crucial for establishing the existence of both primitives and entire tunes. In this system, *tone units* (or *tunes* or *tone groups*) can span entire utterances but are also decomposed into smaller parts, the *pre-head*, *head*, *nucleus* and *tail* (this is the division proposed by Crystal, 1972, and O'Connor & Arnold, 1973; for a review of additional analyses, see Ladd, 1978: 16). The nucleus, defined as the pitch movement on the stressed syllable of the most important word of the utterance, is the only required element of a tune. The F0 of any unstressed syllables following it is the *tail*, while the F0 stretch covering all syllables from the first stressed syllable to the nucleus is the *head*; the F0 of any unstressed syllables preceding the head forms the *pre-head*. As can be surmised, not all melodies include all four primitives, while particular primitives can span arbitrary lengths of an utterance. For instance, the short utterance in Figure 1 consists only of a nucleus, while the two utterances in Figure 3, include a pre-head (*that's*), a high head (*twenty* / *really*) a low-falling nucleus (the stressed syllable of *dollar* / *awesome*) and a tail (the unstressed syllable of *dollar* / *awesome*). These elements do not combine entirely freely. As an example, O'Connor & Arnold (1973) distinguish seven nuclear tones, four types of heads and two types of pre-heads, but only twenty types of tone groups, instead of the 56 that all possible combinations of primitives would produce (tails follow the movement of the nucleus so they do not enter into the calculation).

The use of primitives of unconstrained length forces analyses of the British school into an artificial distinction between *simple* and *compound* rise-falls and fall-rises (e.g. Halliday, 1970; O'Connor & Arnold, 1973). The compound tunes fit uneasily into the system since they are said to contain two nuclear tones. Furthermore, most practitioners admit that it is hard to distinguish simple tunes from their compound counterparts on the basis of meaning, a serious drawback for a system in which the role of meaning is central (e.g. Crystal, 1972; O'Connor & Arnold, 1973: chap. 1). This problem is illustrated by the contours is Figure 1: the tune of *Me?!* is more plausibly analyzed as a simple rise-fall, but that of *A ballgown designer?!* can only be treated as a compound tune, though their meaning affinity is evident (note, however, that these are American English tunes with no exact analogue in British English).

Finally, like gestalt models, the British school analyses face problems with their treatment of intonational meaning, which is hard to pinpoint, yet must be determined if

the elements of the system are to be defined. As a result, the analyses of different authors disagree on the meaning and number of tunes, including the number and shape of nuclei; e.g. Halliday and Crystal recognize one type of fall, while O'Connor & Arnold distinguish between a high fall and a low fall. To the extent that meanings can be determined, they tend to be vague and occasionally contradictory and to a large extent dependent on grammatical aspects of the utterance. For example, O'Connor & Arnold (1973: 78-79) describe the "Jackknife" (simple rise-fall) as showing that the speaker is "*impressed*, perhaps *awed*", but they also add that the speaker can use it to sound "*complacent*, *self-satisfied*, even *smug*" or as "*shrugging aside any involvement*". These meanings apply if the Jackknife is used with statements, but change if it is used with wh-questions, polar questions or commands (for discussions of the problems of the British school approach to intonational meaning see Liberman, 1978: 88ff., and Pierrehumbert & Hirschberg, 1990).

2.4. Superpositional models

In several configurational models contours are said to be composed of two elements, a general trend and local perturbations which "ride" on this overall movement. This conception of intonation was also espoused by Bolinger , who distinguished *accentuation* from *intonation*, using accentuation to refer to pitch movements (*accents*) on stressed syllables, and intonation to refer to the general course of F0, "the rise and fall of pitch as it occurs along the speech chain" (Bolinger, 1986: 194).

The IPO system is one such superpositional model in that its primitives and contours are seen as localized movements superposed on a larger declination component which is taken to be largely automatic and due to the drop in subglottal pressure (declination reset and local movements, on the other hand, are seen as actively controlled by the speaker; 't Hart et al., 1990, chap. 5). The exact role of declination and its physiology are still a matter of debate (see, e.g., Pierrehumbert & Beckman, 1988, chap. 3; Gussenhoven, 2004, chap. 6), though evidence such as that provided for Japanese downtrends by Pierrehumbert & Beckman (1988) does not support the idea of declination playing as important a role as the IPO scholars envisioned.

Superpositional models have also been presented by Fujisaki (1983, 2004), Gårding (1983, 1987) and Thorsen (1980, 1985, 1986). Simplifying somewhat, in Fujisaki's system a *phrase command* results in the rising-falling course of F0 throughout an utterance (or a part thereof), with *accent commands* being responsible for more localized perturbations. Gårding (1983, 1987) posits grids (quasi parallel lines) within which most local F0 minima and maxima can be fitted; the overall range and direction of the grid (rising or falling or a combination thereof) reflect functional differences between utterances, such as the distinction between statements and questions. Similarly, Thorsen (1980: 1022) suggests that the rate of F0 drop in Danish is directly related to utterance function: "falling intonation contours are associated with declarative, intermediate contours with nonfinal, and flat contours with interrogative sentences." Due to the connection between communicative functions and overall F0 trends, the models of Thorsen and Gårding face similar issues to gestalt models with respect to meaning. On

the other hand, Fujisaki's model, which does not rely on meaning distinctions, must resort to counterintuitive solutions – such as negative accent commands, and phase commands that span linguistically arbitrary stretches – in order to adequately describe the course of F0 in languages other than Japanese (e.g. Fujisaki, Ohno & Yagi, 1997, on Greek; Fujisaki et al., 2005, on Mandarin; Gu, Hirose & Fujisaki, 2007, on Cantonese; for a discussion of these problems, see Ladd, 2008: 23 ff.; Arvaniti & Ladd, 2009).

3. Pitch levels as primitives

3.1. Early level-based models

Descriptions of intonation by means of level tones date from the American structuralists (Pike, 1945; Trager & Smith, 1951; Hockett, 1955; Trager, 1961). In these systems, intonation is analyzed by means of four levels, extra-high, high, mid and low. The level tones of these analyses are meant to be phonological abstractions equivalent to phonemes; as such, they are said to be defined relative to each other, rather than representing a specific pitch range each.

