Greek wh-questions and the phonology of intonation*

Amalia Arvaniti

University of California, San Diego

D. Robert Ladd

University of Edinburgh

Abstract

The intonation of Greek wh-questions consists of a rise-fall followed by a low plateau and a final rise. With acoustic data, we show (1) that the exact contour shape depends on the length of the question, and (2) that the position of the first peak and the low plateau depends on the position of the stressed syllables, and shows predictable adjustments in alignment depending on the proximity of adjacent tonal targets. Models that specify the F0 of all syllables, or models that specify F0 by superposing contour shapes for shorter and longer domains, cannot account for such fine-grained lawful variation except by using ad-hoc tonal specifications, which, in turn, do not allow for phonological generalizations about contours applying to utterances of greatly different lengths. In contrast, our findings follow easily from an autosegmental-metrical approach to intonational phonology, according to which melodies may contain long F0 stretches derived by interpolation between specified targets associated with metrically strong syllables and prosodic boundaries.

* The research reported here was supported by the U.K. 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 Ineke Mennen as Research Associate. Special thanks are due to Ineke Mennen for her invaluable assistance with the recording and measurement of the present data. We are most grateful to Mary Beckman for making her algorithmic determination of elbows program available to us, for her generous hospitality to the first author during the initial preparation of this manuscript and for her encouragement. Thanks are also due to our speakers for their cooperation, to the guest editors of this volume for their support, and to Pilar Prieto and an anonymous reviewer for their helpful comments on an earlier version of this paper. Finally we wish to thank Norman Dryden, Mike Bennett, Eddie Dubourg, Cedric Macmartin, and Stewart Smith for technical assistance.

1 Introduction

1.1 Phonetic detail and phonological generalization in the description of intonation

Phonologists have been aware for some time that detailed phonetic data may be relevant to phonological questions. For example, there is a considerable literature, based on instrumental phonetic work, addressing the issue of whether assimilation is a phonological process (viz. the categorical substitution of one phonological element for another) or part of phonetic realization (often modelled as the presence of greater or lesser overlap between two articulatory gestures; e.g. Browman & Goldstein 1990, 1992, Nolan 1992, Zsiga 1994, 1997, Holst and Nolan 1995). The very existence of this literature is based on an agreement that the detailed description of phonological phenomena needs to take account of two distinct kinds of factors, linguistic specifications on the one hand and mechanisms of speech production and perception on the other. In the case of coronal-to-dorsal assimilation in English, for instance, there is now widespread agreement that at some level of description there is a coronal element (segment, feature, gesture, etc.) adjacent to a dorsal element, but that timing adjustments and/or changes in gesture magnitude during their articulation may result in an acoustic outcome that is interpreted by listeners as the deletion of the coronal element (Nolan 1992). In this case, the mechanisms of speech production and perception seem to play a central role in explaining something that appears at first glance to be describable in strictly phonological (abstract, symbolic) terms. Yet there seems to be no basis for privileging one type of explanation over the other; how the two kinds of factors interact is an empirical question that can only be answered on the basis of phonetic data specific to the language and assimilation pattern one wishes to examine (Jun 1996, Zsiga 1997). It is in order to settle such empirical questions that phonologists have increasingly found themselves drawn into the phonetics laboratory.

In the study of intonation, however, the distinction between broadly phonological and broadly phonetic factors has yet to be generally accepted. Intonation is still sometimes assumed to be a phonetic property that can usefully be investigated without any reference to phonology at all, and there is a long tradition of instrumental studies of F0 in which the primary aim is to model – in the strict sense of approximating reality – plots of F0 against time. Some of the quantitative variables of such models are defined in terms of biomechanical and acoustic effects like the

decline of subglottal pressure during the course of an utterance (cf. Maeda 1976) or limitations on the speed with which the speaking apparatus can execute pitch changes (e.g. Sundberg 1979, Xu & Sun 2002). Other variable parameters are loosely attributed to functional effects like 'focus' or 'degree of emphasis', as in the Parallel Encoding and Target Approximation model (PENTA) developed by Xu and colleagues (e.g. Xu 2005, Xu & Xu 2005); still others are simply varied freely in order to optimize the model's fit, as happens with some of the parameters of the command-response model developed by Fujisaki and colleagues (e.g. Fujisaki 1983, 2004, Fujisaki et al. 2005, Gu et al. 2007).

In our view, this purely phonetic approach merely sidesteps questions of intonational phonology; it cannot avoid them altogether. Consider declination, the tendency of F0 to decline gradually over the course of a phrase or utterance, which has been a topic of research for at least forty years since the term was coined by Cohen & 't Hart (1967). In order to describe declination, Fujisaki's model superposes accent commands of variable height and duration on a declining phrase component, which is quantitatively modelled as the response of the F0 production mechanism to a precisely localized phrase command. In this model, that is, declination is conceived of as an automatic consequence of the way the speech production mechanism works. The accent commands in a single prosodic phrase are realized with their accentual F0 peaks scaled progressively lower even if they are all 'intended' as equivalent and are therefore quantitatively specified with the same underlying 'height'. This, though, entails the linguistic hypothesis that the local accent commands do not contribute systematically to declination. If the decline in the height of the F0 peaks in a series can be entirely accounted for by some physical factor in speech production, there is no need to assume that there is anything linguistically systematic about the height of successive accent commands.

This view was challenged by Pierrehumbert (1980) and subsequent work (e.g. Liberman & Pierrehumbert 1984, Beckman & Pierrehumbert 1986), which introduced the notion of downstep into the description of declination data. Pierrehumbert argued that declination did not result from the automatic workings of the speech production system, but from the repeated occurrence of downstep at successive accented syllables – an explicitly phonological effect that (in Fujisaki's terms) should be specified quantitatively in the size of the accent commands. Like Fujisaki, that is, Pierrehumbert assumed that an adequate theory should account for measurable intonational

phenomena, but she argued that declination in particular is a consequence of how specific phonological sequences are realized and should not be assumed to be present as a kind of automatic backdrop. In fact, subsequent research (e.g. Grabe 1998, Arvaniti 2003, Arvaniti 2007b) suggests that there are probably both linguistic and biomechanical factors involved in declination, though the details are far from clear. But what is important in the present context is that the idea of modelling declination in terms of repeated downstep is part of a general theory that explicitly acknowledges the need for a phonological level of description in any adequate model of intonation.

In their own terms, phonetic models like Fujisaki's or Xu's are often extremely successful. The most obvious criterion for evaluating a model is how accurately it reproduces or generates the phonetic detail of F0. This accuracy can be assessed by using standard mathematical characterizations of model fit, such as correlation or root-mean-square (RMS) error; with the advent of speech synthesis by rule, a model can also be assessed by using it to generate actual synthetic utterances and evaluating how natural they sound. By either measure there is no doubt that models like Fujisaki's or Xu's score highly; in addition, such models often incorporate an up-to-date understanding of biomechanical and acoustic influences on F0, which seems to provide further evidence of their scientific adequacy (Fujisaki 2004, Xu & Sun 2002, Xu 2005). Moreover, because the kinds of linguistic distinctions signalled by intonation can be readily described by ordinary notions (such as emphasis or questioning) that do not appear to require theoretical elaboration, explicit recourse to phonology can be portrayed as unnecessary or even faintly absurd (e.g. Xu & Xu 2005: 161ff.). Indeed, partly because F0 contours are so simple, it is possible, as pointed out by Kochanski & Shih (2003), to model them in great detail without providing any insight into their linguistic aspects.

Yet while accurate generation of phonetic detail is a sort of irreducible minimum requirement for any model, and while physical and biological plausibility is ultimately an essential consideration, we contend that any complete theory of intonation also needs an abstract description that accounts for the linguistic aspects of the system and allows for predictions and generalizations based on this description. The goal of this paper is to back up this contention with experimental evidence. While it is certainly true that the value of positing phonological abstractions is not readily quantified in terms of RMS error, we maintain that the issue of whether to invoke an

explicit conception of phonology in describing intonation is nevertheless an empirical one, which can ultimately be assessed in terms of our ability to model phonetic data while taking a broader context into account. Specifically, we aim to demonstrate that if we assume the existence of intonational phonology we acquire a superior ability to *generalize* – to make accurate empirical predictions about phonetic form across an objectively greater range of cases than is possible if one's goal is simply to approximate F0 contours.

1.2 Sparse tonal specification

A special problem for modelling intonation comes from the fact that a given melody can be applied to utterances of hugely varying lengths; e.g. a wh-question may be as short as *Where?* but as long as *Where did you say you were going for spring break?*. We certainly want to be able to model the contours of these two questions as physical events, and account for locally conditioned phonetic detail, but at the same time we need to model them in terms of specifiable shared properties that embody their functional equivalence. In other words, we want to be able to recognize that the pitch contours on pairs of sentences like *Where?* and *Where did you say you were going for spring break?* are 'the same' at some level of description, like two different tokens of the same phoneme.

This property of intonation presents unfamiliar problems for phonetic modelling. The idea that functional equivalence normally involves shared phonetic properties is more or less taken for granted in segmental phonology, and causes no great difficulty. Phonological abstractions such as /m/ or /i/ are phonetically quite concrete, and their realizations vary within a relatively small range. Much of the contextual variation in the phonetic realization of such phonological abstractions depends on very local factors – in particular, the nature of adjacent segments – and at least part of it can be explained on the basis of acoustic and biomechanical properties of the vocal tract. There are obviously language-specific ways in which even these physical constraints are manifested (cf. e.g. Cohn 1993), as well as effects that cannot be explained in physical terms, but they are mostly fairly straightforward (e.g. positional neutralization and allophony) and, again, quite local. In intonation, by contrast, the length of the domain to which a melody is applied represents an important source of conspicuous contextual variation that is unlikely to have a purely physical explanation.

The influential autosegmental-metrical (henceforth AM) theory of intonational phonology, based on the work of Bruce (1977), Pierrehumbert (1980) and others, has generally dealt with this problem by adopting what we might call a target-and-interpolation approach to phonetic modelling. Specifically, AM assumes that intonation contours consist phonologically of strings of High and Low tones, which are phonetically realized as tonal targets, i.e. as specific points in the F0 contour, such as local minima and maxima. (This should not be taken to mean that local minima and maxima are equivalent to phonological tones, but only that the realization of phonological tones gives rise to minima and maxima; this is comparable to saying that the second formant maximum in the word *Maya* is an important, easily measurable aspect of the phonetic realization of the phoneme /j/, but it is not itself the phoneme /j/ and nor is it necessarily even the main manifestation of this phoneme.) Given a string of such targets, a contour can be modelled by describing the phonetic properties of the targets and then interpolating line segments (not necessarily straight lines) from one target to the next. In the version of this approach pioneered by Bruce (1977), the principal phonetic dimensions characterizing the target points are their F0 level (scaling) and their temporal coordination with the segmental string (alignment), which reflect various kinds of linguistic specifications (accent, tone, relative prominence, discourse effects, etc.). Scaling and alignment can also be affected by non-linguistic factors related to speech production constraints, such as undershooting of targets that involve too-rapid pitch changes (Bruce 1977: chapter 5). In short, the theory takes it for granted that any adequate phonetic model of pitch contours – like any model of segmental phonetic detail – must involve both factors related to speech production and linguistic specifications. Unlike Fujisaki, Xu, and others, AM researchers have devoted considerable attention to the question of what these linguistic specifications might be.

