Available online at www.sciencedirect.com

SCIENCE DIRECT®

Speech Communication 48 (2006) 667–696

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Phonetic effects of focus and "tonal crowding" in intonation: Evidence from Greek polar questions

Amalia Arvaniti *, D. Robert Ladd, Ineke Mennen [1]

Linguistics and English Language, School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh EH8 9LN, UK

Received 15 April 2005; received in revised form 29 September 2005; accepted 30 September 2005

Abstract

This paper deals with the intonation of polar (yes/no) questions in Greek. An experiment was devised which systematically manipulated the position of the focused word in the question (and therefore of the intonation nucleus) and the position of the last stressed syllable. Our results showed that all questions had a low level stretch associated with the focused word and a final rise–fall movement, the peak of which aligned in two different ways depending on the position of the nucleus: when the nucleus was on the final word, the peak of the rise fall co-occurred with the utterance-final vowel, irrespective of whether this vowel was stressed or not; when the nucleus was on an earlier word, the peak co-occurred with the stressed vowel of the last word. In addition, our results showed finely-tuned adjustments of tonal alignment and scaling that depended on the extent to which tones were "crowded" by surrounding tones in the various conditions we set up. These results can best be explained within a model of intonational phonology in which a tune consists of a string of sparse tones and their association to specific elements of the segmental string.

Keywords: Intonation; Focus; Tonal alignment; Phrase accent; Tonal crowding

1. Introduction

1.1. Background

This paper presents a detailed study of the intonation of polar (yes–no) questions in Modern Greek. Our central empirical goal is to provide controlled instrumental data on the realization of question intonation under differences of word stress (position of stress in word) and focus (position of nuclear accent in sentence). Quantitative data of this sort contribute to the description of Greek prosody,

* Corresponding author. Present address: Department of Linguistics, University of California, San Diego, 9500 Gilman Drive # 0108, La Jolla, CA 92093-0108, USA. Tel.: +1 858 534 8409; fax: +1 858 534 4789.

E-mail addresses: amalia@ling.ucsd.edu (A. Arvaniti), bob.ladd@ed.ac.uk (D.R. Ladd), imennen@qmuc.ac.uk (I. Mennen).

[1] Present address: Speech and Language Sciences, Queen Margaret University College, Clerwood Terrace, Edinburgh EH12 8TS, UK.

0167-6393/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.specom.2005.09.012

and can also be relevant for practical applications such as speech synthesis; as Kochanski and Shih (2003, p. 345) point out, "lessons learned from controlled experiments may help us to find the right model [of intonation for use in TTS]". Such an application is the use in recent Greek TTS systems (e.g. Zervas et al., 2004; Xydas et al., 2004, 2005) of the GRToBI prosodic analysis (Greek Tones and Break Indices; Arvaniti and Baltazani, 2000, 2005), which in turn is based on quantitative data of the type presented here.

Although our data have a descriptive purpose and may also be of practical relevance, our more fundamental theoretical goal is to show that the acoustic effects of stress and focus are best understood if we draw on the ideas of what Ladd (1996) termed the "autosegmental-metrical" (henceforth AM) approach to intonation, based on the work of Bruce (1977) and Pierrehumbert (1980). Our basic empirical finding is that there are both gross differences of contour shape as a function of differences in focus position and fine differences of tonal realization as a function of word stress location. We argue that this provides evidence for the AM assumption that at some abstract phonological level of description, an intonation contour is best represented as a string of tonal elements associated with specific points in the (prosodically organized) segmental string, such as prominent syllables and phrasal boundaries.

This and related AM assumptions have motivated a great deal of research in the past decade or so, such as work on the alignment of "tonal targets" with the segmental string (among many, Prieto et al., 1995 on Spanish; Arvaniti and Ladd, 1995; and Arvaniti et al., 1998 on Greek; Grabe, 1998 on British English and German; Frota, 2002 on European Portuguese; Atterer and Ladd, 2004 on German; Arvaniti and Garding, in press on American English). Similarly, the burgeoning ToBI family of prosodic analyses and labelling schemes for intonation in a variety of languages is solidly based on AM ideas (for a range of languages covered, see Jun, 2005). Despite the predominance of the AM approach that this body of work demonstrates, we believe that there are still good reasons for providing further evidence for its validity. First, there are still fundamental disagreements among researchers over the best way of analyzing certain types of distinctions, disagreements that reflect our imperfect understanding of how the abstract tonal elements of an AM description correspond to $F0$ targets in speech production (for relevant discussions, see Arvaniti et al., 2000, in press; Prieto et al., in press). New detailed data may help to shed light on the link between abstract *tones* and concrete *targets*. The second reason is that even dominant assumptions can be faulty, and therefore it is important to pay attention when such assumptions are challenged.

One such key "autosegmental" assumption of the AM approach is that there is no necessary one-to-one correspondence between elements of the tonal string and segments or syllables. For example, pitch accents (pitch movements at prominent syllables) may consist of multiple tones all associated with the same syllable, and conversely, some syllables may have no tonal specification at all. This assumption leads naturally to a target-and-transition model of pitch realization, in which tones are realized as specific $F0$ points (tonal targets), but the fundamental frequency of at least some syllables is determined merely as part of a transition from one target to the next. Striking evidence for this general idea was provided by the work of Pierrehumbert and Beckman (1988) on Japanese, who showed that the $F0$ contour on a sequence of unaccented syllables (traditionally regarded as having High tone on each syllable) actually behaves like a transition from one word-initial Low–High tonal sequence to the next.

A consequence of breaking tunes down into independent elements is that overall contour shapes and movements do not have a special status in the AM framework: these shapes result from the interaction between the combination of tones used and the segmental material these associate with, so that if the same tones associate with different segmental materials (e.g. a one-syllable word or a long utterance) the shape of the contour can change without a concomitant change of meaning. A case in point is the (rise-)fall–rise tune of traditional British descriptions of intonation. Simplifying somewhat, this tune is sometimes realized as a unit, in that the falling and final rising part co-occur with the most prominent stressed syllable of the utterance, but at other times the fall and the rise appear on different words. Fall-rises of the latter kind are treated as compound tunes (among many, Crystal, 1969; Halliday, 1970; O'Connor and Arnold, 1973), but such analyses have serious drawbacks, since they mark compound fall–rises as exceptional, and cannot account for the fact that they function pragmatically in a way similar to simple fall–rises. In contrast, in the AM framework the behavior of fall–rises follows naturally from their structure: the fall is seen as the transition between a high pitch accent followed by a low phrase accent and a high boundary tone: since each of these entities is independent of the others, they may appear close together in some utterances, giving the impression they form a unit, but quite far apart in others (for a discussion of these tunes from various perspectives, see Ward and Hirschberg, 1985; Hirschberg and Ward, 1992; Ladd, 1996, pp. 215–216; Grice et al., 2000).

Finally, an additional consequence of treating tunes as strings of independent categorically distinct units is that the functions of intonation (e.g. conveying pragmatic meaning or indicating focus) are not expected to be *directly* reflected by quantitative phonetic parameters, such as pitch range or average *F*0. Rather, it is expected that such functions will be mediated by phonological structure, i.e. by the string of abstract tones that makes up an intonational tune at the phonological level; this string is phonetically realized in different context-dependent ways that affect, among many things, the range and span of the *F*0 curve and its alignment with the segmental string.

Recent work has called all of these basic aspects of the AM model into question. First, Xu (2005), and Xu and Xu (2005) have challenged the target-and-transition model. Specifically, these authors argue that each syllable must be tonally specified, not only in tone languages like Mandarin, in which most syllables are specified for tone in the lexicon, but also in languages like English in which AM claims tones are sparse. Xu's proposal was developed largely on the basis of Mandarin data, but it has recently been expanded to English (Xu and Xu, 2005). Specifically, Xu and Xu (2005) examine the realization of focus in English declaratives and claim that the *F*0 of syllables seen as tonally unspecified in the AM framework cannot be derived by interpolation between specified targets. Instead Xu and Xu (2005) proposed that non-accented syllables should be specified as having a [mid] target, which Xu (2005) suggests might be "near the level of the habitual pitch" and so actively controlled (p. 240).

The idea that contour shapes are just the byproduct of interaction between tones and the segmental material they are associated with is questioned in the work of Grabe et al. (2005), and Xu (2005), albeit from different perspectives. Grabe et al. argue that the global shape of a pitch contour can be usefully quantified in terms of continuous phonetic parameters, such as average *F*0 and slope, modeled as a best-fit sum of Legendre polynomials; more importantly, Grabe et al. argue that such continuous phonetic parameters can be used to directly signal the pragmatic force of an utterance. For example, Grabe et al. suggested that average *F*0 is the main contributor to differences between questions and statements in (some of the) dialects of British English they examined; similarly *F*0 slope is said to distinguish declaratives from declarative questions (questions not syntactically marked as such). (Such proposals are by no means new; closely related claims have been made by e.g. Thorsen, 1979; Grønnum, 1983; Cooper et al., 1985; Eady and Cooper, 1986.) In Xu (2005) and Xu and Xu (2005), on the other hand, the building blocks of intonation are melodic segments that consist of four *melodic primitives*, the pitch target (i.e. the shape and height of each syllable's *F*0), pitch range, strength and duration. According to these authors, pitch range and span are used to indicate focus, and by doing so in different ways that depend on the communicative function of the utterance (e.g. statement vs. question) they signal utterance type. Specifically, Xu (2005) argues that in all languages, independently of prosodic type, declaratives favor a "sharp *F*0 decline as early as possible within the focused item" (Xu, 2005, p. 235), while questions usually show a "significant *F*0 increase [...] from the focused element onward" (p. 238), though other types of pitch range manipulation are said to be possible (these are briefly discussed in Section 4.3).

It is clear that the models discussed above deal with problematic aspects of the AM framework. The relation between AM tonal specifications and pragmatic function is a matter of considerable debate within the AM literature (see e.g. Pierrehumbert and Hirschberg, 1990; Ladd, 1996, chapter 6; Grabe et al., 1998; Steedman, 2000). Similarly, the status of pitch range in the AM framework does require reexamination and we briefly touch upon this issue in Section 4 (see also Ladd, 1996, Section 4.4.2 and chapter 7 for discussion). Nevertheless, we believe that the main AM assumptions about the structure and role of intonation are fundamentally correct. Thus, we argue here that it is important to maintain a distinction between the abstract specification of a tune in terms of distinct tonal categories and the continuous phonetic parameters that affect the realization of the resulting *F*0 contour.[2]

Evidence in favor of the fundamental AM assumptions discussed above comes from the present study, which examines the intonation of Greek polar questions. As mentioned earlier, in this study, we examine two types of differences between contours: first, global differences of overall contour shape due to changes in focus position and consequently to the association of tones to segments; second, fine differences of phonetic detail due to *tonal crowding*, that is small scale differences within gross overall shapes due to the variable spacing of tones. We show that the two types of variation we find can be described simply and elegantly if we assume that at the phonological level the tune has an abstract tonal specification. In keeping with AM assumptions, this tonal specification is independent of the segmental material to which it is attached, and shows no necessary correspondence between the number of tones and the number of syllables. These two characteristics of the tune and the way it relates to the segments have two important consequences. First, they drive the adjustments of the fine phonetic detail that we report here, since in some cases the tones are crowded by being associated with the same segmental material, while in other cases they are further apart. Second, as the contours in Fig. 3 show, they result in large differences in global shape depending on focus position, since focus position dictates the association of the tones with the segments. In turn, the existence of such large scale differences suggests that the realization of focus cannot be reduced to a manipulation of pitch range as proposed by Xu (2005) and Xu and Xu (2005), however sophisticated such manipulation might be; rather, the realization of focus by means of intonation may involve the use of different accent types and may result in radical reorganization of text-to-tune associations. Further, the differences we found between questions with focus in different positions show that there is no strong correspondence between pragmatics and pitch contour—as suggested e.g. by Grabe et al. (2005)—since the polar questions we investigate show superficially very different contours (cf. panels (a) and (b) of Fig. 3) that can nevertheless be related in a simple manner to the same abstract tonal representation.

1.2. Greek polar question intonation

Polar or "yes/no" questions in Greek, as in some other European languages such as Spanish and Italian, are syntactically identical to statements, from which they are only distinguished by intonation. Yet, if we consider the polar question contour and the default statement contour as gross contour shapes, both could be said to involve a falling–rising pitch movement. Since they are clearly distinct, more detailed examination of the two tunes is called for. Indeed, such examination suggests that the structure of the polar question intonation requires reference to the location of lexical stresses and the location of the most prominent word in the utterance, as both these factors play a crucial part in the realization of the tune. As an illustration of the difference between statement and question and the effects of stress and focus location in questions we discuss below three examples involving one word utterances (Fig. 1(a)–(c)).

Panels (a) and (b) of Fig. 1 show the statement and question intonation of a sentence consisting of one word with antepenultimate stress. As can be seen, both contours are characterized by a rise–fall. However, the location of the peak relative to the stressed antepenult /mo/, and the steepness and temporal extent of the subsequent fall are quite different: in statements, *F*0 is rising on the stressed syllable, begins to drop towards the very end of that syllable, and remains low thereafter; in polar questions, *F*0 remains rather low until well past the stressed syllable, while both the *F*0 peak and the following fall occur on the word's last syllable. It is also essential to note that the statement tune does not involve a low level *F*0 stretch before the rise–fall; in contrast, the question tune clearly does so.

Further, the realization of the question contour partly depends, as mentioned, on the position of the stressed syllable. The difference can be seen if one compares panel (b) of Fig. 1 to panel (c), which illustrates a question consisting of a single word with final stress: first, in panel (b) the low *F*0 stretch co-occurs with the stressed syllable [mo], while in panel (c) it continues well into the final stressed syllable [lo]; second, in panel (b) the rise–fall occupies the entire duration of the last two syllables [ʝe] and

[2] This division of labor between what is traditionally seen as phonology and phonetics respectively has been extensively questioned in the past ten years or so. A detailed discussion is beyond the scope of this paper, but see Beckman et al. (2004) for arguments in favor of maintaining separate phonetic and phonological levels in the grammar.

