Tonal association and tonal alignment: evidence from Greek polar questions and contrastive statements

Amalia Arvaniti,[a] D. Robert Ladd,[b] and Ineke Mennen[c]

[a]University of California, San Diego

[b]University of Edinburgh

[c]Queen Margaret University College, Edinburgh

SHORT TITLE: Tonal association and tonal alignment in Greek

Corresponding author: Amalia Arvaniti

Dept. of Linguistics, UC San Diego, 9500 Gilman Drive #0108, La Jolla, CA 92093, USA.

E-mail: amalia@ling.ucsd.edu Phone: +1-858-534-8409 Fax: +1-858-534-4789

Acknowledgments

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 Mennen as Research Associate. This support is gratefully acknowledged. We would also like to thank Mary Beckman for her advice and encouragement through the long gestation of this work, Sun-Ah Jun and Janet Pierrehumbert for discussion of the data presented here, and Norman Dryden, Mike Bennett, Eddie Dubourg, Cedric Mac Martin, and Stewart Smith for technical support. Thanks are also due to our speakers, and to Associate Editor Martine Grice, Esther Grabe and an anonymous reviewer for their helpful comments on an earlier version of this paper.

ABSTRACT

This paper compares the production and perception of the rise-fall contour of contrastive statements and the final rise-fall part of polar questions in Greek. The results show that these superficially similar rise-falls exhibit fine phonetic differences in the alignment of tonal targets with the segmental string, and that these differences can be used by native speakers under experimental conditions to identify the two contour types. It is further shown here that the observed differences in alignment are best attributed to differences in the overall tonal composition of these contours, which results in different degrees of crowding for the targets involved. This analysis accounts for the differences in phonetic detail between the two contours, while obviating the need to posit distinct secondary associations for the peak of the rise-fall. It is suggested that differences in phonetic alignment should be formalized by means of the secondary association mechanism only if simpler analyses and explanations have been considered and shown not to account effectively for the data. Finally, the perceptual results suggest that even small alignment differences like those observed here have a role in perception and should therefore be specified in a full description of the phonetic implementation of tunes.

Keywords: tonal association, tonal alignment, tonal crowing, secondary association

INTRODUCTION

Background

This paper compares the production and perception of phonetic detail in the intonation of polar questions and contrastive statements in Greek. What makes this comparison of some interest is that, although the two contour types are clearly distinct pragmatically and presumably phonologically, they may both exhibit very similar rise-fall F0 movements in the vicinity of a lexically stressed syllable. Our aim was to discover whether there are any differences in the phonetic realization of these rise-fall movements, and if so, whether the differences are perceptually relevant. Answers to these questions may provide us with a better idea of the most appropriate phonological representation of these contours, and a better understanding of the relationship between phonological representation and phonetic realization in the general domain of intonation.

In keeping with the autosegmental-metrical (henceforth AM) approach to intonational phonology (Bruce, 1977; Pierrehumbert, 1980; Beckman and Pierrehumbert, 1986; Pierrehumbert and Beckman, 1988; Ladd, 1996) we assume that utterance melodies are the realization of abstract sequences of L(ow) and H(igh) tones which are associated with the segmental string in various structure-dependent ways. In many European languages, for instance, tones appear to be organized into pitch accents—melodic features associated with certain stressed syllables—and phrase accents and boundary tones—melodic features associated with certain phrasal boundaries. The links between tones (H and L and combinations thereof) and tone bearing units (i.e. stressed syllables and mora) or metrical positions (such as phrasal boundaries) are represented as association lines between different tiers in the phonological representation.

It is hypothesized that the abstract sequence of H and L tones is phonetically realized as a string of tonal targets, the most obvious acoustic manifestations of which are local F0 turning points in the signal (maxima, minima, and "elbows", i.e. points where F0 changes slope but not direction).[1] The timing of F0 turning points is loosely determined by their phonological association, on the assumption that "links between [phonological] elements constrain them to overlap, as they are produced in time" (Silverman and Pierrehumbert, 1990, p. 72). However, the precise nature of the relation between the presumed phonological representation and F0 turning points is still poorly understood. Even in well-studied languages with lexical tone, where the "correct" analysis of the tonal string is seldom in serious dispute, much has recently been discovered about the phonetic details of how tones are realized and how observable F0 turning points are related to abstract H and L tones (see e.g. Laniran, 1992, and Laniran and Clements, 2003, on Yoruba; Gandour, Potisuk and Dechongkit, 1994, on Thai; Myers, 1999, on Chichewa; Myers, 2003, on Kinyarwanda; Xu, 1994, 1997, 1998, 1999, 2001, and Xu and Wang, 2001, on Mandarin).

Within non-tonal languages—mostly European languages—the past decade has seen a great deal of theoretical and empirical work on these issues. On the theoretical side, various authors (e.g.

[1] In response to a comment by an anonymous reviewer, we emphasize that F0 turning points are merely "the most obvious acoustic manifestations" of the tonal targets. We do not mean to suggest that the turning points themselves are the goals of the articulatory gestures that are used to implement the tonal targets: indeed, the articulatory targets may be dynamic in some way, and, as the reviewer points out, "pitch change takes time" (Sundberg, 1979; Xu and Sun, 2002). In the long run, articulatory studies (such as Mücke and Grice 2004) should provide us with a more detailed understanding of how tones are phonetically realized and how their phonetic realizations are coordinated with each other and with the gestures that realize the segmental string. However, since this line of research is still very new and published results rather preliminary, in this paper we use more readily obtainable acoustic data in order to shed light on the relationship between phonological representation and phonetic realization. This is standard practice in other areas of phonetics, and does not in any way imply that acoustic data are primary, or articulatory data irrelevant.

Grice, 1995, Ladd, 2000, Arvaniti, Ladd and Mennen, 2000) have discussed such concepts as "starredness" (the idea that one tone in a bitonal pitch accent, such as L*+H, is primary and governs the alignment of the pitch accent with the segmental string), and have commented on the fact that in many AM analyses there are F0 turning points that are not assumed to reflect phonological tones, and phonological tones that are not manifested by any F0 turning point. On the empirical side, a large number of studies on non-tonal languages have appeared which examine the details of how turning points (and putative phonological tones) are linked to the elements of the segmental string (among many, Silverman and Pierrehumbert 1990, for American English; Arvaniti and Ladd, 1995, and Arvaniti, Ladd and Mennen, 1998, on Greek; Prieto, van Santen and Hirschberg, 1995, on Mexican Spanish; Grabe, 1998, on English and German; Grabe et al., 2000, on varieties of British English; Wichmann, House and Rietveld, 2000, on Standard British English; Frota, 2002, on European Portuguese; Ladd and Schepman, 2003, on British English; Atterer and Ladd, 2004, on varieties of German; Arvaniti and Garding, to appear, on varieties of American English). The present paper contributes to this line of research.

Phonetically, we may define F0 turning points along two dimensions, their scaling, i.e. their F0 value, and their alignment, i.e. their temporal relation with the segmental string. Although scaling has long been recognized as variable across speakers and conditions (due to intrinsic differences between speakers and to paralinguistic effects such as emotion), it has only relatively recently become clear that alignment may exhibit interestingly complex regularities that depend on factors such as speaking rate (Fougeron and Jun, 1998), phonological weight (Ladd, Mennen and Schepman, 2000) and prosodic context (Prieto et al., 1995). Extensive empirical work on these topics, including the papers cited in the preceding paragraph, has now made it clear that the original AM assumption about the temporal overlap of tonal and segmental elements (as stated in the quote from Silverman and Pierrehumbert above) is at best an oversimplification. This then raises the question: what exactly are the phonetic reflexes of autosegmental “association” and of the other representational devices of AM intonational phonology?

Of particular interest here are recent studies showing that pitch accents may exhibit cross-dialectal variation in phonetic realization. Atterer and Ladd (2004) report that German L+H pitch accents show later alignment of both the L and the H tone in Southern than in Northern varieties. Arvaniti and Garding (to appear) report similar results for Southern California and Minnesota English, with speakers of the former variety showing later alignment of the H tone of L*+H than speakers of the latter. On the basis of such results both studies conclude that it is important not to create phonological representations for tones simply in order to reflect phonetic detail; rather, as in all other areas of phonology, phonological distinctions should be established on the basis of meaning and contrast and after taking into consideration the entire system under examination. This position is in line with Silverman and Pierrehumbert’s early contention that “[…] a transcription, in which the coordination between words and melody is mediated through the prosodic structure, is very nonspecific about exactly how segments in the two tiers are overlapped during pronunciation. And this is as it should be, because the phonological representation should capture only linguistic contrasts. Phonetic implementation rules, by contrast, have to specify how the phonological structure is realized in actual speech” (Silverman and Pierrehumbert, 1990, p.73).

Although small differences in alignment like those reported by Atterer and Ladd (2004) or Arvaniti and Garding (to appear) do not need to be phonologically represented, research has shown that the alignment of certain tones requires additional phonological specification. Such additional specification is provided by the mechanism of *secondary association*, first proposed in Pierrehumbert and Beckman (1988: 121 ff.), and used to account for the behavior of certain phrasal tones which, under certain conditions, do not appear at the very edge of the phrasal constituent they are associated with, but co-occur with a non-peripheral tone bearing unit (henceforth TBU). For example, Pierrehumbert and Beckman (1988) proposed that the phrasal H of Japanese has a primary association to the left edge of the accentual phrase and a secondary association to the phrase's second mora; this dual association means that the H aligns with the second mora, if this mora is not already associated to a HL (lexical) pitch accent; if it is, the H aligns with the phrase's first mora. The notion of secondary association has been used to account for the dual patterning of phrasal tones in several other languages (see e.g. Grice, Ladd and Arvaniti, 2000, on Greek, German, English, Hungarian and Romanian; Gussenhoven, 2000, on Roermond Dutch; Welby, 2004, on French; Lickley, Schepman and Ladd, 2005, on Dutch).

Recently Prieto, D'Imperio and Gili Fivela (to appear) have proposed that the phonological representation of certain pitch accents may also require the use of secondary association. Specifically, secondary association is used by Prieto et al. (to appear) to phonologically represent small differences in phonetic alignment between accent types, when the tonal inventory does not allow for distinct representations on the basis of tonal composition and/or differences of starredness. For example, in prenuclear position in broad focus statements, Catalan uses rising accents with a delayed peak (i.e. a peak that appears on the post-tonic syllable); in prenuclear position in imperatives, on the other hand, Catalan uses rising accents in which the peak occurs at the end of the accented syllable. Prieto et al. show that the only possible representation for both accents is L+H*, since they are in contrast with a L*+H and a H* pitch accent. In order to distinguish between the two L+H* accents, Prieto et al. propose that the L+H* of statements has primary association to the accented syllable, and its peak is aligned, by default, with the post-tonic; the L+H* of imperatives, on the other hand, has the same primary association, but also a secondary association of its H tone to the right edge of the accented syllable, and thus it peaks earlier.

Although Prieto et al.'s proposal is appealing in that it provides a principled account of the differences found in their data, we see two potential problems with this use of secondary association. One concerns the nature of the device itself. Pierrehumbert and Beckman's original proposal, like those of Grice et al. (2000), Gussenhoven (2000), Welby (2004) and Lickley et al. (2005), involves taking a tonal element that is primarily associated with a prosodic edge and giving it an *alternative* association to a specific TBU; this alternative association overrides the primary association of the tonal element, if the TBU is available. Prieto et al. (to appear) are doing something subtly but perhaps significantly different: they are taking a tonal element that is primarily associated with a specific TBU and giving it a *supplemental* association with a constituent edge. Crucially, this supplemental association always works in conjunction with the primary association of the tonal element: the primary association provides the docking site for the pitch accent and the secondary association provides additional details. We will term this use of secondary association s*upplemental association* whenever it is necessary to distinguish it from the original use of the device. We return to this issue at the end of the paper.

Our second concern with Prieto et al.'s proposal is essentially a matter of explanatory parsimony. We believe that the mechanism of secondary association should only be invoked when independently motivated principles cannot adequately account for observed alignment differences. For example, the Greek cases we discuss here are quite similar to those discussed by Prieto et al., in that they involve clear differences of alignment between apparently distinct accent types; nevertheless, we show that the Greek cases can be accounted for by differences of phonological context, without recourse to supplemental association. In other cases, it may be possible to subsume certain alignment differences under phonological representations that differ in tonal composition or starredness (a possibility we briefly discuss in the General Discussion section). Clearly, the most appropriate analysis can be chosen only after careful consideration of the meaning and form of the tonal events under investigation, the tonal contexts in which they typically appear, and the contrasts they enter into in a given intonational system. The goal of this paper is to show that by taking all these factors into consideration it is possible in at least some cases to account for distinctive differences of alignment without recourse to supplemental association. We should stress that our goal in bringing these concerns to the fore is not to reject outright Prieto et al.'s proposal, which is interesting and useful. Rather, we aim to show that the proposal can and should be constrained, so as to ensure that its theoretical machinery is not used merely to account for phonetic effects—for which, following Occam's Razor, simpler analyses should be preferred—but is employed only when genuine phonological differences need to be encoded in the grammar.

Greek Polar Questions and Contrastive Statements

Our study is based on production and perception data from two superficially similar melodies of Greek, that of polar questions and that of contrastive statements. We begin with a brief description of these melodies.

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 distinguished only by intonation. As we show in detail elsewhere (Grice, Ladd and Arvaniti, 2000; Arvaniti, Ladd and Mennen, in press), the melody of polar questions in Greek involves a low level F0 stretch associated with the word in focus, and a rise-fall pitch movement which aligns in two different ways with the segmental string, depending on the position of the focused word: if the word in focus is the last one of the question, the rise-fall movement co-occurs with the last syllable of the question; if the word in focus appears earlier, the rise-fall co-occurs with the *stressed* syllable of the *last word* of the question, without, however, signaling prominence or focus on that word.

On the basis of extensive quantitative data, Arvaniti, et al. (in press) conclude that the most appropriate autosegmental representation of the polar question melody is L* L+H- L%, where L* represents the nucleus of the question (i.e. the F0 of the stressed syllable of the word in focus), and the rise-fall movement is analyzed as a L+H- phrase accent followed by a L% boundary tone. Crucially for our purposes, the two modes of alignment of the L+H- phrase accent are interpreted as the result of its phonological association: L+H- has a primary association to the phrasal right edge and a secondary association to the last stressed syllable of the question. When the last word is in focus, the nuclear accent L* is associated with the last stressed syllable; thus, the secondary association of the L+H- phrase accent cannot take effect, and the L+H- is realized at the phrasal edge by virtue of its primary association. When the L* nuclear accent is on an earlier word, the stressed syllable of the final word is not associated with another tone, and thus the secondary association of the L+H- can take effect: in this case, the L+H- phrase accent co-occurs with the last stressed syllable. The two modes of alignment of the L+H- are illustrated in Figure 1 (for similar data, see also Baltazani and Jun, 1999; Grice et al., 2000; Arvaniti et al., in press).[2]

//////////////////////////////////// FIGURE 1 APPROXIMATELY HERE //

What is of interest here is that, as can be seen in Figure 2, Greek statements with contrastive focus (henceforth *contrastive statement*s) also show a rise-fall movement very much like the polar question rise-fall in Figure 1a. In this case, however, the rise-fall always occurs on the stressed syllable of the word with contrastive focus, independently of this word's position in the utterance (Botinis, 1989; Botinis, Bannert and Tatham, 2000). The alignment of this rise-fall is illustrated in panels (a) and (b) of Figure 2, which show contrastive accents on a final and a non-final word respectively. Thus, this rise-fall is best analyzed as a L+H* pitch accent followed by a L- phrase accent and a L% boundary tone (Arvaniti and Baltazani, 2000, 2005; Baltazani, 2002; ch. 2), where the L- L% configuration represents the low level F0 stretch that follows the contrastive accent.

