The timing of pre-nuclear pitch accents in Persian

Vahid Sadeghi

Department of English and Linguistics, Faculty of Humanities,
Imam Khomeini International University
vsadeghi@hum.ikiu.ac.ir

This paper examines the phonetic realization of rising pre-nuclear pitch accents in Persian. In a first experiment, the alignment of f0 valleys and peaks in pre-nuclear pitch accents was analyzed in controlled speech materials as a function of the syllable structure (open vs. closed) and vowel type (short vs. long) of the accented syllable. The results revealed that in words with antepenultimate stress, both the L and the H tones are anchored to specific segmental landmarks irrespective of syllable structure or vowel type. In particular, the L is consistently aligned with the onset of the accented syllable, and the H is placed with similar consistency in the vicinity of the first post-accentual vowel. In a second experiment, the variability in the timing and scaling of L valleys and H peaks was examined as a function of the proximity of the word boundary and of the following accent. The results revealed that while the alignment of the L was unaffected by changes in stress conditions, H peaks were significantly retracted as the syllable approached the end of the word. However, the proximity of the following accent did not produce a significant effect on H alignment. In addition, no significant differences were found on L and H scaling in different stress or tonal crowding conditions. Overall, the results contribute to a growing body of evidence that in the absence of upcoming prosodic pressure, the alignment of pitch targets is specified relative to segmental positions. A comparison between these findings and empirical findings from other languages reveals fine phonetic differences of segmental anchoring that are less likely to be interpreted in terms of distinct association-based phonological representations, and suggests that some aspects of segmental anchoring need to be explained in terms of continuous language-specific alignment rules.

1 Introduction

In the autosegmental-metrical model of intonational phonology, f0 contours are composed of low and high tones (L and H, respectively) associated with prosodic boundaries or prominent, i.e. stressed syllables; these Ls and Hs are phonetically realized as a sequence of local maxima and minima, known as f0 turning points or targets. Under this approach, pitch rises and falls are essentially regarded as transitions from one turning point to another, i.e. local f0 movements are not primitives of the linguistic analysis, but are defined in terms of their beginning and ending points (Arvaniti, Ladd & Mennen 1998, Ladd, Mennen & Schepman 2000).

In recent years, a number of studies have examined the scaling and alignment of putative tonal targets on the basis of this assumption. Many of these studies point to the conclusion that L and H tones behave as static targets, and that they are timed consistently with respect to the segmental string (Silverman & Pierrehumbert 1990, Prieto, van Santen & Hirschberg 1995, Arvaniti et al. 1998, Ladd et al. 2000). A conclusion that has figured prominently in

them is that while L turning points are very consistently anchored to the onset of the accented syllable, H targets are found to have more variable alignment (Caspers & van Heuven 1993 for Dutch; Prieto et al. 1995 for Spanish; Arvaniti et al. 1998 for Greek; Xu 1998 for Mandarin Chinese; Ladd, Faulkner, Faulkner & Schepman 1999, Ladd et al. 2000 and Ladd 2003 for English; Estebas-Vilaplana 2000 for Catalan; Dehé 2010 for Icelandic).

Some of these studies have examined the role of prosodic conditions such as speech rate and proximity of the following word boundary or pitch accent in determining the alignment of H peaks (Silverman & Pierrehumbert 1990, Caspers & van Heuven 1993, Prieto et al. 1995, Prieto 2005, Prieto & Torreira 2007). For example, Silverman & Pierrehumbert (1990) found that a pre-nuclear, i.e. non-final H* pitch accent is aligned earlier in a syllable immediately preceding a nuclear, i.e. final accented syllable than in a syllable not followed by an accented syllable. Similarly, Prieto et al. (1995) showed that while the location of the L target in the accentual rises in Mexican Spanish is consistently aligned before the onset consonant of the accented syllable, the H location is retracted before an upcoming word boundary or a pitch accent. However, research has suggested that H tones show consistent stability once the tonal pressure on a pitch accent from upcoming events is controlled (Arvaniti et al. 1998 for Greek, Ladd et al. 1999 for English, Xu 1999 for Chinese, Atterer & Ladd 2004 for German, Schepman, Lickley & Ladd 2006 for Dutch). Arvaniti et al. (1998) found that in the Greek pre-nuclear L*+H pitch accent, the H is aligned consistently with the onset of the first post-accentual vowel. Specifically they showed that the beginning and end of a pitch movement in Greek are anchored to specific locations in segmental structure, while the slope and duration of the pitch movement vary with respect to the segmental material with which it is associated. This finding later became known as 'segmental anchoring hypothesis', namely that all pitch targets are anchored to segmental landmarks. The finding of segmental anchoring clearly suggested that slope and duration are not the key identifying features of pitch movements, as claimed by the fixed rise time hypothesis (Fujisaki 1983, 't Hart, Collier & Cohen 1990 and others), but rather that slope and duration depend on the scaling and alignment of tonal targets. Similarly, Prieto & Torreira (2007) showed that in Spanish, H peaks align with the right edge of the accented syllable when there are at least two unaccented syllables after it. However, they also found a significant effect of syllable structure, contrary to the predictions of the segmental anchoring hypothesis (Arvaniti et al. 1998): in CV syllables, the peak occurred at the end of the accented syllable, but in CVC syllables the peak was retracted and occurred earlier within the coda consonant. Similar effects of syllable structure have been found on H alignment in Neapolitan Italian (D'Imperio 2000), Pisa and Bari Italian (Gili-Fivela & Savino 2003), Egyptian Arabic (Hellmuth 2005, 2006) and Icelandic (Dehé 2010).