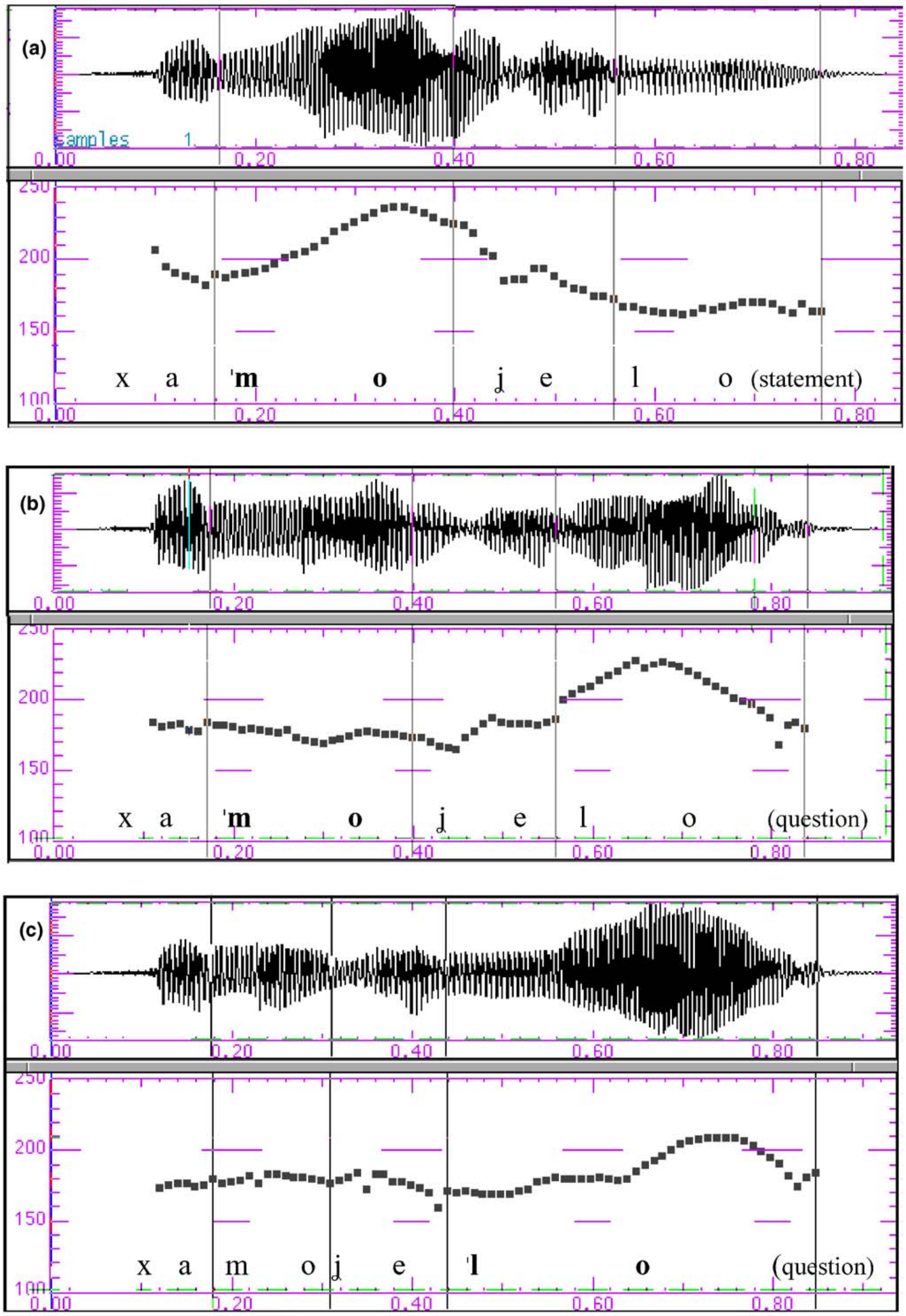


Fig. 1. Waveforms and $F0$ contours of the word [xa'moʝelo] "[a] smile", uttered as a statement (panel a) and as a question (panel b), and for the word [xamoʝe'lo] "I am smiling", uttered as a question (panel c). The vertical lines indicate syllable boundaries.

[lo], while in panel (c) the rise–fall occurs at the very end of the last syllable [lo]. The comparison between the realization of the rise–fall in the statement (panel a) with the realization of a similar pitch movement in the two questions (panels b and c) suggests that in one-word polar questions the stressed

syllable must remain low (to the extent this is possible), while at the same time, it is the end of the question that is marked by the rise–fall. Note that this is different from what happens in the statement contour in which the rise–fall always co-occurs with the stressed syllable independently of this syllable's position in the word (Arvaniti et al., 1998, 2000; Baltazani and Jun, 1999; Baltazani, 2002, chapter 2; Arvaniti and Baltazani, 2000, 2005). Thus, as a first approximation, the polar question contour can be decomposed into two elements: a low level stretch and a right-peripheral rise–fall movement.

This description needs some refining when extended to questions with more than one word. Consider a question like [me to ɣa'lazɟo 'forema] "with the blue dress?": the focus can be either on ['forema] "dress", which makes the question relatively neutral, or on [ɣa'lazɟo] "blue", which contrasts "blue" with some other color. This change of focus is expressed in Greek, as in English, using intonation (see also Baltazani and Jun, 1999; Baltazani, 2002, chapter 2). Specifically, as can be seen in Fig. 2(a), in order to emphasize [ɣa'lazɟo] "blue" *F*0 is *low* on [ɣa'lazɟo] and the final rise–fall starts on the stressed syllable of the following word, i.e. the ['fo] of ['forema]. Note that in this case native speakers do not feel there is particular prominence on the word ['forema] despite the fact that it carries the most obvious pitch movement of the whole tune (this is the intuition of the speakers in our experiment; for similar conclusions see also Baltazani and Jun, 1999; Arvaniti and Baltazani, 2000, 2005; Baltazani, 2002, chapter 2; for an extensive discussion of similar contours in other languages see Ladd, 1996; Grice et al., 2000). Now, as shown in Fig. 2(b), in order to emphasize ['forema] "dress", *F*0 must remain low *throughout the stressed syllable* of ['forema], and then rise and fall again on the last syllable of that word. This is exactly the pattern we saw when lexical stress occurs on the antepenultimate syllable of a one-word question (Fig. 1(b)). This confirms the observation made for one-word utterances, which can be generalized to the following: the stressed syllable of the most prominent word in a Greek polar question is marked by low pitch, while the end of the utterance is marked by a rise–fall; the exact location of the peak of the rise–fall depends on the position of the most prominent word.

Within the AM theory, we would analyze the low pitch as a L*, i.e. as a low pitch accent. In fact, since this accent occurs on the most prominent word of the question it is by definition the *nucleus* or most prominent pitch accent and typically the last accent of the utterance (Pierrehumbert, 1980). In turn, the rise–fall appears to behave in two different ways depending on the location of the nucleus: when the nucleus is on the last word, the peak of the rise–fall occurs on the final syllable of the question (whether stressed or not); when the nucleus is on an earlier word, the peak occurs on the stressed syllable of the final word in the question. The fact that the rise–fall can appear on an unstressed syllable, and is not felt by native speakers to add prominence to the word with which it co-occurs (even when it appears on this word's stressed syllable) supports the view that this rise–fall movement is not a pitch accent, but a boundary phenomenon, i.e. a non-prominence lending pitch movement occurring at the right edge of the question. Furthermore, the fact that the position of the peak relative to the end of the subsequent fall is extremely variable (cf. Figs. 1(b) and (c), 2(a) and (b)) also suggests that the peak and the fall are not a single unit (i.e. a H+L tone of some sort), but two independent elements that sometimes appear closer to each other and sometimes further apart, depending on prosodic conditions. If that is indeed the case, then in an AM analysis the rise would be seen as the reflex of a H− phrase accent (realized as the peak), and the fall as a L% boundary tone (realized as the final *F*0 of the question).

1.3. Goals and hypotheses

Our experiment was designed to provide instrumental data bearing on the impressionistic description just sketched, and in particular to test whether the claims below hold in a large corpus of Greek polar questions in which the position of the nucleus and that of the final stressed syllable are systematically varied. Our hypotheses were as follows:

- the stressed syllable of the most prominent word in the question is low in pitch;
- the rise–fall is not a holistic movement but consists of two elements, such that:
  - the peak of the rise–fall following the nuclear accent occurs on the question's last syllable if the nuclear accent is on the last word;
  - the peak of the rise–fall occurs on the stressed syllable of the last word if the nuclear accent is on an earlier word;

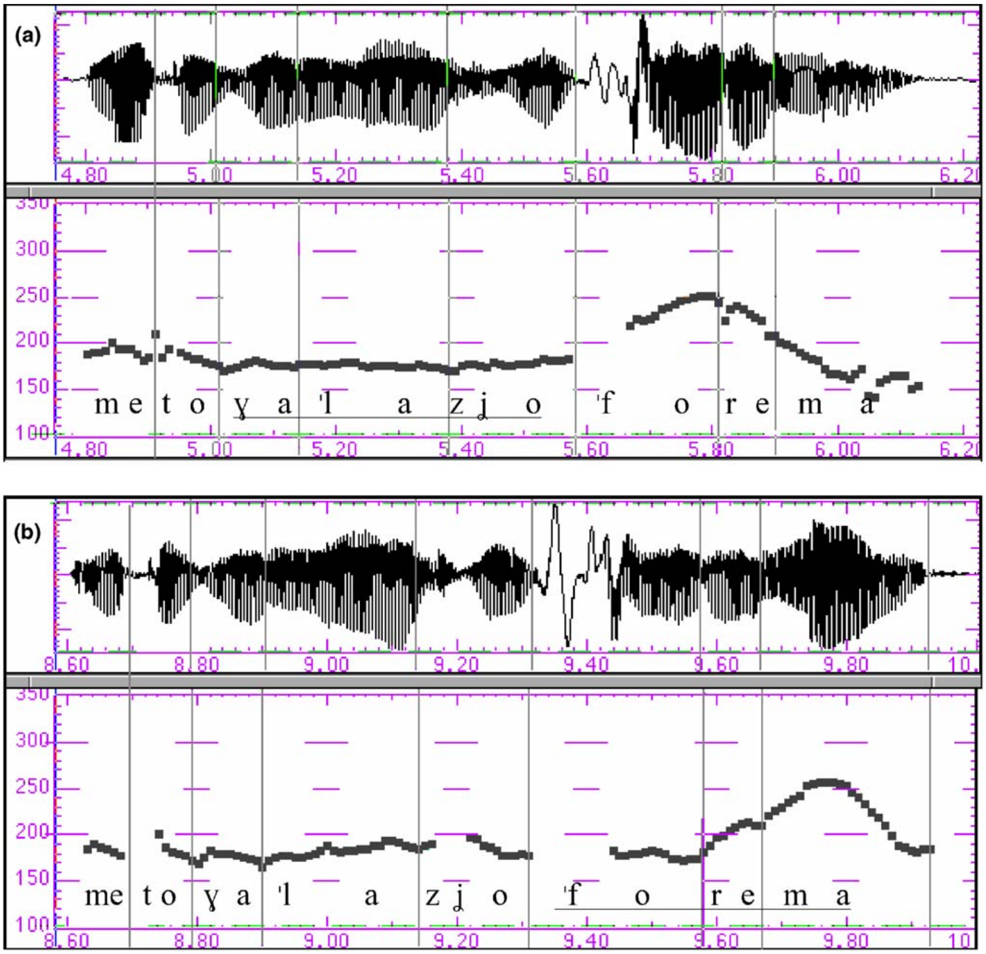


Fig. 2. In panel (a), waveform and *F*0 contour of the phrase [me to ɣa'lazʝo 'forema] "with the blue dress?" with nucleus on the first content word [ɣa'lazʝo] "blue" (early nucleus condition); in panel (b), waveform and *F*0 contour of the same sentence with nucleus on the last content word ['forema] "dress" (final nucleus condition). The vertical lines indicate syllable boundaries.

– in all cases the peak is followed by a fall, which is completed at the very end of the utterance.

On the basis of these hypotheses, we expected to find evidence for three tonal targets, a local dip or short stretch of low *F*0 associated with the nucleus (henceforth NL for *Nuclear Low*), a peak associated with the rise–fall (henceforth H for *High*), and another low point at the end of the utterance (henceforth FL for *Final Low*). Furthermore, we expected each of these targets to show stable alignment and scaling relative to the syllables or boundaries they associate with. These alignment landmarks were as follows: for NL, the stressed syllable of the word in focus; for FL, the end of the utterance; for H, either the stressed syllable of the last word (if the nucleus is early) or the last syllable of the question (if the nucleus is on the last word of the question).

In contrast to the three tonal targets mentioned above, we did not expect that the *elbow*, that is the low turning point marking the beginning of the rise–fall movement, would be a target with specific scaling and alignment. Rather, on the basis of impressionistic data similar to those in Figs. 1 and 2, we expected to find monotonic interpolation between NL and the following H. However, informal inspection of the *F*0 traces once the recordings were completed made it very clear that the interpolation between NL and H is not monotonic but shows a clear inflection point between a relatively level stretch of low *F*0 that most often extends beyond the nuclear syllable and a sharp rise that starts shortly before the peak (this pattern can be most clearly observed in Figs. 2(a) and 3(a)). We therefore decided to measure the elbow and report the findings here, even though we had not designed the materials for this purpose, and the elbow is consequently sometimes obscured by segmental perturbations and voiceless stretches. Similarly, although we expected that FL would align with the very end of the utterance, initial inspection of the data suggested that on many occasions the questions end in a short low *F*0 stretch (see e.g. Fig. 3(a)), the duration of which appears to be related to the position of the last stress. Thus, we decided to measure the alignment of the onset of this low *F*0 stretch from the end of the utterance to see the extent to which it is affected by tonal crowding.