The central phonological notion to emerge from the use of target-and-interpolation models in AM theory is that of sparse tonal specification. This idea is implicit in Bruce (1977), and also informs Fujisaki's model in the sense that phrase and accent commands occur only where needed to model pitch movements. But it was first made into an explicit phonological claim by Pierrehumbert and Beckman (1988: 13ff.), who argued that there is no need to specify the tone of every syllable in a Japanese unaccented word. They proposed instead that there is only one high target point in the word, associated with the second mora, and showed that the pitch of any subsequent syllables can be determined by simple interpolation between that single high target

and the low target at the beginning of the following word. They drew the implication for phonology that languages might allow significant mismatches between the number of syllables and the number of intonational targets, and that syllables can be phonologically unspecified for pitch. Obviously, any syllable with voicing has F0, but the factors that determine F0 may involve nothing more than interpolation from an earlier pitch target to a later one. Similar notions are of course widely accepted in segmental phonology and phonetics as a way of thinking about intrusive stops in words like *tense* or *prince* (cf. Ohala 1974, Browman & Goldstein 1990, Gick 1999), transitional vowels preceding coda liquids (cf. Gick & Wilson 2006), and so on.

Most AM work on intonation accepts the idea of sparse tonal specification, treating the utterance contour as a string of tonal events (pitch accents, boundary tones, etc.) that may be associated with the syllable string in a variety of ways depending on a number of structural, metrical and pragmatic factors. Not every syllable has to have a specification for pitch; conversely, pitch targets may also occur clustered in twos or threes on a single syllable. This makes it possible to describe the contours on *Where?* and *Where did you say you were going for spring break?* as phonologically identical while still modelling the phonetic detail that results from the difference in utterance length: the one-syllable utterance has the same intonational targets as the longer utterance, but the targets may be realized differently because they are crowded together in the former and widely spaced in the latter. In our view, sparse tonal specification is the key to combining accurate phonetic modelling with the expression of linguistic equivalence of intonation contours of markedly different lengths.

The idea of sparse tonal specification has been directly challenged by Xu (2005: 233; cf. also Xu & Wang 2001, Xu & Xu 2005). Basing himself largely on a model of F0 in Mandarin, Xu claims that pitch contours must be specified syllable-by-syllable in all languages. That is, he proposes that all syllables have underlying pitch specifications, not only in languages like Mandarin in which such specifications are linguistically contrastive and determined in the lexicon, but also in languages like English or Greek in which they clearly aren't. Xu justifies this proposal by emphasizing the phonetic accuracy of his model and its compatibility with research on biomechanical constraints on pitch production (Xu 2005). He does not consider the problem of linguistic equivalence between contours of different lengths. Our main goal in this paper is to show that sparse tonal specification can be empirically evaluated on the basis of detailed

phonetic predictions about what happens to linguistically equivalent intonation patterns when they are applied to sentences of different lengths. We show that phonetic models that do not allow for sparse tonal specification are in principle incapable of dealing with the realization of certain intonation contours, particularly those of short utterances.

1.3 Greek wh-question intonation

Our experimental study is based on the intonation of wh-questions in Greek.[1] In these questions the wh-word is normally utterance-initial and is felt by native speakers to be the most prominent word of the utterance (for a discussion see Baltazani 2002). Thus, the pitch accent of this word should be considered the nuclear accent of the question (unlike in English and most Western European languages, in which the nuclear accent in wh-questions normally goes on the rightmost content word, as in declaratives; Ladd 2008: 224ff). Examples of Greek wh-questions are shown in (1)-(3); the contours of these questions are illustrated in Figure 1.

(1) [ˈpu ˈzi]

where live.3SG

'where does s/he live?'

(2) [ˈpu ˈmenune]

where stay.3PL

'where are they staying?'

(3) [apoˈpu ˈmilaje tu ˈmenelu]

from where speak.3SG.PAST the Menelos.GEN

'Where was s/he speaking to Menelos from?'

[1] The reference to 'the intonation of wh-questions' should be interpreted only as a convenient way to refer to the melody under investigation; it is now established that the same melody is also used with negative declaratives (Baltazani 2002, 2006, Arvaniti & Baltazani 2005).

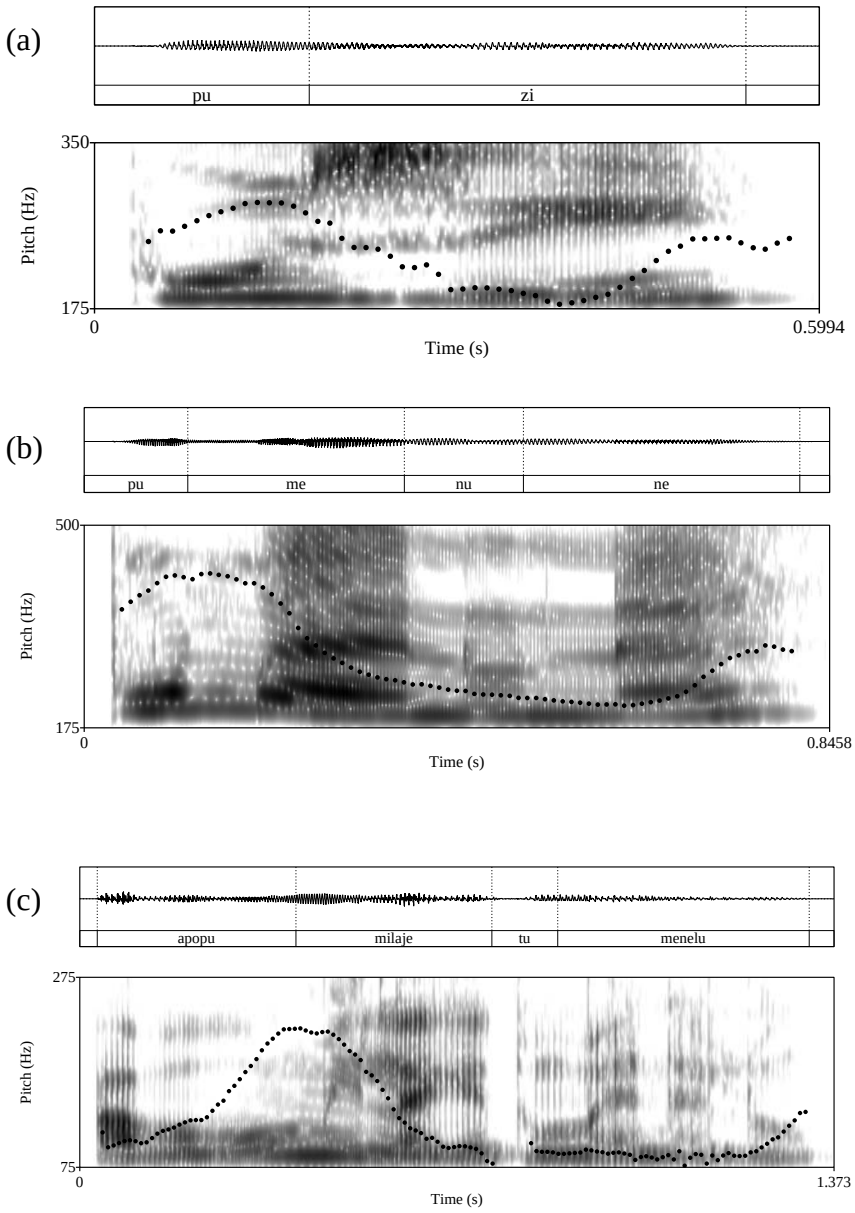

Figure 1: Waveforms, spectrograms and F0 contours of [ˈpu ˈzi] ‘where does s/he live?’ (speaker AA) in panel (a), [ˈpu ˈmenune] ‘where are they staying?’ (speaker DA) in panel (b), and [apoˈpu ˈmilaje tu ˈmenelu] ‘Where was s/he speaking to Menelos from?’ (speaker KP) in panel (c).

As can be deduced from Figure 1, the simplest impressionistic description of Greek wh-question intonation is that it is a fall-rise. However, as the contours in this figure amply illustrate, this description must be qualified depending on the length of the wh-word and of the question overall. As shown in panels (a) and (b), when the wh-word is monosyllabic, the contour starts with a shallow rise; if it consists of more syllables, however, as in panel (c), then the contour starts with a rise from a low F0 point; in both cases, the peak roughly coincides in time with the stressed vowel of the wh-word. In addition, while the contour of the shortest question (panel a) shows a rather brief trough, in the longer questions (panels b and c) the early peak is followed by a long low plateau. Finally, the rise at the very end of all three examples is the most typical pattern, though wh-questions may also end low (Arvaniti & Baltazani 2005).

One obvious measure of the adequacy of any phonetic model is how well it can account for such details. At the same time, however, the phonetic model should be reckoned superior if it is linked to a linguistic description that clearly treats contours like those shown in Figure 1 as predictable variants of the same basic intonational type. We show that this two-pronged task is successfully accomplished by a phonetic model based on an AM-style phonological description. We also show that the task is difficult for any model, like Xu's PENTA, that attempts to account for utterance contours by specifying the phonetic detail of each syllable, or for any model, like Fujisaki's command-response model, that does not recognize the lawful predictability of the scaling and alignment of local pitch movements and thus allows the parameters governing them to vary arbitrarily. The modelling of the instrumental phonetic data from Greek wh-questions therefore constitutes evidence against the implicit phonological assumptions of such models, and in favour of an explicitly phonological approach, thus further developing the line of argument we have presented in earlier work (Arvaniti et al. 2006a, 2006b, Arvaniti, 2007a). It also contributes to our overall aim of demonstrating that phonetic data can be used to shed light on phonological questions not only with respect to segmental phenomena, but in the realm of intonation as well.