[2] Readers whose native language is English or most other Western European languages might think that the occurrence of the nuclear accent on an early word in questions would be a relatively unusual occurrence. However, in Greek, as in many languages in Eastern Europe, the nuclear accent in a polar question falls by default on the verb (Ladd 1996: ch. 5; Grice et al., 2000), so that it is actually quite common to have the pattern of association shown in Figure 1a, with both transitive and intransitive verbs (in which the word order is typically (S)V0 and VS respectively).

//////////////////////////////////// FIGURE 2 APPROXIMATELY HERE //

As is evident from the above discussion, the rise-fall part of polar questions with early nucleus and the rise-fall of contrastive statements have distinct phonological representations in our analysis: the former is represented as a L+H- phrase accent which is preceded by a L* pitch accent and followed by a L% boundary tone; the latter is represented as a L+H* pitch accent followed by a L- phrase accent and a L% boundary tone. However, it is not clear whether we should expect these two representations to correspond to different phonetic realizations, particularly as far as the rise portion is concerned, since the representational difference between L+H* and L+H- is primarily a structural one, namely the difference between a pitch accent and a secondarily-associated phrase accent; AM theory is essentially silent on this matter. The goal of our experiments was to look for potential differences between the two rise-fall movements both in production and in perception, and to consider the implications for phonological representation of any differences we discovered.[3]

EXPERIMENT 1

Experiment 1: Method

<u>Speech Materials</u>. Our materials consisted of contrastive statements and polar questions. Both the questions and the statements came from a corpus of materials recorded for use in two studies, the one reported here and a large-scale study of question intonation (Arvaniti et al., in press).

[3] An anonymous reviewer questions the appropriateness of our comparison on the basis that questions with non-final focus and contrastive statements with final focus are items from different classes. This comment misses the fact that (a) the parts of the pitch contours we compare are final in both the questions and the statements; (b) we are not comparing the realization of *focus* across statements and questions, since as we discuss in more detail in the main text, polar questions have a low (L*) nucleus, while contrastive statements have a rising (L+H*) nucleus.

Although we recorded a large number of both questions and statements, our intention was to base the comparisons reported here on two 20-sentence subsets of the larger corpus, designed to ensure maximum comparability between statements and questions. The questions contained a maximum of two content words, the first of which carried nuclear accent. The statements were also short sentences in which a particular word was expected to bear contrastive accent. In both data-sets, the intonation was manipulated by setting the test sentences in short dialogues, as in examples (1) and (2) below (for the full set of sentences, see Appendix I). In addition, in the statements we underlined the word that was intended to bear contrastive accent, as shown in example (2).

(1) Polar Question

A: [ksanaˈpiɣate stin ˈeɟina] "Have you visited Aegina before?"

B: [ˈoçi aˈfti ˈine i ˈpɾoti foˈɾa] "No, this is the first time."

(2) Contrastive Statement

A: [ap tin aˈθina ðen ˈine] "Isn't he/she from Athens?"

B: [ap tin ˈeɟina] "From <u>Aegina!</u>"

In order to make the two sets of sentences as comparable as possible, we took three criteria into account in constructing the materials. Two of these criteria were straightforward: first, the same test-words were used in both sets—that is, as final words in questions and as the words in focus in contrastive statements—to the extent this was feasible (18 out of 20 pairs had the same test-word); second, all test-words had antepenultimate stress so as to keep any "crowding" effects on the realization of tones to a minimum (we discuss these effects in more detail in the General Discussion section; see also Arvaniti et al., 2000, in press). The third criterion was more difficult to satisfy: on the one hand, it was important for the test-words to be in similar position in the utterance, but on the other it was important for the contrastive statements to sound natural, and in general Greek prefers to place words that bear contrastive focus at the beginning of the sentence (see Keller and Alexopoulou, 2001, and references therein). For this reason, half of the statements contained only one content word, which was often preceded by stressless function words but was not followed by any additional material; where this was not possible, we preferred to have the word with contrastive focus first, followed by at most one more content word, which was always unaccented and had low F0 (Botinis, 1989; Arvaniti and Baltazani, 2005; see Figure 2b). This compromise meant that certain potential comparisons (such as the alignment of the low level F0 stretch after the peak of the rise-fall) were not meaningful across some of our materials. In addition, since the time we recorded the materials discussed here, other work by the first author has suggested that words in different positions in the sentence can show systematic differences in scaling (Arvaniti, 2003). Consequently, we decided to focus our statistical analysis on a group of nine questions and nine contrastive statements which have the same test-words in utterance-final position (items A1-A9 and B1-B9, in Appendix I), though it should be noted that most of the measurements show the same effects even when we consider the original 20-sentence sets; we briefly present results from the rest of our materials in the General Discussion section.

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).[4] Our results are based on

[4] Some of the details in this and the following Method subsections of Experiment 1 are taken almost verbatim from the report of our larger study of question intonation (Arvaniti et al., in press).

data from seven of these speakers. BG's speech rate was so fast and his pitch range so narrow that it was virtually impossible to make meaningful measurements of F0 peaks and valleys; for this reason, we decided to discard his data.

All speakers except VP 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 (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 substantially affected the variables under investigation here. All the speakers were in their twenties or thirties. At the time of the recording all 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 reported any speech or hearing problems and all of them, except AA, were naïve as to the purpose of the experiment. AA's data were included after preliminary statistical analysis had shown that they were not different from those of the naïve speakers.

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 what intonation they should use, though as mentioned the words under focus in contrastive statements were underlined on the cards read by the speakers. The recordings were monitored by the third author (a fluent speaker of Greek) and the speakers were asked to repeat any misread dialogue. In order to avoid "listing" effects, the cards with the test dialogues—which were read at the same session as the materials for the larger experiment—were interspersed with cards which contained filler mini-dialogues of similar structure. Each speaker's recording session lasted about an hour.

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 the rare tokens that were produced with an intonation contour or an accent placement other than the one intended (e.g., polar questions produced as statements), or with a stress placement other than the one indicated by the spelling (some verb forms can be stressed either on the penult or the antepenult in Greek; e.g. [aˈɣorasan] and [aɣoˈrasan] "they-bought" are both possible). The third author also discarded any disfluent utterances and utterances with long stretches of creaky voice which, as the F0 display showed, made it impossible to get reliable measurements of the "Final Low" (see below). After the data had been measured, a few sentences 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. Overall, approximately 85% of the sentences were used for statistical analysis, i.e. 34 utterances per speaker, or a total of 238 utterances (34 sentences × 7 speakers).

Measurements and statistics. Measurements of both duration and F0 were obtained for the test words using the facilities of ESPS Waves+ on a Sun SPARC workstation. F0 in particular was extracted using the pitch-tracking facility of Waves+ (with a 49 ms cos.[4] window moving in 10 ms frames). An illustration of our measurements in provided in Figure 3.

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]). F0 points thought to represent the tonal targets were also located manually. Measurements were made by the third author, except for L (see below) in the data of the four female speakers, which was measured by the second author.

The F0 points measured were the following:

- L (Low): The "elbow" created by the rise from low F0 to the peak of the rise-fall; L was labeled by eye, and was taken to be the point at which the low level stretch turned into a clear rise.
- H (High): The highest extracted F0 value at the peak of the rise-fall movement; in the rare cases in which there was a series of points with the same F0, the first one was chosen for measurement.
- FL (Final Low): The lowest extracted F0 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); this point typically coincided with the beginning of a low level F0 stretch, since such low F0 stretches were either flat or slightly rising (see Figures 1-3), but not falling.

The frequency of L, H and FL was measured in Hz and converted to the ERB scale (Glasberg and Moore, 1990). In addition, the distance of the targets was calculated from specific segmental landmarks. The following temporal distances were obtained:

- SC–L: the distance in ms between the consonant onset of the stressed syllable (SC) in the test word and the L target.
- SV–H: the distance in ms between the onset of the stressed vowel (SV) in the test word and the H target.
- FL–END: the distance in ms between the lowest F0 point after the peak (FL) and the end of the utterance.

//////////////////////////////////// FIGURE 3 APPROXIMATELY HERE ///

In addition, we measured the duration of the segments in the accented syllable, because we wanted to examine the effect of segment duration on tonal alignment. In particular, a known effect of emphasis is segmental lengthening (e.g. Klatt, 1976; Cummins, 1999), and it was reasonable to expect that such lengthening might occur on the accented words in our contrastive statements. If such lengthening did occur, then it could be problematic to measure the alignment of the H from the onset of the stressed vowel, as is typically done; previous research has shown that there is a complicated interplay between peak alignment and segmental duration (Silverman and Pierrehumbert, 1990; Schepman, Lickley and Ladd, 2006). Because of this potential confound, we also measured the alignment of H in two more ways:

- SVoffset–H: the distance of H from the offset of the test-word's stressed vowel.

- H prop: this was a measure of proportional alignment, defined as the distance of the H from the onset of the stressed vowel divided by the duration of this vowel.

As a check on the reliability of our measurement procedures, the third author re-measured the L target in 20 sentences from the set measured by the second author (12.5% of this set). The absolute difference between the two L measurements was calculated, and means and standard deviations were computed; the absolute mean difference was 1 ms (s.d. 33). The large standard deviation was most probably due to the fact that locating this target was not easy, as F0 sometimes showed a smooth rise from L to H without a clear inflection; thus the results for L should be treated with some caution.[5]

The measurements were subjected to analyses of variance with one repeated-measures factor, UTTERANCE TYPE, with two levels, polar question and contrastive statement. The data in the dependent variables were the means of each speaker's data.

Experiment 1: Results

Segment duration. As can be seen in Figure 4, the segments of the stressed syllable were longer in contrastive statements than in questions: the consonant of the stressed syllable in our target words was longer in contrastive statements by on average 17 ms [$F(1,6)=42.04$, $p<0.001$], while the vowel was longer by on average 24 ms [$F(1,6)=44.3$, $p<0.001$]; both differences are above JND (Klatt, 1976) and therefore we would expect them to be perceptible.

[5] Although these differences may cast some doubt on the validity of the L data, it should be noted that Arvaniti, Ladd and Mennen (in prep.), who performed both algorithmic determination and estimation by eye of a similar elbow, found that the two methods did not lead to substantially different conclusions.

////////////////////////////////////// FIGURE 4 APPROXIMATELY HERE ///

<u>Alignment of targets</u>. L showed no statistically significant effect of UTTERANCE TYPE: as shown in Figure 5(a), the average alignment of L was slightly before the consonant onset in questions, and slightly after it in statements, but this difference was not significant [$F(1,6)=2.09$, n.s.]. FL, on the other hand, was aligned later in questions, i.e. closer to the end of the utterance, than in statements [$F(1,6)=7.6$, $p<0.03$], as shown in Figure 5(b); this means that the low F0 stretch was longer in contrastive statements than in questions.

////////////////////////////////////// FIGURE 5 APPROXIMATELY HERE ///

As mentioned earlier, H alignment was expressed in three different ways, because, as we had anticipated, our data showed durational differences across utterance types due to the emphasis associated with contrastive focus. All three measurements revealed a consistent pattern, namely later peak alignment in questions than in statements, though for one of the measures the statistical comparison narrowly missed significance. As illustrated in Figure 6a, when H alignment was expressed with respect to the stressed vowel onset, it appeared to be aligned later in questions than in statements, though this result was not, strictly speaking, statistically significant [$F(1,6)=5.7$, $p<0.054$]. Nevertheless, the validity of the result is corroborated by the comparisons based on the other two measures. As shown in figure 6b, when the alignment of H was measured from the *offset* of the vowel, it was clear that H aligned much closer to the end of the vowel in questions than in contrastive statements [$(F1,6)=26.02$, $p<0.01$]. Similarly, proportional alignment results, illustrated in Figure 6c, showed that H occurs in the middle of the vowel in contrastive statements, while in questions it occurs roughly at three quarters of the vowel's duration [$F(1,6)=13.6$, $p<0.01$].

////////////////////////////////////// FIGURE 6 APPROXIMATELY HERE //

<u>*Scaling of targets.*</u> The scaling of the three targets showed no significant differences across UTTERANCE TYPE [for L, $F<1$; for H, $F<1$; for FL, $F(1,6)=1.6$, n.s.]. These results are shown in Figure 7.

////////////////////////////////////// FIGURE 7 APPROXIMATELY HERE //

Experiment 1: Summary and Interim Discussion

In summary, our results showed small but consistent differences of tonal alignment between the contrastive statement rise-fall and the rise-fall portion of polar question contours: both H and FL aligned later in questions than in statements, and only L showed no statistically significant alignment difference between the two types of utterance. In addition, our results showed strong effects of UTTERANCE TYPE on the duration of segments, with both the consonant and the vowel of the accented syllable in contrastive statements being longer than in questions. Tonal scaling, on the other hand, did not show a statistically significant effect of UTTERANCE TYPE.

Overall these results are in agreement with results previously reported for similar data. First, effects of emphasis on segmental duration are in agreement with previously reported results (e.g. Klatt, 1976; Cummins, 1999; Arvaniti and Garding, to appear). Similarly, the alignment results for L and H are compatible with other Greek data and data from other languages, which show that the alignment of L tones tends to be stable, while that of H tones is much more variable and influenced by several factors, such as speaking rate, proximity to other tones and phonological structure (on L tones, see e.g., Arvaniti and Ladd, 1995, and Arvaniti et al., 1998, for Greek; Ladd and Schepman, 2003, for British English; Arvaniti and Garding, to appear, for American English; for H tones, see Silverman and Pierrehumbert, 1990, for English; Prieto et al., 1995, for Spanish; Arvaniti et al., 1998, for Greek; Ladd et al., 2000, for Dutch).

However, it is important to note that in the present data the H is aligned *earlier* in the contrastive statements, which show longer vowel duration, and not in the questions, which have shorter stressed syllables. This pattern runs counter to previously reported results, which show that peaks tend to align later when the segments with which they associate are lengthened (Silverman and Pierrehumbert, 1990), especially if the lengthening is related to the presence of contrastive focus or emphasis (Ladd and Morton, 1997; Arvaniti and Garding, to appear). This discrepancy between our results and previous reports on peak alignment strongly suggests that the alignment difference we uncovered between polar questions and contrastive statements in Greek cannot simply be attributed to durational differences between the two utterance types.

Finally, the data for FL are in agreement with other reported data which show that some L tones may not be realized as local F0 minima, but as low F0 stretches (cf. Arvaniti and Garding, to appear; Arvaniti et al., in press). Most importantly, the longer duration of these stretches in contrastive statements supports our analysis of their melody as involving two L tones after the H, a L- phrase accent and a L% boundary tone, and our analysis of the questions as involving only one L tone after the H, a L% boundary tone. This difference is crucial for our interpretation of our experimental results, and we return to it in the General Discussion section.