In general, we expected the alignment and scaling of the tonal targets to be influenced by tonal crowding, since it has often been reported that in such conditions targets are regularly either undershot in scaling or shifted in alignment (e.g. Bruce, 1977; Silverman and Pierrehumbert, 1990; Arvaniti et al., 1998, 2000; Nibert, 2000; Prieto, 2005; Schepman et al., in press). Specifically we expected that the position of the last stress would provide information about the effects of tonal crowding. Accord-

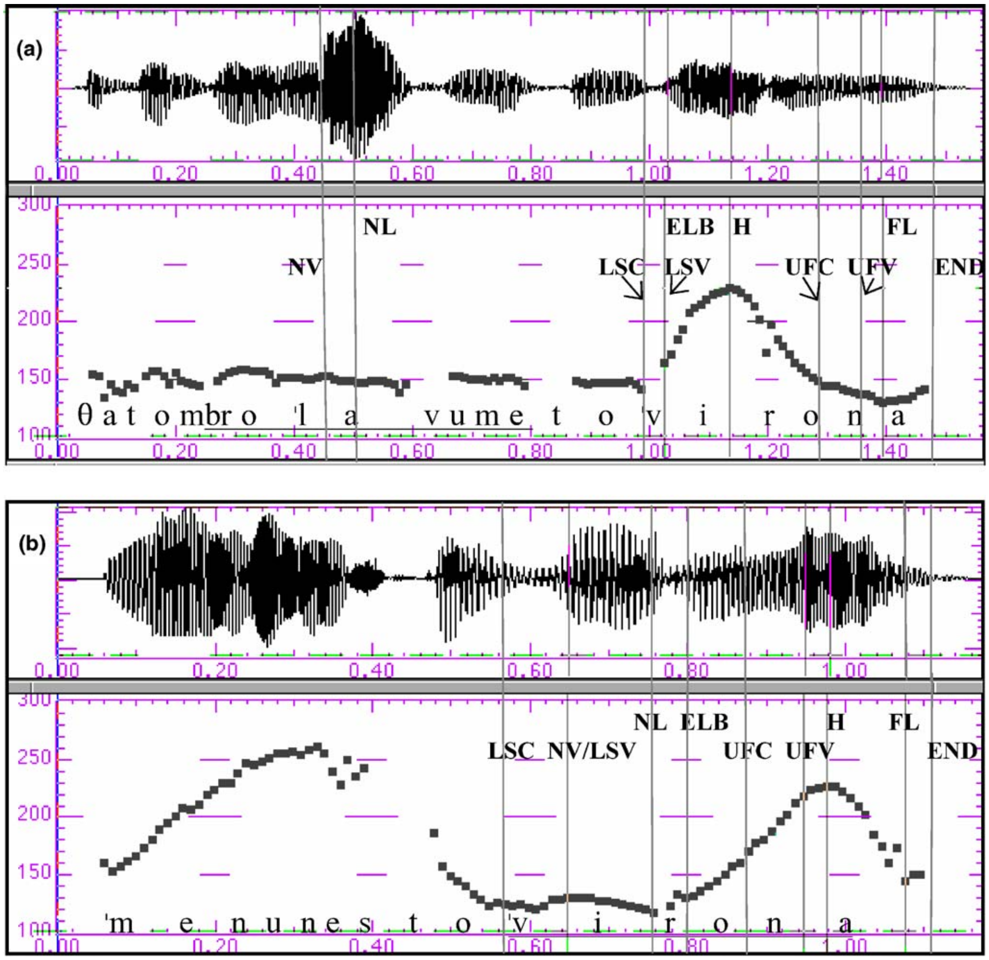


Fig. 3. In panel (a) waveform and *F*0 contour of the sentence [θa tom bro'lavume to 'virona] *lit*. 'will him catch1pl. the Vironas' "[Do you think] we will get there before Vironas [leaves]?" with nucleus on the first content word [bro'lavume] (Early Nucleus condition); this question is preceded by "Gee! We're late!" In panel (b), waveform and *F*0 contour of the sentence ['menune sto 'virona] *lit*. 'live3pl. in-the Vironas' "Do they live in Vironas [area of Athens]?" with nucleus on the final content word ['virona] (Final Nucleus Condition); this question is a response to "Filippos and Stella are neighbors of yours." The vertical lines indicate segmental landmarks and the position of tonal targets.

ing to our hypothesis, if the nucleus is early, the peak of the rise–fall will co-occur with the final stressed syllable; it follows that if this stressed syllable is utterance-final, the whole of the rise–fall pitch movement will have to be realized on that final syllable. This results in tonal crowding, since there are two targets (H and FL) competing for alignment with the same segmental material. The situation will be exacerbated by final nucleus placement and penultimate final stress, as in these conditions all three tones (NL, H, FL) will have to be realized within two syllables. If the nucleus is final and stress is on the last syllable of the last word, we anticipated the most extreme case of tonal crowding, since all three tones would have to co-occur with that last syllable. Our experience with other melodies of Greek (Arvaniti et al., 1998, 2000) suggested that under conditions of tonal crowding Greek does not resort to *truncation* (the non-realization of tones), as happens in some linguistic varieties (among several, Palermo Italian (Grice, 1995), Standard Hungarian (Ladd, 1996), and Belfast English (Grabe et al., 2000)), or even varieties of Greek other than the Standard, such as Cypriot Greek (Arvaniti, 1998). We expected instead to find undershoot of the relevant tones, i.e. adjustments to their scaling (such as lower scaling of the H and higher scaling of the FL) and shifts in alignment, such as tones moving to the left or right of their canonical position in order to make room for other tones.

2. Method

2.1. Speech materials

We designed a corpus of 120 test sentences (see Appendix A), all of which were short polar questions containing a maximum of two content words. In this corpus, the placement of the nucleus and the location of stress in the final word were systematically varied.

Regarding the placement of the nucleus, the test sentences fell into two groups of 60 sentences each. The first group included questions in which the speakers were expected to place the nuclear accent on the final content word of the utterance (henceforth *Final Nucleus*). The second group included questions in which the nuclear accent was expected on the first content word (henceforth *Early Nucleus*). To achieve the desired placement of the nuclear accent, we established appropriate contexts by embedding the test sentences in short dialogues, as in the examples below in which the focus of the question naturally falls on [vori'no] "North-facing" and [çe'retises] "you-said-hello" in examples 1 and 2, respectively. (Readers whose native language is English or another Western European language should be aware that in many cases in Greek, as in example 2 below, the nuclear accent in questions is on the verb, not on the last content word; the interested reader is referred to Grice et al. (2000) and Ladd (1996, chapter 5), for more detailed discussion of nucleus placement in Greek and other languages from Central and Eastern Europe.)

1. Final Nucleus
   A: [to çi'mona ipo'fero ɟa'ti to 'spiti mu 'ine po'li 'krio] ("In the winter I suffer because my house is very cold")
   B: ['ine <u>vori'no</u>] ("Is-it <u>North-facing?</u>")

2. Early Nucleus
   A: [ti <u>çe'retises</u> ti 'nina] ("Did you <u>say hello</u> to Nina?")
   B: ['oçi ðe di 'broseksa] ("No! I didn't notice her.")

In each of the two nucleus placement conditions there were 20 questions ending in a word stressed on the antepenult, 20 questions ending in a word stressed on the penult (e.g. ['nina] in example 2 above) and 20 questions ending in a word stressed on its final syllable (e.g. [vori'no] in example 1 above). The aim of this manipulation was to test the validity of our description, according to which the rise–fall co-occurs with the last stressed syllable in Early Nucleus but not in Final Nucleus questions. The position of the last stress would also provide information about the effects of tonal crowding, since, as mentioned earlier, tonal crowding was expected to be more extreme (a) when nucleus is final and/or (b) the last stress is on the penultimate or final syllable of the question.

On the other hand, in the construction of our materials we wanted to avoid tonal crowding between the targets we intended to investigate and any preceding targets, as such crowding would not allow us to observe the effects of the tonal crowding we had controlled for independently of other similar effects, and thus could confound our results. For this reason, the Early Nucleus questions were designed so that the stressed syllable of the final word was always separated by at least two unstressed syllables from the preceding (nuclear) stress (this did not apply to Final Nucleus questions, in which the last stressed syllable was also the one that carried the nuclear accent). Finally, in order to be able to determine the location of *F*0 maxima and minima with relative ease and accuracy, we chose test words containing mostly sonorants in the relevant syllables (the only exception being, as mentioned, the elbow, which we had not anticipated measuring).

2.2. Speakers

The materials were recorded by eight educated native speakers of Modern Greek, four females (DA, KA, VP, and AA) and four males (BG, AH, TV, and KP). With the exception of VP, all speakers had been brought up in Athens and spoke standard Athenian Greek. VP had been brought up in the Peloponnese and her accent differed slightly from that of the other speakers. Specifically, she used palatal [ʎ] and [ɲ] before the front vowels [i] and [e], instead of the standard alveolars, [l] and [n]. Despite these differences, we have no reason, on the basis of her results, to believe that her background affected the variables under investigation here.

At the time of the recording the speakers were all in their twenties or thirties and had been resident in Edinburgh for periods ranging from a few months to four years (with the exception of AA, the first author, who was on a working visit). None of the speakers had any speech or hearing impairment and all of them, except AA, were naive as to the purpose of the experiment. AA's data were included once it was clear that they were not different from those of the naïve speakers (see Appendices B and C).

The reports that follow are based on the recordings of seven of the eight speakers. BG's speech rate was so fast and his pitch range so narrow that it was virtually impossible to make meaningful measurements of *F*0 peaks and valleys; for this reason, we decided to discard his data.

2.3. Recording and analysis

The recordings were made on digital audio tape (DAT) on professional equipment in the recording studio of the Department of Linguistics, University of Edinburgh. The mini-dialogues containing the test sentences were typed in Greek on cards, one dialogue per card. The speakers read through the entire set of cards once, in random order, speaking both parts of each dialogue aloud. They were instructed to read the dialogues as naturally as possible and take as much time as needed to produce a natural result. No explicit instructions were given to the speakers about where to place the nuclear accent, but in the vast majority of cases speakers placed the nucleus where we expected. The recordings were monitored by the third author (a fluent speaker of Greek), and the speakers were asked to repeat any misread dialogues. In order to avoid "listing" effects the cards with the dialogues were interspersed with cards that contained filler dialogues and materials for another experiment (also in the form of mini-dialogues).

The recordings were digitized at a 16 kHz sampling rate with appropriate low-pass prefiltering. The third author listened to the digitized data of each speaker and discarded any problematic tokens. On average 19 sentences (out of 20) were considered acceptable for each speaker and each accent and stress combination. Of the 45 sentences that were discarded, 11 (1.3%) were produced with a non-intended contour (e.g. they were uttered as statements or as questions showing surprise), 25 (3%) were uttered with either accent or stress placement on the wrong syllable, and 9 (1%) were discarded because of disfluencies (6 sentences) or extensive creak that made reliable FL measurements impossible (3 sentences). After the data had been measured, another 81 sentences (10%) were removed from the statistical files because they were missing values for several measurements, and we thought it best to base the statistical analysis of all variables on the same tokens. Thus, the results we report here are based on 17 sentences per speaker and accent and stress combination, i.e. on a total of 714 sentences.

2.4. Measurements

Measurements of both duration and *F*0 were obtained for the test words, that is, the nuclear word and the final word in Early Nucleus questions, and the final word in Final Nucleus questions. These measurements were obtained using the facilities of ESPS Waves+ on a Sun SPARC workstation; *F*0 in particular was extracted using the pitchtracking facility of Waves+ (with a 49 ms cos.[4] window moving in 10 ms frames).

The test words were manually segmented and labeled, by simultaneous inspection of waveforms and wide-band spectrograms and following standard criteria of segmentation (Peterson and Lehiste, 1960), supplemented by additional criteria for segments found in Greek and not dealt with by Peterson and Lehiste (e.g. the alveolar tap [ɾ]). *F*0 points thought to represent the tonal targets were also located manually. Generally speaking, when measuring *F*0, obviously spurious values and obvious microprosodic effects (such as markedly high *F*0 values after a voiceless stop) were discarded. Most measurements were made by the third author, though some measurements (specifically, NL, ELB, and the onset of the nuclear vowel NV, in the data of the four female speakers) were made by the second author.

The *F*0 points measured, shown in Fig. 3, were the following:

- NL: The lowest extracted *F*0 value within the stretch of the nuclear syllable; this value was taken to be the reflex of the L* nuclear pitch accent of our phonological analysis. As mentioned earlier, we had expected that the stressed syllable of the word in focus would either show a dip in *F*0 or be partly low; instead we found that in most cases, *F*0 remained low until well after that syllable. For this reason, this target is operationally defined in the way stated here.
- ELB: The elbow created by the rise from low *F*0 to the peak of the final rise–fall; ELB was labeled by eye, and was taken to be the point at which the low level stretch turned into a clear rise. As mentioned earlier, since we had not anticipated the presence of this elbow, the materials were not optimally designed for measuring this point. In cases where *F*0 information for the elbow was missing, e.g. due to the presence of a voiceless segment, the first *F*0 point of the rise was measured instead.
- H: The highest extracted *F*0 value at the peak of the final rise–fall movement and taken to represent the putative H– phrase accent; in the rare cases in which there was a series of points with the same *F*0, the first one was selected for measurement.
- FL: The lowest extracted *F*0 value at the end of the utterance that did not appear to be spurious (i.e. ignoring marked outliers due to pitch-halving or doubling and regions of creaky voice). If a low stretch of *F*0 was present, FL coincided with its beginning, since such low *F*0 stretches were either flat (see e.g. Fig. 2(b)) or slightly rising (see e.g. Fig. 3(a)), but never falling.

The above mentioned *F*0 measurements reflect the scaling of the various tonal targets. These measurements were in Hz and were later expressed in units of the ERB scale, using the equation proposed in Glasberg and Moore (1990). All statistical results refer to the ERB-transformed values (though it should be noted that statistical tests run on the original *F*0 measurements in Hz lead to the same conclusions as those run on the ERB transformed values).

In addition, we measured the *alignment* of the *F*0 targets, that is, their time of occurrence relative to the onset (or offset) of specific segments with which we hypothesized that these tonal targets co-occurred. Fig. 3 illustrates how the various measurements were made. The following alignment measurements were obtained:

- NVtoNL, the distance between the onset of the nuclear vowel (NV) and NL.
- LSCtoELB, the distance between the consonant onset of the last stressed syllable (LSC) and the elbow.
- UFCtoELB, the distance between the onset of the utterance-final syllable's consonant (UFC) and the elbow.
- LSVtoH, the distance between the onset of the last stressed vowel (LSV) and H.
- UFVtoH, the distance between the onset of the utterance-final vowel (UFV) and H.
- FLtoEND, the distance of the lowest *F*0 point at the end of the question (FL) and the actual end of the utterance.

As a check on the reliability of our measurement procedures, the third author re-measured 3.5% of the data measured by the second author. These duplicate measurements were subjected to *t*-tests, all of which yielded non-significant results. Further, the absolute difference between measurements was calculated, and means and standard deviations were computed. The comparison showed good inter-labeler agreement: mean absolute differences ranged from 7 to 12 ms (s.d.s from 6 to 13). The only case of some discrepancy related to the measuring of ELB (absolute mean difference 31 ms, s.d. 69), a result that was expected due to the fact that locating ELB was not as easy as locating the other targets; thus these results should be treated with some caution.[3]

2.5. Statistical analysis

The scaling and alignment measurements were statistically analyzed by means of two-way repeated measures analyses of variance (ANOVAs) with NUCLEUS and LAST STRESS as the independent variables (where NUCLEUS refers to the position of the nucleus—early or final—and LAST STRESS refers to the position of the final stressed syllable of the question—antepenultimate, penultimate or final). Differences between levels of LAST STRESS, and NUCLEUS × LAST STRESS interactions were investigated by means of planned comparisons since these interactions were expected.

Statistical testing was based on mean values for each speaker rather than raw data. This was done in order to avoid two serious problems associated with the use of raw data (for a thorough discussion of these problems see Max and Onghena, 1999). First, using mean values for each speaker, and thereby treating each speaker as one experimental unit, avoids violating one of the main assumptions underlying analysis of variance, namely the independence of observations (also referred to as *independent error effects* in repeated measures designs). Second, by using means, a substantial increase in type I error, that is the incorrect rejection of the null hypothesis, is avoided (type I error can increase dramatically if raw data are used as experimental units, due to the large number of degrees of freedom for the error term, also referred to as the *sum of squares residual* in repeated measures designs). Readers interested in inter-speaker variation, which obviously cannot be investigated if means are used for the statistical analysis, can consult Appendices B and C which provide means and standard deviations for each speaker's alignment and scaling data, respectively.