2 The instrumental study

2.1 Background and experimental questions

An autosegmental representation of the Greek wh-question melody would be as follows: the peak on the wh-word would be analyzed as a high or rising pitch accent (notated H* or L+H* or

L*+H; see further below); the fall and the low plateau would be attributed to the presence of a low phrase accent (notated L-); and the final rise would be the reflex of a high boundary tone (notated H%). On the basis of our preliminary impressionistic observations and prior AM research, we expected to find evidence for the following targets: a low and a high target manifesting the pitch accent on the wh-word (Arvaniti & Baltazani 2005); two low targets – the beginning and end of the low plateau – manifesting the low phrase accent (Grice et al. 2000); and a final high target manifesting the boundary tone (Arvaniti 2001; Arvaniti & Baltazani 2005). In the following paragraphs we present these expectations in more detail.

Pitch accent***:*** We expected that the pitch accent on the stressed syllable of the wh-word would display the same phonetic behaviour as the 'prenuclear' (non-final) pitch accent investigated in studies of Greek declaratives (Arvaniti & Ladd, 1995; Arvaniti et al. 1998, 2000). In this earlier work, these accents are described as bitonal L+H accents[2] on the basis of evidence showing stable alignment and scaling of two targets: a low (L) target aligned just at the end of the syllable preceding the accented syllable, and a high (H) target aligned on average 10-20 ms into the vowel of the syllable following the accented syllable. This alignment of the L and H targets entails the presence of unstressed syllables before and after the nuclear syllable. This condition is not always met, as shown in examples (1) and (2), and our study was designed to investigate what happens in such cases. We predicted that if the wh-word begins with a stressed syllable, the L target would be truncated and the utterance-initial F0 level would be higher than if the wh-word begins with unstressed syllables. In effect, the actual starting F0 level could be predicted on the basis of a virtual target preceding the onset of phonation (an idea first suggested in Bruce 1977).

In PENTA, the rise that AM attributes to a pitch accent would be viewed as the reflex of focus and would be modelled as high or rising F0 on the stressed syllable of the wh-word. Predictions based on virtual targets are in principle impossible, because pitch specifications attach to

[2] The prenuclear accent of Greek is described as L*+H in Arvaniti & Ladd (1995), and as L+H* in Arvaniti et al. (1998). The reasons for this difference are thoroughly discussed in Arvaniti et al. (2000). More recent results (Arvaniti et al. 2006b) favour the L*+H analysis, though the details are not relevant here; for further discussion see Arvaniti et al. (2000; 2006b), Arvaniti & Baltazani 2005, and §3.2 below.

syllables. In Fujisaki's model, initial truncation can readily be modelled, but only because the precise timing of phrase commands and accents commands relative to syllables is a freely variable parameter in his model. No principled basis is provided for expecting initial truncation to occur.

Phrase accent and low plateau: Grice et al. (2000) analyze the low plateau of Greek wh-questions as the reflex of a low phrase accent that exhibits phonological properties of both a boundary tone and an ordinary pitch accent. The key evidence for this claim is that, given enough segmental material after the nuclear accent, the phrase accent will seek to associate with metrically prominent (e.g. lexically stressed) syllables rather than being manifested phonetically at the edge of the phrase. Our goal was to examine this stress-seeking behaviour by testing the prediction that the beginning and end of the low plateau, which we take to be the manifestation of the L-phrase accent, are affected by the location of lexically stressed syllables following the wh-word.

Once again, such regularities would be difficult to handle in purely phonetic models. In Fujisaki's model, a lengthy negative accent command or a negative phrase command would be necessary to account for the low plateau, a modelling necessity not predicted for Greek (Fujisaki, Ohno & Yagi 1997; Fujisaki 2004). PENTA would attribute the low plateau to the presence of focus on the wh-word and would also predict a steep fall from the peak to the plateau as a way of highlighting the focused item (Xu 2005, Xu & Xu 2005).

Effects of tonal crowding and of sentence length: The primary motivation for this manipulation was to determine whether targets are undershot and/or displaced when they are subject to tonal crowding (i.e. when two or more tones are associated with the same tone-bearing unit or with adjacent units). The effects of tonal crowding have been documented in several studies (e.g. Silverman & Pierrehumbert, 1990; Prieto et al. 1995; Prieto 2005) and can be inferred from the fact that tonal realization shows adjustments when crowding is present, but remains stable once tones are more than two syllables apart (Arvaniti & Ladd 1995; Arvaniti et al. 1998, 2000, 2006a, 2006b). Past findings suggest that (a) certain targets tend to be undershot in scaling when affected by tonal crowding (e.g. the L tone of the prenuclear pitch accent is scaled higher when immediately preceded by another such accent; Arvaniti et al. 1998, 2000), but (b) most targets tend to show adjustments in alignment rather than scaling (e.g. the L of the L* nuclear accent in

yes-no questions is aligned considerably earlier when it occurs on a phrase-final syllable than otherwise; Arvaniti et al. 2006a). In the present study, we expected that tonal crowding effects would be especially noticeable when a lexically stressed syllable immediately follows the wh-word, as in examples (1)-(3) above, and that the effects would be more striking in short questions, such as (1) and (2), in which pressure on some targets is exerted from both preceding and upcoming targets: this should affect especially the realization of the low plateau between the accentual peak on the wh-word and the final rise.

Note once again that tonal crowding effects cannot easily be accommodated in either Fujisaki's or Xu's models except by ad hoc adjustments of parameters, such as the height and duration of accent commands (command-response model), or the strength and identity of targets (PENTA). Note also that PENTA assumes that only carry-over adjustments are possible, and thus does not predict any anticipatory effects.

A second reason for the manipulation of question length was to determine whether this would have any effect on target scaling. Specifically, some declination models predict that F0 is affected by utterance length, so that in a longer sentence F0 starts higher (e.g. Cooper and Sorensen 1981) or ends lower (e.g. Fujisaki 1983). However, the evidence so far is inconclusive (for a brief review see Ladd & Johnson 1987). Since there are a number of ways that such effects might be incorporated into a phonetic model, this issue is not central to our main point; our investigation of this question was purely exploratory, and we report the results here primarily for completeness.

2.2 Method

Our study was based on the acoustic analysis of speech materials read aloud under laboratory conditions by four native speakers of Greek (see Appendix for details). The materials consisted of wh-questions embedded in mini-dialogues (so as to make the speakers' task as natural as possible); e.g.

(4) [mu ˈleɣane pos θa ˈðun ti baˈrelasi ˈfetos]

me say.3PL.PAST that FUT see3PL.SUBJ the parade this-year

'They were telling me that they are going to watch the parade this year.'

[apoˈpu ˈlene na ti ˈðun]

from where say.2PL to it see.3PL.SUBJ

‘Where are they thinking of seeing it from?’

The materials were based on controlled variation of the following four parameters, in keeping with our experimental goals.

(i) The length of the question: in ‘long’ questions, the wh-word was followed by two content words as in (4), while in ‘short’ wh-questions, the wh-word was followed by only one content word, as in (5); e.g.

(5) [apoˈpu na mu miˈlaɣane]

from where to me speak. 3PL.PAST.SUBJ

‘Where could they have been talking to me from?

The purpose of this manipulation was to examine the extent to which the nuclear H and following L targets would show scaling and alignment adjustments to the greater tonal crowding present in the short questions, and to look for evidence that the scaling would be influenced by the greater amount of declination that some models would predict with the long questions.

(ii) The number of unstressed syllables between the stressed syllable of the wh-word and the first postnuclear stressed syllable (henceforth *interstress interval*): this number was zero, two or three unstressed syllables; for instance, in example (5) it is three syllables, but in example (4) it is zero. This variable was manipulated in order to test the hypothesis that the beginning of the low plateau (the assumed reflex of the L- phrase accent) seeks to align with the first postnuclear stressed syllable. If this hypothesis is correct, both the nuclear H and following L should be ‘trying’ to occur on the same syllable when interstress interval is zero, giving rise to tonal crowding effects.

(iii) The distance of the last stressed syllable from the end of the question: Lexical stress on the last word in the question fell on the ultima, penult or antepenult; for instance, in example (4), the last word is stressed on the ultima, while in (5), it is stressed on the antepenult. (In short questions the last stressed syllable was also the first postnuclear stressed syllable.) The aim of

this manipulation was to test the hypothesis that the end of the low plateau (again, assumed to be a target reflecting the L- phrase accent) seeks to align with the last stressed syllable of the question.

(iv) The length of the wh-word: the wh-word was either [ˈpu] 'where' or [apoˈpu] 'from where'.[3] The purpose of this manipulation was to see if the F0 level at the vowel onset of [ˈpu] would be higher than the F0 level at the onset of [a] in [apoˈpu] because of the hypothesized virtual target.

Our analyses were based on the scaling and alignment of the following hypothesised targets (see Figure 2): (i) Initial Low (IL), defined as the lowest non-spurious F0 point at the onset of the utterance; (ii) Nuclear Low (NL), defined as the F0 level at the onset of the nuclear vowel of the wh-word (for [pu], IL and NL are the same point); (iii) the first peak (NH for nuclear high), defined as the highest F0 point in the contour (typically located in the vicinity of the wh-word's stressed syllable); (iv) the first elbow (L1), defined as the point that showed a clear change in slope between the fall after the nuclear peak and the low plateau; (v) the second elbow (L2), defined as the point that showed a clear upward inflection between the low plateau and the utterance-final rise; (vi) the final H (FH), defined as the highest non-spurious F0 value at the end of the utterance-final rise. L2 and FH were not measured on utterances with no final rise.

All F0 data were converted to ERB units (Glasberg & Moore 1990) before being used for statistical analysis, since this arguably gives a better approximation of perceptual distance between F0 levels and allows more easily for the comparison of male and female data.

In addition to measuring F0 at the target points just listed, we computed the following alignment measurements (positive measurements indicate that the tonal target appeared after the segmental landmark from which its alignment was measured; negative measurements indicate that the tonal target occurred before the relevant segmental landmark): (i) the temporal interval between the onset of the nuclear vowel (NV) and the nuclear H; (ii) the interval between the onset of the first postnuclear vowel (PNV) and Nuclear H; (iii) the interval between the onset of the first

[3] Although [apo] 'from' is spelt with an accent in Greek, it is rarely stressed in speech, and certainly not in this case where it forms one phonological word with [pu] 'where' (see Arvaniti & Baltazani 2005; Arvaniti 2007c).

postnuclear stressed vowel (PNSV) and the first elbow (note that PNV and PNSV refer to the same point in utterances with an interstress interval of zero); (iv) the interval between the onset of the utterance-final vowel (FV) and the second elbow. This measurement was different from what we had anticipated because preliminary inspection of the data showed that L2 was aligned closest to the vowel of the last syllable in the question, independently of whether this vowel was stressed or not.