EXPERIMENT 2: PERCEPTION

Experiment 2: Method

Stimuli. The stimuli were fragments extracted from the data of four of the speakers who had taken part in Experiment 1, DA, VP, KA (females) and AH (male). There were sixteen fragments from speakers DA, VP and AH, and eight fragments from speaker KA (whose data were used for a practice session). Half of the fragments for each speaker were extracted from polar questions and the other half from contrastive statements. The fragments were either single nouns (e.g., [ˈlivano] "Lebanon [accusative]") or complete constituents, such as [tin ˈiɾiða] "Iris [woman's name, accusative]"; all of them could be used as complete utterances. The same materials were used for all four speakers as far as this was feasible (90% of the stimuli involved the same materials across speakers). The fragments were excised from the original digital speech files using the standard Waves^{+} editor. When preparing the stimuli care was taken to avoid conspicuous transition effects at the beginning of the fragments, so that they would sound as natural as possible.

For one part of the experiment, each stimulus was resynthesized using its original F0 file. For the other part, the stimuli were resynthesized with an F0 contour that declined from 210 Hz to 190 Hz for the female speakers and from 125 Hz to 115 Hz for the male speaker (henceforth, we refer to these stimuli as *monotonized* although strictly speaking they had falling F0; for our purposes, what is important is that they showed no pitch inflection on the stressed syllable).

Resynthesis was done using a Waves$^+$-compatible pulse-excited linear prediction resynthesis program (written by Diego Molla Aliod and Steve Isard). The purpose of resynthesizing not only the stimuli with altered contours but also those with unmodified F0 was to ensure that the stimuli were comparable in overall quality.

The stimuli were recorded onto two cassette tapes in which speech from different speakers appeared in two different orders: DA, AH and VP in order A, and AH, VP and DA in order B. Each tape included a total of 112 stimuli (8 practice stimuli and 48 test stimuli, 16 per speaker, in each of the two parts). For each tape, the order of speakers was the same for the resynthesized and the monotonized part, and in both tapes, the latter followed the former, because we were concerned that counterbalancing might make the task too difficult for those who would hear the monotonized stimuli first. In all conditions contrastive statements and polar question fragments were presented in random order.

Both tapes started with the same instructions (see Appendix II, part A) and the same practice session consisting of the eight stimuli from KA mentioned above. In the beginning of the instructions, the similarities between contrastive statements (described as "emphatic" in the instructions) and the last part of polar questions were brought to the attention of the subjects by means of examples from our data. The subjects were told that they would hear words or short phrases that were part of longer contrastive statements or questions, and that their task was to choose for each fragment whether it was a contrastive statement or *the last part* of a question; they were asked to give an answer even if they were not sure of their choice. After hearing the stimuli with unmodified intonation, the listeners heard further instructions that prepared them for the monotonized stimuli (see Appendix II, part B). In these instructions they were told the sentences they were about to hear would sound monotonous ("as if spoken by a computer"), but their task was the same, i.e. to choose whether the fragments they heard were excised from questions or statements and to provide an answer even if they were not certain it was correct.

In both parts of the experiment the instructions were followed by 6 s of silence, after which the listeners heard a warning (two 100 Hz tones of 300 ms duration, separated by 100 ms and followed by 400 ms of silence). This warning was used both at the beginning of the practice session and at the beginning of each speaker's data. Each stimulus was preceded by a 100 Hz warning tone of 300 ms duration followed by 1 s of silence. There were 3 s of silence between the end of one stimulus and the beginning of the warning tone for the next stimulus.

Subjects. The experiment took place in the 36th Lyceum, in Athens. The listeners were a class of 26 fifteen-year-olds and a class of 18 seventeen-year-olds, who heard Order A and Order B respectively from a loudspeaker connected to a portable SONY analog tape recorder. In the class that heard order A, some of the subjects did not take their task very seriously and as a result eight response sheets contained several empty cells. These eight sheets were not included in the statistical analysis. Thus, the results reported here are based on the data from two groups of eighteen listeners each.

Procedures and statistical analysis. The listeners were provided with a questionnaire in Greek, with separate sheets for each speaker's stimuli. Each sheet presented two sentences for each stimulus, a polar question (i.e. a sentence ending in a question mark) and a contrastive statement (in which the word in focus was underlined); questions were on the left and contrastive statements on the right of a table. These sentences were not the same as the test sentences in Experiment 1 from which the stimuli were extracted; e.g. the fragment [tin ˈeɟina] was extracted from the question [ksanaˈpiɣate stin ˈeɟina] “Have you visited Aegina before?”, and from the contrastive statement [ap tin ˈeɟina] “From Aegina!”; however, the contexts provided in the perceptual experiment were [ɟiˈɾisane ap tin ˈeɟina] “Have they returned from Aegina?” and [ˈoçi | ap tin ˈeɟina] “No, from Aegina!” respectively. The listeners were told that each stimulus had been extracted from one of the two sentences on their sheet, and they were asked to tick or circle the sentence from which they thought each stimulus had been extracted.

The responses were statistically analyzed by means of a repeated-measures ANOVA with GROUP (subjects who heard Order A vs. subjects who heard Order B) as a between-subjects factor, and SPEAKER (DA, AH, VP), MANIPULATION (unmodified or monotonized F0) and STIMULUS ORIGIN (stimulus extracted from a question or a statement) as the within-subjects factors; the dependent variable was the percentage of correctly identified stimuli per listener.

Experiment 2: Results and Discussion

The results showed that the stimuli with unmodified F0 were identified correctly in a greater proportion of cases than the monotonized stimuli [$F(1,34)=96.8$, $p<0.001$]. This effect of MANIPULATION can be seen in Figure 8. The monotonized stimuli were correctly identified slightly better than chance (56%), suggesting that part of what distinguishes contrastive statements from (the final part of) polar questions may be durational and other differences.

However, pitch differences must play a larger role, since identification drops by 16% when the stimuli are monotonized. This result strongly suggests that the alignment differences discovered in our production experiment are perceptually relevant.

////////////////////////////////////// FIGURE 8 APPROXIMATELY HERE ///

The analysis showed no statistically significant difference between the two GROUPS of listeners [$F<1$], and no effect of STIMULUS ORIGIN [$F<1$] (i.e. stimuli extracted from questions and stimuli extracted from statements were identified equally well). SPEAKER, on the other hand, did have an effect [$F(2,68)=11.4$, $p<0.001$], but also interacted with STIMULUS ORIGIN [$F(2,68]=12.4$, $p<0.001$]; there was also a three-way interaction of SPEAKER, STIMULUS ORIGIN and MANIPULATION [$F(2,68)=11.4$, $p<0.001$]. Tukey HSD post-hoc tests on the SPEAKER main effect did not show any statistically significant differences in the identification rate of stimuli from different speakers. Similarly, the investigation of the STIMULUS ORIGIN × SPEAKER interaction showed no significant differences between stimuli extracted from questions and stimuli extracted from statements for any of the speakers. The three-way interaction, on the other hand, showed that on two occasions the difference in identification rate between unmodified and monotonized stimuli was not statistically significant; this was the case for DA's stimuli extracted from questions, and for VP's stimuli extracted from contrastive statements.

Overall, the results show that listeners can identify a rise-fall as pertaining to a polar question or a contrastive statement on the basis of F0 information, suggesting that the alignment differences detected between the two types of utterance in our production experiment are perceptually relevant. Nevertheless, it is clear that the identification rate is not high and that on occasion identification fails altogether. It is also undeniable that factors additional to F0—such as segmental duration, and perhaps even differences in segmental quality (cf. Fourakis, Botinis and Katsaiti, 1999; Erickson, 2002), and articulation (Erickson, 1998; Eriskson, Fujimura and Pardo, 1998) associated with emphasis—also contribute to identification, but these additional clues play a smaller part than F0.

GENERAL DISCUSSION

Our production experiment showed that the rise-fall of contrastive statements and the rise-fall portion of question contours in Greek, although superficially similar, exhibit consistent differences in alignment, particularly regarding the alignment of the F0 peak relative to the lexically stressed syllable, and the duration of the final low F0 stretch. Furthermore, our perception experiment showed that these differences can be used by listeners to identify a stretch of speech as being part of a question or a contrastive statement. While it may seem unsurprising that questions and statements should be distinguishable, it should be borne in mind that we were not comparing whole contrastive statements with whole questions, but whole contrastive statements with *parts* of questions. The motivation for our study was to see whether the difference between the hypothesized AM representations of these rise-fall contour fragments corresponds to any consistent phonetic difference, and whether any phonetic difference between them could be related to differences of phonological representation in a principled way.

In the introduction we discussed the proposal of Prieto et al. (to appear) to use what we called supplemental (secondary) association to distinguish between the phonological representations of pragmatically distinct but phonetically similar accent types in Catalan and other languages. One conceivable way of incorporating our findings into the phonological representation of the Greek contours under investigation would make use of this kind of secondary association. For example, we could define the alignment of the L+H* pitch accent in contrastive statements as reflecting the default primary association of the pitch accent to the stressed syllable of the word in focus, and we could use supplemental association to the right edge of the last stressed syllable to account for the later peak alignment of the L+H- phrase accent in the question contour. However, we think that a better explanation of our findings would be based on the different tonal context in which the L+H- and L+H* are found: the former occurs in the question tune L* L+H- L%, where the H tone is followed by one L% boundary tone, whereas the latter occurs in the contrastive statement tune L+H* L- L%, where the H tone is followed by two L tones, L- and L%.

The presence of the L tones affects the course of F0 after the peak (as shown, the low F0 stretch is longer in the statements, where it is the reflex of two L tones), but may also affect the position of the peak itself by causing "tonal crowding." Tonal crowding effects have long been reported in intonation studies (e.g. Bruce, 1977, ch. 5; Silverman and Pierrehumbert, 1990; Arvaniti et al., 1998, 2000; Nibert, 2000; Prieto, 2005; Schepman et al., 2006). They are typically observed when the number of tones in a tune exceeds the number of TBUs these tones can associate with. In such cases, several adjustments to the realization of tonal targets are possible (though the exact mechanisms that lead to these adjustments are not clearly understood, as Silverman and Pierrehumbert, 1990, point out): in some languages certain tones are truncated (e.g. Palermo Italian, Grice, 1995; Standard Hungarian, Ladd, 1996; Cypriot Greek, Arvaniti, 1998; Belfast English, Grabe et al., 2000), while in others, all tones are realized, but their phonetic realization is modified in some way, typically by either undershooting F0 levels or by adjusting tonal alignment in one direction or the other (e.g. Swedish, Bruce, 1977, ch. 5; Greek, Arvaniti et al., 1998, 2000; Catalan, Prieto, 2005). In the Greek data under investigation, tonal crowding affects the contrastive statements, in which four tones (L+H* L- L%) must co-occur with three TBUs (the last three syllables of the utterance). Tonal crowding is less of a problem in questions, in which only three tones (L+H- L%) must co-occur with the last three syllables. Because of this difference, it is possible that the earlier alignment of the peak in contrastive statements is not due to a difference in phonological association but is more simply the result of crowding from the tones following the peak.

If this alternative interpretation of our results is correct, then we should expect to find a difference between our two subsets of contrastive statements: those statements in which the nuclear accent is not utterance-final should show later peak alignment than those in which the nuclear accent is final, because in the former subset, the phrasal tones (L-L%) do not need to appear immediately after the L+H* accent, and therefore do not exert pressure on the accentual peak to align early. In order to test this hypothesis, we conducted repeated-measures analyses of variance on our entire set of contrastive statements using NUCLEAR WORD POSITION (final or non-final nuclear word) as the repeated-measures factor (i.e. we compared items B1-B9 to items B10-B20 in Appendix I); the alignment of the L and the H tones were the dependent variables. The results of these analyses support our tonal crowding interpretation: although no differences were found in the alignment of L, the H of L+H* aligns later when the nuclear word is not final than when it is [$F(1,6)=9.3$, $p<0.02$]; mean alignment from vowel onset is 59 ms when the nuclear word is final, and 78 ms when it is not. This difference is confirmed by the two other measures we used for H alignment: H aligns closer to the vowel offset and proportionately later when the nuclear word is not final [for alignment from vowel offset, $F(1,6)=15.3$, $p<0.01$; mean alignment is 70 ms for utterance-final nuclear words, and 39 ms for non-final nuclear words; for proportional alignment, $F(1,6)=14.8$, $p<0.01$; mean alignment is 46% for utterance-final nuclear words, and 66% for non-final nuclear words]. In short, the peak of L+H* is aligned earlier relative to the accented vowel onset when L+H* is immediately followed by the phrasal tones, L-L%, than when these tones occur later in the utterance.

This difference in the alignment of the L+H* peak raises the following question: if the default peak alignment for the L+H* accent is the one observed in contrastive statements in which the nucleus is on a non-final word—a reasonable expectation, since in such cases there is no tonal crowding involved—then could it be the case that the accentual L+H* of statements is not aligned any differently from the phrasal L+H- of questions when neither is crowded by following tones? In order to provide an answer, we ran further analyses of variance on the alignment of H comparing contrastive statements with non-utterance-final nuclear words with the polar questions that contained the same words in utterance-final position using UTTERANCE TYPE (question or statement) as a repeated-measures factor (i.e. we compared items B10-B20 to items A10-A20 in Appendix I). In this case there were no statistically significant differences between contrastive statements and questions regarding either the absolute or the proportional alignment of the H tone.

In short, our original results and the additional results reported here strongly support an analysis in which the peak alignment differences found between polar questions and contrastive statements in utterance-final position are due to tonal crowding effects. This in turn suggests that we do not need to use supplemental association to account for the differences between L+H* and L+H-, since these two tonal entities are realized in similar fashion under comparable prosodic conditions, i.e. when minimal tonal crowding is involved. Further, our results suggest that the structural difference between the phrasal L+H- of questions and the accentual L+H* of contrastive statements need not underlie a difference in the phonetic realization of these tonal entities, when they occur in comparable contexts and they are both aligned with stressed syllables (as happens when the secondary association of the phrasal L+H- takes effect). On the other hand, our results show that even though we are dealing here with a context-dependent difference in the realization of L+H* and L+H-, this type of phonetic detail is clearly perceptually relevant. In particular, it is not unlikely that small-scale phonetic effects, like those observed in our data, are used during speech processing and contribute to the naturalness of speech (cf. Whalen et al., 1997). Thus, details like those discovered here cannot be dismissed as noise, but should be specified in a full description of the phonetic implementation of tunes.

Although the foregoing discussion makes it clear that it is possible to treat the alignment differences between the L+H- of questions and the L+H* of contrastive statements without recourse to supplemental association, it is worth considering how our analysis fits into the overall intonational system of Greek. As argued in the Introduction, taking the whole system into consideration is important for any analysis, but this is particularly so in our case because the Greek intonational system includes another L+H sequence, the default prenuclear pitch accent discussed by Arvaniti et al. (1998, 2000). This accent also consists of a L and H tone, with the L aligning at or just before the beginning of the stressed syllable (like the two patterns discussed in this paper) but with the H aligning early in the vowel of the post-tonic syllable. Since this accent aligns differently and is pragmatically distinct from the contrastive statement accent (in that it cannot be used to signal narrow focus; Arvaniti and Baltazani, 2005), it would appear to require a different phonological representation. One such possible representation would make use of the supplemental association mechanism. In this case, the prenuclear accent could be analyzed as a L+H* with a supplemental association of its H tone to the left edge of the first post-tonic mora. This representation would work well with our analysis of the contrastive statement L+H* and the analysis of the question L+H-. Nevertheless, the most parsimonious representation would be to represent the prenuclear accent as L*+H, thereby using the "star" notation as originally intended by Pierrehumbert (1980), to signal a phonological distinction of alignment within a single language (this analysis of the prenuclear and contrastive accents of Greek as L*+H and L+H* respectively was originally proposed by Arvaniti and Baltazani, 2000, 2005; a similar analysis is also adopted by Beckman et al., 2002, for Spanish, which has rising accents comparable to the prenuclear and contrastive accent of Greek).