These early analyses were heavily criticized by configurationalists, most notably, Bolinger (1951) who questioned the claim that the four levels are relative, pointing out that if this assertion is taken at face value, it is not possible to distinguish combinations such as 123 from 234, although, theoretically, such combinations should be distinct. Thus, Bolinger concluded that level tones cannot be relative but must "rove each in its own bailiwick" (p. 200) and set out to test this hypothesis by recording utterances in various *ways* (as Bolinger termed them, i.e. melodies differing in various aspects) and having listeners judge them for similarity or appropriateness for a given purpose, such as appeasing a child. His results showed that contours analyzed as contrastive in the system of Trager & Smith (whom he particularly targeted) are perceived as similar by listeners, while others, analyzed as allophones of the same basic melody, are considered by listeners to be contrastive. Bolinger used his results to argue that a system with four levels can be at the same time too powerful and not powerful enough to capture contrastive and allophonic variations in the intonational system of English. His results led him to reject the level tone analysis as untenable and to propose instead that melodies are gestalts.

Bolinger's critique reflects the assumptions of concreteness and biuniqueness prevalent at the time. It is clear, for example, that Bolinger expected the different levels to faithfully represent the entire course of an utterance's F0, and to do so in such a way that the pitch range of each level did not overlap with that of others at any point in the utterance. Further, his comment that F0 forms "a continuous line that can be traced on a piece of paper" (Bolinger, 1951: 206) coupled with his distinction between monotones, which he accepts, and level tones which he does not (e.g. Bolinger, 1986: 29) suggest that Bolinger expected a level-based representation to be phonetically realized as a series of sustained pitch levels. It is obvious that if these assumptions are adopted, a level based analysis is unworkable on both phonological and phonetic grounds (on the latter, see Xu & Sun, 2002, and Dilley & Brown, 2007).

Although Bolinger's critique was well accepted, it is fair to note that many of the assumptions he attributes to the structuralists are not found in their works. Pike (1945), Hockett (1955) and Trager (1961) all note that levels represent "only those points in the contour crucial to the establishment of its characteristic rises and falls" (Pike, 1945: 26); with the exception of terminal junctures, these points associate with stressed syllables. Similarly, the structuralists noted that absolute pitch levels are not significant as such, and recognized the existence of both level tones and contours (e.g. Trager & Smith's terminal juncture phonemes). Further, Pike (1945) discusses at length the fact that level tones need not be realized as a series of sustained pitches but can be realized as glides especially when they are found close to each other, as happens, e.g., in short utterances. Nevertheless, as a result of configurationalist critiques, level-based analyses were largely abandoned in the following decades. Research within the generative framework focused primarily on the description of tone languages and no theoretical position was strongly taken either in favor of levels or of configurations with respect to intonation.

Intonational analyses by means of level tones adopting many of the principles of the early structuralist accounts appear again in the late 1970s, in Goldsmith (1976), Leben (1976) and Liberman (1978). Based on the idea of Leben (1973), who first conceptualize tones as distinct tonal segments rather than features of particular tone bearing units (typically syllables or moras), Goldsmith represented tones as *autosegments* residing on a separate tier, and analyzed English intonational melodies as sequences of H and L level tones. Liberman (1978), on the other hand, proposed an analysis of English intonation with the traditional four levels represented by means of two features, [± high] and [± low], though, informally, he also uses autosegmental representations closer to those of Leben (note that [+hi +lo] is possible in Liberman's system and represents a high-mid level).

In addition to the obvious differences in terms of formalism and the overall conception of phonology, the early autosegmental models depart from those of the early structuralists in that they assume that all syllables in an utterance are associated with some tone (something that may be accomplished, e.g., by tone copying or spreading). However, as discussed in more detail in section 3.2, the assumption that contours should be fully specified leads back to the problems that Bolinger (1951) first noted.

3.2 The autosegmental-metrical model of intonational phonology

A major breakthrough in our understanding of intonation come from Bruce's (1977) dissertation, a phonetic investigation of Swedish tonal structure. Bruce showed that the difference between the two lexical pitch accents of Swedish, accent I and accent II, does not lie in the shape of the accent, which is a fall in both cases, but in its *timing* with respect to the accented syllable: for accent I the fall is timed early, and for accent II it is timed late. A corollary of this difference is that a large part of the fall is truncated when a syllable with accent I is utterance initial, while it is fully present in words with accent II (giving rise to a peak preceding the fall). Further, Bruce provided an explanation for the second peak of accent II words, the seemingly erratic presence of which had been a long-standing puzzle: he showed that this peak is not part of the lexical accent at all but part of

the utterance's intonation, which he analyzed as a sequence of a *sentence accent* and a *terminal juncture*.

Several key points emerged from Bruce's work. First, it demonstrated the importance of *turning points*, F0 minimal and maxima which temporally align with particular elements of the segmental string. Bruce showed that these timing relations are regular in production and salient in perception. Crucially, they are also sufficient for modeling a contour without specifying the F0 in between (that is, by interpolating between salient points). Second Bruce showed that an F0 contour can be composed of elements of different origins within the grammar: in Swedish, some parts of the contour, the pitch accents, are lexically specified, while others, the sentence accents and terminal junctures, are phrasal elements, the result of intonation and phrasing. Similar distinctions were hinted at in other models – e.g. in Trager & Smith's terminal junctures, and in the *prominence lending* vs. *non-prominence lending* distinction between pitch movements espoused by the IPO – but they had not been clearly demonstrated before. Finally, Bruce showed that phrasal and lexical tonal elements simply concatenate, rather than forming two distinct systems superimposed on each other, and that this concatenation can result in lawful, context-dependent variation in the realization of tones.

The insights of the early autosegmentalists regarding the representation of tone together with Bruce's insights about the structure of tunes and their phonetic realization were applied to the analysis of English intonation in Pierrehumbert's dissertation (1980). Her model was further developed, particularly in Beckman & Pierrehumbert (1986) and Pierrehumbert & Beckman (1988), into the model currently known as the *autosegmental metrical model of intonational phonology* (a term coined by Ladd, 1996). The autosegmental metrical model (henceforth *AM*) has since been applied, with various modifications, to a series of unrelated languages with diverging prosodic systems (e.g. Jun, 2005; see also Arvaniti, to appear, and D'Imperio, to appear, for reviews).