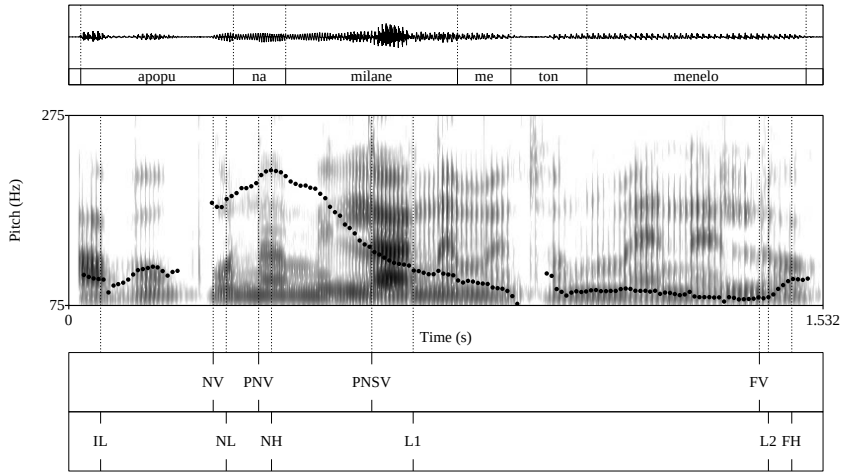

Figure 2: Waveform, spectrogram and F0 contour of [apoˈpu na miˈlane me to ˈmenelo] ‘Where could they be speaking to Menelos from?’ from the data of speaker KP, illustrating the measurements taken on the F0 contour and relevant segmental onsets.

The scaling and alignment measurements were statistically analyzed by means of repeated-measures analyses of variance (ANOVAs), which involved one or more of the following factors: QUESTION LENGTH (short or long question); INTERSTRESS INTERVAL (zero, two or three unstressed syllables); FINAL STRESS (final, penultimate or antepenultimate stress on the last word); WH-WORD (long or short wh-word); TONE TYPE (NH, FH, L1, L2). Possible differences among levels of one factor and expected interactions were explored using planned comparisons; post-hoc Tukey HSD tests were used to explore unexpected interactions. All reported differences are significant at $p < 0.05$. Unless otherwise stated, statistical results refer to comparisons between sets B and C in the Appendix.

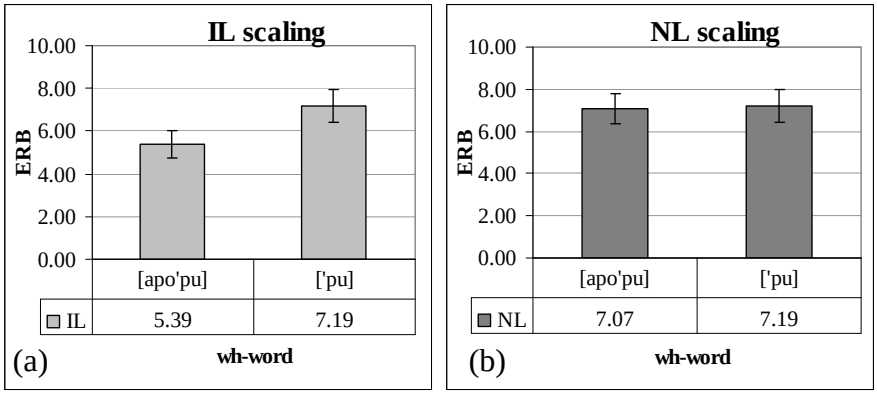

Figure 3: On the left, mean scaling and standard errors for Initial Low (IL); on the right, mean scaling and standard errors for Nuclear Low (NL, the F0 at the onset of the nuclear vowel of the wh-word).

2.3 Results

2.3.1 Initial L and Nuclear L

The hypothesis that the initial L (IL) is truncated (i.e. becomes a virtual target) in questions with short wh-words was tested by means of an ANOVA with WH-WORD as the repeated-measures factor. The results, illustrated in Figure 3(a), showed that, as expected, IL is lower when the wh-word is [apoˈpu] 'where from' than when it is [ˈpu] 'where' [$F(1, 3) = 20.5$]. The hypothesis was further supported by comparing the scaling of the nuclear L (the F0 level at the onset of the stressed vowel of [pu]) in short and long wh-words. In this case, the data showed that there was no difference between questions with short and long wh-words, i.e. the F0 at the onset of their stressed vowel was the same in both cases (see Figure 3(b)).

2.3.2 Scaling and alignment of nuclear H

The scaling of nuclear H (NH) was investigated by means of an ANOVA with QUESTION LENGTH and INTERSTRESS INTERVAL as repeated-measures factors. Neither factor affected NH scaling and there was no interaction (see Table 1). This strongly suggests that the scaling of the peak is a target that speakers aim to achieve accurately regardless of other phonetic pressures, a point we return to in the discussion.

Table 1. Mean scaling in ERB and standard errors (in brackets) for Nuclear H and L1.

	NH scaling		L1 scaling	
interstress interval	long question	short question	long question	short question
0 syllables	7.9 (0.8)	7.7 (0.7)	5.68 (0.7)	5.27 (0.7)
2 syllables	8.1 (0.8)	7.9 (0.7)	5.97 (0.8)	5.56 (0.8)
3 syllables	8.1 (0.8)	8.0 (0.8)	5.97 (0.7)	5.68 (0.7)

In contrast, the alignment of NH with respect to the onset of the nuclear vowel showed effects of tonal crowding, in that alignment was affected by INTERSTRESS INTERVAL [$F(2, 6) = 27.9$]. Specifically, NH occurred earlier when the interstress interval was zero syllables than in the other two conditions, between which there was no difference [for 0 vs. 2 syllables, $F(1, 3) = 42.06$; for 0 vs. 3 syllables, $F(1, 3) = 23.02$]. These effects, illustrated in Figure 4(a), suggest that the alignment of the peak is stable except when there is extreme tonal crowding.

As can be seen in Figure 4(a), when there is no stress clash, NH appears 114-120 ms after the onset of the nuclear vowel. Since the nuclear vowel is 59 ms in duration on average, this result suggests that NH could be aligned close to the onset of the first postnuclear vowel. In order to test for this possibility, we examined the distance of NH from the onset of the first postnuclear vowel by means of an ANOVA with QUESTION LENGTH and INTERSTRESS INTERVAL as repeated-measures factors. The results for this way of measuring the alignment of NH corroborate those of measuring NH from the onset of the nuclear vowel. Specifically, the data showed an interaction between QUESTION LENGTH and INTERSTRESS INTERVAL [$F(2, 6) = 9.2$]. Tukey HSD tests showed that in both short and long questions the alignment of NH was primarily affected by the length of the interstress interval: NH aligned earlier when the interstress interval was zero than in the other two conditions, between which there was no difference. This in effect means that NH aligned before the onset of the postnuclear vowel when interstress interval was zero (see the negative values in Figure 4(b)), but around the onset of this vowel when the interstress interval was longer. The effect of the zero interstress interval was more pronounced in short questions in which tonal crowding is more extreme. Recall that in the short questions with stress clash, the pressure on NH is not only local, from the upcoming L1, but also exerted by the following targets, L2 and FH: in these short sentences, all four targets (NH, L1, L2 and FH) must be

realized within a segmental stretch that is just one or three syllables long. Because of this added pressure, when the interstress interval was zero, the peak appeared much earlier in short than in long questions, and in fact aligned with the nuclear vowel itself; in contrast, in long questions, in which the pressure on NH comes only from the following L1, the peak co-occurred with the onset consonant of the postnuclear syllable. There were no such differences in alignment between short and long questions when the interstress interval increased to two or three syllables.

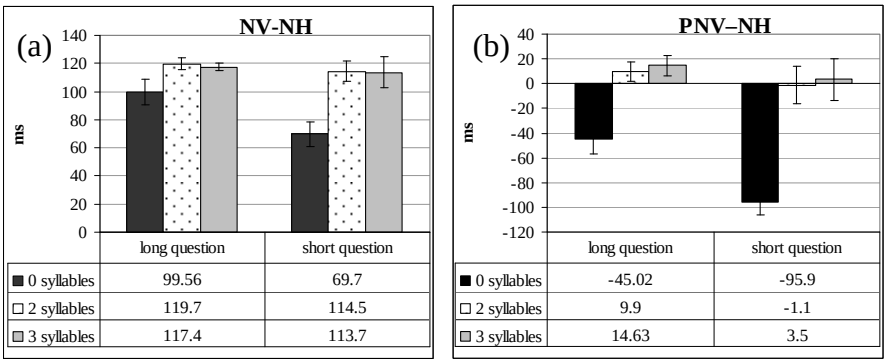

Figure 4: On the left, mean alignment and standard errors for NH with respect to the onset of the nuclear vowel of the wh-word; on the right, mean alignment and standard errors for NH with respect to the first postnuclear vowel; in both panels, data are presented as a function of QUESTION LENGTH and INTERSTRESS INTERVAL.

2.3.3 The first elbow: scaling, alignment and F0 slope

The scaling of the first elbow (L1) was investigated by means of an ANOVA with QUESTION LENGTH and INTERSTRESS INTERVAL as repeated-measures factors. The results showed that the scaling of L1 was not affected by either factor and there was no interaction between the two (see Table 1).

On the other hand, the alignment of L1 from the onset of the first postnuclear stressed vowel showed both effects of QUESTION LENGTH [$F(1, 3) = 40.5$] and of INTERSTRESS INTERVAL [$F(2, 6) = 91.7$]. As can be seen in Figure 5(a), the alignment of L1 was clearly affected by the position of the postnuclear stressed vowel. When there was no tonal crowding – that is, when the interstress interval was two or three syllables – L1 appeared before the onset of the postnuclear stressed vowel, aligning either with or slightly before the onset of the postnuclear stressed

syllable. When there was tonal crowding, however – that is, when the postnuclear stressed syllable immediately followed the wh-word's stressed syllable – L1 co-occurred with the postnuclear vowel itself in short questions or aligned after it in long questions [for 0 vs. 2 syllables, $F(1, 3) = 59.4$; for 0 vs. 3 syllables, $F(1, 3) = 111.8$; for 2 vs. 3 syllables, $F(1, 3) = 142$]. In other words, L1 late alignment was observed only in the stress clash condition, and in this case, it was more extensive in long questions in which pressure on L1 was exercised mostly by the preceding NH. Under the same circumstances, in short questions, in which pressure was also exercised by the upcoming targets, the alignment of L1 was earlier than in long questions. Despite the differences, it is important to note that in most cases, L1 temporally coordinated with the postnuclear stressed syllable in such a way that the vowel of that syllable had low F0.