Similarly, alternative analyses that rely on tonal composition and differences in starredness alone may also apply to some of the data of Prieto et al. (to appear). According to these authors, the two rising accents of Pisan Italian have distinct phonological representations, H* (or L+H*) for the accent with late peak, and H*+L or (L+H*+L) for the accent with early peak, so their differences in alignment could follow from their representations and need not be encoded in the phonology by means of supplemental association. In addition, in Neapolitan Italian the accent represented as L*+H does not appear to be in contrast with another L*+H accent in this system and thus should not require a supplemental association of its L* to the first mora of the accented syllable.

CONCLUSION

Our production and perception data on the rising-falling contours of Greek polar questions and contrastive statements have shown that these superficially similar configurations exhibit small-scale differences in production, which, however, are sufficiently large to be perceptually relevant. The differences we found are consistent with the AM analyses we proposed for these contours, L+H* L-L% for contrastive statements and L+H- L% for polar questions, in which the phrasal L+H- has a secondary association to the last stressed syllable. These autosegmental analyses and our results provide an explanation for the peak alignment differences we found by showing that they can plausibly be attributed to the greater tonal crowding that the presence of the two phrasal L tones exerts on the peak in contrastive statements. Thus, we conclude that the autosegmental representation of these data does not require use of a supplemental (secondary) association, as suggested for similar cases in Italian and Catalan by Prieto et al. (to appear). Our results lead us to propose that the autosegmental representations of intonation need not be more phonetically precise and transparent than the phonological representation of segments. Thus, differences of phonological representation, if they are posited for the primary purpose of expressing matters of phonetic detail, should be used with caution, and only when other interpretations that take into account the context in which the accents appear and the overall intonational system they are part of can be shown not to adequately represent a given accentual contrast.

REFERENCES

ARVANITI, A. (1998). Phrase accents revisited: comparative evidence from Standard and Cypriot Greek. *Proceedings of the 4th International Conference on Spoken Language Processing*, vol. 7, pp. 2883-2886. Sydney.

ARVANITI, A. (2003). Peak scaling in Greek and the role of declination. *Proceedings of 15th International Congress of Phonetic Sciences*, pp. 2269-2272. Barcelona: Universidad Autónoma de Barcelona.

ARVANITI, A., & BALTAZANI, M. (2000). GREEK ToBI: a system for the annotation of Greek speech corpora. *Proceedings of the Second International Conference on Language Resources and Evaluation*, vol. II, pp. 555-562. Athens: European Language Resources Association.

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

ARVANITI, A., & GARDING, G. (to appear). Dialectal variation in the rising accents of American English. In J. Hualde and J. Cole (Eds.), *Laboratory Phonology 9*. Berlin: Mouton de Gruyter.

ARVANITI, A., & LADD, D. R. (1995). Tonal alignment and the representation of accentual targets. *Proceedings of the 13th International Congress of Phonetic Sciences*, vol. 4, pp. 220-223. Stockholm.

ARVANITI, A., LADD, D. R., & MENNEN, I. (1998). Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics,* 26, 3-25.

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

ARVANITI, A., LADD, D. R., & MENNEN, I. (in press). Phonetic effects of focus and "tonal crowding" in intonation: evidence from Greek polar questions. *Speech Communication*.

ARVANITI, A. & LADD, D. R. & MENNEN, I. (in preparation). Scaling and alignment of pitch targets in Modern Greek wh-question intonation.

ATTERER, M., & LADD, D. R. (2004). On the phonetics and phonology of "segmental anchoring" of F0: evidence from German. *Journal of Phonetics*, 32, 177-197.

BALTAZANI, M. (2002). *Quantifier Scope and the Role of Intonation in Greek*. Unpublished Ph.D. dissertation, University of California, Los Angeles.

BALTAZANI, M., & JUN, S. (1999). Focus and topic intonation in Greek. *Proceedings of the 14th International Congress of Phonetic Sciences*, pp. 1305-1308. San Francisco.

BECKMAN, M. E., DÍAZ-CAMPOS, M., TEVIS MCGORY, J., & MORGAN, T. A. (2002). Intonation across Spanish, in the Tones and Break Indices framework. *Probus,* 14, 9-36.

BECKMAN, M. E., & PIERREHUMBERT, J. (1986). Intonational structure in English and Japanese. *Phonology Yearbook,* 3, 255-310.

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

BOTINIS, A., BANNERT, R., & TATHAM, M. (2000) Contrastive tonal analysis of focus perception in Greek and Swedish. In A. Botinis (Ed.), *Intonation: Analysis, Modelling and Technology* (pp. 97-116). Dordrecht: Kluwer Academic Publishers.

BRUCE, G. (1977). *Swedish Word Accents in Sentence Perspective*. Gleerup: Lund.

CUMMINS, F. (1999). Some lengthening factors in English speech combine additively at most rates. *The Journal of the Acoustical Society of America,* 105, 476-480.

ERICKSON, D. (1998). Effects of contrastive emphasis on jaw opening. *Phonetica*, 55, 147-169.

ERICKSON, D. (2002). Articulation of extreme formant patterns for emphasized vowels. *Phonetica*, 59, 134-149.

ERICKSON, D., FUJIMURA, O., & PARDO, B. (1998). Articulatory correlates of prosodic control: emotion and emphasis. *Language and Speech*, 41, 399-417.

FOUGERON, C., & JUN, S. (1998). Rate Effects on French Intonation: Prosodic Organization and Phonetic Realization. *Journal of Phonetics*, 26, 45-69.

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

FOURAKIS, M., BOTINIS, A., & KATSAITI, M. (1999). Acoustic characteristics of Greek vowels, *Phonetica*, 56, 28-43.

GANDOUR, J., POTISUK, S., & DECHONGKIT, S. (1994). Tonal coarticulation in Thai, *Journal of Phonetics*, 22, 477-492.

GLASBERG, B. R., & MOORE, B. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47, 103-138.

GRABE, E. (1998). Pitch accent realization in English and German. *Journal of Phonetics*, 26, 129-143.

GRABE, E., POST, B., NOLAN, F., & FARRAR, K. (2000). Pitch accent realization in four varieties in British English. *Journal of Phonetics*, 28, 161-185.

GRICE, M. (1995). Leading tones and downstep in English. *Phonology*, 12, 183-233.

GRICE, M., LADD, D. R., & ARVANITI, A. (2000). On the place of “phrase accents” in intonational phonology. *Phonology*, 17, 143-185.

GUSSENHOVEN, C. (2000). The boundary tones are coming: On the non-peripheral realization of boundary tones. In M. Broe & J. Pierrehumbert (Eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon* (pp. 132-151). Cambridge: Cambridge University Press.

KELLER, F. & ALEXOPOULOU, T. (2001). Phonology competes with syntax: experimental evidence for the interaction of word order and accent placement in the realization of Information Structure. *Cognition*, 79, 301-372.

KLATT, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, 59, 1208-1221.

LADD, D. R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.

LADD. D. R. (2000). Bruce, Pierrehumbert, and the elements of intonational phonology. In M. Horne (Ed.), *Prosody, Theory and Experiment: Studies presented to Gösta Bruce* (pp. 37-50). Dordrecht: Kluwer.

LADD, D. R., MENNEN, I., & SCHEPMAN, A. (2000). Phonological conditioning of peak alignment in rising pitch accents in Dutch. *The Journal of the Acoustical Society of America*, 107, 2685-2696.

LADD, D. R., & MORTON, R. (1997). The perception of intonational emphasis: continuous or categorical?. *Journal of Phonetics,* 25, 313-342.

LADD, D. R, & SCHEPMAN, A. (2003). “Sagging transitions” between high pitch accents in English: Experimental evidence. *Journal of Phonetics*, 31, 81–112.

LANIRAN, Y. O. (1992). *Intonation in Tone Languages: The Phonetic Implementation of Tones in Yoruba*. Unpublished Ph.D. dissertation, Cornell University.

LANIRAN, Y. O., & CLEMENTS, G. N. (2003). Downstep and high raising: Interacting factors in Yoruba tone production. *Journal of Phonetics*, 31, 203-250.

LICKLEY, R. J., SCHEPMAN, A., LADD, D. R. (2005). Alignment of "phrase accent" lows in Dutch falling rising questions: Theoretical and methodological implications. *Language and Speech*, 48, 157-183.

MÜCKE, D., & GRICE, M. (2004). Gestural anchors for F0 targets in German speech: an EMA-study. Paper presented at the Tone and Intonation in Europe Conference, Santorini, Greece 9-11 September, 2004.

MYERS, S. (1999). Tone association and F0 timing in Chichewa. *Studies in African Linguistics*, 28, 215-239.

MYERS, S. (2003). F0 Timing in Kinyarwanda. *Phonetica*, 60, 71-97.

NESPOR, M., & VOGEL, I. (1986). *Prosodic Phonology*. Dordrecht: Foris.

NIBERT, H. J. (2000). *Phonetic and Phonological Evidence for Intermediate Phrasing in Spanish Intonation*. Unpublished Ph.D. dissertation, University of Illinois, Urbana-Champaign.

PETERSON, G. E., & LEHISTE, I. (1960). Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America*, 32, 693-703.

PIERREHUMBERT, J. (1980). The Phonology and Phonetics of English Intonation. PhD Dissertation, MIT.

PIERREHUMBERT, J., & BECKMAN, M. (1988). *Japanese Tone Structure*. Cambridge, MA: The MIT Press.

PRIETO, P. (2005). Stability effects in tonal clash contexts in Catalan. *Journal of Phonetics*, 33, 215-242.

PRIETO, P., van SANTEN, J., & HIRSCHBERG, J. (1995). Tonal alignment patterns in Spanish. *Journal of Phonetics*, 23, 429-451.

PRIETO, P., D'IMPERIO, M., & GILI FIVELA, B. (to appear). Prenuclear and nuclear rising accents in Romance: tonal alignment and primary and secondary associations with metrical structure. *Language and Speech.*

SCHEPMAN, A., LICKLEY, R., LADD, D. R. (2006). Effects of vowel length and "right context" on the alignment of Dutch nuclear accents. *Journal of Phonetics,* 34, 1-28.

SILVERMAN, K., & PIERREHUMBERT, J. (1990). The timing of prenuclear high accents in English. In J. Kingston and M. E. Beckman, M. E. (Eds.), *Papers in Laboratory Phonology I: Between the grammar and physics of speech* (pp. 72-106). Cambridge: Cambridge University Press.

SUNDBERG, J. (1979). Maximum speed of pitch changes in singers and untrained subjects. *Journal of Phonetics*, 7, 71-79.

WELBY, P. (2004). The structure of French intonational rises: A study of text-to-tune alignment. *Proceedings of the Speech Prosody 2004 Conference*, pp. 127-130. Nara, Japan.

WHALEN, D. H., BEST, C. T., & IRWIN, J. R. (1997). Lexical effects in the perception and production of American English /p/ allophones. *Journal of Phonetics*, 25, 501-528.

WICHMANN, A., HOUSE, J., & RIETVELD, T. (2000). Discourse constraints on F_0 peak timing in English. In A. Botinis (Ed.), *Intonation: Analysis, Modelling and Technology* (pp. 163-182). Dordrecht: Kluwer Academic Publishers.

XU, Y. (1994). Production and perception of coarticulated tones. *The Journal of the Acoustical Society of America*, 95 (4), 2240-2253.

XU, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, 25, 61-83.

XU, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica*, 55, 179-203.

XU, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics*, 27, 55-105.

XU, Y. (2001). Fundamental frequency peak delay in Mandarin. *Phonetica*, 58, 26-52.

XU, Y., & WANG, Q. E. (2001). Pitch targets and their realization: evidence from Mandarin Chinese. *Speech Communication*, 33, 319-337.

XU, Y., & SUN, X. (2002). Maximum speed of pitch change and how it may relate to speech. *The Journal of the Acoustical Society of America*, 111 (3), 1399-1413.

APPENDIX I: EXPERIMENT 1 MATERIALS

Test materials are presented in broad IPA transcription, together with a gloss. For reasons of space, only the yes/no questions and contrastive statements are presented, not the entire dialogs they were part of. The words with contrastive accent are underlined and they are shown in small capitals in the gloss.

A: Polar questions

1. [tin anaɣnoˈɾisate tin ˈiɾiða] “Did you recognize Iris?”
2. [ksanaˈpiɣate stin ˈeʝina] “Have you visited Aegina before?”
3. [ti ɣnoˈɾizete tin ˈelena] “Have you met Elena?”
4. [ton ˈiðes to xaˈɾiðimo] “Have you seen Charidimos?”
5. [aˈliθça to noˈmizune] “Do they really believe that?”
6. [θa ˈfiɣun apo to ˈlivano] “Will they leave Lebanon?”
7. [aɾˈʝi to laðoˈlemono] “Is the oil-and-lemon sauce going to take long?”
8. [θa me siˈkosi to moˈnoziɣo] “Will the horizontal bar bear me?”
9. [siˈniɣoɾo se ˈvalane] “Why are you defending them?” [*lit.* Did they appoint you their councilor?]
10. [su aˈɾesun ta ksiˈnomila] “Do you like crab-apples?”
11. [peˈɾisepse ɾiˈzoɣalo] “Is there any rice-pudding left?”
12. [ma oɾiˈmasan ta koˈɾomila] “But are the plums ripe?”
13. [ˈpame tuɾkoˈlimano] “Shall we go to Turkolimano?”

14. [ma ˈeɾikse çoˈnoneɾo] “But was it raining sleet?”

15. [θa ˈftasi to ɾoˈðoneɾo] “Is there enough rosewater?”

16. [θa ton pɾoˈlavume to ˈviɾona] “Will we get there before Vironas?”

17. [taˈɣoɾasan ta oˈmoloɣa] “Did they buy the bonds?”

18. [to θiˈmase to ˈnumeɾo] “Do you remember the number?”

19. [ton epiˈskevasan to neˈɾomilo] “Did they repair the watermill?”

20. [aˈvɣa su kaθaˈɾizune] “What’s so funny about it?” [*lit*. Is someone peeling eggs for you?]

B: Contrastive statements

1. [me tin ˈiɾiða] “With IRIS!”

2. [ap tin ˈeɟina] “From AEGINA!”

3. [tin ˈelena] “ELENA!”

4. [xaˈɾiðimo] “CHARIDIMOS!”

5. [noˈmizune] “So they-THINK!”

6. [sto ˈlivano] “In LEBANON!”

7. [laðoˈlemono] “OIL-AND-LEMON SAUCE!”

8. [sto moˈnoziɣo] “On-the HORIZONTAL BAR!”

9. [me ˈvalane] “I was APPOINTED!” [*lit*. I was put]

10. [ksiˈnomila ˈine] “They-are CRAB-APPLES!”

11. [ɾiˈzoɣalo ˈitane] “It-was RICE-PUDDING!”

12. [ko'ɾomila 'ine] "They're PLUMS!"

13. [sto turko'limano 'ine] "It's in TURKOLIMANO!"

14. [ço'nonero 'rixni] "It's raining SLEET!"

15. [ɾo'ðonero 'vazo] "I-put ROSEWATER!"

16. [to 'viɾona ka'le] "VIRONAS silly!"

17. [o'moloɣa θa 'pername] "We would buy BONDS!"

18. ['numeɾo 'ine] "He-is RIDICULOUS!" [*lit*. number he-is]

19. [ane'momilos 'ine] "It's a WINDMILL!"