In Pierrehumbert (1980) melodies are represented as strings of H and L tones on an autosegmental tier. Crucially, the purpose of this string of tones is not to trace or *transcribe* the course of F0, but, rather to represent the linguistically significant parts of the melody; that is, intonational representations are underspecified. The tones associate with the segmental string indirectly, via associations with the metrical tree (in Pierrehumbert & Beckman, 1988, this is formalized as association between tones and prosodic trees, specifically between tones on the one hand and feet and phrasal boundaries on the other). Thus, like Bolinger and Bruce, Pierrehumbert adopts the distinction between tones that associate with stressed syllables (i.e. the heads of feet) and tones that associate with the edges of phrasal constituents. The former, following Bolinger, are *pitch accents* notated with an asterisk (e.g. H*, a notation first used by Goldsmith, 1976); the latter, known as *boundary tones*, are notated with a percent (e.g. H%). Pierrehumbert also noted that tunes included a pitch movement between the boundary tone and the preceding *nuclear accent* (by definition, the last pitch accent of an utterance, often referred to in other literature as *sentence stress*). Pierrehumbert analyzed these pitch movements between the nuclear accent and following boundary tone as floating tones dubbed *phrase accents*. In Beckman & Pierrehumbert (1986), phrase

accents are instead analyzed as phrasal tones that associate with the right edge of *intermediate phrases* (*ip*s), a prosodic constituent larger than the prosodic word and smaller than the intonational phrase (IP), the nature of which is formalized in Pierrehumbert & Beckman (1988).

In Pierrehumbert (1980), the H and L tones are said to be realized as tonal targets, typically peaks and dips respectively, although the relationship between phonological tones and phonetic realization is not always transparent. First, phonological tones may not be realized as F0 minima or maxima; e.g. in Pierrehumbert's original analysis the H*+L- accent is not realized as a fall from a high to a low F0 level; rather the L tone is said to trigger downstep on a following H tone. Realizations are also context-dependent; e.g. after a L- phrase accent, L% is realized as a drop to the bottom of the speaker's range, but after a downstepped H- phrase accent it is realized as sustained mid-level pitch. The scaling of targets is computed on the fly, and is determined by metrical strength and context (see also Liberman & Pierrehumbert, 1984; Pierrehumbert & Beckman, 1988; Prieto, Shih & Nibert, 1996). Further, following Goldsmith (1976), Pierrehumbert assumed that tones co-occur (*align*) with the segmental material they are (indirectly) associated with: pitch accents align with associated stressed syllables, and boundary tones with phrase final syllables; phrase accents, being floating tones, are realized in a less precise manner (for evidence that this view cannot fully account for tonal alignment cross-linguistically, see, inter alia, Arvaniti, Ladd & Mennen, 2000; Grice, Ladd & Arvaniti, 2000; Gussenhoven, 2000).

Since Bruce (1977) and Pierrehumbert (1980) several phonetic studies have demonstrated the crucial role of local F0 minima and maxima in the production of intonation. Such tonal targets have been shown to be consistently aligned with the segmental string and to show lawful, variation based on a variety of factors, such as speaking rate (Fougeron and Jun, 1998; Prieto & Torreira, 2007), phonological weight (Ladd, Mennen & Schepman, 2000), tonal crowding (Silverman & Pierrehumbert, 1990; Prieto, 2005; Arvaniti & Ladd, 2009) and dialectal differences (Arvaniti & Garding, 2007; Ladd et al., 2009). Results attesting to the regularlity of turning points have been reported for languages with very different prosodic systems, including not only languages without lexical tone but also tone languages – such as Mandarin (Xu, 1999), Kinyarwanda (Myers, 2003) and Thai (Moren & Zsiga, 2006) – pitch accent languages – such as Roermond Dutch (Gussenhoven, 2000), Basque (Hualde et al., 2002; Elordieta & Hualde, 2003) and Serbian (Smiljanić, 2006) – and languages with hybrid systems, such as Chickasaw (Gordon, 2008; for a comprehensive review, see Arvaniti, to appear, and D'Imperio, to appear).[1]

[1] Despite the overwhelming evidence in favor of level tones being realized as highly localized minima and maxima, it is clear that other realizations are also possible. First, in some cases turning points may be just indirect reflexes of tones that are convenient to identify and measure in a signal that presents as a continuous curve; e.g. Arvaniti & Ladd (2009) show that the turning points that define the low-level stretch in the contours of Figure 2 are not necessarily targets themselves, but quite possibly the easy-to-measure outcome of the tune's requirement for a low-level stretch, the reflex of the L- phrase accent. Results like these suggest that tones should not be posited as phonological entities simply because a turning point in an F0 contour has been noted. Second, tones may not be realized as instantaneous events (among many, Pierrehumbert, 1980; Silverman, 1987; D'Imperio, 2000; Arvaniti et al., 2006a; Knight & Nolan, 2006;

The search for highly localized tonal targets is in part a corollary of one of the most important tenets of the AM model, namely that contours are underspecified not only phonologically, but phonetically as well: while H and L tones are realized as targets with specific scaling and alignment, the F0 between them is determined by interpolation. Undespecification was empirically documented by Pierrehumbert & Beckman (1988, chap. 2), who showed that the fall between the initial phrasal H of unaccented accentual phrases in Japanese and the final L% of the phrase is a fall that varies in steepness depending on the number of moras between the two tones. This type of realization is incompatible with fully specified representations, whether such specifications are present from the beginning (as, e.g., in Xu's PENTA) or achieved at a later stage through tone spreading.

The role of underspecification cannot be underestimated: underspecification leads to a clear understanding that the number of tones need not match the number of tone bearing units (TBUs), so that both strings of tonally unspecified TBUs and instances of several tones associating with the same TBU are possible. Thus, in the AM framework, the two English contours in Figure 1, are both analyzed as L*+H L- H%. By using the same representation for these two contours, AM captures the fact that they are instantiations of the same melody, thereby generalizing beyond surface form. AM can also account for the systematic differences between contours like those in Figure 1 which, as mentioned earlier, were particularly problematic for the British school. In *Me?!*, all three tonal events must co-occur with the only syllable of this utterance; hence the obvious lengthening of *me* (720 ms) and the swift movement from one tonal target to the next. In *A ballgown designer?!*, L*+H is associated with the metrically strongest syllable in the utterance, i.e. *ball*, and it co-occurs with it (showing the peak delay expected for this accent; e.g. Pierrehumbert & Steele, 1989; Arvaniti & Garding, 2007). The H% is realized on the last syllable which is the one showing a rise. The L-, which is associated with the ip boundary, spreads between the L*+H and H% accounting for the fall and low-level stretch of F0 (for details on the realization of the L- see Grice et al., 2000; Barnes et al., 2006). A similar analysis applies to the Greek wh-questions shown in Figure 2, analyzed as L*+H L- !H% (where !H refers to a downstepped H tone; Grice et al., 2000; Arvaniti & Baltazani, 2005; Arvaniti & Ladd, 2009).