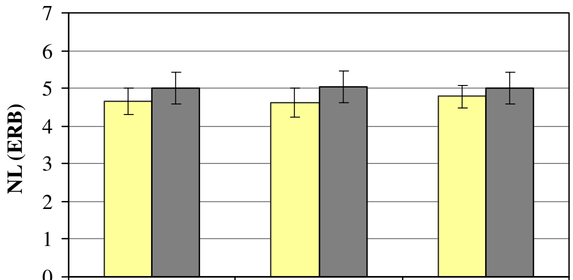
[3] Although these greater differences may cast some doubt on the validity of the ELB data, it should be noted that in a related study (Arvaniti and Ladd, in preparation), algorithmic determination of the elbow did not yield substantially more consistent results than estimation by eye.

3. Results

3.1. Scaling and alignment of Nuclear Low (NL)

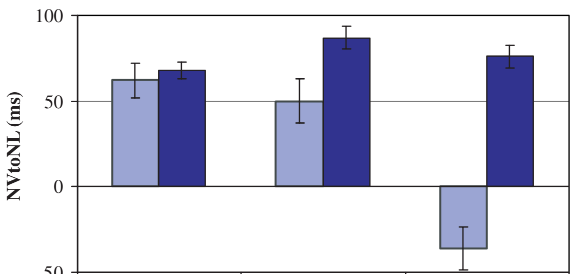
As can be seen in Fig. 4, the scaling of NL was higher in Early than in Final Nucleus [$F(1,6) = 7.2$, $p < 0.04$], but showed no effect of LAST STRESS [$F(2,12) = 0.9$, n.s.], and no interaction between NUCLEUS and LAST STRESS [$F(2,12) = 1.97$, n.s.].

The alignment of NL, i.e. its distance from the onset of the nuclear vowel (NVtoNL) was affected by both NUCLEUS [$F(1,6) = 103.7$, $p < 0.001$] and LAST STRESS [$F(2,12) = 17.9$, $p < 0.001$], and also showed interaction between the two [$F(2,12) = 18.8$, $p < 0.001$], illustrated in Fig. 5. The investigation of this interaction showed that NL aligned earlier in Final than in Early Nucleus when LAST STRESS was final or penultimate [$p < 0.001$ for final stress; $p < 0.01$ for penultimate stress], but there was no



	antepenult	penult	final
FN	4.65	4.62	4.78
EN	4.99	5.04	4.99

Fig. 4. Means and standard errors of Nuclear Low (NL) scaling in ERB, as a function of NUCLEUS and LAST STRESS.



	antepenult	penult	final
FN	62	50	-36
EN	68	87	76

Fig. 5. Means and standard errors of the distance of Nuclear Low (NL) from the onset of the Nuclear Vowel (NV) in ms, as a function of NUCLEUS and LAST STRESS. Negative values mean that NL occurs before the onset of the nuclear vowel.

difference between Early and Final Nucleus when LAST STRESS was antepenultimate (i.e. when tonal crowding in Final Nucleus was minimal). As can be seen in Fig. 5, in Early Nucleus, there was a small effect of LAST STRESS such that alignment was later with penultimate stress than in the other two levels [for final vs. penultimate stress, $p < 0.02$; for penultimate vs. antepenultimate stress, $p < 0.001$]; in Final Nucleus, on the other hand, NL aligned before the NV onset when LAST STRESS was final, but during NV in the other two stress conditions (between which there was no difference) [$p < 0.01$ for both final vs. penultimate stress, and for final vs. antepenultimate stress].

3.2. Scaling and alignment of High (H)

As can be seen in Fig. 6, H was scaled lower in Final than in Early Nucleus [$F(1,6) = 45.2$, $p < 0.001$], and also showed lower scaling with each level of LAST STRESS [$F(2,12) = 28.14$, $p < 0.001$; for antepenultimate vs. penultimate stress, and for antepenultimate vs. final stress, $p < 0.001$; for penultimate vs. final stress, $p < 0.02$]. There was no interaction between NUCLEUS and LAST STRESS [$F(2,12) = 0.39$, n.s.].

As we had hypothesized, H showed two modes of alignment, depending on the position of the nucleus, appearing closer to the onset of the last stressed vowel in Early Nucleus and closer to the onset of the final vowel in Final Nucleus (cf. Fig. 7(a) and (b)). Specifically, the distance of H from the onset of the last stressed vowel (LSVtoH) was affected by both NUCLEUS [$F(1,6) = 185.75$, $p < 0.001$] and LAST STRESS [$F(2,12) = 200.86$, $p < 0.001$] and showed interaction between these two factors

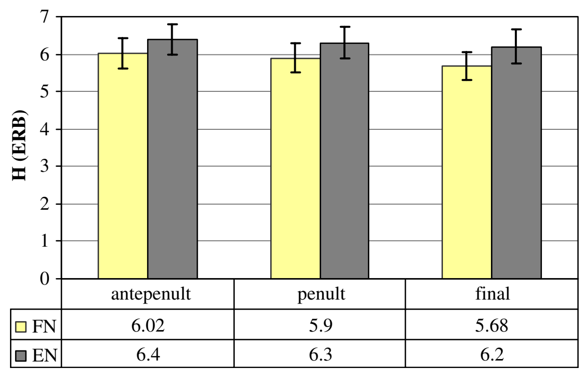


Fig. 6. Means and standard errors of High (H) scaling in ERB, as a function of NUCLEUS and LAST STRESS.

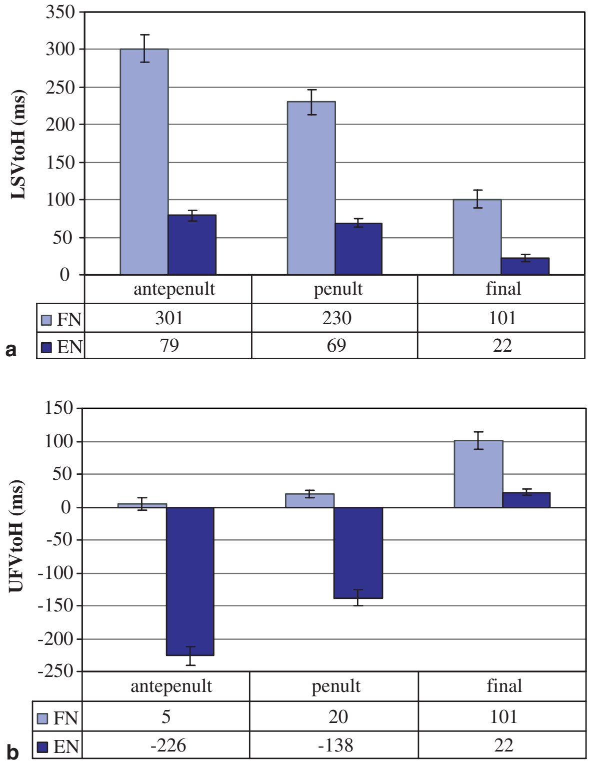


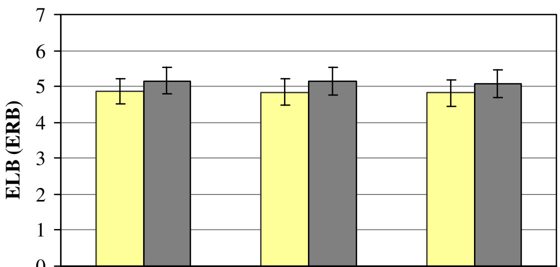
Fig. 7. Panel (a): Means and standard errors of the distance of H from the onset of the last stressed vowel (LSVtoH) in ms, as a function of NUCLEUS and LAST STRESS. Panel (b): Means and standard errors of the distance of H from the onset of the utterance-final vowel (UFV) in ms, as a function of NUCLEUS and LAST STRESS. Negative values mean that H occurs before the onset of the utterance-final vowel.

[$F(2,12) = 122.01$, $p < 0.001$]. As can be seen in Fig. 7(a), LSVtoH was overall substantially shorter in Early than in Final Nucleus irrespective of LAST STRESS [$p < 0.001$ for all pair-wise comparisons]; the short duration of LSVtoH in Early Nucleus shows that H aligned close to the onset of the last stressed vowel in this condition. In addition LSVtoH was affected by LAST STRESS: in both Early and Final Nucleus, H appeared increasingly further away from the onset of the last stressed vowel (and increasingly closer to the utterance end), the further away from the end of the utterance the last stressed vowel was [$p < 0.001$ for all pair-wise comparisons, except Early Nucleus antepenultimate vs. penultimate stress, $p < 0.01$]; as can be seen in Fig. 7(a), this effect was much more pronounced in Final Nucleus.

When H alignment was measured with respect to the onset of the utterance-final vowel (UFVtoH), it was the H in Final Nucleus that occurred closer to the chosen segmental landmark (see Fig. 7(b)). UFVtoH showed effects of NUCLEUS [$F(1,6) = 166.9$, $p < 0.001$] and FINAL STRESS [$F(2,12) = 207.5$, $p < 0.001$] and, as expected, interaction between these two factors [$F(2,12) = 100.14$, $p < 0.001$]. As can be seen in Fig. 7(b), in Final Nucleus H occurred within the utterance final vowel and occurred later than in Early Nucleus at all LAST STRESS levels [for antepenultimate and penultimate stress, $p < 0.001$; for final stress, $p < 0.01$]. In addition, in Final Nucleus, H occurred close to the onset of the utterance-final vowel, except when LAST STRESS was final, presumably because of pressure from the preceding NL which must also co-occur with this vowel and "pushes" the H rightward [$p < 0.001$ for all pair-wise comparisons, except penultimate vs. final stress, $p < 0.01$]. In Early Nucleus, in contrast, H aligned increasingly closer to the onset of the utterance-final vowel, as LAST STRESS got closer to the end of the question [$p < 0.001$ for all pairwise comparisons]; in Early Nucleus, however, H and the utterance-final vowel co-occurred only when LAST STRESS was final (i.e. when the utterance-final vowel was the same as the last stressed vowel).

3.3. Scaling and alignment of the Elbow (ELB)

As can be seen in Fig. 8, ELB was scaled higher in Early than in Final Nucleus [$F(1,6) = 11.2$, $p < 0.02$], but showed no effect of LAST STRESS level [$F(2,12) = 2.6$, n.s.]. There was also no interaction between NUCLEUS and LAST STRESS [$F(2,12) = 0.24$, n.s.].

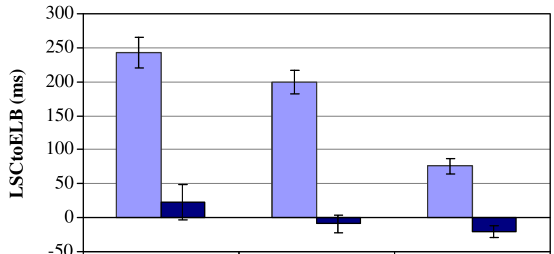


	antepenult	penult	final
FN	4.88	4.84	4.82
EN	5.16	5.15	5.08

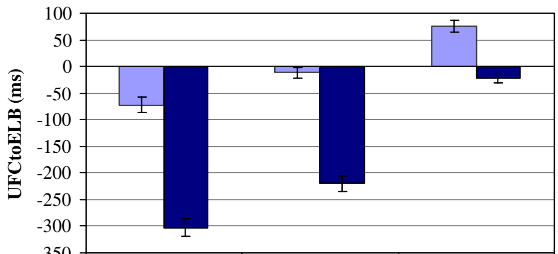
Fig. 8. Means and standard errors of Elbow (ELB) scaling in ERB, as a function of NUCLEUS and LAST STRESS.

Like the alignment of the H, the alignment of ELB depended on the position of the nucleus. When ELB alignment was measured from the consonant onset of the last stressed syllable (LSCtoELB), it showed NUCLEUS [$F(1,6) = 70.5$, $p < 0.001$] and LAST STRESS [$F(2,12) = 36.9$, $p < 0.001$] effects, and also interaction between these two factors [$F(2,12) = 10.9$, $p < 0.01$]. As illustrated in Fig. 9(a), in Early Nucleus ELB aligned very close to this consonant onset; in this condition, LAST STRESS did not affect ELB alignment. In Final Nucleus, on the other hand, ELB generally appeared well after the consonant onset, and aligned increasingly later as LAST STRESS moved further from the end of the utterance [$p < 0.001$ for all pair-wise comparisons, across NUCLEUS levels, and across LAST STRESS levels within Final Nucleus].

Similar effects were in place when ELB alignment was measured with respect to the onset of the utterance-final consonant (UFCtoELB).



a	antepenult	penult	final
FN	243	199	76
EN	22	-9	-21



b	antepenult	penult	final
FN	-72	-11	75
EN	-303	-220	-22

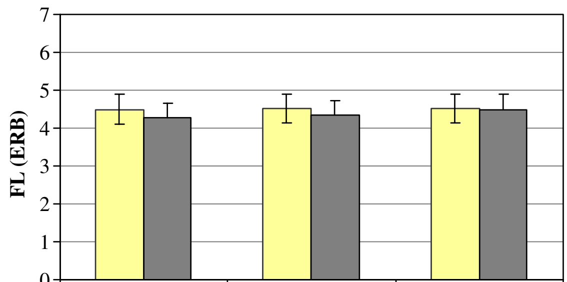
Fig. 9. Panel (a): Means and standard errors of the distance of the Elbow (ELB) from the consonant onset of the last stressed syllable (LSC) in ms, as a function of NUCLEUS and LAST STRESS. Panel (b): Means and standard errors of the distance of ELB from the onset of the utterance-final consonant (UFC) in ms, as a function of NUCLEUS and LAST STRESS. Negative values mean that ELB is realized before the onset of the relevant segmental landmark.

UFCtoELB showed effects of NUCLEUS [$F(1,6) = 91.9$, $p < 0.001$] and LAST STRESS [$F(2,12) = 277.2$, $p < 0.001$], and also the expected interaction between these two factors [$F(2,12) = 32.04$, $p < 0.001$]. As can be seen in Fig. 9(b), ELB aligned close to the utterance-final consonant in Final Nucleus, but well before it in Early Nucleus; in both nucleus conditions, ELB retracted from the utterance-final consonant onset the further away for the end of the utterance LAST STRESS was. Interestingly, there was no statistically significant difference between Early Nucleus with final stress and Final Nucleus with penultimate stress, two levels that show similarly moderate tonal crowding [$p < 0.001$ for all other pair-wise comparisons].