20. ['liʝi 'ɾiɣani 'vazo] "I just use some OREGANO!"

APPENDIX II: INSTRUCTIONS FOR EXPERIMENT 2

The instructions are translated from the Greek original. Transcriptions in square brackets indicate that a spoken example was heard; bracketed text in italics explains what these examples were; other text in italics consists of transliterations of Greek examples.

A: Unmodified stimuli

When we speak, it is often the case that a single word or a very short phrase can replace a complete sentence. We can, for example, use somebody's name to answer or pose a question. In many cases, we can even tell a question from a statement only by the way a word is pronounced. We say, e.g., [tin ˈelena] [*default declarative intonation*] to answer a question, but [tin ˈelena] [*polar question*] when we want to ask a question. Sometimes, however, such a short phrase can be the last part of a larger utterance, the beginning of which we have missed (perhaps because somebody else was speaking at the same time, or because there was a problem with the telephone line etc.) In such cases, the phrase we hear may be pronounced differently from a phrase that is by itself a whole utterance. This is particularly true of questions. If, for example, we want to use the same phrase, *tin Elena*, on its own in order to ask a question, then we pronounce it like this: [tin ˈelena] [*polar question*]. But when the same word is found at the end of a larger question, then we utter it as follows: [tin ˈelena] [*end of an early nucleus question from the data*]. You will better understand what we mean if you first listen to the whole question and then to the last word on its own: [ti ɣnoˈrizete tin ˈelena] [*entire question*] – [tin ˈelena] [*last part of the same question*]. On its own, the last word of this question does not sound the same as a one-word question. In reality, it sounds a bit like the end of an emphatic statement, the kind of sentence we use, e.g., when we correct somebody, as in *Ochi Eleni, Elena!* "(Her name is) not Eleni, (it's) Elena" [ˈelena] [*fragment of a contrastive statement from the data*].

In this experiment, you will hear a series of isolated words or short phrases. In reality they are all part of larger sentences the first part of which has been cut off, so you will hear only their last part. What we ask you to do is to decide if each word or phrase that you hear is the end of a question or an emphatic statement. The questionnaire in front of you shows two sentences for each word or phrase you will hear, a question and an emphatic statement from which the word or phrase has been excised. In order to choose, please underline the sentence from which you think the word or phrase you heard has been extracted. Try to answer even if you are not sure of your choice. Even though you may think you are responding at random, it is likely that this is not happening. You have to provide an answer to every question or phrase that you hear. We will start with a short practice session.

B: Monotonized stimuli

You have now finished the first part of the experiment. In the second part we ask you to perform a somewhat more difficult task. We have used exactly the same words and phrases as before, only this time we have manipulated them in such a way that they sound monotonous (as if spoken by a computer). For example, a phrase like [tin ˈelena] [*contrastive statement from the data*] has been turned into [tin ˈelena] [*same phrase, monotonized*]. Your task is the same as before, that is to decide whether the word or phrase you hear was originally part of a question or emphatic statement. As in the first part, in each case we provide you with two possible sentences. What you need to do is underline the sentence from which you think the phrase you heard came. As in the first part of the experiment, you must provide an answer to each word or phrase you will hear, even if you are not certain of your choice.

FIGURE TITLES

Figure 1. In panel (a), spectrogram and F0 contour of the sentence [θa ˈfiɣun apo to ˈlivano] "Will they LEAVE Lebanon?" with nucleus on the first content word [θa ˈfiɣun] "(they) will leave"; in panel (b), spectrogram and F0 contour of the sentence [ˈvazis ðedroˈlivano] "Do you use ROSEMARY?" with nucleus on the last content word [ðedroˈlivano] "rosemary." The vertical lines indicate syllable boundaries. The autosegmental representation of the tunes is given in the figures (note that in panel (b) there is an additional prenuclear L*+H accent on the first content word [ˈvazis]). The data in panel (b) are from the corpus of Arvaniti et al. (in press).

Figure 2: In panel (a), spectrogram and F0 contour of the sentence [sto ˈlivano] "In LEBANON!" in which the nuclear accent L+H* is immediately followed by L-L% phrasal tones; in panel (b), spectrogram and F0 contour of the sentence [sto turkoˈlimano ˈine] "It's in TURKOLIMANO [district of the city of Piraeus]!" in which the nuclear accent is on the first word of the utterance. The vertical lines indicate syllable boundaries. The autosegmental representation of the tunes is given in the figures.

Figure 3. In panel (a) spectrogram and F0 contour of [tin ˈiɾiða] "Iris [woman's name, accusative]" uttered as the final part of a polar question; in panel (b) spectrogram and F0 contour of the same phrase uttered as a contrastive statement (note that the coda [n] of the article [tin] resyllabifies [Nespor and Vogel, 1986], so it is considered here—and in other similar cases in our data—to create a CV syllable with the stressed vowel [ˈi]). The vertical lines indicate the segmental landmarks used for measuring alignment (i.e. the segment boundaries in the stressed

syllable and the end of the utterance). Arrows indicate the position of the tonal targets, L, H and FL.

Figure 4. In panel (a), means (in ms) and standard errors of the duration of the consonant at the beginning of the stressed syllable in contrastive statements (CS) and polar questions (PQ); in panel (b), means (in ms) and standard errors of stressed vowel duration in the two utterance types.

Figure 5: In panel (a), means (in ms) and standard errors of L (Low) alignment with respect to the onset of the stressed syllable in contrastive statements (CS) and polar questions (PQ); in panel (b), means (in ms) and standard errors of FL (Final Low) alignment with respect to the end of the utterance in the two utterance types.

Figure 6: In panel (a), means (in ms) and standard errors of H (High) alignment with respect to the onset of the stressed syllable's vowel, in contrastive statements (CS) and polar questions (PQ); in panel (b), means (in ms) and standard errors of H alignment with respect to the offset of the stressed syllable's vowel; in panel (c), means (in ms) and standard errors of H alignment with respect to the stressed vowel onset expressed as a percentage of the vowel's duration.

Figure 7: In panel (a), means (in ERB) and standard errors of L (Low) scaling in contrastive statements (CS) and polar questions (PQ); in panel (b), means (in ERB) and standard errors of H (High) scaling; in panel (c), means (in ERB) and standard errors of FL (Final Low) scaling.

Figure 8: Percentages and standard errors of correctly identified stimuli, as a function of MANIPULATION.