Overall, the AM model avoids several pitfalls of previous analyses. First, by formally separating stress from intonation and providing a mechanism for their interaction, the AM model incorporates the insights of Bolinger about pitch accents without requiring distinct accentual and phrasal components to account for pitch contours. In addition, the use of only H and L tones avoids the problems noted by Bolinger (1951) with respect to level tones. At the same time, by treating the issue of pitch range as a matter of phonetic

---

Arvaniti & Garding, 2007). Thus, although it is convenient to measure tonal targets as F0 minima and maxima, tones may also have duration, a type of realization that may be perceptually enhancing (e.g. D'Imperio, Terken & Pitermann, 2000; D'Imperio, Gili Fivela & Niehbuhr, 2010). Finally, recent results by Barnes et al. (2010) have also shown that F0 transitions from one tone to another can be perceptually relevant and aid in the identification of particular pitch accents (at least in experimental settings in which transitions are the only information listeners have for tonal identification).

realization, AM avoids the problems that plagued the British analyses which were due to the confounding of linguistic and paralinguistic aspects of pitch range (cf. the disagreements regarding whether high falls and low falls are distinct entities). Further, by making explicit the separation between the phonetics and phonology of intonation, the AM model provides a principled account of the context-dependent variation of tones, a point that was not explicitly addressed in previous models that mostly confounded contours and representations. Finally, the use of underspecification provides a parsimonious and elegant way of capturing both the similarities of melodies and the differences in phonetic realization that arise from the properties of the metrical structure with which a melody associates. In this way, the model can account for both local phonetic detail and abstract phonological form, something that configurational and full specification models cannot do (for extensive discussions of this point, see Pierrehumbert & Beckman, 1988; Arvaniti et al. 2006a; Arvaniti & Ladd, 2009). Finally, since the degree of underspecification can vary, AM can account both for languages with dense tonal specifications, such as Mandarin, as well as for languages with more sparse specifications, such as English. In short then, AM not only provides an answer regarding the nature of intonational primitives but crucially addresses the even more fundamental question of what should be represented phonologically when it comes to intonation, an issue that most other theories have not tackled by focusing exclusively on faithful representations of entire F0 curves.

3.3. AM and meaning

AM has not grappled systematically with meaning though several models of information structure have relied on AM analyses to understand the role of intonation particularly with respect to focus marking (Steedman, 2000; Büring, 2007, inter alia). Although research on intonational meaning is not extensive, it so far suggests that the principles of AM are more likely to lead to an understanding of intonational meaning than configurational approaches, since AM analyses are compositional and thus in principle more flexible than configurational approaches in dealing with the complex relationship between intonational meaning and form both within and across languages.

Perhaps the best known treatment of meaning within AM is Pierrehumbert & Hirschberg (1990) who developed a theory of intonational meaning specifically for English (the principles of which are, however, applicable to intonational systems at large). According to this model, each pitch accent, phrase accent and boundary tone is seen as a morpheme that has its own pragmatic meaning; the meaning conveyed by an entire melody is therefore compositional and depends on contributions from all tones. Further, Pierrehumbert & Hirschberg suggest that each tone's meaning is to be interpreted with respect to the phonological domain with which it is associated: specific lexical items in the case of pitch accents, ips in the case of phrase accents, and IPs in the case of boundary tones. The advantage of a system like this is that it can account for similarities in meaning conveyed by identical components of different melodies, such as the use of the same pitch accent with different phrasal tones, or the use of a given boundary tone with different pitch and phrase accents. In addition, since the system does not to rely on a natural or biologically determined relationship between tones and meaning, it is more

easily amenable to an understanding of cross-linguistic and cross-dialectal variation in the relationship between meaning and form.

Crucially, Pierrehumbert & Hirschberg take pains to explain that the meaning of a tune is not to be directly interpreted; e.g. H* L- L%, often used with declaratives, is not to be interpreted as *S* (the speaker) *believes x*; rather, the speaker's belief in *x* will be inferred from all the tones that make up a tune and the context in which they are used. This conception of intonational meaning could account for the different interpretation of the same tune in *That's twenty dollars* and *That's really awesome* shown in Figure 3. Simplifying somewhat, the downstepped !H* accent on the second word (*dollar*, *awesome*) does not impart salience on the accented item (note that in the analysis of Pierrehumbert, 1980, and that of Pierrehumbert & Hirschberg, 1990, this accent is analyzed as H*+L); rather, it implies that this item should be inferable by the hearer. This is expected for *dollar* in a context in which dollars are the only currency in which purchases can be made; but when used with *awesome* the inferred predictability implies that the speaker is either ritualistically using an expression associated with excitement (and therefore is not excited) or deliberately avoids the more plausible H* pitch accent in order to convey sarcasm.

Despite the existence of this framework, it is fair to say that many aspects of intonational meaning remain unclear. For example, Pierrehumbert & Hirschberg themselves note that similarities between the meanings of bitonal accents (L*+H and L+H*, and H*+L and H+L*) are evident in English but not easy to account for in their model. Further, models similar to that developed by Pierrehumbert & Hirschberg (1990) have not been as developed for other languages, so the framework has not been extensively tested for cross-linguistic validity. Research has also shown that certain tonal combinations are more frequent than others both in English (e.g. Dainora, 2001, 2006) and in other languages (e.g. Arvaniti & Baltazani, 2005, for Greek). Observations of this sort abound and have given rise to hypotheses that perhaps intonational meaning depends on strings of intonational primitives rather than being strictly compositional; in this case, the importance of the melody from the nuclear accent onwards (akin to the traditional British *nucleus*) is emphasized (for a review and discussion, see Gussenhoven, 2004: chap. 7). More recently, Calhoun (2010) has provided a metrical analysis for the importance of the nuclear pitch accent, suggesting that the presence of prenuclear accents is metrically motivated and that they do not contribute to information structure, contra Pierrehumbert & Hirschberg's position that all accents contribute to meaning (for arguments similar to Calhoun's, see also Büring, 2007).

As even this brief discussion indicates, much remains to be done before the relationship between melodies and their contribution to information structure is fully understood. It is clear, however, that crude distinctions such as that between statement vs. question or focus vs. lack thereof are not sufficient to explain intonational meaning and that a more sophisticated understanding, quite possibly following the main principles of the treatment provided in Pierrehumbert & Hirschberg (1990) is likely to be more successful.