3.4. Scaling and alignment of Final Low (FL)

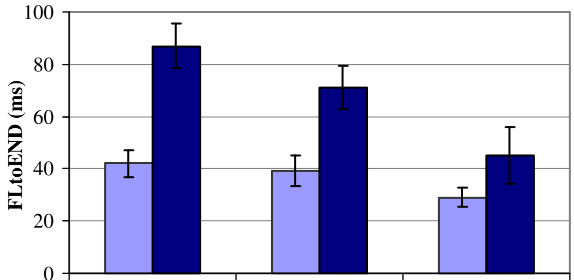
As can be seen in Fig. 10, the scaling of FL was not affected by either NUCLEUS [$F(1,6) = 5.6$, n.s.] or LAST STRESS [$F(2,12) = 1.14$, n.s.], and showed no interaction between these two factors [$F(2,12) = 3.08$, n.s.].

On the other hand, the alignment of FL with respect to the end of the utterance was affected by both NUCLEUS [$F(1, 6) = 37.2$, $p < 0.001$] and LAST STRESS [$F(2,12) = 29.65$, $p < 0.001$], and showed interaction between the two [$F(2,12) = 9.45$, $p < 0.01$]. As can be seen in Fig. 11, FL was closer to the end of the utterance in Final Nucleus than in Early Nucleus [$p < 0.001$ for all pair-wise comparisons within stress levels]. In addition, in Early Nucleus, FL was aligned closer to the end of the utterance, the closer to the utterance end LAST STRESS was [$p < 0.001$ for all pair-wise comparisons].



	antepenult	penult	final
FN	4.5	4.51	4.52
EN	4.29	4.36	4.48

Fig. 10. Means and standard errors of Final Low (FL) scaling in ERB, as a function of NUCLEUS and LAST STRESS.



	antepenult	penult	final
FN	42	39	29
EN	87	71	45

Fig. 11. Means and standard errors of the distance of Final Low (FL) from the end of the utterance (END) in ms, as a function of NUCLEUS and LAST STRESS.

A similar but much weaker effect was observed in the Final Nucleus data, in which FL appeared closer to the utterance end when LAST STRESS was final than when it was penultimate or antepenultimate, but showed similar alignment between these two levels [for antepenultimate vs. final stress, $p < 0.02$; for penultimate vs. final stress, $p < 0.04$].

4. Discussion and conclusion

4.1. The autosegmental representation of Greek polar question intonation

The results of the experiment presented here largely supported the hypotheses we set out to test. First, our results show that the stressed syllable of the nuclear word has low $F0$ (and as we further show in Section 4.4 it is often associated with the lowest $F0$ in the entire question, barring FL). Second, our results show that Greek polar questions also end in low $F0$, as we had anticipated. Most importantly for the description of this melody, our data confirmed our hypothesis that the peak of the rise–fall pitch movement is aligned in two different ways, depending on which word in the question is the nucleus. When the nucleus is early, H co-occurs within the last stressed vowel of the question; when the nucleus is on the last word, H occurs within the final vowel of the question, irrespective of the position of the last stressed syllable.

On the other hand, our experiment provided us with data that cannot be easily accommodated by our initial hypotheses. First, the scaling and alignment properties of ELB, which we originally did not consider a target, strongly suggest that it must be the reflex of a L tone and therefore a tonal target. Our results show that the alignment of this L tone, just like the alignment of H, depends on the position of the nucleus. They also show that the L precedes the H tone in a consistent way: the L tone aligns with a segmental landmark close to H, the consonant onset before the vowel with which the H aligns. This suggests to us that this L tone is closely connected to the H, and most probably forms a bitonal L+H entity with it. Within the AM framework, the dual patterning of this L+H entity can best be accounted for if L+H is analyzed as a phrase accent (i.e. L+H−), a type of tonal unit that has been shown to exhibit two modes of alignment in several languages (Grice et al., 2000). Specifically, according to Grice et al. (2000), phrase accents are boundary tones that can have two types of association, a primary association to a phrasal boundary, and a secondary association to a specific tone-bearing unit (such as the penultimate syllable of the utterance-final word, as in Hungarian polar questions). Phrase accents are thus realized in two different ways: when the tone-bearing unit to which they have a secondary association is available (i.e. not associated with another tone), they co-occur with it; when it is not, their primary association takes over and they appear at a phrasal edge.

In the case of yes–no questions in Greek, the tone-bearing unit to which the phrase accent has a secondary association is the last stressed syllable. If it is available, i.e. if it is not already associated with the nuclear L* pitch accent, then L+H− co-occurs with it. If the secondary association is not possible—as happens when the nucleus is on the final word, in which case the last stressed syllable is already associated with L*—then L+H− appears on the final syllable of the utterance. This happens even when there are two syllables available after the nucleus, as is the case in questions with Final Nucleus and antepenultimate stress (see Fig. 7(b)). Although the co-occurrence of L+H− with the penultimate syllable in this case would avoid crowding the following L% boundary tone, this is not the alignment we observe; rather, L+H− still aligns with the final syllable. This alignment with the utterance-final syllable provides further support for analyzing L+H as a tonal unit with a primary association to a boundary, i.e. a phrase accent. Thus, on the basis of our results we conclude that the appropriate AM analysis of the Greek polar question intonation is L* L+H− L%.

4.2. Accounting for differences in gross contour shape

As mentioned in the introduction, we believe that our results have repercussions not just for the AM analysis of Greek polar question intonation but for the more general validity of the basic AM principles about the structure of intonation and the relation between its phonetics and its phonology. Specifically, taken all together our results suggest that although it is possible to describe the basic melody of Greek yes–no questions as a rise–fall pitch movement that starts with low pitch, such a characterization would be too simplistic to adequately account for the patterns uncovered by this study. First, it is clear that the contour is not a unit that can be stretched to accommodate any number of words or syllables; rather, it is composed of separate elements, the exact realization and location of which depends on the position of the nucleus (final or not) and the position of the last stressed syllable in the utterance (if the nucleus is not on the final word). Thus, for example, the low level *F*0 stretch that roughly starts on the nuclear syllable can be as short as half that syllable—when the nucleus is on the final word and this word is stressed on the ultima—or it may stretch for several syllables—if the nucleus is on a non-final word and the last stressed syllable appears much later.

The differences in the contours of Greek polar questions are difficult to accommodate in a model that treats contours holistically, that is as overall shapes, such as Grabe et al. (2005). Even in short questions, like those discussed here, it is clear that the shape of the tune is strongly affected by the position of the nucleus and of the last stressed syllable, as Fig. 3 amply demonstrates. And such differences are likely to proliferate in longer questions because of additional rising prenuclear accents on content words before the nucleus (Baltazani and Jun, 1999; Baltazani, 2002; Arvaniti and Baltazani, 2005). Thus, in a framework like that of Grabe et al. one would most likely be forced to conclude that—at the very minimum—questions with the same segmental material but different nucleus location have different melodies. Reaching such a conclusion, however, would not only make the analysis lose a significant generalization and become more complex, but would also be incorrect in that it would go strongly against the intuitions of native speakers that all polar questions share the same basic tune. Thus, although a quantification of pitch contours along the lines proposed by Grabe et al. may be useful for the phonetic modeling of intonation and for practical applications, it cannot obviate the need for an abstract representation.

4.3. The realization of focus

Our results also have consequences for our understanding of the intonational marking of focus, by showing that the realization of focus in Greek polar questions cannot be analyzed as simply a matter of pitch range manipulation, as some have suggested. For example Xu and Xu (2005) argue that English declaratives have a constant tune in which three pitch range areas are distinguished: a neutral prefocal area, an "on-focus" pitch range expansion and a post-focal pitch range suppression (p. 189). Within this scheme, parts of the same constant contour may expand or shrink but they remain otherwise the same. According to Xu and Xu (2005, p. 186) the *F*0 drop that is expected after focus "is largely *equated* with focus itself [emphasis in the original]". As mentioned in the introduction, Xu (2005) takes this evidence one step further and suggests that this drop in *F*0 early in the focused item is not only found in English and Mandarin, but drives pitch modifications in languages of all prosodic types. Further, Xu (2005) argues that a similar cross-linguistic pattern is in place for questions, which are said to show rising pitch from the focused element and beyond.

Our results clearly show that this view of focus is too simplistic to account for our data. (Incidentally, it does not seem to be appropriate for all tonal languages either: e.g. Pan (in press) shows that in Taiwanese Min duration is used more consistently to signal focus than expanded *F*0 range.) First, it is clear that if we were to follow the method of Xu and Xu (2005) and superimpose the contours of Greek questions with early and final nucleus onto each other, we would not see the same contour with different local ranges, but two quite different contours (cf. panels (a) and (b) of Figs. 2 and 3). This alone suggests that much more than just pitch range is necessary to account for the patterns shown here.

We do not of course dispute that pitch range may be part of what signals focus; e.g. as we discuss below, unlike other tones, the nuclear L* does not appear to be undershot under conditions of tonal crowding, suggesting that pitch span may be locally expanded during the realization of the nucleus. On the other hand, however, our data also show that the *F*0 of the item in focus is low, and that the overall *F*0 level is substantially the same for a large part of the contour both before and after the nucleus. This in turn suggests that the stressed syllable of the word in focus is not a pivotal point regarding pitch range changes, as Xu and Xu (2005) and Xu (2005) suggest. In addition, all questions end with very low *F*0, as the scaling of FL indicates (cf. Figs. 4, 8 and 10). This is equally inconsistent with the idea of a raised pitch range that Xu (2005) associates with questions: if FL, which is post-focal, was produced within a high pitch range, we would expect it to be higher than NL, but in fact it is substantially lower [$F(1,6) = 15.7, p < 0.01$]. To be sure, Xu (2005) does suggest that languages that do not use pitch raising to show focus in questions may be using a reduced (and low) post-focal range, like Neapolitan Italian (D'Imperio, 2001). But this appears inconsistent with our Greek data as well, given that the post-focal stretch includes a pitch peak (the proposed L+H− phrase accent) that is typically the highest in the sentence.

Most importantly, what the foregoing discussion shows clearly is that the relationship between focus and pitch range is not the same in every language; different combinations of pitch range and pitch span in different subdivisions of the sentence can be used to indicate focus in different languages. This suggests that the relationship between pitch, focus and utterance type is in some sense *part of the grammar* of each language. This means that we cannot universally equate pitch range factors with focus as Xu and Xu (2005, p. 186) do. As far as the intonation of the Greek polar questions is concerned, this means that we must in addition to any pitch range effects assume the presence of a specific tone, in this case a L* tone, that must co-occur with the stressed syllable of the word in focus.

4.4. Fine differences in tonal realization

As already mentioned, in addition to the differences in gross contour shape due to the position of nucleus and the last stressed syllable, our results show fine-grained differences in the realization of all targets across different prosodic conditions, suggesting that tonal targets exhibit scaling and alignment adjustments to phonetic effects such as declination and tonal crowding. We believe that our AM analysis captures not only the gross differences in contour shape due to the location of the nucleus, but can also account for these fine-grained differences in the scaling and alignment of the targets.

Regarding scaling, most targets (NL, ELB and H) show effects of declination, in that all are scaled lower when the nucleus is final—that is when they all appear later in the utterance—than when the nucleus is early. On the other hand, not all targets are influenced by tonal crowding in the same fashion. NL shows no scaling undershooting due to tonal crowding (this is consistent with the claim of Xu (2005) that focus—i.e. the nucleus—is realized with a widened pitch range). In contrast, H is scaled lower in both Early and Final Nucleus when the last stressed syllable is final than when it is penultimate or antepenultimate, i.e. when tonal crowding is more acute. Similarly, ELB is scaled lower in Final Nucleus, i.e. when it appears closer to another L target, NL.

The only exception to this trend for scaling adjustments is FL, which was stable across all conditions. This result contrasts with the initial hypothesis that FL would be undershot in cases of tonal crowding. Although we cannot exclude the possibility of small effects that our statistical analysis did not detect, this finding corroborates reports from other languages, which show scaling stability of utterance-final L tones (e.g. Menn and Boyce, 1982 and Liberman and Pierrehumbert, 1984 for English; Connell and Ladd, 1990 for Yoruba; Ladd and Terken, 1995 for Dutch). Further, the present results on the scaling of FL provide solid quantitative evidence against a widespread misconception—which appears to date back to Waring's (1976) impressionistic description of Greek intonation—that the Greek yes–no question intonation ends with a "fall to mid pitch" (Mackridge, 1990, 90ff.; Holton et al., 1997). Indeed, a comparison of the scaling of NL, ELB and FL shows that FL is the lowest of the three low targets [for FL vs. NL, $F(1,6) = 15.7$, $p < 0.01$; for FL vs. ELB, $F(1,6) = 60.4$, $p < 0.001$].

In contrast to the rather small effects on scaling, our data show significant adjustments in tonal alignment under conditions of tonal crowding. Thus, in Early Nucleus, which results in minimal crowding, the leftmost targets, NL and ELB, show stable alignment with the nuclear vowel and the last stressed consonant respectively. In Final Nucleus, on the other hand, both ELB and NL show alignment adjustments. NL occurs earlier when the nucleus is on the penultimate syllable of the utterance than when it is on the antepenult, and even earlier when the nucleus is on the final syllable, the only condition in which NL appears before the nuclear vowel. Similarly, in Final Nucleus, ELB is generally aligned with the utterance-final consonant, but its alignment is not stable: ELB aligns very close to the onset of this consonant when last stress is antepenultimate, i.e. when tonal crowding is not extreme, but it aligns closer to the consonant's offset when last stress is final, i.e. when tonal crowding is at its worst. It is also worth noting that in conditions of comparable crowding, these targets show similar alignment across nucleus positions; e.g. in Final Nucleus questions with antepenultimate stress, which result in little tonal crowding for NL, this target aligns in a similar fashion to Early Nucleus questions; likewise, ELB alignment is similar between Early Nucleus and Final Nucleus when last stress is final and penultimate respectively, both conditions in which the last syllable of the utterance must accommodate three tones, the L and H of the L+H− phrase accent and the following L% boundary tone.

The adjustments in the alignment of NL and ELB suggest effects from the upcoming tones on NL, and from the preceding NL on ELB. This interplay between anticipatory and carryover effects is most evident in the alignment of H, which is affected both by the preceding NL and ELB and by the upcoming FL. In Early Nucleus, in which pressure is due to the upcoming FL, H aligns earlier when the last stress is final than in penultimate and antepenultimate stress, which do not create tonal crowding. In contrast, in Final Nucleus, in which the pressure is mainly due to the preceding NL and ELB, H aligns later when the last stress is final than in the other two conditions.

In addition, the Greek data show that tonal crowding influences not only the scaling and exact alignment of targets, but may also affect other aspects of their realization. This is the case, e.g., with FL, which is realized as a low level stretch in the Early Nucleus condition, but as a low $F0$ point in Final Nucleus (i.e. when it is more crowded by preceding tones). As mentioned in Section 4.2, similar effects are observed for NL, which is always realized as a low $F0$ stretch, but with a duration that varies depending on the degree of tonal crowding.