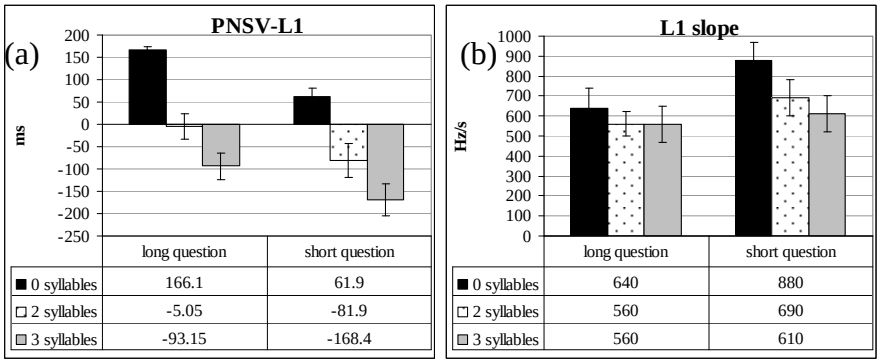

Figure 5: On the left, mean alignment and standard errors for the first elbow (L1) with respect to the onset of the first postnuclear stressed vowel; on the right, mean slope (in Hz/s) and standard errors for the fall from NH to L1; in both cases, results are presented as a function of QUESTION LENGTH and INTERSTRESS INTERVAL.

The effect of the different amount of pressure that other targets exercised on L1 is also reflected in the F0 slope from NH to L1 (calculated on the basis of Hz values, i.e. Hz/s). An ANOVA with QUESTION LENGTH and INTERSTRESS INTERVAL showed effects of both factors and also interaction between the two [for QUESTION LENGTH, $F(1, 3) = 10.4$; for INTERSTRESS INTERVAL, $F(2, 6) = 45.6$; for QUESTION LENGTH × INTERSTRESS INTERVAL, $F(2, 6) = 5.4$]. Specifically, when the interstress interval was two or three syllables, there was no statistically significant difference in slope between short and long questions. The only difference between them pertained to questions with zero interstress interval: in this case, the slope was steeper in short than in long questions

[F(1, 3) = 10.8], as was to be expected by the fact that in short questions L1 has to be reached earlier so that there is room for the following targets (recall that a similar effect, i.e. greater pressure from upcoming targets, was observed in NH and resulted in its earlier than canonical alignment, §2.3.2). In addition, in the short questions only, there was an overall effect of INTERSTRESS INTERVAL: as can be seen in Figure 5(b), the slope became steeper as the interstress interval decreased [for 0 vs. 2 syllables, F(1, 3) = 67.7; for 0 vs. 3 syllables, F(1, 3) = 83.01; for 2 vs. 3 syllables, F(1, 3) = 11.3].

2.3.4 Scaling and alignment of the second elbow

The scaling of the second elbow (L2) was investigated by means of a three-way ANOVA, with QUESTION LENGTH, INTERSTRESS INTERVAL and FINAL STRESS as repeated-measures factors. QUESTION LENGTH did not affect the scaling of L2, but INTERSTRESS INTERVAL and FINAL STRESS did [F(2, 6) = 7.2, and F(1, 3) = 55.6 respectively]. For FINAL STRESS, the results showed higher L2 scaling when the last word was stressed on the final syllable than when it was stressed on the antepenult (see Figure 6). In addition, there was interaction between QUESTION LENGTH and INTERSTRESS INTERVAL [F(2, 6) = 6.9], such that INTERSTRESS INTERVAL did not affect the scaling of L2 in long questions (Figure 6a) but did have an effect in short questions (Figure 6b). Specifically, in short questions, L2 was scaled somewhat higher when interstress interval was zero, i.e. when there was more tonal crowding, than in the other two conditions, between which there was no difference [for 0 vs. 2 syllables, F(1, 3) = 25.2; for 0 vs. 3 syllables, F(1, 3) = 9.3, $p < 0.055$, a result that narrowly missed significance].

The alignment of L2 with respect to the final vowel in the question was investigated by means of a three-way ANOVA, with QUESTION LENGTH, INTERSTRESS INTERVAL and FINAL STRESS as repeated-measures factors. The analysis showed only effects of INTERSTRESS INTERVAL [F(1, 3) = 11.32] and FINAL STRESS [F(1, 3) = 95.1]. Specifically, as can be seen in Figure 7(a), L2 aligned earlier in long wh-questions than in short questions (in which tonal crowding is more extreme). In addition, in both short and long questions, L2 occurred after the onset of the final vowel, when this vowel was stressed, but slightly before it, when stress was on the antepenult; in the latter case, L2 co-occurred with the consonant of the question's last syllable.

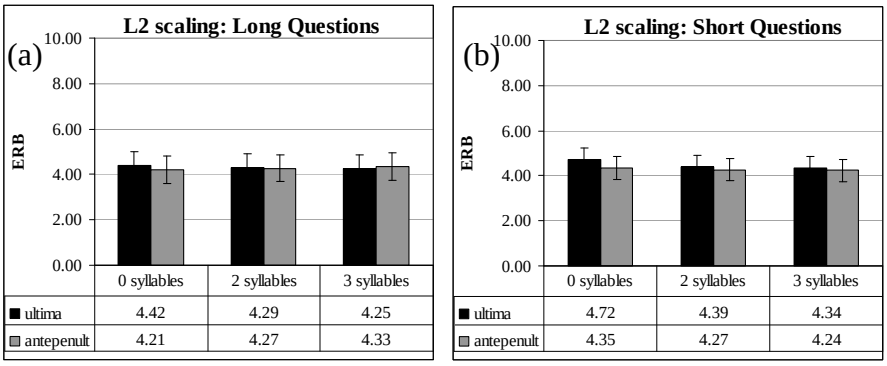

Figure 6: Mean scaling and standard errors for the second elbow (L2) as a function of INTERSTRESS INTERVAL and FINAL STRESS; data are presented separately for long and short questions in panels (a) and (b) respectively.

Because the difference in the alignment of L2 was relatively large between questions ending in a stressed final syllable and those in which the last stress was on the antepenult, we also investigated the alignment of L2 in short questions for which the corpus also included final words with penultimate stress (see sets A and B in the Appendix). These data were investigated by means of an ANOVA with WH-WORD LENGTH (short ([ˈpu] 'where' or long [apoˈpu] 'from where'), INTERSTRESS INTERVAL (0 or 3 syllables) and FINAL STRESS (stress on the ultima, penult or antepenult) as repeated-measures factors. In these data, the distance of L2 from the onset of the final vowel was affected by INTERSTRESS INTERVAL [$F(1, 3) = 28.8$] and, as before, by FINAL STRESS [$F(2, 6) = 61.03$]. In particular, L2 occurred earlier with respect to the final vowel onset when the interstress interval was three rather than zero syllables (see Figure 7(b)), that is when tonal crowding was greater. In addition, while L2 co-occurred with the onset of the final vowel when the last word was stressed either on the penult or the antepenult, it occurred half-way through the final vowel when this vowel was stressed [for antepenultimate vs. final stress, $F (1, 3) = 69.4$; for penultimate vs. final stress, $F(1, 3) = 89.8$]. These results are in agreement with the results of the main dataset which also showed alignment of L2 with the onset of the final vowel when this vowel is not stressed. They also agree with the main dataset in terms of the scaling of L2, which was affected only by FINAL STRESS [$F(2, 6) = 5.7$]: L2 was scaled higher when the final vowel was stressed than when stress was on the antepenult [$F(1, 3) = 10.8$]; the same pattern was observed in the comparison of final and penultimate stress, though in this case the difference narrowly failed to reach significance [$F(1, 3) = 7.8$, $p < 0.07$].

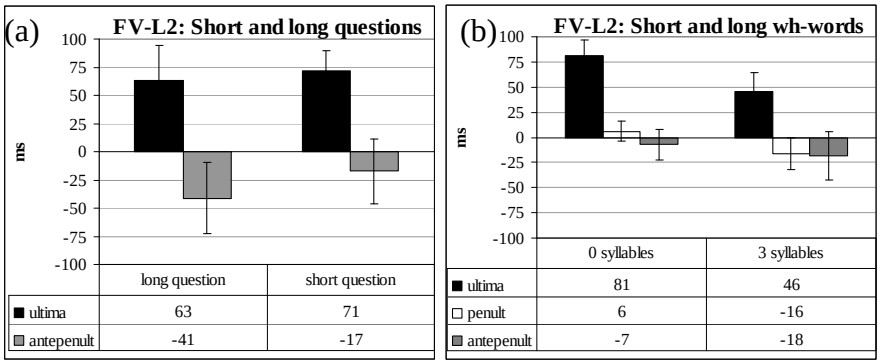

Figure 7: In panel (a), means and standard errors for the alignment of the second elbow (L2) with respect to the onset of the last vowel, as a function of QUESTION LENGTH and FINAL STRESS (results based on datasets B and C); in panel (b), the same results for datasets A and B.

2.3.5 Scaling of final H

The scaling of the final H (FH) was investigated by means of an ANOVA with QUESTION LENGTH and FINAL STRESS as repeated-measures factors. The results showed that neither factor affected the scaling of this target, and there was no interaction between the two.

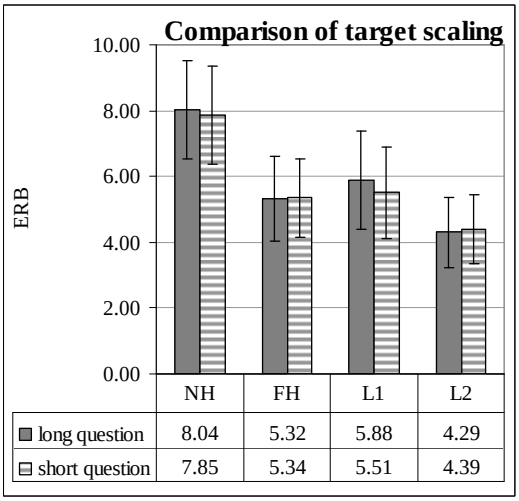

Figure 8: Means and standard errors of NH, FH, L1 and L2 scaling separately for short and long questions.

In addition, the scaling of FH was compared to that of NH, L1 and L2 by means of an ANOVA with TONE TYPE (NH, FH, L1, L2) and QUESTION LENGTH as repeated-measures factors. The

purpose of this comparison was to see whether the scaling of FH was indeed lower than that of NH and whether it was comparable to or higher than that of the two L targets, since impressionistically FH appears to be half-way between NH and the L targets in scaling. The results showed only an effect of TONE TYPE [$F(3,9) = 24.5$]: FH was scaled lower than NH [$F(1, 3) = 19.2$] and higher than L2 [$F(1, 3) = 37.6$], but not differently from L1 (see Figure 8). In contrast, the scaling of the targets was similar in long and short questions.

3 Discussion

3.1 Lawful variability in tonal alignment and scaling

The experimental data just presented provide ample evidence that adjustments for tonal crowding can systematically and significantly affect the realization of intonation contours. The very notion of tonal crowding is difficult to reconcile with a model in which each syllable is synchronized with its own tonal specifications, and our data are therefore prima facie difficult to accommodate in a syllable-based model like Xu's PENTA.