4. The phonetic realization of tones: level or dynamic tones?

With the noted exceptions of Bolinger and the IPO, the choice between level and dynamic tones was not always explicitly motivated in the various models discussed here. It is of course easy to consider this a mundane issue: after all, as Bolinger (1986: 225) put it "it makes no difference, in describing a movement, whether one says 'first you are going to be up and then you are going to be down' or 'you are going to go down'." Choosing the right answer, however, is neither trivial nor a matter of taste, as the answer has empirically testable consequences.

This issue was addressed in Pierrehumbert & Beckman (1988: chapters 2, 3 and 4) who discuss extensively how the context-related scaling differences of tones they observed in Japanese cannot be elegantly accounted for in models using dynamic tones as primitives. Arvaniti, Ladd & Mennen (1998) specifically compared the predictions of IPO to those of AM, by examining the realization of the rising pitch accents found in prenuclear position in Greek declaratives. Arvaniti et al. found that the timing, duration and speed (rate of change) of these rises were not invariable, as advocated by the IPO, but depended on syllable duration. On the other hand, the timing and scaling of the onset and offset of the rise were held constant: the initial dip – interpreted as the reflex of a L tone – coincided with the onset of the accented syllable, while the peak – interpreted as the reflex of a H tone – was reached about 10 ms after the onset of the first post-accentual vowel.

At the very minimum, the results of Arvaniti et al. (1998) suggest that the IPO notion of dynamic tones does not apply to all languages. More generally, they pose the question of how dynamic primitives such as rises and falls can be determined if none of the properties that may define them is stable. Finally, it is not clear how the notion of an indivisible unit can be defended for the Greek accents at all, since the beginning and ending points of the rise do not behave as one. Their relative autonomy is demonstrated by the fact that they align independently of each other and are not similarly affected by tonal crowding, which typically results in the undershooting of the L, while the realization of the H remains largely unaltered (Arvaniti et al., 2000). This pattern is difficult to account for if the rise is a unit, in which case one would more plausibly expect a curtailment of the entire pitch movement.

It is thus clear that dynamic tones cannot account for some of the attested patterns. On the other hand, level tones can be used not only for the representation of loosely defined rises and falls, as in Greek, but also for rises and falls that are more closely knit. Such units have been reported by Frota (2002) for European Portuguese. Specifically, Frota found that in the falling accent indicating broad focus, the H and L are timed with respect to distinct segments (similarly to the Greek case), but the fall of the accent indicating narrow focus shows a constant timing relationship between the H and L tones (similar to that discussed by Pierrehumbert, 1980, for L*+H and L+H* in English). This difference between the two accents of European Portuguese can be represented by means of a hierarchical representation of tones shown in (1), as first proposed by Grice (1995a, b) and adopted by Frota, or it can be treated as an issue of phonetic realization, as argued, in Arvaniti et al. 2006b). Either way, it is clear that while level tones can adequately

describe all attested tonal patterns, level tones cannot. In short then both the empirical evidence and phonological considerations of parsimony and descriptive adequacy make a theory based on level tones preferable.

(1)

5. Conclusion

The original debate about levels vs. configurations (Bolinger 1951) focused on the issue of whether melodies are gestalts or should be seen as composites of primitives. Despite this distinction, Bolinger's views, espoused by many before him and since, are based on the idea that intonational contours should be represented in their entirety, either as a series of primitives, or as a "line […] on a piece of paper". Current understanding suggests that couching the problem in these terms is misleading, as neither type of representation is likely to be correct: as shown, gestalt approaches cannot account in a satisfactory manner for either intonational meaning or intonational form; yet representations that fully specify the course of F0, either in terms of dynamic tones or in terms of levels, do not fare much better. Due to the particularities of intonation, especially the fact that its realization depends on the metrical structure of the utterance with which it co-occurs, significant generalizations about melodies and their phonetic variation are best captured if it is recognized that only certain parts of F0 contours are linguistically relevant and should be represented phonologically and phonetically. Finally, empirical evidence as well as considerations of representational parsimony strongly suggest that these linguistically relevant aspects of F0 contours are best represented as levels rather than movements.

References

Armstrong, Lilias E. & Ida Ward. 1926. *Handbook of English Intonation*. Cambridge: Heffer.

Arvaniti, Amalia. 2007. On the relationship between phonology and phonetics (or why phonetics is not phonology). Special Session: Between Meaning and Speech: On the Role of Communicative Functions, Representations and Articulations, *Proceedings of ICPhS XVI*, pp. 19-24.

Arvaniti, Amalia. To appear. Prosodic representations (Part I: Segment-to-Tone Association). In A. Cohn, C. Fougeron & M. Huffman (to appear), *The Oxford Handbook of Laboratory Phonology*. Oxford: Oxford University Press.

Arvaniti Amalia & Evangelia Adamou. In press. Focus expression in Romani. *WCCFL 28* (2010).

Arvaniti, Amalia & Mary Baltazani. 2005. Intonational analysis and prosodic annotation of Greek spoken corpora. In Jun (2005a), pp. 84-117.

Arvaniti, Amalia & Gina Garding. 2007. Dialectal variation in the rising accents of American English. In J. Cole & J. H. Hualde (Eds.), *Papers in Laboratory Phonology* 9, pp. 547-576. Berlin, New York: Mouton de Gruyter.

Arvaniti, Amalia & D. Robert Ladd. 2009. Greek wh-questions and the phonology of intonation. *Phonology* 26: 43-74.

Arvaniti, Amalia, D. Robert Ladd & Ineke Mennen. 1998. Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics* 26: 3-25.

Arvaniti, Amalia, D. Robert Ladd & Ineke Mennen. 2000. What is a starred tone? Evidence from Greek. In M. B. Broe & J. B. Pierrehumbert (2000), pp. 119-131.

Arvaniti, Amalia, D. Robert Ladd & Ineke Mennen. 2006a. Phonetic effects of focus and "tonal crowding" in intonation: Evidence from Greek polar questions. *Speech Communication* 48: 667-696.

Arvaniti, Amalia, D. Robert Ladd & Ineke Mennen. 2006b. Tonal association and tonal alignment: evidence from Greek polar questions and contrastive statements. *Language and Speech* 49: 421-450.

Baltazani, Mary. 2006. Focusing, prosodic phrasing, and hiatus resolution in Greek. In Goldstein, Whalen & Best (2006), pp. 473-493.

Barnes, Jonathan, Stefanie Shattuck-Hufnagel, Alejna Brugos &Nanette Veilleux. 2006. The domain of realization of the L- phrase tone in American English. *Proceedings of Speech Prosody 2006*, Dresden, May 2-5, 2006.