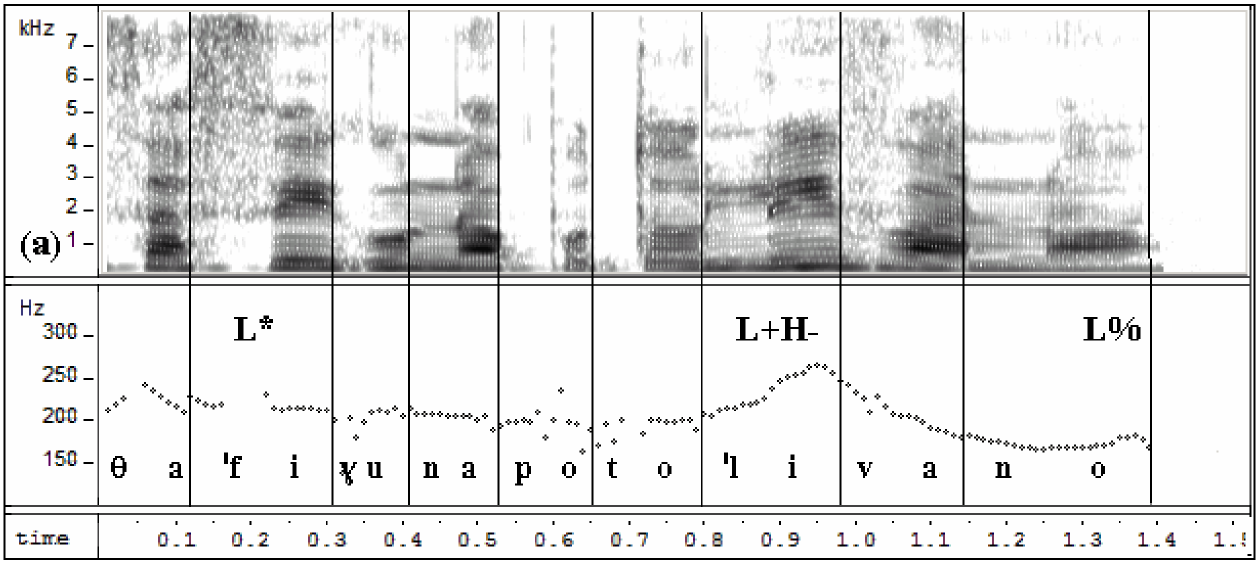

FIGURE 1a

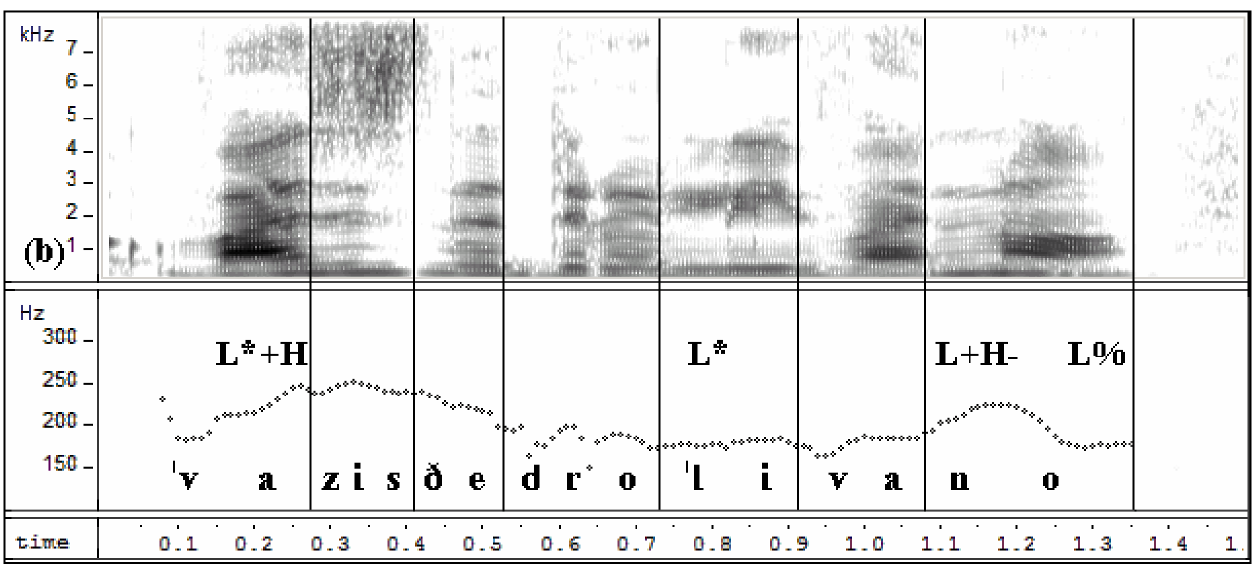

FIGURE 1b

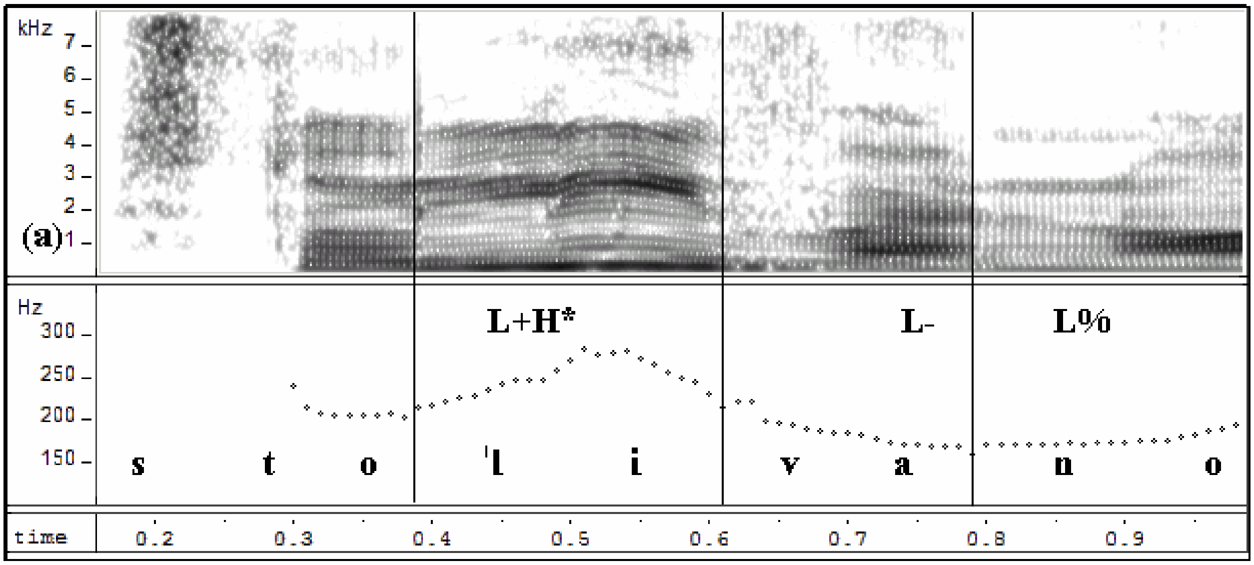

FIGURE 2a

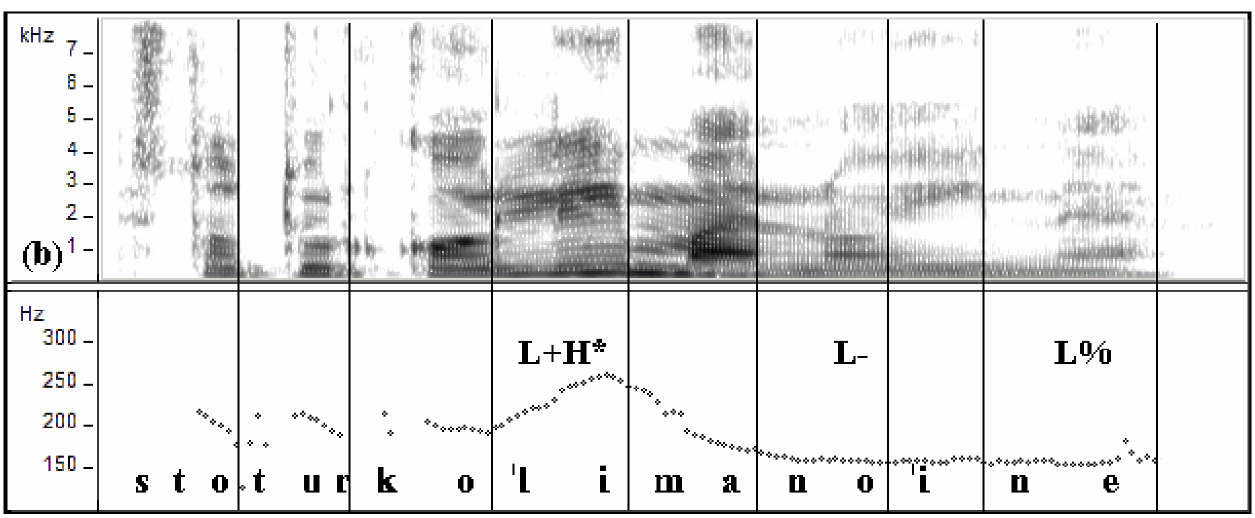

FIGURE 2b

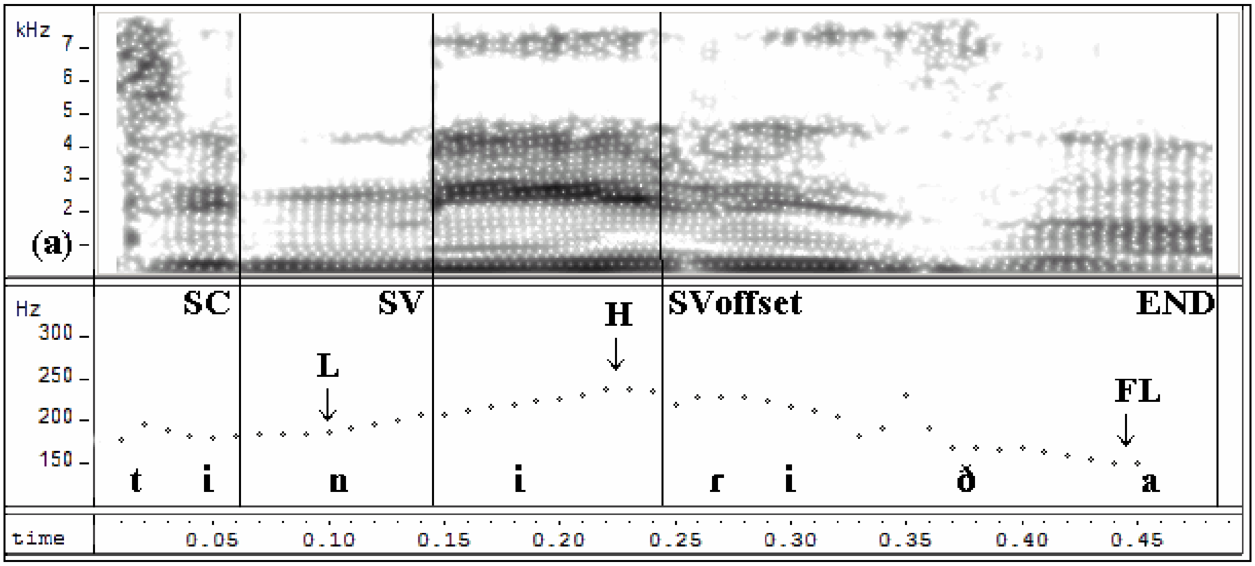

FIGURE 3a

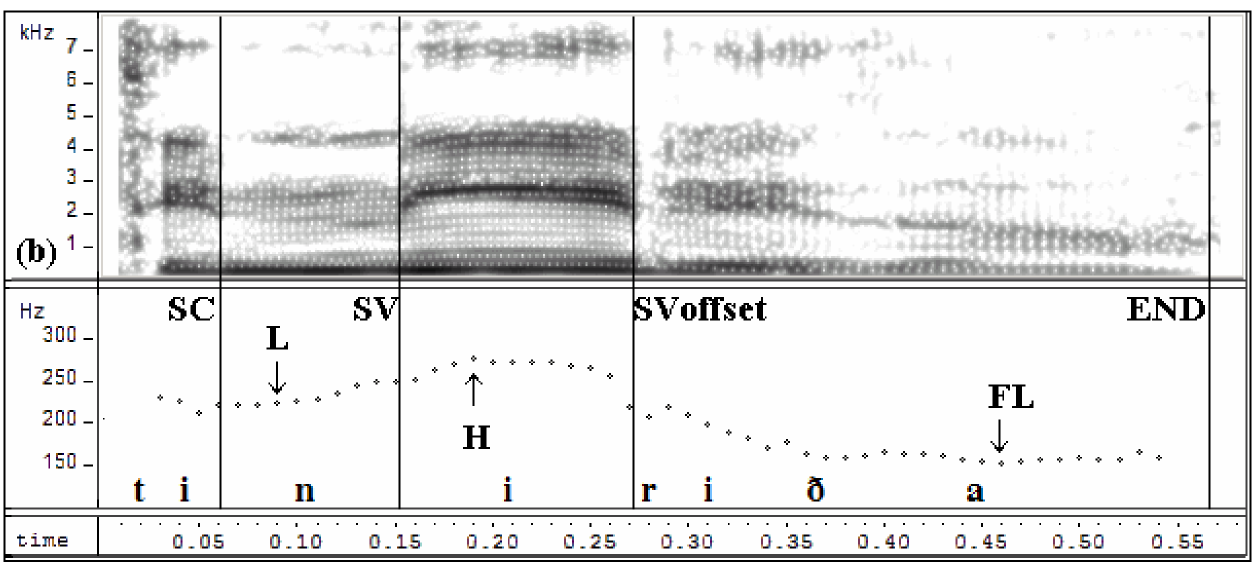