First, our data support previous findings of precise effects on the scaling and alignment of specific targets. We have replicated previous findings that the scaling of peaks is not significantly affected by tonal crowding, whereas the scaling of lows appears to be more susceptible to such effects (Arvaniti et al. 1998; 2000; 2006a; Prieto 2005). For example, L2 in our data shows some evidence of undershooting (i.e. higher scaling) when it is too close to the final rise, while the scaling of the nuclear peak remains unaffected by tonal crowding.

Second, we have also replicated findings that adjustments to alignment are substantial, both for L and H targets. In the present data, for example, NH aligns within the nuclear vowel when there is tonal crowding (i.e. when interstress interval is zero), but with the first postnuclear vowel when pressure from tonal crowding is removed. Similarly, both the extent and the exact alignment of the low plateau demarcated by the two elbows L1 and L2 are affected by tonal crowding: L1 generally aligns with the first stressed syllable after the nucleus when there is no tonal crowding, but may occur after this syllable when tonal crowding is present. Crucially the alignment of L1 also shows clearly that pressure can be exercised by both preceding and following targets: in long questions, where there is sufficient segmental material to support the low plateau and the

upcoming FH, L1 aligns later than in short questions, in which there is pressure not only from the preceding NH but from the upcoming targets as well. Similarly, L2 shows different patterns of alignment, depending on the position of the final stressed syllable and the length of the question: as with L1, the alignment adjustments for L2 are more dramatic in short questions in which tonal crowding is more extreme.

Third, our results suggest that the canonical realization of some targets may take precedence over that of others, creating a subtle interplay. For instance, we found that in long questions with zero interstress interval, L1 occurred after the postnuclear stressed syllable (with which it aligned in the rest of our data); this would suggest that realizing the NH with peak delay is more important than realizing the postnuclear stressed syllable with low F0. Nevertheless, the NH peak delay does not take precedence over the low F0 stretch altogether: in questions where peak delay could result in outright loss of the low plateau, speakers retract the nuclear peak instead. Overall, these fine-grained adjustments provide strong evidence in favour of viewing intonational contours as consisting of a string of tonal targets whose alignment with specific syllables is phonologically governed and can phonetically vary within limits.

These same adjustments also argue against the view of melodies as a string of syllable-specific contours, as advocated, e.g., by PENTA. This is so, not only because it is simpler to describe the adjustments if we assume a notion of syllable-independent tonal target, but also because certain of our findings actually run counter to specific principles of Xu's PENTA model. First, as mentioned earlier, Xu has often suggested that tonal coarticulation only ever involves carry-over effects; e.g. Xu & Wang (2001: 329) argue that "[w]hen two pitch targets occur next to each other, if the offset of the first one is different from the onset of the second one, the second one will appear as if it has been assimilated or partially assimilated to the first one" (for an extensive discussion of this point, see Xu 2005: 227ff. and 245 ff.). Yet our data show clearly that some of the coarticulatory effects of tonal crowding involve anticipatory retraction of targets, most notably the alignment of the NH when the wh-word's stressed syllable is followed immediately by another stressed syllable. (For similar evidence of anticipatory effects in tonal coarticulation in Mandarin, see Shih et al. 2007 and references therein.)

Second, Xu has also repeatedly suggested (especially in Xu & Sun 2002) that pitch changes are in general executed at close to maximum speed, and that this constraint 'plays a significant role in shaping the f0 contours in speech' (Xu & Xu 2005: 164). Indeed, as discussed in §2.1, PENTA predicts that such a steep change would characterize the (post-focal) fall from NH to L1 in our data. These claims are conspicuously at odds with our findings about the fall from NH to L1, which exhibits variable duration and slope, depending on the degree of tonal crowding, but little or no difference in the size of the pitch drop (a result in line with those reported in Beckman & Pierrehumbert 1988 for Japanese; see §1.2). To be sure, Xu (2005) does note that changes in the rate of F0 change may take place to accommodate different functional needs; but in our data all tokens fulfil the same function. Clearly, then, the variation in slope we have uncovered is neither physiological, nor functional, but linguistic.[4]

Third, Xu assumes, based on his data from Mandarin, that in some sense alignment takes precedence over scaling: "the implementation of an underlying tonal segment seems to start at the onset of the host syllable and end at the offset of the syllable …. [S]uch synchrony is often achieved at the expense of full implementation of the tonal targets" (Xu 2005: 224). He further suggests that there is a biophysical basis for the synchrony and hence for the precedence of alignment (Xu & Sun 2002). Our data make clear that, in general, Greek tonal realization reverses the priorities Xu reports for Chinese: scaling of most of the targets we have considered is essentially unaffected by tonal crowding, while alignment varies substantially. It thus seems unlikely that any effects of tonal coarticulation and realization can be attributed exclusively to physics rather than phonology.

A more general problem for Xu is that our data show clearly that syllables do not exhibit stable F0 properties. This makes sense if we make the phonological assumption that the F0 of many syllables is derived from the position of particular targets and the pressures that affect them, but it is difficult to reconcile with a model based on syllable-by-syllable specification of contours unless these specifications can be changed ad hoc. Informal attempts to model utterance contours similar to those in our dataset, using the online PENTA model at

[4] A similarly functional explanation is offered by PENTA for the final rise, which is attributed to the question function of the utterances. Recall, however, that wh-questions need not end in a rise and that, as discussed in footnote 1, the melody under consideration here is also used with negative declaratives (Baltazani 2002, 2006, Arvaniti & Baltazani 2005).

http://www.phon.ucl.ac.uk/home/yi/qTA/, provide a clear illustration of this problem. To give but one example, for utterances beginning with a monosyllabic wh-word, F0 on the wh-word is best approximated if it is specified as falling (and only if an ad hoc stipulation is added that the starting F0 level must be high rather than the typical PENTA specification of middle level for the start of the utterance; cf. Xu & Xu 2005). By contrast, for a polysyllabic wh-word beginning with unstressed syllables, F0 is better approximated if all syllables in the wh-word are specified as rising and as having weak strength (note that the weak strength specification for the stressed syllable in particular is inconsistent both with the functional and phonetic salience of the wh-word and with PENTA principles regarding the connection between stress and articulatory force). It is of course true that such finely tailored adjustments to phonetic parameters are compatible with PENTA's goal of modelling phonetic detail; however, as we have argued throughout the paper, the issue is not simply whether contours can be approximated, but whether the model can be made compatible with generalizations about intonational form. Such generalizations are plainly necessary if speakers are to extrapolate from melodies that look different superficially (such as the three contours in Figure 1) to a more abstract representation that can be used in conjunction with segmental material of varying lengths and metrical patterns. But these generalizations cannot be extracted if individual syllables change tonal specification from one instantiation of a contour to another. In effect, as argued by Arvaniti (2007a), Xu's model is both more powerful than it needs to be to account for phonetic variation, and not powerful enough to account for generalizations beyond phonetic form.

Similar problems arise with Fujisaki's model. An obvious but trivial problem is the fact that in this model the presence of accent commands is directly linked to lexical prominence (e.g. Fujisaki et al. 1997), so that the final rise in wh-questions ending in an unstressed syllable must be arbitrarily stipulated to occur after the last stressed syllable. This problem is trivial in the sense that the restriction of local pitch movements to prominent syllables is based on the model's origins as a model of Japanese; a more universally applicable model of local pitch movements could simply abandon that restriction, albeit at the expense of added arbitrariness (this is, e.g., the direction taken in Fujisaki 2004). A more serious problem is that the physical characteristics of such commands – their height and duration – are specified in the model in ways that cannot be derived by some linguistic principle. For example, the phonetic details of the low plateau of the wh-question contours depend on the extent and slope of the initial fall and (to a lesser extent) the

final rise, which, as we have seen, vary systematically depending on the position of the stressed syllables. Specifying this variation will have to be achieved by arbitrarily adjusting the parameters of the two accent commands. To be sure, Fujisaki's model allows for this variation and can thus create excellent approximations of observed F0 contours, but it provides no way of making the variation depend systematically on the position of the stressed syllables. Since Fujisaki and his colleagues explicitly state that a phonologically plausible phonetic model is the ultimate goal of their work (e.g. Fujisaki 2004, Gu et al. 2007), this arbitrariness must be reckoned a significant shortcoming.

With regard to long-domain effects that are central to superposition models such as Fujisaki's, our data show no evidence that tonal scaling is affected by declination, since targets were not scaled differently in the short and long wh-questions (see Figure 8). Since the phrase component in Fujisaki's model involves an invariant impulse response that decays over time, this result is unexpected, and the lack of scaling differences between long and short sentences would have to be expressed in terms of arbitrary adjustments to the scaling parameters of the accent commands. A more traditional model of declination (e.g. 't Hart et al. 1990) would have to account for the lack of effect of sentence length as involving the selection of drastically different declination rates, 220 Hz/s for long questions and 450 Hz/s for short questions (these rates were calculated by dividing the difference in Hz between NH and L2 by the temporal distance between them). Although such preplanning effects have been discussed in the literature (e.g. Sternberg et al. 1980, Liberman & Pierrehumbert 1984, Ladd 1988), it seems more parsimonious to assume that fixed tonal targets are the principal basis of scaling, and that most of what has been called declination is primarily due to phonological factors such as downstep rather than biomechanical ones. This more parsimonious assumption is obviously consistent with our findings here. As noted in the introduction, though, the issue of whether there are time-dependent declination effects is still current, and specifically with regard to Greek there is some evidence (Arvaniti 2003; Arvaniti & Godjevac 2003; Arvaniti et al. 2006a) that such effects do occur. Our present data therefore make clear that the issue of time-dependent declination is still unresolved.

3.2 The autosegmental representation of the wh-question tune

Since the experimental results clearly seem to favour a sparse-specification approach to modelling intonational phonetics, we now briefly consider the relevance of our findings for the AM analysis of the Greek wh-question melody. In the introduction, we offered a tentative representation of the wh-question melody as (L+)H L- H%. This was largely confirmed by the experimental data, but the results warrant further discussion of this analysis.