Barnes, Jonathan, Nanette Veilleux, Alejna Brugos & Stefanie Shattuck-Hufnagel. 2010. The effect of global F0 contour shape on the perception of tonal timing contrasts in American English intonation. *Proceedings of Speech Prosody 2010*, Chicago, May 11-14, 2010.

Beckman, Mary E. & Janet Pierrehumbert. 1986. Intonational structure in English and Japanese. *Phonology Yearbook* 3: 255-310.

Bishop, Judith & Janet Fletcher. 2005. Intonation in six dialects of Bininj Gun-wok. In Jun (2005), pp. 331-361.

Bolinger, Dwight. 1951. Intonation: Levels versus configurations. *Word* 7: 199-210.

Bolinger, Dwight. 1964. Around the edge of language: Intonation. *Harvard Educational Review* 34(2): 282-293. Reprinted in Bolinger (1972), pp. 19-29.

Bolinger, Dwight (Ed.). 1972. *Intonation: Selected Readings*. Penguin Books.

Bolinger, Dwight. 1986. *Intonation and its Parts: Melody in Spoken English*. Stanford, CA: Stanford University Press.

Bruce, Gösta. 1977. *Swedish Word Accents in Sentence Perspective*. Lund: Gleerup.

Botinis, Antonis. 1989. *Stress and Prosodic Structure in Greek: A Phonological, Acoustic, Physiological and Perceptual Study*. Lund: Lund University Press.

Broe, Michael B. & Janet B. Pierrehumbert (Eds.). 2000. *Papers in Laboratory Phonology V: Acquisition and the Lexicon*. Cambridge: Cambridge University Press.

Bryant, Gregory A. & Jean E. Fox Tree. 2005. Is there an ironic tone of voice? *Language and Speech* 48: 257-277.

Büring, Daniel. 2007. Intonation, semantics and information structure. In G. Ramchand & C. Reiss (Eds.), *The Oxford Handbook of Linguistic Interfaces,* pp. 445-473. Oxford: Oxford University Press.

Calhoun, Sasha. 2010. The centrality of metrical structure in signaling information structure: A probabilistic perspective. *Language* 86: 1-42.

Chen, Yiya. 2006. Durational adjustment under corrective focus in Standard Chinese. *Journal of Phonetics* 34: 176-201.

Chen, Yiya. To appear. Post-focus F0 compression - now you see it, now you don't. *Journal of Phonetics.*

Cohen, Antonie & Johan 't Hart. 1968. On the anatomy of intonation. *Lingua* 19: 177-192.

Cohn, Abigail, Cécile Fougeron & Marie Huffman (Eds.). To appear. *The Oxford Handbook of Laboratory Phonology*. Oxford: Oxford University Press.

Cooper, William & John Sorensen. 1981. *Fundamental frequency in sentence production.* Heidelberg. Springer.

Crystal, David. 1972. The intonation system of English. In D. Bolinger (ed.), *Intonation*, pp. 110-136. Penguin.

Dainora, Audra. 2001. *An empirically based probabilistic model of intonation in American English*. Chicago: University of Chicago Ph.D. dissertation.

Dainora, Audra. 2006. Modeling intonation in English: A probabilistic approach to phonological competence. In Goldstein, Whalen & Best (2006), pp. 107-132.

de Jong, Kenneth J. 1995. The supraglottal articulation of prominence in English: Linguistic stress as localized hyperaerticulation. *Journal of the Acoustical Society of America* 97(1): 491-504.

Dilley, Laura C. & Meredith Brown. 2007. Effects of pitch range variation on f_0 extrema in an imitation task. *Journal of Phonetics* 35 (4): 523-551.

D'Imperio, Mariapaola. 2000. *The role of perception in defining tonal targets and their alignment*. Unpublished Ph.D. dissertation, The Ohio State University.

D'Imperio, Mariapaola. To appear. Prosodic representations (Part II: Tonal alignment). In A. Cohn, C. Fougeron & M. Huffman (Eds.).

D'Imperio, Mariapaola & David House. 1997. Perception of questions and statements in Neapolitan Italian. *Proceedings of Eurospeech* (Rhodes), pp. 251-254.

D'Imperio, Mariapaola, Jacques Terken and Michel Pitermann. 2000. Perceived tone "targets" and pitch accent identification in Italian. *Proceedings of the 8th International Conference on Speech Science & Technology* (*SST'2000*), pp. 206-211.

D'Imperio, Mariapaola, Barbara Gili Fivela & Oliver Niebuhr. 2010. Alignment perception of high intonational plateaux in Italian and German. *Proceedings of Speech Prosody 2010*.

Elordieta, Gorka & José Ignacio Hualde. 2003. Tonal and durational correlates of accent in contexts of downstep in Lekeitio Basque. *Journal of the International Phonetic Association* 33(2): 195-209.

Fletcher, Janet, Esther Grabe & Paul Warren. 2005. Intonational variation in four dialects of English: the high rising tune. In Jun (2005), pp. 390-409. Oxford: Oxford University Press.