FIGURE 3b

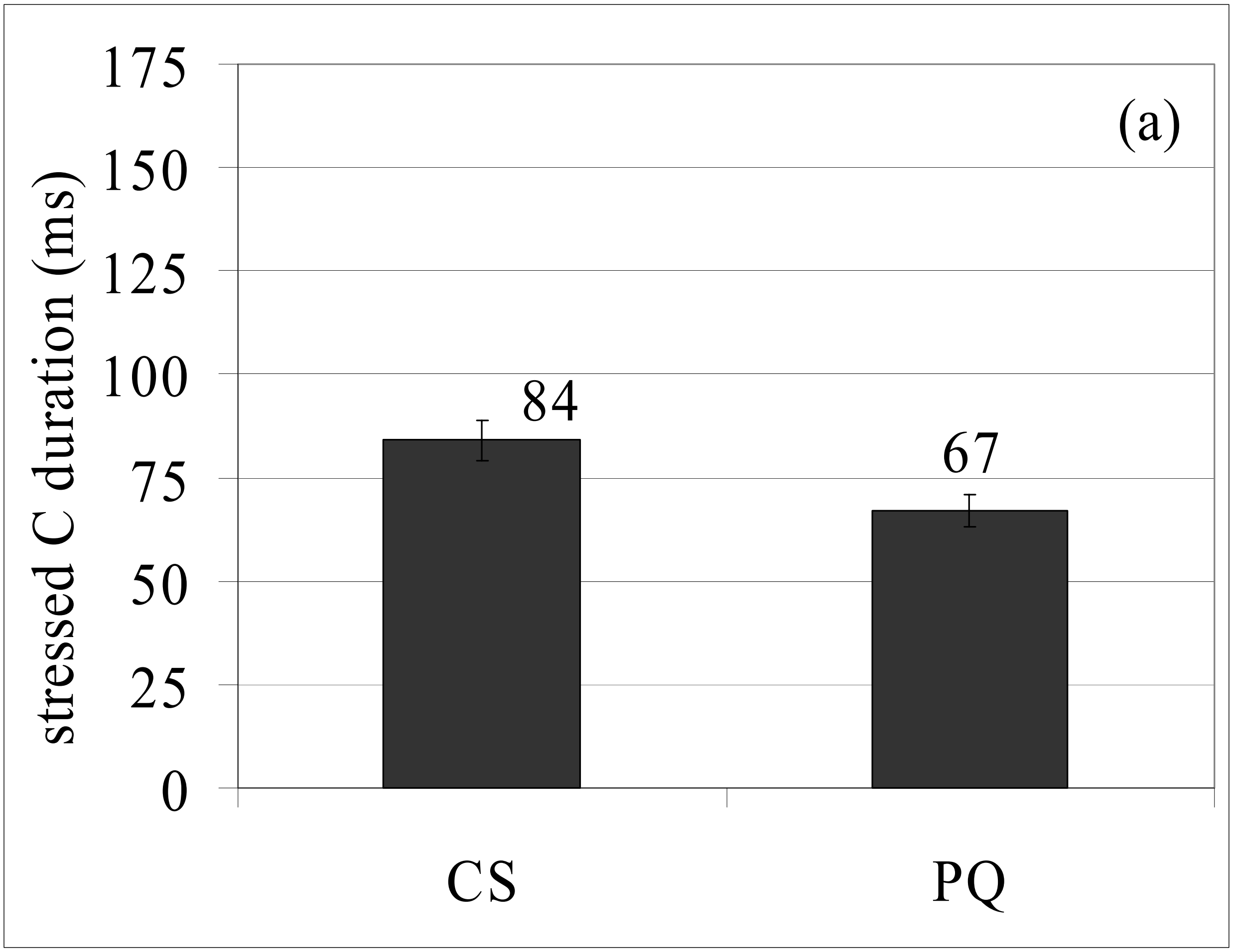

FIGURE 4a

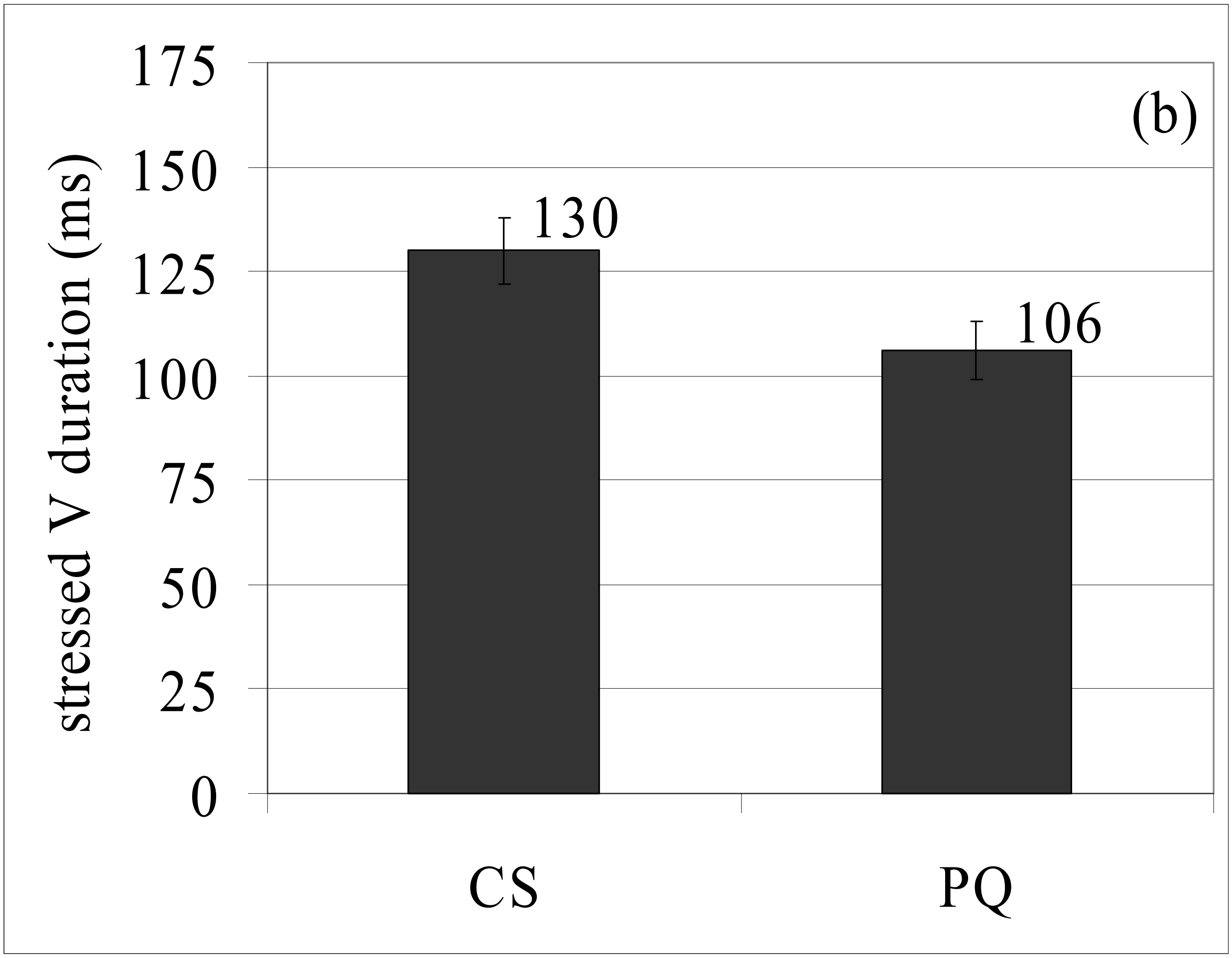

FIGURE 4b

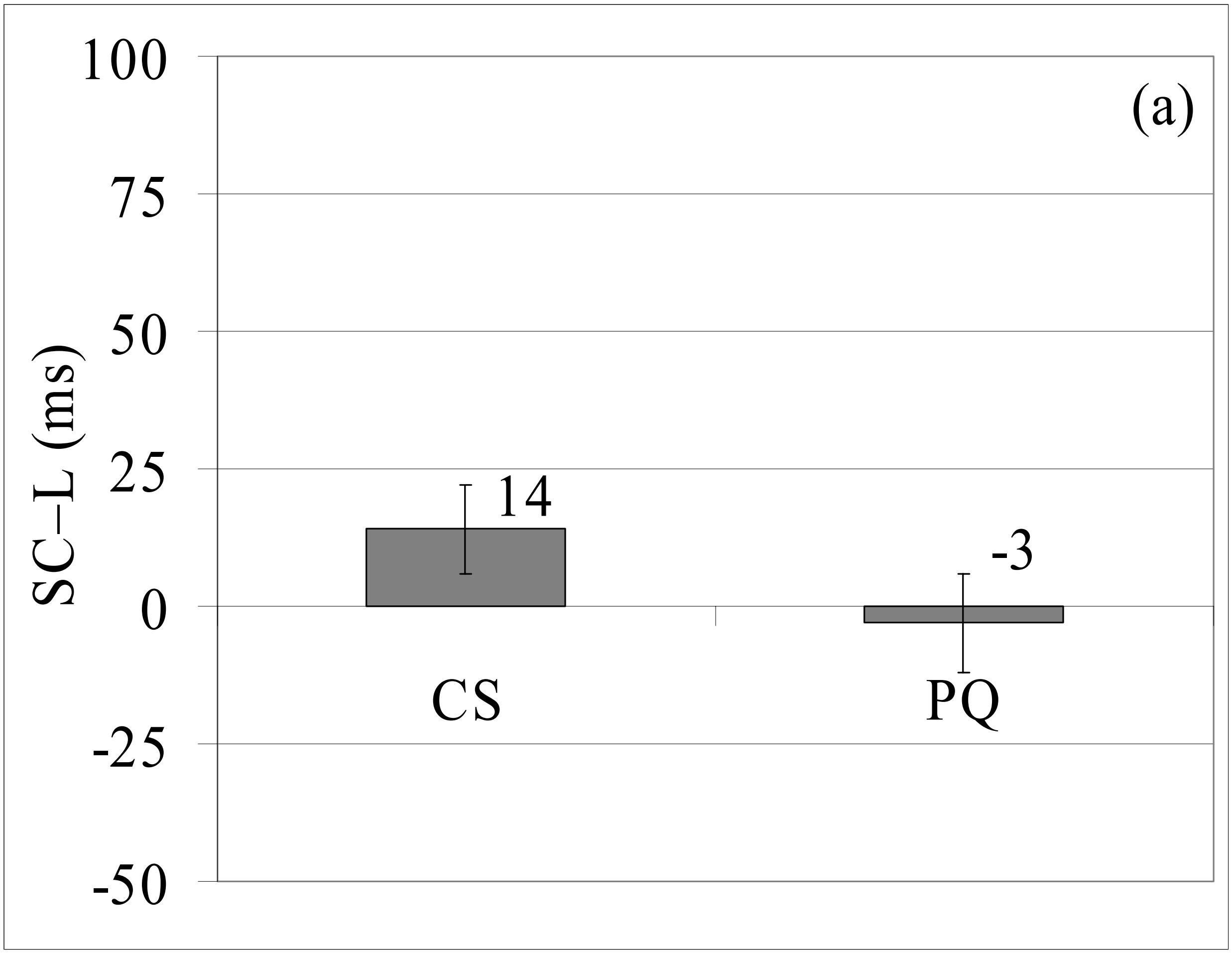

FIGURE 5a

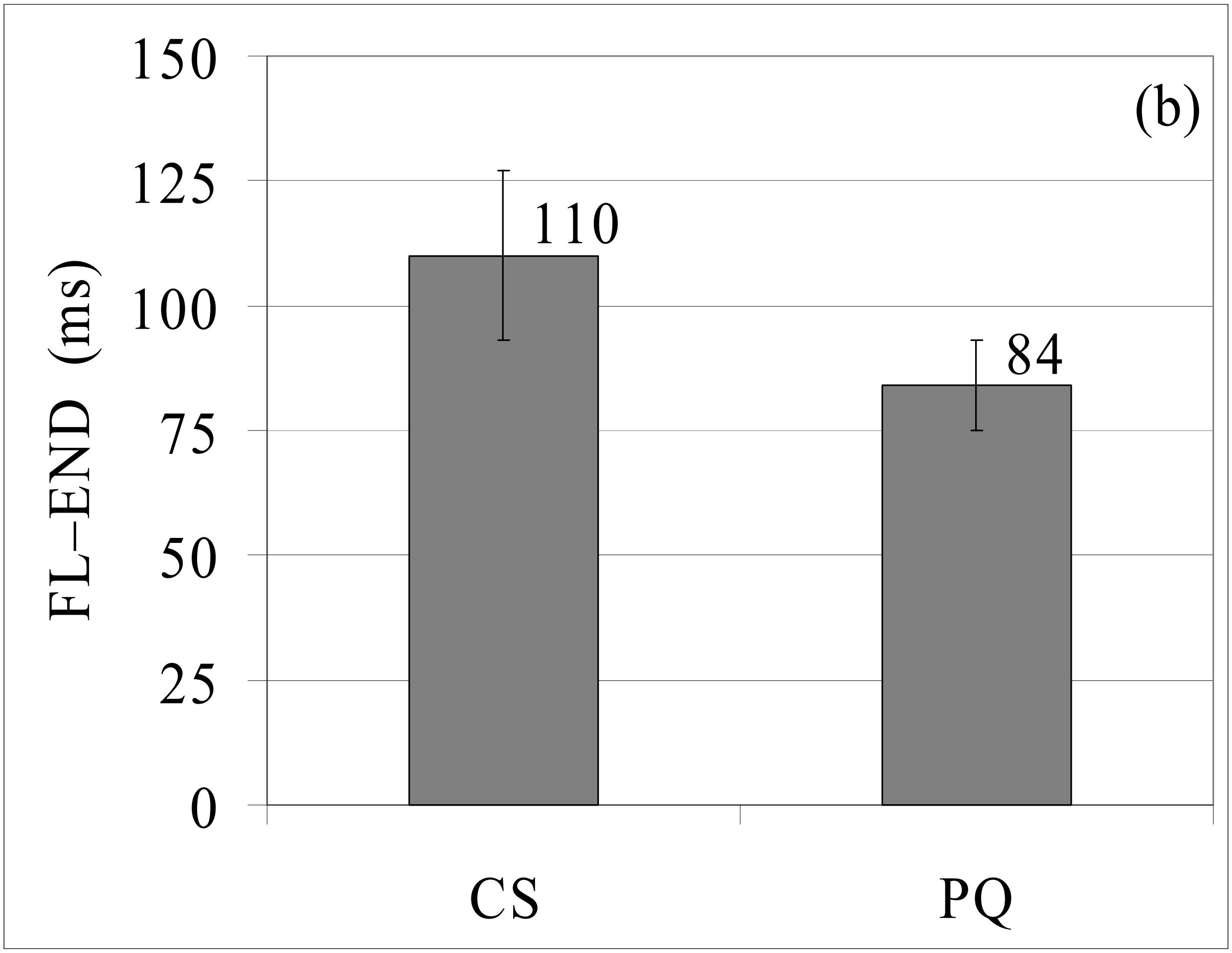

FIGURE 5b

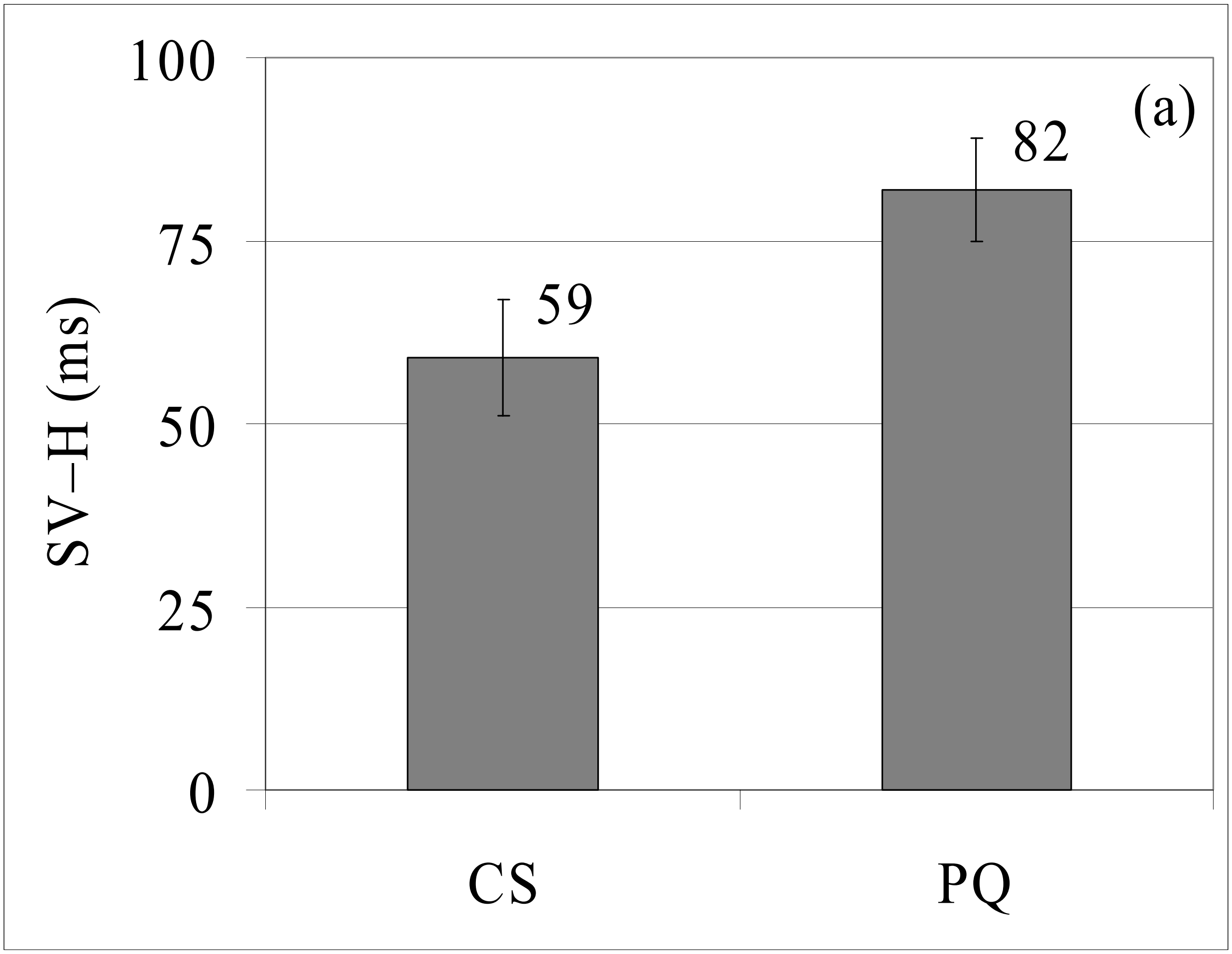

FIGURE 6a