The main goal of the present paper is to contribute to this line of research by providing results from a less well documented language – Persian. Pitch accents in Persian are characterized as instances of a bitonal L+H*. Previous studies have found that the target H* tone, corresponding to such accentual rises, shows variable alignment, while the preceding L tone is produced with a high degree of stability. However, little quantitative data of alignment have been presented in support of this claim. The current study is intended to examine the effects of various prosodic factors such as syllable structure and vowel quantity, as well as proximity of the following word boundary and accent on the scaling and alignment of pre-nuclear rising accents in Persian, in order to shed light on our understanding of the production of the tonal targets and their coordination with the segmental material.

2   Stress and intonation in Persian

Persian is an Iranian language which belongs to the Indo-Iranian sub-branch of the eastern branch of the Indo-European language family. Persian contains six vowels and twenty-three consonants. The Persian vowel system consists of the front vowels /iː/, /e/ and /æ/, and the back vowels /uː/, /o/ and /aː/. /iː/, /uː/ and /aː/ are phonologically long or bimoraic, while /æ/, /e/

and /o/ are phonologically short or monomoraic (Hayes 1979, Windfuhr 1979, Toosarvandani 2004). Syllables in Persian follow a CV(C)(C) template (Windfuhr 1979).

The majority of words in Persian are stressed on the final syllable (Ferguson 1957). The word-final stress pattern applies to nouns, adjectives, most adverbs, and simple verbs. Derivational as well as nominal inflectional suffixes receive stress (e.g. /keta:b-ˈi:/ 'bookish' and /keta:b-ˈha:/ 'books'). However, verbal inflectional suffixes and enclitics do not attract stress (e.g. /neˈveʃt-æm/ 'I wrote' and /keˈta:b-æʃ/ 'his book'), leaving the stress pattern of the stem unaffected (Ferguson 1957, Kahnemuyipour 2003). Prefixes in inflected verbs attract stress (e.g. /ˈmi:-nevis-æm/ 'bookish'), causing stress to be retracted from the end of the word. Moreover, in complex verbs, stress is located on the final syllable of the non-verbal element[1] (/tæˈmiz kærd/ [clean did] 'he cleaned' /ˈhærf zædænd/ [talk hit] 'they talked') (Kahnemuyipour 2003).

The intonational structure of Persian is said to involve two levels of the prosodic hierarchy, the accentual phrase (AP) and the intonational phrase (IP) (Sadat-Tehrani 2007, 2009). An accentual phrase normally consists of one content word with its possible enclitic(s). One or more accentual phrases are immediately dominated by an intonational phrase which corresponds to a mono-clausal sentence[2] (Sadat-Tehrani 2007). The tonal pattern of an accentual phrase involves a pitch accent, followed by an edge tone marking the end of the accentual phrase (Sadat-Tehrani 2007, 2009). The pitch accent is described as a bitonal L+H* accent, characterized by a Low followed by a High tone associated with the stressed syllable (Eslami 2000, Eslami & Bijankhan 2002, Mahjani 2003, Sadat-Tehrani 2009, Abolhasani Zadeh, Gussenhoven & Bijankhan 2010, Taheri Ardali 2010). L+H* is used for polysyllabic words or phrases with non-initial stress. This pitch accent may be realized as an allophonic H* in monosyllabic words or initially-stressed polysyllabic words, when no segmental material is available for the L tone to be realized (Sadat-Tehrani 2007, 2009). The part of an accentual phrase between the pitch accent and the AP end is handled by a phrase accent, which can be high or low (see Figure 1). This part can consist of zero syllables (when there is no enclitic in an AP, and stress is located on its final syllable), in which case the phrase accent is realized on the stressed syllable. It may also consist of several unstressed syllables due to the presence of one or more enclitics, in which case the phrase accent spreads over all the post-accentual unstressed syllables up to the L+H*.