Fougeron, Cécile & Sun-Ah Jun. 1998. Rate effects on French intonation: prosodic organization. *Journal of Phonetics* 26: 45-69.
Frota, Sónia. 2002. Tonal association and target alignment in European Portuguese nuclear falls. In Gussenhoven & Warner (2002), pp. 387-418.
Fujisaki, Hiroya. 1983. Dynamic characteristics of voice fundamental frequency in speech and singing. In P. F. MacNeilage (ed.) *The production of speech*, pp. 39-55. Heidelberg: Springer-Verlag.
Fujisaki, Hiroya. 2004. Information, prosody and modeling – with emphasis on tonal features of speech. *Proceedings of Speech Prosody 2004*, Nara, Japan, March 23–26, 2004. Available at http://www.isca-speech.org/archive
Fujisaki, Hiroya, Sumio Ohno & Takashi Yagi. 1997. Analysis and modelling of fundamental frequency contours of Greek utterances. *Proceedings of Eurospeech '97*, pp. 465–468.
Fujisaki, Hiroya, Changfu Wang, Sumio Ohno & Wentao Gu. 2005. Analysis and synthesis of fundamental frequency contours of Standard Chinese using the command–response model. *Speech Communication* 47: 59–70.
Gårding, Eva. 1983. A generative model of intonation. In A. Cutler & D. R. Ladd (Eds.), *Prosody: Models and Measurements*, pp. 11-25. Heidelberg: Springer.
Gårding, Eva. 1987. Speech act and tonal pattern in standard Chinese – constancy and variation. *Phonetica* 44: 13-29.
Goldsmith, John A. 1976. *Autosegmental Phonology*. Cambridge, MA: MIT Ph.D. Dissertation. Reprinted 1979, Outstanding Dissertations in Linguistics. New York & London: Garland Publishing.
Goldstein, Louis, Douglas H. Whalen & Catherine T. Best (Eds.). 2006. *Laboratory Phonology 8*. Berlin, New York: Mouton de Gruyter.
Gooden, Shelome, Kath-Ann Drayton & Mary Beckman. 2009. Tone inventories and tune-text alignments: Prosodic variation in 'hybrid' prosodic systems. *Studies in Language* 33(2): 354-394.
Gordon, Matthew. 2005. Intonational phonology of Chickasaw. In Jun (2005), 301-330.
Gordon, Matthew. 2008. Pitch accent timing and scaling in Chickasaw. *Journal of Phonetics* 36(3): 521-535.
Grabe, Esther, Brechtje Post, Francis Nolan & Kimberley Farrar. 2000. Pitch accent realization in four varieties of British English. *Journal of Phonetics* 28: 161-185.
Grabe, Esther, Greg Kochanski & John Coleman. 2003. Quantitative modelling of intonational variation. *Proceedings of Speech Analysis and Recognition in Technology, Linguistics and Medicine 2003*. [Available at <http://www.phon.ox.ac.uk/oxigen/publications.php>].
Grice, Martine. 1995a. *The Intonation of Interrogation in Palermo Italian: Implications for Intonational Theory*. Tübingen: Niemeyer.
Grice, Martine. 1995b. Leading tones and downstep in English. *Phonology* 12(2):183-233.
Grice, Martine, D. Robert Ladd & Amalia Arvaniti. 2000. On the place of "phrase accents" in intonational phonology. *Phonology* 17:143-185**.**
Gu, Wentao, Keikichi Hirose & Hiroya Fujisaki. 2007. Analysis of tones in Cantonese Speech based on the command-response model. *Phonetica* 64: 29–62.
Gussenhoven, Carlos 1999. Discreteness and gradience in intonational contrasts. *Language and Speech* 42: 283-305.

Gussenhoven, Carlos. 2000. The boundary tones are coming: On the non-peripheral realization of boundary tones. In M. Broe & J. Pierrehumbert (2000), pp. 132-151.
Gussenhoven, Carlos. 2004. *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.
Gussenhoven, Carlos & Natasha Warner (Eds.) 2002. *Laboratory Phonology 7*. Berlin, New York: Mouton de Gruyter.
Halliday, M.A.K. 1967. *Intonation and Grammar in British English*. The Hague & Paris: Mouton.
Halliday, M.A.K. 1970. *A Course in Spoken English: Intonation*. Oxford: Oxford University Press.
Harrington, Jonathan, Janet Fletcher & Mary E. Beckman. 2000. Manner and place conflicts in the articulation of accent in Australian English. In Broe & Pierrehumbert (2000), pp. 40-51.
't Hart, Johan & Antonie Cohen. 1973. Intonation by rule: A perceptual quest. *Journal of Phonetics* 1: 309-327.
't Hart, Johan & René Collier. 1975. Integrating different levels of intonation analysis. *Journal of Phonetics* 3: 235-255.
't Hart, Johan, René Collier & Antonie Cohen. 1990. *A Perceptual Study of Intonation: An Experimental-Phonetic Approach to Speech Melody*. Cambridge: Cambridge University Press.
Hayes, Bruce & Aditi Lahiri. 1991. Bengali intonational phonology. *Natural Language and Linguistic Theory* 9: 47-96.
Hirschberg, Julia & Gregory Ward. 1992. The influence of pitch range, duration, amplitude and spectral features on the interpretation of the Rise-Fall-Rise intonation contour in English. *Journal of Phonetics*, 20 (2): 241-251
Hirst, Daniel & Albert Di Cristo. 1998. A survey of intonation systems. In D. Hirst & A. Di Cristo (Eds.), *Intonation Systems a Survey of Twenty Languages*, pp.1-44. Cambridge: Cambridge University Press.
Hirst, Daniel, Albert Di Cristo & Robert Espesser. 2000. Levels of representation and leves of analysis for the description of intonation systems. In M. Horne (Ed.), *Prosody: Theory and Experiment, Studies presented to Gösta Bruce*, pp. 51-87. Dordrecht: Kluwer Academic Publishers.
Hockett, Charles F. 1955. *A Manual of Phonology*. Bloomington: Indiana University Publications in Anthropology and Linguistics, vol. 11.
Hualde, José I., Gorka Elordieta, Inaki Gaminde & Rajka Smiljanić. 2002. From pitch accent to stress accent in Basque. In C. Gussenhoven & N. Warner (2002), pp. 547-584.
Jones, Daniel. 1972. *An Outline of English Phonetics* (9th edn.). Cambridge: Cambridge University Press.
Jun, S. 2005. Korean intonational phonology and prosodic transcription. In Jun (2005), pp. 201-229.
Jun, Sun-Ah (Ed). 2005. *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.
Kingdon, Roger. 1958. *The groundwork of English Intonation*. London: Longmans, Green & Co.