The variability in the realization of the FL and NL suggest that some low tones are not realized as

*F*0 points but as stretches of low pitch. This is the case with FL and NL here; similar evidence comes from recent results of Arvaniti and Garding (in press) which show that the L tone of the American English L*+H pitch accent is also realized as a low level stretch. On the other hand, it is clear that other L tones are simply not realized in this way, but rather in a manner similar to H tones, that is as local *F*0 points; this is, for example, the case of ELB in the present data, of the L in L+H* in American English (Arvaniti and Garding, in press), and of the L of L+H accents in Greek (Arvaniti et al., 1998). It is at present unclear why these differences exist, but their presence suggests that further research is necessary before we can gain a clear understanding of the relationship between phonological tones and their exact phonetic realization.

Although the findings about the realization of L tones indicate that the basic AM assumption that all tones are realized as *F*0 points may be too strong and should be revised with respect to low tones, overall our results are compatible with the AM approach. First, they confirm that high tones are indeed realized as specific *F*0 points, i.e. as high peaks. They also confirm what is shown in much other work (among many, Bruce, 1977 for Swedish; Silverman and Pierrehumbert, 1990 for American English; Prieto et al., 1995 for Spanish; Arvaniti et al., 1998 for Greek; Atterer and Ladd, 2004 for German; Prieto, 2005 for Catalan; Arvaniti and Garding, in press for American English), namely that the position of such peaks is substantially affected by tonal crowding and therefore they may not consistently occur within the syllable they are associated with, or consistently outside it. Thus indirectly our results provide evidence for the position of Arvaniti et al. (1998) that strict alignment between targets and segments cannot be expected, whether one adopts the AM view of the relationship between the two, or espouses the view that *F*0 levels or movements are synchronized with syllables (e.g. Xu, 2005; Xu and Xu, 2005). Data like those presented here show that the relationship between tones, targets and segmental material is much more complex than either view suggests, and in need of further research.

Nevertheless, it should be noted that AM has the advantage that it can predict both anticipatory and carryover effects of tonal coarticulation, and this is exactly what we found here, as amply discussed above. Such combination of both anticipatory and carryover effects in tonal coarticulation has been reported even for tonal languages (in which a close relationship between syllables and *F*0 levels or movements is plausible) such as Taiwan Min (Peng, 1997; Wang, 2002) and Thai (Gandour et al., 1994). These results further support the view that a simple equation between syllable and tone duration does not hold. Although it does appear to be the case that carryover effects are stronger than anticipatory effects, the existence of both seems now fairly well documented (and is hardly surprising since it simply shows that *F*0 gestures are coordinated in the same fashion as other articulatory gestures). Given the above results, any suggestions that anticipatory effects are virtually non-existent in tonal coarticulation (cf. the claims of Xu (2005) and Xu and Xu (2005) about Mandarin and English) are clearly too strong.

4.5. Accounting for fine differences in tonal realization

As shown, the fine adjustments of tonal target realization discussed above are easy to account for within the AM framework; however, they are not so easy to accommodate in other models of intonation. First, holistic quantitative models like Grabe et al. (2005) may not be fine-grained enough to include the kind of small scale differences that we are discussing here, or at least would require much more complicated modeling in order to do so. Moreover, such differences, though significant and linguistically relevant as we show elsewhere (Arvaniti et al., in press), do not stem from the function to which the contours are put, and therefore would be hard to accommodate within a model like that of Grabe et al.

Accounting for our crowding data is equally difficult for full specification models, like Xu (2005) and Xu and Xu (2005). Such models would encounter several problems. First, if all syllables are tonally specified, it is impossible to account for the fact that in some circumstances targets show more adjustments (suggesting greater tonal crowding) than in others, and for the fact that the direction of the adjustments differs depending on context. If all syllables were tonally specified, we should observe similar types of adjustments in all cases, as tonal crowding would be present to the same degree in all conditions. Likewise, the direction of the adjustments should also be the same in all cases. This more balanced adjustment between tones is largely what is observed in languages like Mandarin as shown by Kochanski et al. (2003a,b), Xu (2001, 2005) and Xu and Wang (2001) among others. But there is no a priori reason why intonation, as manifested in a language like Greek, in which tones are not lexically specified, should be equated with the phonetic parameters that affect the realization of lexical tones in languages like Mandarin, in which the tone of practically every syllable is specified in the lexicon.

Note also that the small scale adjustments we report here would not respond well to the notion of *articulatory strength* (Xu, 2005; Xu and Xu, 2005) which "specifies the speed at which a local pitch target is approached" (Xu, 2005, p. 243; see also Kochanski and Shih, 2003 for a similar use of the term *strength*). The problem is that strength affects both the scaling and the alignment of a target and our results clearly show that the two are independent of each other. The NL and FL, which are undershot primarily in terms of alignment but not in terms of scaling, are a case in point.

Finally, Greek data are problematic for full-specification models for another reason: syllables between the nucleus and the post-accentual rise are as low as the nucleus, a pattern that cannot be accounted for if all of these syllables have [mid] pitch, which is actively controlled as Xu (2005) suggests. If it were so, we should at least see a small rise after the end of the nuclear syllable and a return to the same low level just before the beginning of the final rise, but as Figs. 1–3 demonstrate this is not the case. In contrast, the low *F*0 of these syllables falls out naturally from the AM description: these syllables are found between two low targets, L* and the L of L+H−, and thus they remain low because the course of *F*0 is interpolated between the two L tones.

We acknowledge of course than in a complex model like that presented in Xu (2005) and in Xu and Xu (2005) it may be possible to set parameter values for strength, pitch range, pitch span, etc. so as to eventually produce the "right" result, but in our view that may prove only that those models are too powerful. This point is echoed in a comment of Kochanski and Shih (2003) that in their model (Stem-ML) "the same pitch contour can often be approximated many different ways [sic], using different sets of tags, some of which may well be linguistically unreasonable" (p. 344). This may not be an issue for a model primarily intended for synthesis purposes, but unlike Kochanski and Shih (2003), Xu (2005) and Xu and Xu (2005) say that the primitives in their model constitute a symbolic representation. If so, then the question that arises is what type of representation this is. If it is a detailed presentation of phonetic implementation, then it is legitimate to ask whether anything is gained by providing tonal specifications for all syllables if the same result can be achieved without them. Parsimony suggests that we should not increase, beyond what is necessary, the number of entities required to explain a phenomenon, and postulating that all syllables need to be tonally specified in a language without lexical tone goes against this basic principle of modeling and theory building. If, on the other hand, this model is meant to be a phonological representation (or, more likely, a hybrid between phonetics and phonology; see Xu, 2005, 242 ff.), then it is not clear that providing distinct representations for practically each and every tune that an individual will utter in the course of a lifetime is a desirable feature. In other words, even though at some level quantitative information is necessary in order to produce a tune, an abstract representation that allows the language learner to generalize over different tunes with the same pragmatic meaning is still necessary; and this is precisely the kind of generalization that an AM representation provides.

4.6. Conclusion

In conclusion, our experimental data largely supported our original AM analysis of the intonation of yes–no questions in Greek, but also provided us with evidence that lead us to revise this analysis in order to incorporate the presence of the L tone we uncovered. We propose, therefore, that the autosegmental analysis of the Greek yes–no melody is L* L+H− L%. Our data also show that the L+H− phrase accent of this melody has two modes of alignment, depending on nucleus placement. Further, the data show that the alignment and, to a lesser extent, the scaling of targets is subject to adjustments due to tonal crowding. Specifically, it appears that *all* targets are subject to alignment shifts in order to accommodate preceding and following targets, although their scaling proved to be less prone to such modifications. These results have consequences for our understanding of Greek intonation, as well as practical implications for its modeling for speech synthesis. Most importantly, the results are significant from the point of view of the phonology and phonetics of intonation more generally, as they provide robust evidence in favor of the autosegmental-metrical theory of intonation and strongly suggest that measuring the scaling and alignment of specific points in an *F*0 contour is essential for understanding the workings of a contour and its interaction with the segmental string.

Acknowledgements

The research reported here was supported by the UK Economic and Social Research Council through grant no. R000 23 5614 to the University of Edinburgh, with Ladd and Arvaniti as Co-Principal Investigators and Mennen as Research Associate. This support is gratefully acknowledged. We would also like to thank our speakers, and also Mary Beckman, Sun-Ah Jun, and Janet Pierrehumbert for discussion of this material, two anonymous reviewers for their detailed comments, and Norman Dryden, Mike Bennett, Eddie Dubourg, Cedric MacMartin, and Stewart Smith for technical support. A much abbreviated, preliminary version of this paper is published as Arvaniti, A., (2002) The intonation of yes–no questions in Greek. In: M. Makri-Tsilipakou (Ed.), Selected Papers on Theoretical and Applied Linguistics, pp. 71–83, Thessaloniki, Department of Theoretical and Applied Linguistics, School of English, Aristotle University.

Appendix A. Materials

The test materials are presented in IPA transcription, together with a gloss. For reasons of space, only the yes–no questions are presented, not the entire dialogs they were part of. The sentences are divided into six sets of 20 sentences each; the first three sets include Final Nucleus questions; the other three include Early Nucleus questions. Within each group, the first set has questions that end in a word with lexical stress on the antepenult, the second set has questions that end in a word with lexical stress on the penult, and the third set has questions that end in a word with lexical stress on the final syllable.

Final Nucleus, antepenultimate last stress

1.	[ˈleɣane vɾoˈmoloɣa]	"Were they cursing?"
2.	[ˈkanis avɣoˈlemono]	"Are you making egg-and-lemon sauce?"
3.	[paˈtuses sta vroˈmoneɾa]	"Did you step in the sewage?"
4.	[foˈɾesate ce ˈmalina]	"Did you even put on woollies?"
5.	[ˈvazis ðenðroˈlivano]	"Do you use rosemary?"
6.	[ˈpiɣane sto ˈmeɣaɾo]	"Did they go to Megaro [Athens Concert Hall]?"
7.	[ˈmazepse vɾoˈxoneɾo]	"Did it gather rainwater?"
8.	[toɲ ˈɟeɾðises sto ˈdomino]	"Did you beat him at dominoes?"
9.	[ˈfeɾondan aˈnaɣoɣa]	"Were they ill-mannered?"
10.	[θa ˈminune sta ˈʝanena]	"Will they settle in Ioannina [Greek city]?"
11.	[ˈine ap ti ˈmiɾina]	"Is he from Mirina [Greek city]?"
12.	[meˈɣalose sta ˈmeɣaɾa]	"Did he grow up in Megara [Greek city]?"
13.	[na ˈpiʝena ˈmoni mu]	"Should I have gone alone?"
14.	[ˈmenune sto ˈviɾona]	"Do they live in Vironas [area of Athens]?"
15.	[ˈekane kaˈɾubalo]	"Did he get a bump?"
16.	[ɣnoˈɾizis to xaˈɾiðimo]	"Do you know Charidimos?"
17.	[ˈeɾixne çoˈnoneɾo]	"Was it sleeting?"
18.	[se voiˈθa ce sto maˈʝiɾema]	"Does he also help you with the cooking?"
19.	[xriˈazete siˈðeɾoma]	"Does it need ironing?"
20.	[ˈine tu vlaˈðimiɾu]	"Is it Vladimiros's?"

Final Nucleus penultimate last stress

1.	[ˈforaɟe ti ˈɣuna]	"Was she wearing the fur coat?"
2.	[ˈisuna sto ˈbaɲo]	"Were you in the bath?"
3.	[ˈtavale sti ˈɟala]	"Did she put them in the bowl?"
4.	[xtiˈpai to kuˈðuni]	"Is the bell ringing?"
5.	[ˈine θimoˈmeni]	"Is she angry?"
6.	[foˈɾuse ce pleˈɾeza]	"Was she also wearing a mourning veil?"
7.	[ˈtavale maˈzimu]	"Has she taken against me?"
8.	[metaˈkomisan sto ˈðoma]	"Did they move into the loft?"
9.	[ˈpinis lemoˈnaða]	"Are you drinking lemonade?"
10.	[ˈɣnoɾises ti ˈɾena]	"Have you met Rena?"
11.	[sta ˈmesa tu ɟeˈnaɾi]	"In the middle of January?"
12.	[se ˈpiɾakse i ˈɟiɾi]	"Did the pollen affect you?"
13.	[ˈeçi lanoˈlini]	"Does it contain lanolin?"
14.	[ˈvazete leˈvanda]	"Do you use lavender?"
15.	[aˈnevice sto ˈvima]	"Did he go on to the podium?"
16.	[θa se peɾiˈmeni i maɾiˈlena]	"Will Marilena wait for you?"
17.	[ˈpsaxnis ɟa veˈlona]	"Are you looking for a needle?"
18.	[ˈiðes to maˈɟo mu]	"Have you seen my swimsuit?"
19.	[ˈiðes to stiˈlo mu]	"Have you seen my pen?"
20.	[ˈmenune sti ˈɾomi]	"Do they live in Rome?"

Final Nucleus, final last stress

1.	[θa ˈpane sto baˈli]	"Are they going to Bali?"
2.	[ˈpiɾe to proiˈno]	"Did he take the morning [train]?"
3.	[miˈlai sovaˈɾa]	"Is he serious?"
4.	[ˈeçi maidaˈno]	"Does it have parsley?"
5.	[aˈfti me to beˈɾe]	"The one with the beret?"
6.	[ˈine voɾiˈno]	"Is it north-facing?"
7.	[ˈine ɣalaˈna]	"Are they blue?"
8.	[ˈfiɣane maˈzi]	"Did they leave together?"
9.	[ˈevale ɟaˈʎ a]	"Has he started wearing glasses?"
10.	[peɾpaˈtusan aŋgaˈʎ a]	"Were they walking arm in arm?"
11.	[ˈexo melaˈɲa]	"Do I have a bruise?"
12.	[eˈðo aɾisteˈɾa]	"Here to the left?"
13.	[su ˈfenete almiˈɾo]	"Does it seem salty to you?"
14.	[ˈekane ziˈmɲa]	"Did he make a mess?"
15.	[ˈitane kaˈlo]	"Was it good?"
16.	[ˈeçi galeˈɾi]	"Does he own a gallery?"
17.	[na ˈine ɟeɾaˈni]	"[Do you reckon] that they are cranes?"
18.	[ˈeçi ɾandeˈvu]	"Does she have a date?"
19.	[ˈkapço maɣaˈzi]	"[Is it] some [kind of] club?"
20.	[miˈla me tin ˈiɾo]	"Is she talking to Iro?"