The existence of the AP phrase accent is motivated by a comparison between nuclear and pre-nuclear APs. A nuclear AP, which is the last AP in most types of mono-clausal unmarked[3] sentences, carries a low phrase accent, L-, while a pre-nuclear AP is associated with a high phrase accent, H-.

An IP is phonologically associated with a low L% or high H% boundary tone. L% is used for declaratives (SOV or scrambled), leading polar questions (those with the particle /mæge/), *wh*-questions and imperatives; H% is used for polar questions, tag questions, echo questions and conditional structures.

The intonational structure of Persian is illustrated in Example 1 and Figure1. The nuclear AP is underlined. In the figure, each AP is demarcated by dashed vertical lines.

(1) /moˈdi:r-emu:n   na:ˈmæ-mu:n-o   <u>ʔemˈza: kærd</u>   foˈræn/
      manager-our   letter-our-RA   signature   do.PAST.3SG   quickly
      'Our manager signed our letter quickly.'

---

[1] Persian complex verbs consist of a non-verbal element, usually a noun or an adjective, followed by a verbal element.

[2] In complex and compound sentences, each clause usually forms a separate intonational phrase (Sadat-Tehrani 2007).

[3] Unmarked sentences are all-new sentences in the sense of Féry & Kügler (2008), in which no element has been mentioned in the preceding context or was especially prominent in the common ground of the discourse participants.
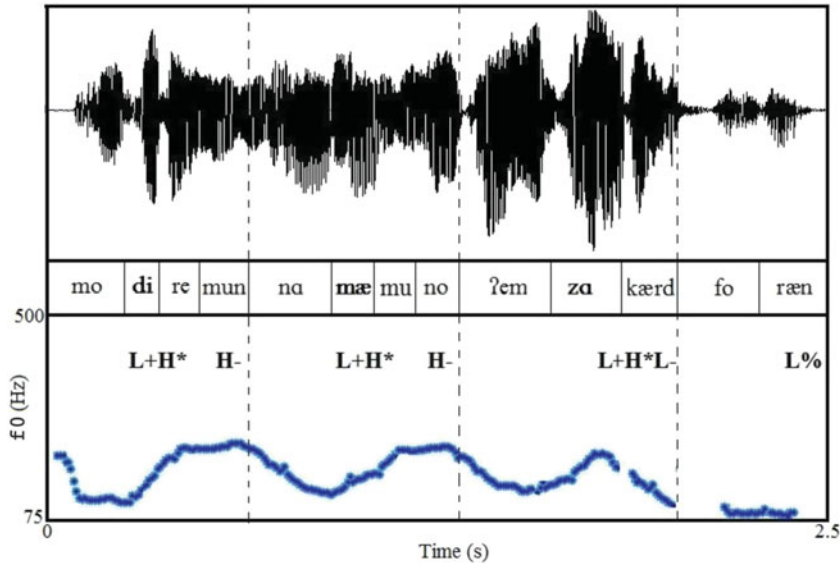
Figure 1 (Colour online) Waveform (upper panel) and f0 contour (lower panel) of the utterance /moˈdiːr-emuːn naˈmæ-muːn-o ʔemzaː kærd foˈræn/ 'Our manager signed our letter quickly' produced by a Persian speaker. Accentual phrases are demarcated by vertical dashed lines. Accented syllables are in boldface.

The utterance consists of three APs: the subject plus its enclitics (/moˈdiːr-emuːn/ 'our manager'), the object plus its enclitics (/naˈmæ-muːn-o/ 'our letter'), and the verb (/ʔemzaː kærd/ 'signed'). All the three APs involve a bitonal pitch accent followed by a phrase accent. The first two pre-nuclear APs carry a high phrase accent and the nuclear one (the verb) a low phrase accent. The adverb (/foˈræn/ 'quickly') is deaccented as it follows the last (nuclear) AP,[4] whose L- phrase accent spreads right up to the IP end. The utterance ends low with a low IP boundary tone (L%) which marks it as a declarative.