Knight, Rachael-Anne & Francis Nolan. 2006. The effect of pitch span on intonational plateau. *Journal of the International Phonetic Association* 36(1): 21-38.
Ladd, D. Robert 1978. *The Structure of Intonational Meaning: Evidence from English*. Bloomington: Indiana University Press.
Ladd, D. Robert. 1996. *Intonational Phonology* (1st edn.). Cambridge: Cambridge University Press.
Ladd, D. Robert. 2008. *Intonational Phonology* (2nd edn.). Cambridge: Cambridge University Press.
Ladd, D. Robert, Ineke Mennen & Astrid Schepman. 2000. Phonological conditioning of peak alignment in rising pitch accents in Dutch. *Journal of the Acoustical Society of America* 107: 2685-2696.
Ladd, D. Robert, Astrid Schepman, Laurence White, Louise May Quarmby, Rebekah Stackhouse. 2009. Structural and dialectal effects on pitch peak alignment in two varieties of British English. *Journal of Phonetics* 37(2): 145-161.
Leben, William R. 1973. *Suprasegmental Phonology*. Cambridge, MA: MIT dissertation.
Leben, William R. 1976. The tones of English intonation. *Linguistic Analysis* 2: 69-107.
Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge, MA: The MIT. Press.
Liberman, Mark Y. 1978. *The Intonational System of English*. Bell Laboratories (reproduced by the Indiana University Linguistics Club).
Liberman, Mark Y. & J. B. Pierrehumbert. 1984. Intonational invariance under changes in pitch range and length. In M. Aronoff & R. T. Oehrle (Eds.), *Language Sound Structure: Studies in Phonology Presented to Morris Halle*, pp. 157-233. Cambridge, MA: The MIT Press.
Moren, Bruce & Elizabeth Zsiga. 2006. The lexical and post-lexical phonology of Thai tones. *Natural Language & Linguistic Theory* 24: 113-178.
Myers, Scott. 2003. F0 Timing in Kinyarwanda. *Phonetica* 60: 71-97.
O'Connor, J. D. & G. F. Arnold. 1973. *Intonation of Colloquial English*. London: Longman.
Ohala, John. 1983. Cross-language view of pitch: an ethological view. *Phonetica* 40: 1-18.
Ohala, John. & William G. Ewan. 1973. Speed of pitch change. *Journal of the Acoustical Society of America* 53: 345.
Pan, Ho-Hsien. 2007. Focus and Taiwanese unchecked tones. In C. Lee, M. Gordon, D. Büring (Eds.), *Topic and Focus: Cross-linguistic Perspectives on Meaning and Intonation*, pp. 195-213. Springer.
Pierrehumbert, Janet B. 1980. *The Phonology and Phonetics of English Intonation*. Cambridge, MA: MIT dissertation. Published 1988, UILC, Bloomington, IN.
Pierrehumbert, Janet & Mary Beckman. 1988. *Japanese Tone Structure*. Cambridge, MA: The MIT Press.
Pierrehumbert, Janel & Julia Hirschberg. 1990. The meaning of intonation in the interpretation of discourse. In P. Cohen, J. Morgan & M. Pollack (eds), *Intentions in Communication*, pp. 271-311. Cambridge MA: The MIT Press.
Pierrehumbert, Janet & Shirley A. Steele. 1989. Categories of tonal alignment in English. *Phonetica* 46: 181-196.
Pike, Kenneth L. 1945. *The Intonation of American English*. Ann Arbor: The University of Michigan Press.

Prieto, Pilar. 2005. Stability effects in tonal clash contexts in Catalan. *Journal of Phonetics* 33(2): 215-242.
Prieto, Pilar, Chilin Shih, & Holy Nibert. 1996. Pitch downtrend in Spanish. *Journal of Phonetics* 24: 445-473.
Prieto, Pilar & Francisco Torreira. 2007. The segmental anchoring hypothesis revisited: Syllable structure and speech rate effects on peak timing in Spanish. *Journal of Phonetics* 35(4): 473-500.
Redi, Laura. 2003. Categorical effects in the production of pitch contours in English. *Proceedings of the 15th International Congress of Phonetics Sciences*, pp. 2921-2924. Barcelona.
Rialland, Annie. 2007. Question prosody: An African perspective. In C. Gussenhoven & T. Riad (Eds.), *Tones and Tunes, Volume 1: Typological Studies in Word and Sentence Prosody*, pp. 35-62. Berlin, New York: Mouton de Gruyter.
Rietveld, Toni & Carlos Gussenhoven. 1995. Aligning pitch targets in speech synthesis: effects of syllable structure. *Journal of Phonetics* 23: 375-385.
Silverman, Kim E. A. 1987. *The structure and processing of fundamental frequency contours*. Ph.D. Dissertation, Cambridge University.
Silverman, Kim & Janet Pierrehumbert. 1990. The timing of prenuclear high accents in English. In J. Kingston, &M. Beckman (Eds.), *Papers in Laboratory Pphonology I: Between the Grammar and Physics of Speech*, pp. 72–106. Cambridge: Cambridge University Press.
Smiljanić, Rajka. 2006. Early vs. late focus: Pitch-peak alignment in two dialects of Serbian and Croatian. In Goldstein, Whalen & Best (2006), pp. 494-518.
Steedman, Mark. 2000. Information structure and the syntax-phonology interface. *Linguistic Inquiry* 31: 649-689.
Sundberg, Johan. 1979. Maximum speed of pitch changes in singers and untrained subjects. *Journal of Phonetics* 7: 71-79.
Swerts, Marc, Emiel Krahmer & Cinzia Avesani. 2002. Prosodic marking of information status in Dutch and Italian: A comparative analysis. *Journal of Phonetics* 30: 629-654.
Thorsen, Nina. 1980. A study of the perception of sentence intonation: evidence from Danish. *Journal of the Acoustical Society of America* 67: 1014-1030.
Thorsen, Nina. 1985. Intonation and text in Standard Danish. *Journal of the Acoustical Society of America* 77: 1205-1216.
Thorsen, Nina. 1986. Sentence intonation in textual context: supplementary data. *Journal of the Acoustical Society of America* 80: 1041-1047.
Trager, George L. & H. L. Smith. 1951. *An Outline of English Structure*. Norman, OK: Battenburg Press.
Trager, George L. 1961. The intonation system of American English. In D. Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott and J. L. M. Trimm (Eds.), *In Honour of Daniel Jones: Papers Contributed on the Occasion of his Eightieth Birthday, 12 September 1961*, pp. 266-270. London: Longman. Reprinted in D. Bolinger (1972), pp. 83-86.
Venditti, Jennifer V., Kikuo Maekawa, & Mary E. Beckman. 2008. Prominence marking in the Japanese intonation system. In S. Miyagawa & M. Saito (Eds.), *Handbook of Japanese Linguistics*, pp. 456-512. Oxford University Press.

Ward, Gregory and Julia Hirschberg. 1985. Implicating uncertainty: The pragmatics of fall-rise intonation. *Language* 61: 747-776.
Xu, Yi. 1999. Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics* 21: 55-105.
Xu, Yi. 2005. Speech melody as articulatorily implemented communicative functions. *Speech Communication* 46(3-4): 220-251.
Xu, Yi & Xuejing Sun. 2002. Maximum speed of pitch change and how it may relate to speech. *Journal of the Acoustical Society of America* 111: 1399–1413.

AMALIA ARVANITI

Amalia Arvaniti received her Ph.D. from the University of Cambridge and has held research and teaching appointments at the Universities of Cambridge, Oxford, Edinburgh and Cyprus. She is currently an Associate Professor in the Department of Linguistics at the University of California, San Diego. Her research focuses on the phonetics and phonology of prosody, with special emphasis on the experimental investigation and formal representation of rhythm and intonation.