Early Nucleus, antepenultimate last stress

1.	[ton epi'skevasan to ne'ɾomilo]	"Did they repair the watermill?"
2.	[a'vɣa su kaθa'ɾizune]	"What's so funny about it?" [*lit*. Is someone peeling eggs for you?]
3.	[tin anaɣno'ɾisate tin 'iɾiða]	"Did you recognize Iris?"
4.	[ksana'piɣate stin 'eɟina]	"Have you visited [the island of] Aegina before?"
5.	[ti ɣno'ɾizete tin 'elena]	"Have you met Elena?"
6.	[su a'ɾesun ta ksi'nomila]	"Do you like crab-apples?"
7.	[ton 'iðes to xa'ɾiðimo]	"Have you seen Charidimos?"
8.	[pe'ɾisepse ɾi'zoɣalo]	"Is there any rice-pudding left?"
9.	[a'liθça to no'mizune]	"Do they really believe that?"
10.	[θa 'fiɣun ap to 'livano]	"Will they leave Lebanon?"
11.	[aɾ'ɟi to laðo'lemono]	"Is the oil-and-lemon sauce going to take long?"
12.	[ma oɾi'masan ta ko'ɾomila]	"But have the plums ripened?"
13.	[ma 'eɾikse ço'noneɾo]	"But was it sleeting?"
14.	[θa 'ftasi to ɾo'ðoneɾo]	"Is there enough rosewater?"
15.	[θa ton pɾo'lavume to 'viɾona]	"[Do you think] we will get there before Vironas [leaves]?"
16.	[to θi'mase to 'numeɾo]	"Do you remember the number?"
17.	[θa me si'kosi to mo'noziɣo]	"Will the horizontal bar hold me?"
18.	[ta'ɣoɾasan ta o'moloɣa]	"Did they buy the bonds?"
19.	['pame tuɾko'limano]	"Shall we go to Tourkolimano [Pireus neighborhood]?"
20.	[si'niɣoɾo se 'valane]	"Did they appoint you their council?"

Early Nucleus, penultimate last stress

1.	[ðe 'bame sto ka'vuɾi]	"Shall we go to Kavouri [Athens suburb]?"
2.	[a'ɣoɾases le'moɲa]	"Did you buy lemons?"
3.	[θi'mase pos ti 'lene]	"Do you remember her name?"
4.	[ti çe'ɾetises ti 'nina]	"Did you say hello to Nina?"
5.	[ma xo'ɾane sti ba'ɲeɾa]	"But do they fit in the bathtub?"
6.	[to 'lavate to 'ðema]	"Did you receive the package?"
7.	[se vo'levi mesi'meɾi]	"Does afternoon suit you?"
8.	[ta 'iðes ta ɣla'ɾoɲa]	"Did you see the sea-gulls?"
9.	[ti ɣno'ɾizis ti ma'ɾina]	"Do you know Marina?"
10.	[su a'ɾesi to 'vazo]	"Do you like the vase?"
11.	[tu 'aɾese to 'ðoɾo]	"Did he like the present?"
12.	[ta ka'θaɾises ta 'mila]	"Did you peel the apples?"
13.	[na 'vɣume sti ve'ɾanda]	"Shall we go out on the veranda?"
14.	[kselabi'kaɾise to ne'ɾo]	"Is the water clear again?"
15.	[ma mas xo'ɾai to di'vani]	"But is there enough room for us on the couch?"
16.	[ta fu'skosan ta ba'loɲa]	"Did they blow up the balloons?"
17.	[me to po'ðilato pi'ɟeni]	"Does he go [there] on bicycle?"
18.	[psi'θikan ta la'zaɲa]	"Is the lasagna done?"
19.	[le'çazi to me'lani]	"Does ink stain?"
20.	[to 'θes to maksi'laɾi]	"Do you want the pillow?"

Early Nucleus, final last stress

1.	[ta 'tɾos ta laðe'ɾa]	"Do you eat [vegetables] cooked in oil?"
2.	[θa bo'ɾusa na se 'ðo]	"Could I see you?"
3.	[na 'katsume e'ðo]	"Shall we sit here?"
4.	[sa'ɾesi to me'nu]	"Do you like the menu?"
5.	[ta 'piɾes ta ɟa'ʎ a]	"Did you get the glasses?"
6.	[ti 'vlepis ti ma'ɟa]	"Can you see the yeast?"
7.	[ðe 'vazis ce ana'na]	"Won't you add some pineapple?"
8.	[θa 'fas kaɾame'le]	"Will you have some crème caramel?"
9.	[to 'kseɾis to ɣu'ði]	"Do you know Goudi [Athens neighborhood]?"
10.	[mu 'ðinis mɲa ɣu'ʎ a]	"Could you give me a sip?"
11.	[ðo'cimases me'ze]	"Have you tried the starters?"
12.	[to 'iðes to mo'ɾo]	"Did you see the baby?"
13.	[θa su 'ftasi to ma'li]	"Will you have enough wool?"
14.	[ksanama'ɟiɾepses ɾa'ɣu]	"Have you made ragout before?"
15.	[sas 'eftase i bo'ɟa]	"Did you have enough paint?"
16.	[ka'θaɾise i vro'mɲa]	"Did the stain wash out?"
17.	[θa 'valete mɲa'lo]	"Will it knock sense into you?"
18.	[ti 'vlepis ti mo'ni]	"Can you see the monastery?"
19.	[ma'zevi to li'no]	"Does linen shrink?"
20.	[ti bo'ʎ asan ti mi'ʎ a]	"Did they graft the apple tree?"

Appendix B

Means and standard deviations of all durational measurements (in ms), for each speaker separately. *FN* stands for Final Nucleus; *EN* stands for Early Nucleus.

Speaker		NVtoNL—Nuclear Vowel to Nuclear Low					
		Final Nucleus			Early Nucleus		
		FN-antepenult	FN-penult	FN-final	EN-antepenult	EN-penult	EN-final
AA	Mean	64.0	61.9	−49.6	83.0	89.0	86.6
	s.d.	59.0	56.5	41.2	31.9	43.7	37.6
AH	Mean	3.6	−20.1	1.1	41.9	57.4	55.8
	s.d.	22.8	37.7	3.7	29.9	33.0	30.1
DA	Mean	85.2	95.8	−64.4	76.3	100.4	84.2
	s.d.	50.7	43.0	43.5	22.9	33.4	27.1
KA	Mean	70.0	55.2	−78.2	47.1	75.7	73.3
	s.d.	35.7	26.4	43.0	22.7	35.1	30.9
KP	Mean	81.4	48.4	−5.7	68.4	77.5	59.9
	s.d.	10.6	41.4	22.2	24.4	22.0	21.0
TV	Mean	74.4	58.2	0.2	81.9	96.5	87.5
	s.d.	54.7	55.2	0.8	39.9	39.6	34.9
VP	Mean	56.3	51.6	−53.8	77.8	103.7	81.2
	s.d.	42.3	37.8	32.9	27.9	29.4	21.7

Speaker		LSCtoELB—Last Stressed Consonant to Elbow					
		Final Nucleus			Early Nucleus		
		FN-antepenult	FN-penult	FN-final	EN-antepenult	EN-penult	EN-final
AA	Mean	266.4	232.2	123.2	−34.7	−44.0	−55.9
	s.d.	32.5	36.2	41.9	62.4	75.8	62.7
AH	Mean	128.4	103.8	46.5	3.1	−3.8	−36.3
	s.d.	44.8	24.3	31.9	26.9	30.8	89.1
DA	Mean	326.7	249.8	99.5	−0.2	0.9	−26.2
	s.d.	44.5	41.8	32.2	34.1	24.2	57.9
KA	Mean	247.1	204.1	44.7	−4.4	14.8	−9.3
	s.d.	49.1	24.3	40.8	24.1	29.3	19.4
KP	Mean	255.7	197.7	72.3	6.4	−60.8	−5.9
	s.d.	52.4	39.3	38.3	25.8	157.6	35.2
TV	Mean	250.9	201.9	84.5	13.5	39.5	16.1
	s.d.	88.8	47.8	48.0	23.0	29.2	33.2
VP	Mean	223.3	206.3	58.5	173.3	−9.1	−32.9
	s.d.	49.2	46.6	34.7	84.4	64.5	57.6

Speaker		UFCtoELB—Utterance-Final Consonant to Elbow					
		Final Nucleus			Early Nucleus		
		FN-antepenult	FN-penult	FN-final	EN-antepenult	EN-penult	EN-final
AA	Mean	−55.4	4.4	123.2	−332.7	−262.2	−55.9
	s.d.	26.8	28.8	41.9	60.4	75.9	62.7
AH	Mean	−129.6	−58.6	45.3	−236.5	−151.2	−38.5
	s.d.	53.5	20.2	32.4	19.9	26.6	91.3
DA	Mean	−25.5	19.1	99.5	−337.9	−242.8	−26.2
	s.d.	41.6	36.2	32.2	37.1	29.7	57.9
KA	Mean	−63.7	−10.0	44.7	−305.5	−206.4	−9.3
	s.d.	44.2	11.7	40.8	48.5	35.4	19.4
KP	Mean	−26.4	12.1	72.3	−248.2	−233.1	−5.9
	s.d.	39.5	34.0	38.3	26.5	161.2	35.2
TV	Mean	−90.1	−22.5	83.9	−319.1	−200.9	17.9
	s.d.	92.7	54.7	49.8	37.8	41.7	33.7
VP	Mean	−111.8	−22.8	58.5	−342.3	−239.9	−32.9
	s.d.	42.4	39.5	34.7	56.4	56.8	57.6

Speaker		LSVtoH—Last Stressed Vowel to High					
		Final Nucleus			Early Nucleus		
		FN-antepenult	FN-penult	FN-final	EN-antepenult	EN-penult	EN-final
AA	Mean	336.2	271.8	164.8	82.3	80.5	34.8
	s.d.	19.1	23.5	28.4	20.8	20.9	17.4
AH	Mean	204.6	143.0	71.4	40.6	37.6	−0.2
	s.d.	35.9	22.1	13.6	15.2	19.8	22.6

(continued on next page)

Appendix B (*continued*)

Speaker		LSVtoH—Last Stressed Vowel to High					
		Final Nucleus			Early Nucleus		
		FN-antepenult	FN-penult	FN-final	EN-antepenult	EN-penult	EN-final
DA	Mean	347.4	253.2	105.7	75.6	62.6	14.0
	s.d.	26.3	32.8	22.5	20.1	19.2	20.1
KA	Mean	281.2	222.7	64.1	86.9	78.7	15.3
	s.d.	23.5	27.8	27.5	21.0	22.8	22.6
KP	Mean	285.9	209.9	84.2	84.8	76.1	23.8
	s.d.	27.5	27.6	17.0	33.5	22.6	16.2
TV	Mean	321.0	238.5	103.4	87.9	78.0	28.7
	s.d.	40.8	25.0	16.8	16.5	25.4	17.9
VP	Mean	328.3	267.5	110.9	92.9	72.2	34.5
	s.d.	37.7	16.0	26.8	23.9	22.2	15.9

Speaker		UFVtoH—Utterance-Final Vowel to High					
		Final Nucleus			Early Nucleus		
		FN-antepenult	FN-penult	FN-final	EN-antepenult	EN-penult	EN-final
AA	Mean	33.8	38.8	164.8	−225.6	−141.3	34.8
	s.d.	22.6	22.6	28.4	15.3	29.8	17.4
AH	Mean	−37.6	−0.9	71.4	−196.5	−111.3	−0.2
	s.d.	37.6	24.8	13.6	26.6	17.3	22.6
DA	Mean	5.8	12.5	105.7	−282.1	−187.1	14.0
	s.d.	30.9	20.6	22.5	35.5	26.7	20.1
KA	Mean	−16.8	2.6	64.1	−219.1	−138.7	15.3
	s.d.	24.3	15.9	27.5	34.2	29.9	22.6
KP	Mean	29.4	34.7	84.2	−165.9	−90.9	23.8
	s.d.	19.2	27.3	17.0	36.2	19.5	16.2
TV	Mean	8.9	20.5	103.4	−248.8	−149.4	28.7
	s.d.	24.9	10.6	16.8	28.0	30.0	17.9
VP	Mean	9.5	34.0	110.9	−241.5	−149.6	34.5
	s.d.	26.1	16.2	26.8	39.4	30.0	15.9

Speaker		FLtoEnd—Final Low to End (of utterance)					
		Final Nucleus			Early Nucleus		
		FN-antepenult	FN-penult	FN-final	EN-antepenult	EN-penult	EN-final
AA	Mean	33.2	23.7	18.3	91.4	56.6	10.6
	s.d.	15.0	9.9	12.9	31.5	33.8	10.4
AH	Mean	25.9	29.6	22.1	53.9	40.2	12.7
	s.d.	12.2	14.9	11.0	17.0	14.3	11.7
DA	Mean	66.0	55.6	39.7	110.8	77.1	74.1
	s.d.	36.0	32.8	16.9	42.2	39.1	32.5
KA	Mean	51.0	63.0	38.2	94.4	95.5	79.3
	s.d.	22.1	25.0	17.8	29.5	29.8	22.2

Appendix B (*continued*)

Speaker		FLtoEnd—Final Low to End (of utterance)					
		Final Nucleus			Early Nucleus		
		FN-antepenult	FN-penult	FN-final	EN-antepenult	EN-penult	EN-final
KP	Mean	31.8	22.1	22.1	53.7	48.2	25.0
	s.d.	14.1	12.9	10.5	31.4	17.3	17.5
TV	Mean	43.9	38.9	40.3	103.5	94.3	63.7
	s.d.	20.3	19.4	23.3	25.1	33.2	19.8
VP	Mean	40.5	40.8	25.6	97.9	82.1	50.1
	s.d.	20.4	29.3	20.6	23.4	37.5	30.4

Appendix C

Means and standard deviations of all *F*0 measurements (in ERB), for each speaker separately. *FN* stands for Final Nucleus; *EN* stands for Early Nucleus.