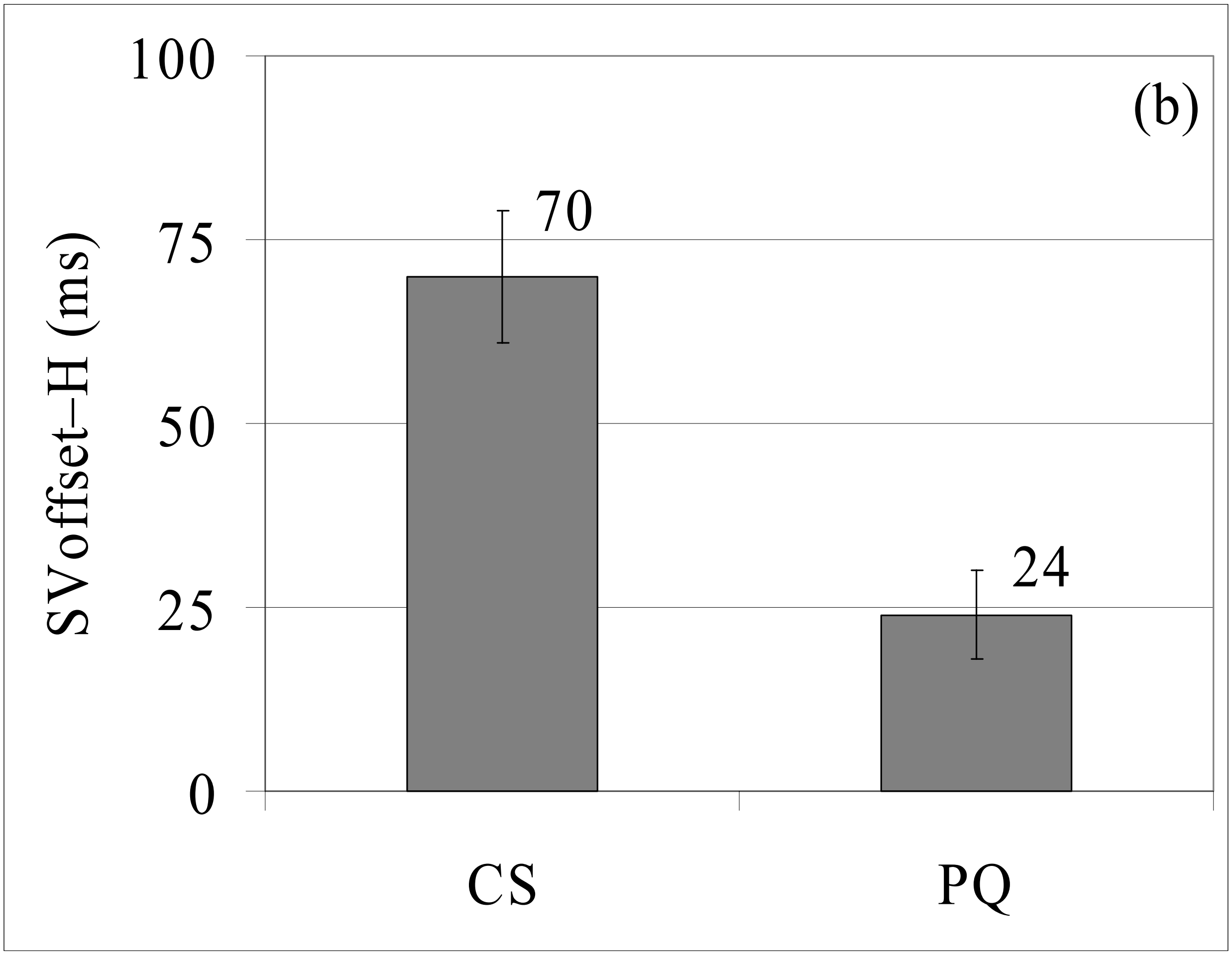

FIGURE 6b

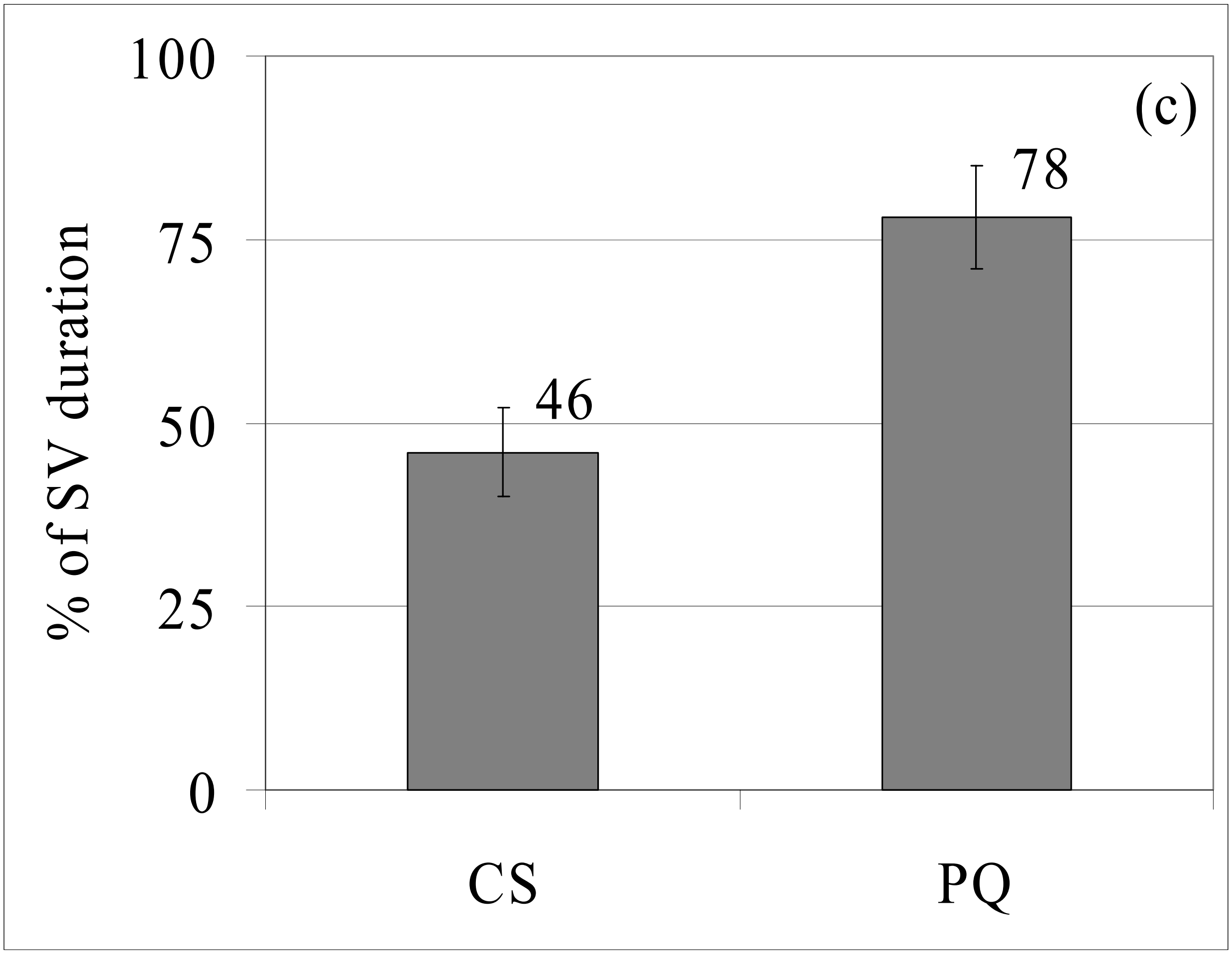

FIGURE 6c

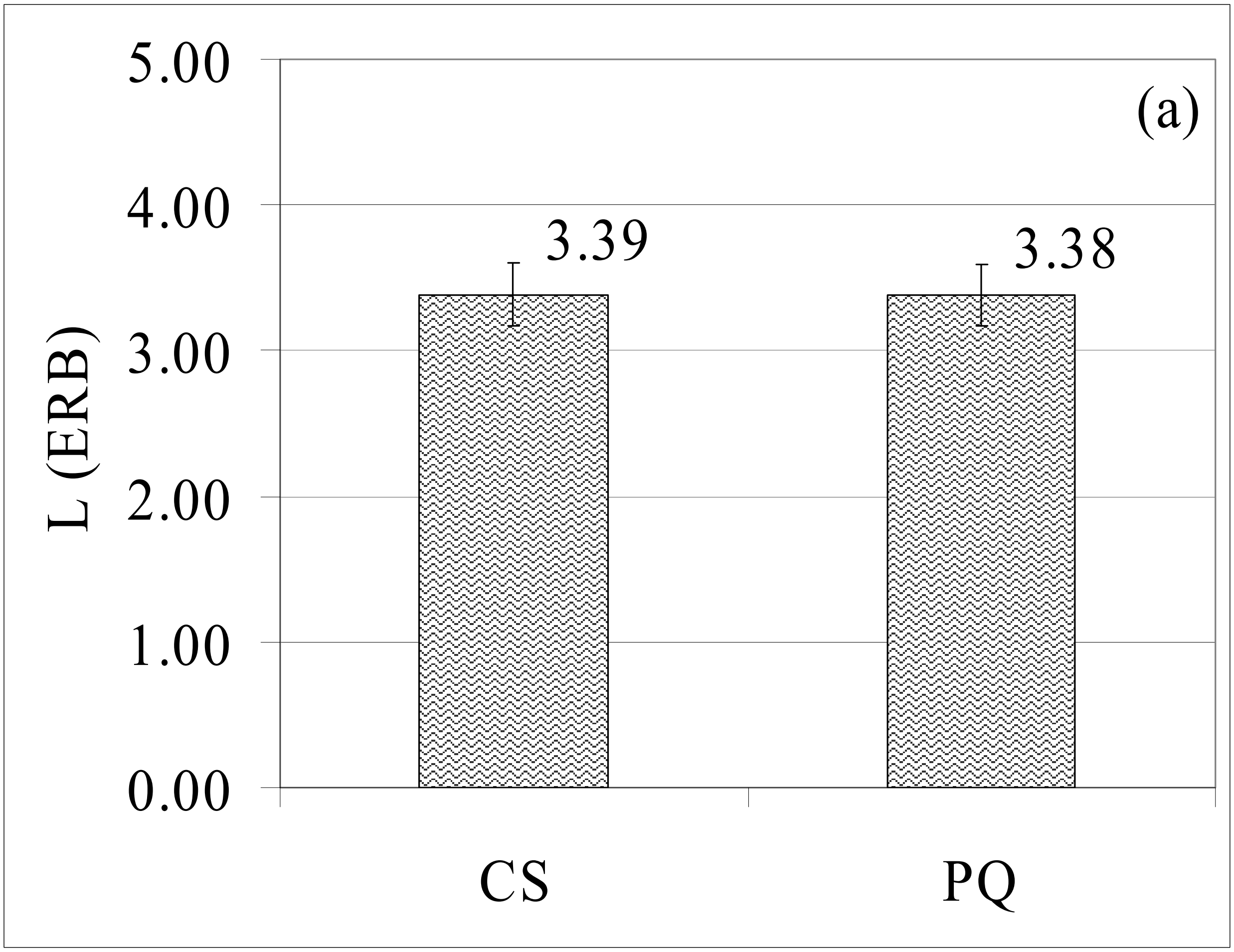

FIGURE 7a

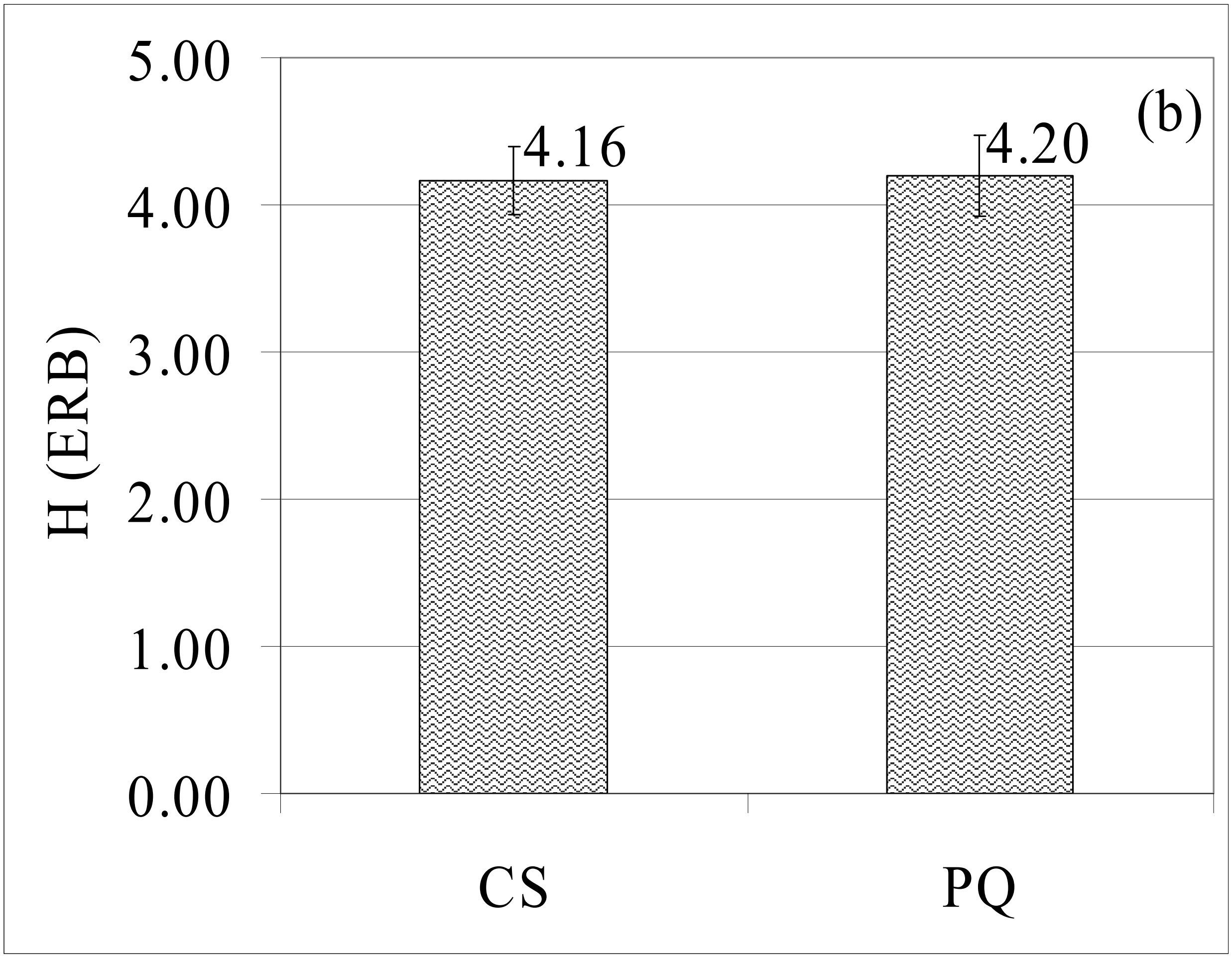

FIGURE 7b

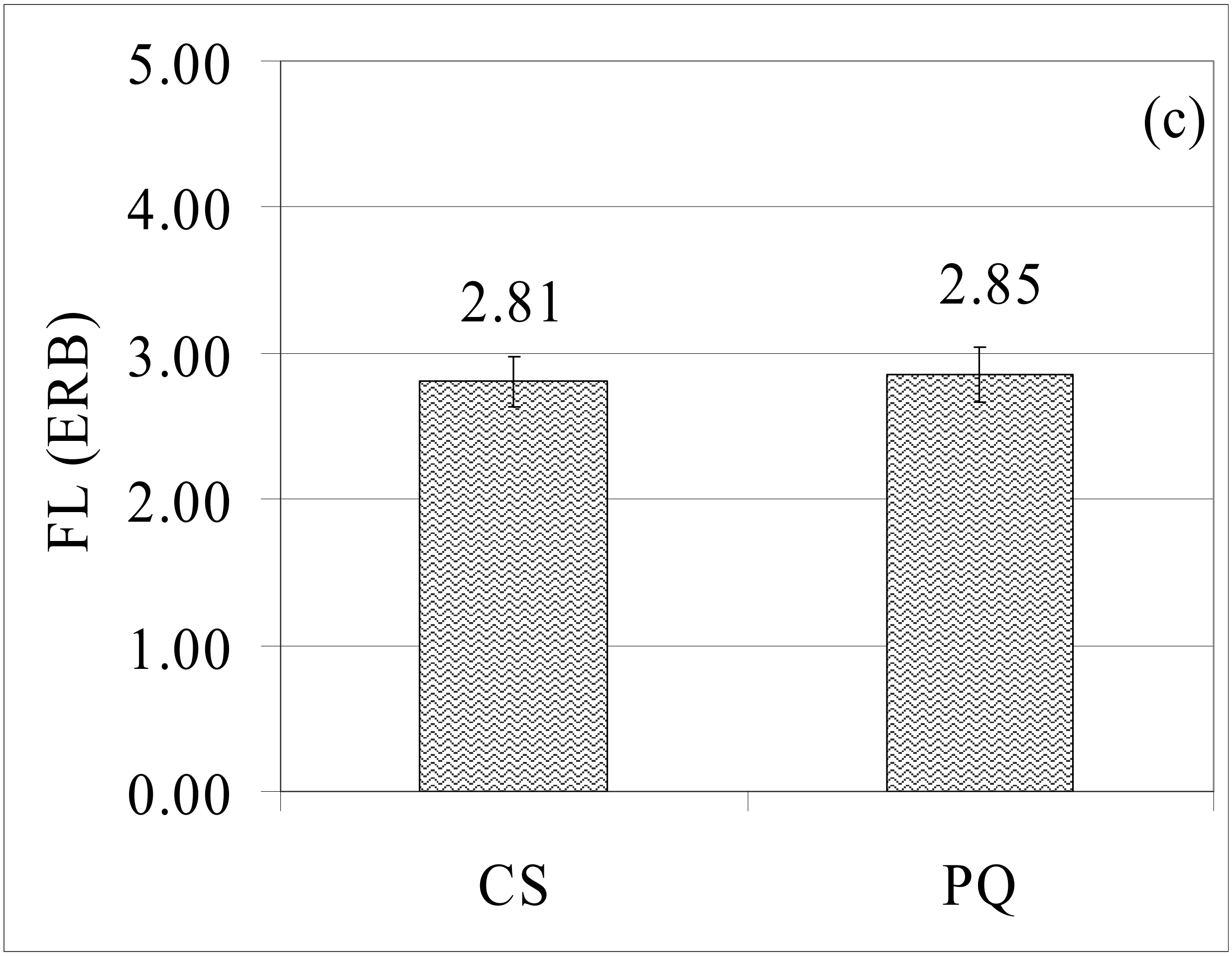

FIGURE 7c

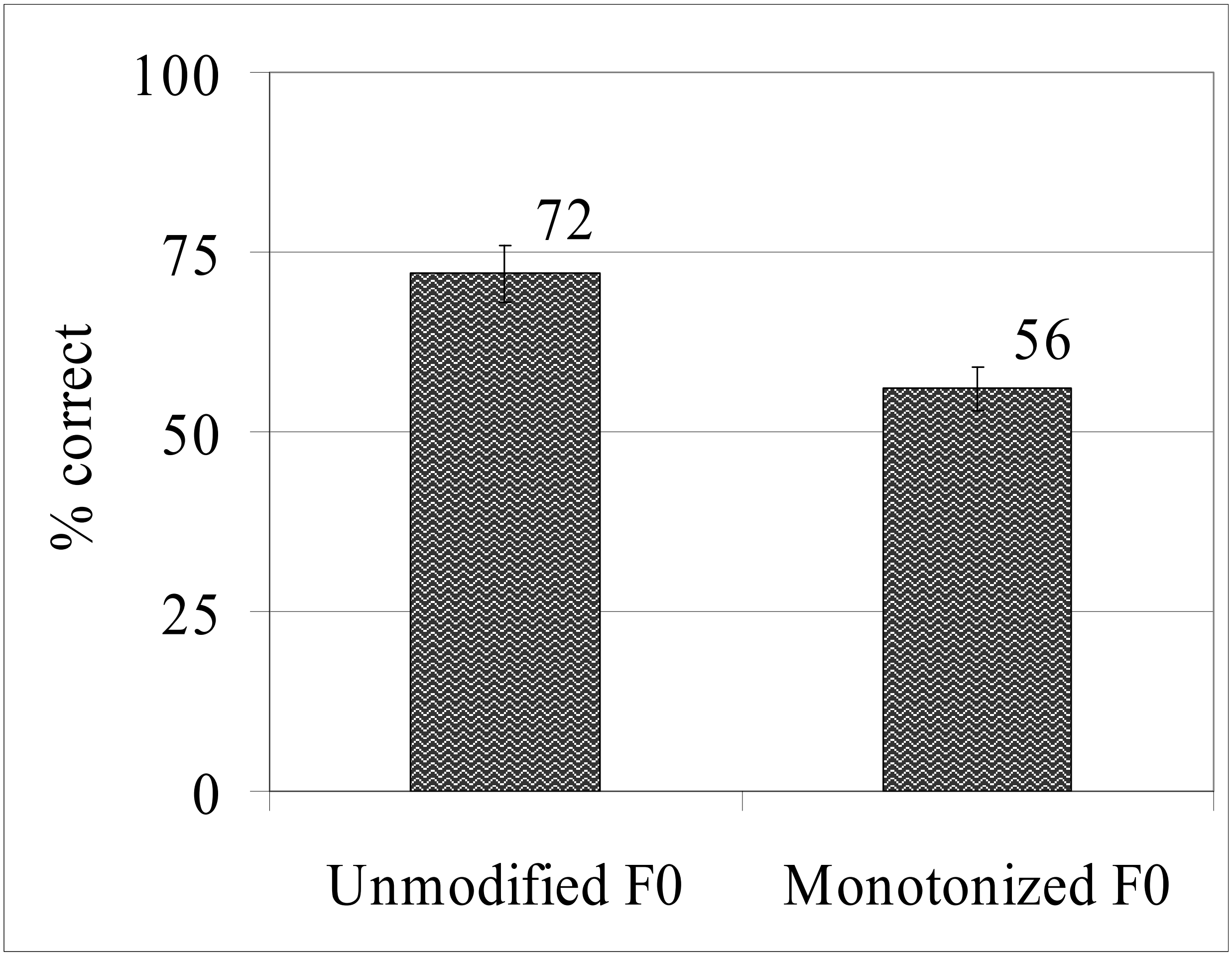

FIGURE 8