Phonetically, as can be seen in Figure 1, pitch accents in pre-nuclear accentual phrases involve a fairly sharp rise in f0, starting at the onset of the accented syllable and ending in the following unaccented syllable. The end of the rise is followed by a plateau-like shape that extends to the end of the accentual phrase domain. In both cases, the plateau slightly rises to a higher f0 point to the right of the accent domain boundary, suggesting that there is a clear turning point at the end of the plateau. Thus, the tonal structure of a pre-nuclear AP in Persian seems to consist of a bitonal pitch accent, as a tonal sequence LH, followed by a H phrase accent aligned with the edge of the accentual phrase.

Very few studies have investigated the alignment and scaling of rising pitch accents in Persian. In an instrumental study, Sadat-Tehrani (2009) examined the patterns of pitch alignment and scaling of rising pitch accents in three types of accentual phrases in Persian, namely pre-nuclear, nuclear and contrastive focus (which is defined as a type of focused accentual phrase that involves a correction of what has previously been said). The results suggested that while accentual L turning points in Persian are consistently anchored to the onset of the accented syllables in extremely consistent ways, accentual H alignment shows variability as a function of the type of the accentual phrase: the peak is located later in pre-nuclear APs than nuclear or contrastive focus ones. Despite this variability, however,

---

[4] It is assumed that adverbs constitute a separate AP unless when they follow the nuclear accentual phrase (most often, the verb), in which case they are deaccented (Sadat-Tehrani 2007).

Sadat-Tehrani showed that the accentual H is aligned past the end of the stressed syllable, i.e. into the following unaccented syllable, irrespective of the AP type. He explained the contrast in accentual H alignment between pre-nuclear AP, on the one hand and nuclear and contrastive focus APs, on the other hand, in terms of the differences in the phrase accents associated with the right edges of the accentual phrases. In particular, he suggested that the high phrase accent following pre-nuclear APs can cause the accentual peak to occur earlier; in effect, the peak is delayed when there is a high phrase accent aligning with the right edge of the accentual phrase (Sadat-Tehrani 2009). Importantly, he proposed that the contrast in the alignment of the accentual H between pre-nuclear and nuclear (and contrastive focus) APs may indeed reflect the occurrence of distinct phonological representations. Furthermore, Sadat-Tehrani's analysis of their scaling data revealed that the accentual H is scaled significantly higher in contrastive focus APs than pre-nuclear and nuclear ones. On the basis of this finding, he suggested a distinct phonological pitch accent for contrastive focus rises, namely L+ˆH*, with the diacritic ˆ denoting extra high and signifying the raised f0 in the contrastive focus AP. The results of Sadat-Tehrani's (2009) study further revealed that the durations of the three types of accentual phrases were significantly different from each other, with the duration of contrastive focus AP being significantly greater than pre-nuclear AP, and with the duration of nuclear AP being between the other two types but not significantly different from either. Similar results are reported by Abolhasani Zadeh, Bijankhan & Gussenhoven (2012) and Taheri Ardali & Xu (2012) for focused APs in Persian.

The study reported here examines the factors that cause the variability of H tones in Persian accentual rises in pre-nuclear APs in order to provide a more accurate phonological interpretation of such rising accents. Choices of the variables in this study were based on previous work on the Persian language. Previous studies have all reported significantly longer duration for closed syllables than for open syllables, and for phonologically long vowels /iː/, /uː/ and /aː/ than for phonologically short vowels /e/, /æ/ and /o/ (Ferguson 1957, Samareh 1977). I first examine the potential effects of syllable structure (open vs. closed) and vowel type (short vs. long) of the accented syllable on accentual L and H timing in pre-nuclear rises. This helps to refine the phonological representation of the accent. In particular, it will determine whether the accentual H occurs at a constant interval after the accentual L or whether the alignment of the H is determined independently from the L. The first possibility was embodied in Pierrhumbert's (1980) early theoretical assumption that in a bitonal pitch accent one tone would be starred (aligned with the accented syllable) and the other tone would merely lead or trail it, occurring at a fixed distance before or after the starred tone. The second possibility was shown to be true of pre-nuclear rising accents in Greek by Arvaniti et al. (1998) and has since been the basis of valid predictions about tonal alignment in other languages like English (Ladd et al. 1999) and Dutch (Ladd et al. 2000).

Two studies were carried out to examine tonal alignment in the data from 12 speakers of Standard Persian. Experiment 1 was specifically designed to examine to what extent the syllable structure and vowel type of the accented syllable may affect the alignment of accentual L and H tones in pre-nuclear rising accents. The second experiment dealt primarily with the effect of lexical stress location on the alignment of the accentual L and the H target.