Speaker		NL—Nuclear Low					
		Final Nucleus			Early Nucleus		
		FN-antepenult	FN-penult	FN-final	EN-antepenult	EN-penult	EN-final
AA	Mean	5.3	5.3	5.3	5.5	5.4	5.4
	s.d.	0.1	0.2	0.2	0.2	0.2	0.3
AH	Mean	3.8	3.7	4.5	3.9	3.9	3.9
	s.d.	0.2	0.2	0.4	0.2	0.2	0.3
DA	Mean	5.4	5.5	5.4	6.2	6.4	6.2
	s.d.	0.2	0.3	0.2	0.3	0.5	0.4
KA	Mean	5.3	5.3	5.1	5.5	5.6	5.7
	s.d.	0.2	0.2	0.1	0.3	0.3	0.2
KP	Mean	3.1	3.1	3.2	3.2	3.2	3.2
	s.d.	0.1	0.2	0.2	0.1	0.2	0.2
TV	Mean	4.3	4.1	4.3	4.7	5.0	4.7
	s.d.	0.2	0.2	0.3	0.2	0.4	0.3
VP	Mean	5.5	5.5	5.6	5.9	5.8	5.8
	s.d.	0.2	0.1	0.2	0.2	0.3	0.2

Speaker		ELB—Elbow					
		Final Nucleus			Early Nucleus		
		FN-antepenult	FN-penult	FN-final	EN-antepenult	EN-penult	EN-final
AA	Mean	5.6	5.5	5.6	5.9	5.8	5.9
	s.d.	0.2	0.2	0.3	0.3	0.3	0.3
AH	Mean	4.2	4.1	4.1	4.9	4.8	4.7
	s.d.	0.4	0.2	0.3	0.3	0.3	0.5
DA	Mean	5.6	5.7	5.5	5.9	5.9	5.7
	s.d.	0.2	0.3	0.2	0.3	0.3	0.3
KA	Mean	5.4	5.4	5.2	5.5	5.6	5.6
	s.d.	0.2	0.2	0.1	0.2	0.3	0.3

(*continued on next page*)

Appendix C (*continued*)

Speaker		ELB—Elbow					
		Final Nucleus			Early Nucleus		
		FN-antepenult	FN-penult	FN-final	EN-antepenult	EN-penult	EN-final
KP	Mean	3.2	3.3	3.2	3.3	3.1	3.2
	s.d.	0.2	0.2	0.1	0.2	0.2	0.3
TV	Mean	4.6	4.3	4.3	4.6	4.7	4.5
	s.d.	0.3	0.3	0.3	0.3	0.3	0.4
VP	Mean	5.6	5.7	5.7	6.0	6.0	6.1
	s.d.	0.2	0.2	0.3	0.3	0.3	0.3

Speaker		H—High					
		Final Nucleus			Early Nucleus		
		FN-antepenult	FN-penult	FN-final	EN-antepenult	EN-penult	EN-final
AA	Mean	7.0	6.8	6.6	7.3	7.2	7.3
	s.d.	0.2	0.3	0.3	0.2	0.2	0.2
AH	Mean	5.6	5.5	5.5	5.8	5.8	5.5
	s.d.	0.5	0.7	0.3	0.3	0.3	0.3
DA	Mean	6.4	6.2	6.0	7.0	6.9	6.6
	s.d.	0.4	0.3	0.3	0.4	0.4	0.5
KA	Mean	6.2	6.2	5.7	6.8	6.3	6.8
	s.d.	0.2	0.2	0.2	0.3	1.8	0.3
KP	Mean	4.0	3.8	3.8	4.4	4.2	4.0
	s.d.	0.4	0.3	0.3	0.4	0.4	0.5
TV	Mean	5.9	5.7	5.5	6.2	6.2	5.7
	s.d.	0.3	0.6	0.6	0.3	0.3	0.5
VP	Mean	7.1	6.9	6.7	7.5	7.5	7.5
	s.d.	0.3	0.2	0.3	0.1	0.3	0.3

Speaker		FL—Final Low					
		Final Nucleus			Early Nucleus		
		FN-antepenult	FN-penult	FN-final	EN-antepenult	EN-penult	EN-final
AA	Mean	5.5	5.5	5.7	5.1	5.2	5.7
	s.d.	0.2	0.2	0.3	0.1	0.1	0.3
AH	Mean	3.7	3.9	4.1	3.3	3.3	3.6
	s.d.	0.2	0.4	0.3	0.1	0.1	0.1
DA	Mean	5.3	5.2	5.2	5.1	5.1	5.2
	s.d.	0.3	0.3	0.2	0.1	0.1	0.2
KA	Mean	4.8	4.8	4.7	4.7	4.5	4.8
	s.d.	0.4	0.2	0.1	0.1	1.3	0.2
KP	Mean	2.8	2.8	2.8	2.9	2.8	2.8
	s.d.	0.2	0.1	0.1	0.1	0.1	0.1
TV	Mean	4.1	4.1	3.8	3.9	4.3	3.9
	s.d.	0.4	0.4	0.3	0.2	0.4	0.3
VP	Mean	5.4	5.4	5.3	5.1	5.2	5.4
	s.d.	0.2	0.2	0.2	0.1	0.1	0.2

References

Arvaniti, A., 1998. Phrase accents revisited: comparative evidence from Standard and Cypriot Greek. In: Proc. 4th Internat. Conf. on Spoken Language Processing, Sydney, pp. 2883–2886.

Arvaniti, A., Baltazani, M., 2000. GREEK ToBI: a system for the annotation of Greek speech corpora. In: Proc. Second Internat. Conf. on Language Resources and Evaluation (LREC 2000), Vol. II, European Language Resources Association, Athens, pp. 555–562.

Arvaniti, A., Baltazani, M., 2005. Intonational analysis and prosodic annotation of Greek spoken corpora. In: Jun, S.-A. (Ed.), Prosodic Typology: The Phonology of Intonation and Phrasing. Oxford University Press, Oxford, pp. 84–117.

Arvaniti, A., Garding, G., in press. Dialectal variation in the rising accents of American English. In: Hualde, J., Cole, J. (Eds.), Laboratory Phonology 9, Mouton de Gruyter, Berlin.

Arvaniti, A., Ladd, D.R., 1995. Tonal alignment and the representation of accentual targets. In: Proc. 13th Internat. Congress of Phonetic Sciences, Stockholm, vol. 4, pp. 220–223.

Arvaniti, A., Ladd, D.R., in preparation. Scaling and alignment of pitch targets in Modern Greek wh-question intonation.

Arvaniti, A., Ladd, D.R., Mennen, I., 1998. Stability of tonal alignment: the case of Greek prenuclear accents. J. Phonetics 26, 3–25.

Arvaniti, A., Ladd, D.R., Mennen, I., 2000. What is a starred tone? Evidence from Greek. In: Broe, M., Pierrehumbert, J. (Eds.), Papers in Laboratory Phonology V: Acquisition and the Lexicon. Cambridge University Press, Cambridge, pp. 119–131.

Arvaniti, A., Ladd, D.R., Mennen, I., in press. Tonal association and tonal alignment: evidence from Greek polar questions and contrastive statements. Language and Speech.

Atterer, M., Ladd, D.R., 2004. On the phonetics and phonology of "segmental anchoring" of *F*0: evidence from German. J. Phonetics 32, 177–197.

Baltazani, M., 2002. Quantifier scope and the role of intonation in Greek, Ph.D. dissertation, UCLA.

Baltazani, M., Jun, S.-A., 1999. Focus and topic intonation in Greek. In: Proc. 14th Internat. Congress of Phonetic Sciences, San Francisco, pp. 1305–1308.

Beckman, M.E., Munson, B., Edwards, J., 2004. Vocabulary growth and developmental expansion of types of phonological knowledge. Paper presented at Laboratory Phonology 9, 24–26 June, University of Illinois, Urbana-Champaign. Available from: <http://www.ling.ohio-state.edu/~edwards/LabPhon9Preprint.pdf>.

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

Connell, B., Ladd, D.R., 1990. Aspects of pitch realisation in Yoruba. Phonology 7, 1–30.

Cooper, W.E., Eady, S.J., Mueller, P.R., 1985. Acoustical aspects of contrastive stress in question–answer contexts. J. Acoust. Soc. Amer. 77, 2142–2156.

Crystal, D., 1969. Prosodic Systems and Intonation in English. Cambridge University Press, Cambridge.

D'Imperio, M., 2001. Focus and tonal structure in Neapolitan Italian. Speech Comm. 33, 339–356.

Eady, S.J., Cooper, W.E., 1986. Speech intonation and focus location in matched statements and questions. J. Acoust. Soc. Amer. 80, 402–415.

Frota, S., 2002. Tonal association and target alignment in European Portuguese nuclear falls. In: Gussenhoven, C., Warner, N. (Eds.), Laboratory Phonology, vol. 7. Mouton de Gruyter, Berlin, pp. 387–418.

Gandour, J., Potisuk, S., Dechongkit, S., 1994. Tonal coarticulation in Thai. J. Phonetics 22, 477–492.

Glasberg, B.R., Moore, B., 1990. Derivation of auditory filter shapes from notched-noise data. Hearing Res. 47, 103–138.

Grabe, E., 1998. Pitch accent realization in English and German. J. Phonetics 26, 129–143.

Grabe, E., Gussenhoven, C., Haan, J., Marsi, E., Post, B.J., 1998. Preaccentual pitch and speaker attitude in Dutch. Language Speech 41, 63–85.

Grabe, E., Post, B., Nolan, F., Farrar, K., 2000. Pitch accent realization in four varieties in British English. J. Phonetics 28, 161–185.

Grabe, E., Kochanski, G., Coleman, J., 2005. Quantitative modelling of intonational variation. In: Proc. SASRTLM 2003 (Speech Analysis and Recognition in Technology, Linguistics and Medicine). Available from: <http://www.phon.ox.ac.uk/~esther/Grabe_Kochanski_Coleman.pdf>.

Grice, M., 1995. The Intonation of Interrogation in Palermo Italian: Implications for Intonation Theory. Niemeyer, Tübingen.

Grice, M., Ladd, D.R., Arvaniti, A., 2000. On the place of "phrase accents" in intonational phonology. Phonology 17, 143–185.

Grønnum, N., 1983. Two issues in the prosody of standard Danish. In: Cutler, A., Ladd, D.R. (Eds.), Prosody: Models and Measurements. Springer-Verlag, Berlin, pp. 27–38.

Halliday, M.A.K., 1970. A Course in Spoken English: Intonation. Oxford University Press, Oxford.

Hirschberg, J., Ward, G., 1992. The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise–fall–rise intonation contour in English. J. Phonetics 20, 241–251.

Holton, D., Mackridge, P., Philippaki-Warburton, I., 1997. Greek: A Comprehensive Grammar of the Modern Language. Routledge, London.

Jun, S.-A. (Ed.), 2005. Prosodic Typology: The Phonology of Intonation and Phrasing. Oxford University Press, Oxford.

Kochanski, G., Shih, C., 2003. Prosody modeling with soft templates. Speech Comm. 39, 311–352.

Kochanski, G., Shih, C., Jing, H., 2003a. Hierarchical structure and word strength prediction of Mandarin prosody. Internat. J. Speech Technol. 6, 33–43.

Kochanski, G., Shih, C., Jing, H., 2003b. Quantitative measurement of prosodic strength in Mandarin. Speech Comm. 41, 625–645.

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

Ladd, D.R., Terken. J., 1995. Modelling intra- and inter-speaker pitch range variation. In: Proc. 13th International Congress of Phonetic Sciences, Stockholm, vol. 2, pp. 386–389.

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

Mackridge, P., 1990. The Modern Greek Language. Ekdoseis Pataki, Athens (in Greek).

Max, L., Onghena, P., 1999. Some issues in the statistical analysis of completely randomized and repeated measures designs for

speech, language, and hearing research. J. Speech Language Hearing Res. 42, 261–270.
Menn, L., Boyce, S., 1982. Fundamental frequency and discourse structure. Language Speech 25, 341–383.
Nibert, H.J., 2000. Phonetic and phonological evidence for intermediate phrasing in Spanish intonation, Ph.D. dissertation, University of Illinois at Urbana-Champaign.
O'Connor, J.D., Arnold, G.F., 1973. Intonation of Colloquial English. Longman, London.
Pan, H., in press. Focus and Taiwanese unchecked tones. In: Lee, C., Gordon, M., Büring, D. (Eds.), Topic and Focus, Intonation and Meaning, Kluwer Academic Publishers.
Peng, S., 1997. Production and perception of Taiwanese tones in different tonal and prosodic contexts. J. Phonetics 25, 371–400.
Peterson, G.E., Lehiste, I., 1960. Duration of syllable nuclei in English. J. Acoust. Soc. Amer. 32, 693–703.
Pierrehumbert, J., 1980. The phonology and phonetics of English intonation, Ph.D. dissertation, MIT.
Pierrehumbert, J., Beckman, M., 1988. Japanese Tone Structure. MIT Press, Cambridge, MA.
Pierrehumbert, J., Hirschberg, J., 1990. The meaning of intonation in the interpretation of discourse. In: Cohen, P., Morgan, J., Pollack, M. (Eds.), Intentions in Communication. MIT Press, Cambridge, MA, pp. 271–311.
Prieto, P., 2005. Stability effects in tonal clash contexts in Catalan. J. Phonetics 33, 215–242.
Prieto, P., van Santen, J., Hirschberg, J., 1995. Tonal alignment patterns in Spanish. J. Phonetics 23, 429–451.
Prieto, P., D'Imperio, M., Gili Fivela, B., in press. Prenuclear and nuclear rising accents in Romance: tonal alignment and primary and secondary associations with metrical structure. Language Speech.
Schepman, A., Lickley, R., Ladd, D.R., in press. Effects of vowel length and "right context" on the alignment of Dutch nuclear accents. J. Phonetics.
Silverman, K., Pierrehumbert, J., 1990. The timing of prenuclear high accents in English. In: Kingston, J., Beckman, M.E. (Eds.), Papers in Laboratory Phonology. I: Between the Grammar and Physics of Speech. Cambridge University Press, Cambridge, pp. 72–106.
Steedman, M., 2000. Information structure and the syntax-phonology interface. Linguist. Inq. 31, 649–689.
Thorsen, N., 1979. Lexical stress, emphasis for contrast, and sentence intonation in advanced standard Copenhagen Danish. In: Proc. Ninth Internat. Congress of Phonetic Sciences, Copenhagen, pp. 417–423.
Wang, S., 2002. The prosodic effects on Taiwan Min tones. Lang. Linguist. 3, 839–852.
Ward, G., Hirschberg, J., 1985. Implicating uncertainty: the pragmatics of the fall–rise. Language 61, 747–776.
Waring, H., 1976. The intonation of Modern Greek, Ph.D. dissertation, University of London.
Xu, Y., 2001. Fundamental frequency peak delay in Mandarin. Phonetica 58, 26–52.
Xu, Y., 2005. Speech melody as articulatorily implemented communicative functions. Speech Comm. 46, 220–251.
Xu, Y., Wang, Q.E., 2001. Pitch targets and their realization: evidence from Mandarin Chinese. Speech Comm. 33, 319–337.
Xu, Y., Xu, C.X., 2005. Phonetic realization of focus in English declarative intonation. J. Phonetics 33, 159–197.
Xydas, G., Spiliotopoulos, D., Kouroupetroglou, G., 2004. Modeling prosodic structures in linguistically enriched environments. Lect. Notes Artificial Intell. 3206, 521–528.
Xydas, G., Spiliotopoulos, D., Kouroupetroglou, G., 2005. Modeling improved prosody generation from high-level linguistically annotated corpora. IEICE Trans. Inform. Systems, E88-D:3, 510–518.
Zervas, P., Xydas, G., Fakotakis, N., Kokkinakis, G., Kouroupetroglou, G., 2004. Evaluation of a corpus based tone prediction in mismatched environments for Greek TtS synthesis. In: Proc. 8th Internat. Conf. on Spoken Language Processing, pp. 761–764.