First, the data strongly suggest that the initial rise to a peak should be treated as part of a bitonal L+H accent, since the level of F0 when there are unstressed syllables before the stressed syllable of the wh-word is significantly lower than when the wh-word is monosyllabic. In fact, as can easily be seen in Figures 1 and 2 it is as low as the low plateau; this observation is supported by ANOVA results showing that IL is not significantly different from L1, and only marginally higher than L2 [$F(2,6)=14.6$]. This suggests that when the wh-word begins with a stressed syllable, we are dealing with a truncated L target. In addition, our data show that, in the absence of tonal crowding, the peak of this accent shows delay and appears early in the postnuclear vowel. This type of late peak alignment is exactly what has been reported for the so-called prenuclear accents of Greek, examined in detail in Arvaniti et al. (1998; 2000). Arvaniti et al. (2006b) provide quantitative data which confirm that Greek also has a different L+H accent, typically used to signal narrow focus, the peak of which occurs roughly in the middle of the nuclear vowel (see also Arvaniti & Baltazani 2005). Taken together, the results of the present study and those of previous work suggest that the nuclear accent in wh-questions must be the same as the accent used in prenuclear position in declaratives. Although as mentioned in the introduction there are problems with the autosegmental representation of this accent, here we follow Arvaniti et al. (2006b) and suggest that the most appropriate representation of this accent within the Greek intonational system is L*+H.

Our results also support the idea that the low plateau is the reflex of an L- phrase accent. In particular the alignment of L1 is consistent with the analysis of Grice et al. (2000), in the sense that L1 exhibits stress-seeking behaviour: as shown in §2.3.3, L1 typically co-occurs with the first stressed syllable after the nucleus, thereby ensuring that this syllable has low F0 to the extent that tonal crowding permits. L2, on the other hand, clearly does not align with the last stressed vowel of the question, as we had expected, but rather with the last vowel independently of stress. However, stress is not without influence: as shown in § 2.3.4, L2 appears well after the

onset of the final vowel when this vowel is stressed, but before it when stress is on the penult or the antepenult (a similar effect of final stress in falling-rising contours in Dutch was found by Lickley et al. 2005). One way to interpret the alignment results of L1 and L2 is to say that the realization of the L- phrase accent is such that the postnuclear stressed syllables must be low, insofar as possible.[5]

Finally, our initial analysis suggested that the final rise is the reflex of an H%. The data, however, clearly showed that the F0 at the end of the questions (FH) is much lower in scaling than the peak on the wh-word (NH). The data also showed that FH is not statistically different in scaling from L1, yet it is clearly a high target, since it is higher than L2 and its absence gives questions a different pragmatic nuance (see Appendix). One possible explanation for the low scaling of this high target compared to the nuclear H could be declination. However, if this were the main reason for the difference in scaling between FH and NH, then the temporal distance between the two should correlate with the value of FH, but our data show no such correlation ($r = 0.06$). This result is corroborated by the fact that neither the scaling of NH and FH, as shown earlier, nor the difference in scaling between them is affected by question length [$F(1, 3) < 1$].

Another possible explanation for the scaling of FH is that it is due to context. Specifically, mid-level targets that cannot be easily classified as H or L have been analyzed as contextual variants: for example, Grice et al. (2005) analyze utterance-final mid-level F0 in German as a H% that is downstepped by a preceding L-, while Beckman & Ayers-Elam (1997) analyze utterance-final mid-level F0 in English as a L% that is upstepped by a preceding H-. Greek, however, has both L- H% and H- L% sequences which do not show effects of upstep or downstep (Arvaniti & Baltazani 2005; Arvaniti et al. 2006a). This means that the mid-level scaling of FH in our data cannot be seen as a contextual effect. Instead, Arvaniti & Baltazani (2005) suggest that there is a phonological distinction in Greek – a meaningful intonational choice – between downstepped and non-downstepped H tones, including boundary tones. This is in keeping with other AM descriptions of intonational phonology in which downstep is seen as an independently selected phonological choice, not merely an aspect of phonetic realisation triggered by specific sequences

[5] In her review, Pilar Prieto suggests that the L2 results are amenable to a different interpretation, namely the presence of a bitonal boundary tone, LH%, the L tone of which is aligned with the onset of the last vowel. Both our analysis, which involves spreading, and the analysis suggested by Prieto are compatible with the present data, and nothing crucial hinges on whether one or the other is adopted.

of tones (e.g. Ladd 1983, Gussenhoven 2004: 307ff.). It therefore seems likely that FH is best analysed phonologically as !H%.[6]

4. Conclusion

This paper has presented an empirically supported autosegmental analysis of the Greek wh-question melody as L*+H L- !H%. It has done so, moreover, on the basis of data that provide strong evidence in favour of the general autosegmental-metrical approach to intonational phonology. Certain points in the Greek wh-question melody show little variability in scaling and predictable variability in alignment, and thus appear to be controlled in production. These phonetic effects cannot be explained by superposition models of intonation, such as Fujisaki's command-response model, which lack the mechanisms to account for effects such as the truncation of targets or asymmetrical adjustments to the larger tonal gestures they postulate. Nor can they be accounted for by models, such as Xu's PENTA, that assume that all syllables are specified for tone: these models are particularly problematic because they cannot account for either phonetic detail or phonological generalization. We conclude that phonetic data like those presented here for the melody of Greek wh-questions strongly argue in favour of a model of intonational phonetics based on the autosegmental-metrical framework of intonational phonology and in particular on the notion of sparse tonal specification.

Appendix: Methodological Details

Materials. The materials are presented in Table 2. Insofar as possible, the sentences were designed so as to avoid segmental effects on F0. Such effects could not be avoided on the wh-words, since the stressed syllables of all Greek wh-words start with a voiceless stop. In order to reduce recording time (since the materials were recorded together with materials for unrelated experiments), some combinations of factors were not included: set C did not include questions ending in words with penultimate stress; such words were included in sets A and B in questions with interstress interval 0 or 3 only (see shadowed sentences); these questions were used for the

[6] Esther Grabe and Carlos Gussenhoven suggested to us that FH may not be the phonetic reflex of a tone at all, but rather the lack of one, i.e. a return to a 'default' mid-level pitch. However, this alternative seems somewhat unsatisfactory given that this final mid-level pitch is stable in scaling and its presence in a wh-question results in a meaningful pragmatic nuance (the rising tune has connotations of being less insistent and more polite than the non-rising tune; Arvaniti & Baltazani 2005; Arvaniti 2007c). More importantly, as mentioned, there is evidence from other phonological contexts to suggest that mid-level pitch in Greek must be analysed as downstepped H (see Arvaniti & Baltazani 2005).

analysis of L2 alignment. In set A, the combination of interstress interval 2 and final stress was inadvertently replaced by a combination of interstress interval 2 and penultimate stress.

Table 2: The experimental materials

Set A: Short questions with short wh-word	Gloss
[ˈpu ˈzi]	'Where does s/he live?'
[ˈpu ˈmeni]	'Where is s/he staying?'
[ˈpu ˈmenune]	'Where are they staying?'
[ˈpu periˈmeni]	'Where is s/he waiting?'
[ˈpu periˈmenune]	'Where are they waiting?'
[ˈpu ne to maˈli]	'Where's the wool?'
[ˈpu me periˈmeni]	'Where is s/he waiting for me?'
[ˈpu me periˈmenune]	'Where are they waiting for me?'
Set B: Short questions with long wh-word	**Gloss**
[apoˈpu ˈles]	'Which way do you think [we should go]?'
[apoˈpu ˈlene]	'Which way do they think [we should go]?
[apoˈpu ˈmilaɟe]	'Where was s/he talking from?'
[apoˈpu na miˈla]	'Where could s/he be talking from?'
[apoˈpu na miˈlaɣane]	'Where were they have been talking from?'
[apoˈpu na mu miˈla]	'Where could s/he be talking to me from?'
[apoˈpu na mu miˈlane]	'Where could they be talking to me from?'
[apoˈpu na mu miˈlaɣane]	'Where could they have been talking to me from?'
Set C: Long questions with long wh-word	**Gloss**
[apoˈpu ˈlene na ti ˈðun]	'Where are they planning to see it from?'
[apoˈpu ˈmilaɟe tu ˈmenelu]	'Where was s/he speaking to Menelos from?'
[apoˈpu na miˈla me ti maˈma]	'Where could s/he be speaking to mom from?'
[apoˈpu na miˈlane me to ˈmenelo]	'Where could they be talking to Menelos from?'
[apoˈpu na mu miˈlaɣane proˈxtes]	'Where could they have been speaking to me from the day before yesterday?'
[apoˈpu na mu miˈlaɣane to ˈsavato]	'Where could they have been speaking to me from on Saturday?'

Speakers. The materials were read by two female and two male educated native speakers of Standard Greek, who were naïve as to the purpose of the experiment (with the exception of the first author, speaker AA, whose data were included once it was clear that they did not differ from those of the naïve speakers). 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 who was on a research visit). None of the speakers reported or was known to have any speech or hearing impairment.

Procedures. The questions were read from cards on which mini-dialogues were typed in Greek orthography. The test dialogues were interspersed with fillers of similar structure (in effect materials for unrelated experiments). The speakers were told to read the sentences as naturally as possible and at a speaking rate that felt comfortable. They were not given any instructions as to which tune to use, though they all used the two variants of the tune described in the introduction (see §1.3). Each speaker read the materials seven times in random order. Six repetitions were selected for measurement by Ineke Mennen (the project's R.A. and an L2 speaker of Greek with near-native competence) on the basis of their naturalness and fluency.

Measurements**.** Measurements were made by Mennen, in consultation with Arvaniti when necessary, using Waves+. Segmental onsets were identified on the basis of wide-band spectrograms and waveforms, following standard criteria of segmentation (Peterson & Lehiste 1960). F0 contours were obtained by means of the Waves+ pitch tracking facility, with a 49 ms $\cos^4$ window moving in 10 ms steps. F0 was converted from Hz to ERB using the equation of Hermes and van Gestel (1991: 97): ERB = 16.7 log (1 + f / 165.4) where f is frequency in Hz.

When selecting F0 points for measurement care was taken to avoid obvious microprosodic perturbations. F0 points were easy to locate, except for the first elbow (L1) for which the F0 slope gradually changed from steep to gentle (see e.g. Figures 1(b) and 2). The results reported here are based on Mennen's impressionistic judgment of the point where this transition occurred. Because of the subjectivity involved in this procedure, we performed limited reliability checks using two algorithmic methods of elbow location. Comparisons of the manually and algorithmically annotated data clearly show that the overall picture does not depend on method. The first algorithmic method applied the techniques of Taylor (2000) for determining the beginning and end of accent-related F0 movements. The second method was that used for locating elbows by D'Imperio (2000), which was based on a program written by Mary Beckman and employed for regression line fitting in Pierrehumbert & Beckman (1988). Specifically, the elbow is located by fitting two straight lines by conventional least-squares methods to the

relevant part of the F0 contour: the elbow is taken to be the intersection of the two lines that yield the smallest total modelling error computed on the basis of linear regressions. Del Giudice et al. (2007) have since shown that this is the most robust algorithmic method for locating elbows and that human intuition correlates well with its results. Finally, we note that 8% of the measured utterances (43/528) ended with a low level F0 stretch (see §1.3 and footnote 6); 85% of them were in the male speakers' data. In these non-rising tokens it was not possible to measure either L2 or the FH; this does not affect the results beyond the fact that in 20 cases (out of a total of 88), the means used for statistical analysis were based on fewer than six measurements.

References

Arvaniti, Amalia (2001). The intonation of wh-questions in Greek. *Studies in Greek Linguistics* **21**. 57–68.

Arvaniti, Amalia (2003). Peak scaling in Greek and the role of declination. *Proceedings of XVth International Congress of Phonetic Sciences*. 2269–2272.

Arvaniti, Amalia (2007a). On the relationship between phonology and phonetics (or why phonetics is not phonology). *Proceedings of XVIth International Congress of Phonetic Sciences*. 19–24.

Arvaniti, A. (2007b). On the presence of final lowering in British and American English. In C. Gussenhoven & T. Riad (eds.) *Tone and Tunes, vol. 2: Experimental Studies in Word and Sentence Prosody*. Berlin and New York: Mouton de Gruyter. 317–347

Arvaniti, Amalia (2007c). Greek Phonetics: The State of the Art. *Journal of Greek Linguistics* **8**: 97–208.

Arvaniti, Amalia & Mary Baltazani (2005). Intonational analysis and prosodic annotation of Greek spoken corpora. In S. Jun (ed.) *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press. 84–117.

Arvaniti, Amalia & Svetlana Godjevac. 2003. The origins and scope of final lowering in English and Greek. *Proceedings of the XVth International Congress of Phonetic Sciences*. 1077–1080.

Arvaniti, Amalia & D. Robert Ladd (1995). Tonal alignment and the representation of accentual targets. *Proceedings of the XIIIth International Congress of Phonetic Sciences*. Vol. 4. 220–223.

Arvaniti, Amalia, D. Robert Ladd & Ineke Mennen (1998). 'Stability of tonal alignment: the case of Greek prenuclear accents'. *JPh* **26**. 3–25.

Arvaniti, Amalia, D. Robert Ladd & Ineke Mennen (2000). 'What is a starred tone? Evidence from Greek'. In M. Broe & J. Pierrehumbert (eds.) *Papers in Laboratory Phonology V*. Cambridge: Cambridge University Press. 119–131.

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

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

Baltazani, Mary (2002). *Quantifier Scope and the role of intonation in Greek*. Ph.D. dissertation, UCLA.

Baltazani, Mary (2006). Intonation and pragmatic interpretation of negation in Greek. *Journal of Pragmatics* **38**. 1658–1676.

Beckman, Mary E. & Gayle Ayers Elam (1997). *Guide to ToBI Labelling*, available at http://www.ling.ohio-state.edu/research/phonetics/E_ToBI/

Beckman, Mary E. & Janet B. Pierrehumbert (1986). Intonational structure of Japanese and English. *Phonology Yearbook* **3**: 255–309.

Browman, Catherine P. & Louis Goldstein (1990). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman (eds.) *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*. Cambridge: Cambridge University Press. 341–376.

Browman, Catherine P & Louis Goldstein. 1992. Articulatory Phonology: An overview. *Phonetica* **49**: 155–180.

Bruce. Gösta (1977). *Swedish word accents in sentence perspective*. Lund: Gleerup.

Cohen, Antonie & Johan 't Hart (1967). On the anatomy of intonation. *Lingua* **19**. 177–192.

Cohn, Abigail C. (1993). Nasalisation in English: phonology or phonetics. *Phonology* **10**. 43–81.

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

del Giudice, Alex, Ryan K. Shosted, Katherine Davidson, Mohammad Salihie & Amalia Arvaniti (2007). Comparing methods for locating pitch "elbows." *Proceedings of the XVIth International Congress of Phonetic Sciences*. 1117–1120.

D'Imperio, Mariapaola (2000). The role of perception in defining tonal targets and their alignment. PhD dissertation, The Ohio State University.

Fujisaki, Hiroya (1983). Dynamic characteristics of voice fundamental frequency in speech and singing. In P. F. MacNeilage (ed.) *The production of speech*. Heidelberg: Springer-Verlag. 39–55.

Fujisaki, Hiroya (2004). Information, prosody and modeling – with emphasis on tonal features of speech. *Proceedings of Speech Prosody 2004*, Nara, Japan, March 23–26, 2004. Available at http://www.isca-speech.org/archive

Fujisaki, Hiroya, Sumio Ohno & Takashi Yagi (1997). Analysis and modelling of fundamental frequency contours of Greek utterances. *Proceedings of Eurospeech '97*. 465–468.

Fujisaki, Hiroya, Changfu Wang, Sumio Ohno, Wentao Gu (2005). Analysis and synthesis of fundamental frequency contours of Standard Chinese using the command–response model. *Speech Communication* **47**. 59–70.

Gick, Bryan (1999). A gesture-based account of intrusive consonants in English. *Phonology* **16**: 29–54.

Gick, Bryan & Ian Wilson (2006). Excrescent schwa and vowel laxing: cross linguistic responses to conflicting articulatory targets. In L. Goldstein, D. H. Whalen & C. T. Best (eds.) *Laboratory Phonology 8*. Berlin/New York: Mouton de Gruyter. 635–659.

Glasberg, B. R. and Brian Moore (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research* **47**: 103–38.

Grabe, Esther (1998). *Comparative intonational phonology: English and German.* Ph.D. dissertation, Max Planck Institute for Psycholinguistics, Nijmegen.

Grice, Martine, D. Robert Ladd & Amalia Arvaniti (2000). On the place of phrase accents in intonational phonology. *Phonology* **17**. 143–185.

Grice, Martine, Stefan Baumann & Ralf Benzmüller (2005). German intonation in Autosegmental-Metrical phonology. In S. Jun (ed.) *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press. 55–83.

Gu, Wentao, Keikichi Hirose & Hiroya Fujisaki (2007). Analysis of tones in Cantonese Speech based on the command-response model. *Phonetica* **64**. 29–62.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

Hermes, Dik & Joost van Gestel (1991). The frequency scale of speech intonation. *JASA* **90**. 97–102.

Holst, Tara & Francis Nolan (1995). The influence of syntactic structure on [s] to [] assimilation. In B. Connell & A. Arvaniti (eds.) *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*. Cambridge: Cambridge University Press. 315–333.

Jun, Jongho (1996). Place assimilation is not the result of gestural overlap: evidence from Korean and English. *Phonology* **13**: 377–407.

Kochanski, Greg P. & Chilin Shih (2003). Prosody modeling with soft templates. *Speech Communication* **39**. 311–352.

Ladd, D. Robert (1983). Phonological features of intonational peaks. *Lg*. **59**. 721–59.

Ladd, D. Robert (1988). Declination 'reset' and the hierarchical organization of utterances. *JASA* **84**. 530–44.

Ladd, D. Robert (2008). *Intonational Phonology*. 2nd edn. Cambridge: Cambridge University Press.

Ladd, D. Robert & Catherine Johnson (1987). 'Metrical' factors in the scaling of sentence-initial accent peaks. *Phonetica* **44**. 238–245.

Liberman, Mark & Janet Pierrehumbert (1984). Intonational invariance under changes in pitch range and length. In M. Aronoff & R. Oerhle (eds.) *Language sound structure*. Cambridge, MA: MIT Press. 157–233.

Lickley, Robin J., Astrid Schepman & D. Robert Ladd (2005). Alignment of 'phrase accent' low in Dutch falling rising questions: Theoretical and methodological implications. *Language & Speech* **48**. 157–83.

Maeda, Shinji (1976). A characterization of American English intonation. PhD dissertation, MIT.

Nolan, Francis (1992). The descriptive role of segments: evidence from assimilation. In G. J. Docherty & D. R. Ladd (eds.) *Papers in Laboratory Phonology 2*: Cambridge University Press. 261–280.

Ohala, John (1974). Experimental historical phonology. In J. M. Anderson & C. Jones (eds.) *Historical Linguistics II: Theory and Description in Phonology.* Amsterdam: North Holland. 353–389.

Peterson, Gordon, E. & Ilse Lehiste (1960). Duration of syllable nuclei in English. *JASA* **32**. 693-703.

Pierrehumbert, Janet B. (1980). *The phonetics and phonology of English intonation*. Ph.D. dissertation, MIT. Distributed by Indiana University Linguistics Club.

Pierrehumbert, Janet B. & Mary E. Beckman (1988). *Japanese Tone Structure*. Cambridge, Mass.: The MIT Press.

Prieto, Pilar (2005). Stability effects in tonal clash contexts in Catalan. *JPh* **33**. 215–242

Prieto, Pilar, Jan van Santen, & Julia Hirschberg (1995). Tonal alignment patterns in Spanish. *JPh* **23**. 429–451.

Shih, Chilin, Greg Kochanski & Su-Youn Yoon (2007). The missing link between articulatory gestures and sentence planning. *Proceedings sof XVIth International Congress of Phonetic Sciences*. 35–38.

Silverman, Kim & Janet Pierrehumbert (1990). The timing of prenuclear high accents in English. In J. Kingston & M. E. Beckman (eds.) *Papers in Laboratory Phonology I: Between the grammar and physics of speech*. Cambridge: Cambridge University Press. 72–106.

Sternberg, S., C. E. Wright, R. L. Knoll & S. Monsell (1980). Motor programs in rapid speech: additional evidence. In R. A. Cole (ed.) *Perception and production of fluent speech.* Hillsdale NJ: Lawrence Erlbaum. 507–534.

Sun, Xuejing (2002). *The determination, analysis, and aynthesis of fundamental frequency*. Ph.D. dissertation, Northwestern University.

Sundberg, Johan (1979). Maximum speed of pitch changes in singers and untrained subjects. *JPh* **7**: 71–79.

Taylor, Paul (2000). Analysis and synthesis of intonation using the Tilt model. *JASA* **107**. 1697–1714.

't Hart, John, René Collier & Antonie Cohen (1990). *A perceptual study of intonation.* Cambridge: Cambridge University Press.

Xu, Yi (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication* **46**. 220–251.

Xu, Yi & Xuejing Sun (2002). Maximum speed of pitch change and how it may relate to speech. *JASA* **111**. 1399–1413.

Xu, Yi & Q. Emily Wang (2001). Pitch targets and their realization: evidence from Mandarin Chinese. *Speech Communication* **33**. 319–337.

Xu, Yi & Ching X. Xu (2005). Phonetic realization of focus in English declarative intonation. *JPh* **33**. 159–197.

Zsiga, Elizabeth C. (1994). Acoustic evidence for gestural overlap in consonant sequences. *JPh* **22**. 121–140.

Zsiga, Elizabeth C. (1997). Features, gestures, and Igbo vowels: An approach to the phonology-phonetics interface. *Lg* **73**: 227–274.