3 Experiment 1

The goal of the first experiment was to examine the potential effects of syllable structure and vowel type of the accented syllable on the alignment of accentual L and H tones in pre-nuclear rising accents (accents in APs that are not utterance-final) in Persian. The study addresses the validity of two hypotheses:

Hypothesis A
The two tones of the Persian pre-nuclear accents are separated by a constant temporal interval. If so, then the manipulation of syllable structure and vowel duration should not modify the distance of the accentual L and H with respect to each other.

Hypothesis B
The two targets are aligned with respect to particular segmental positions and independently of each other. If this hypothesis is correct, then changes in syllable structure and vowel duration should not have an effect on the alignment of the accentual L and the H with respect to their anchoring points. It is further assumed that if the two tones are aligned independently from each other, then the duration of the rise (normalized L-to-H interval) should correlate with the distance in time between the accentual L and the onset of the post-accentual vowel, while there should not be a strong correlation between pitch excursion and rise time. These are the predictions of the segmental anchoring hypothesis (Arvaniti et al. 1998).

3.1 Method

The general method was that used in the studies by Arvaniti, Ladd and their colleagues discussed in the introduction (e.g. Arvaniti et al. 1998).

3.1.1 Speakers

Twelve speakers of standard Persian (six females and six males) participated in Experiment 1. The subjects were all university students, aged 19–35 years. Their participation was voluntary and did not involve compensation.

3.1.2 Materials

A corpus of 16 sentences was designed to test the hypotheses. The speakers read the test sentences six times at a normal speaking rate, and the first four repetitions for each speaker were used for analysis. Thus, the materials under study consist of 768 utterances (16 target sentences × 12 speakers × 4 repetitions). Since the effects examined in Experiment 1 were those due to syllable structure and vowel type of the accented syllable rather than the number of unaccented syllables following the accent, only words with stress on the antepenult (henceforth proparoxytones) were chosen as test words. Following Arvaniti et al. (1998), it is assumed that two unaccented post-accentual syllables provide sufficient prosodic space to allow for the full realization of tonal targets and thereby make for more stable and interpretable results. The test sentences used in the experiment met the following criteria:

 (i) The test word was generally either a subject (or a genitive noun inside a subject) or an object noun followed by a predicate (e.g. a verb or an adjective) in a simple syntactic pattern (e.g. /maː tæˈlæb-emuːn vuːsuːl ʃod/ [we claim-our (subject) paid become (PAST-3SG)] 'Our claim was paid'). This is the most unmarked word order of Persian that normally ensures that a pre-nuclear rising accent is placed on the test word followed by a nuclear accent on the following predicate. In addition, the accented syllable in the test word was always preceded by one unstressed syllable to allow for the full realization of the accentual L target.

(ii) To test for syllable structure and segmental effects of the accented syllable, the test syllables included two syllable structures, i.e. open vs. closed, as well as two different vowel types, i.e. short: /e/, /o/ and /æ/ vs. long: /iː/, /uː/ and /aː/. There were four sentences for each of the conditions described above. This resulted in 16 sentences (2 syllable structures × 2 vowel types × 4 repetitions) per speaker. If the accentual H target is aligned with respect to a particular segmental position according to Hypothesis B, we would expect the closed syllables and long vowels to delay the occurrence of the accentual H tone with respect to the onset of the accented syllable, and the open syllables and short vowels to bring it forward.

(iii) In order to avoid any effect of durational differences in consonants of the sort found in Greek by Arvaniti et al. (1998),[5] as well as to minimize micro-prosodic perturbations and ensure a smooth f0 contour, the target sequences of the test syllable and post-accentual syllables always contained the nasals /m/ or /n/ in onset or coda positions; the only exception was the onset position following the test syllable which was filled by a voiced stop (namely /b/, /d/ or /g/). The examples in (2) provide sample sentences from the corpus, with the test words underlined (a full list of the test sentences used in Experiment 1 can be found in Appendix A).

(2) a. /maː koˈmod-emuːn χæraːb ʃod/   (open short)
          we   closet-our   damaged become.PAST.3SG
          'our closet tore apart'

     b. /maː poˈmaːd-emuːn pejdaː ne-mi-ːʃe/   (open long)
          we   lotion-our   found   not-DUR-become.PRES.3SG
          'our lotion is not available'

     c. /kaːrt-e huːʃˈmænd-emuːn saːder næ-ʃode/   (closed short)
          cart-GEN intelligent-our   issued   not-become.PRES PERF.3SG
          'our intelligent card has not been issued'

     d. /maː pæsˈmaːnd-emuːn ziːjaːd niːst/   (closed long)
          we   waste-our   too.much not.be.PRES.3SG
          'we do not have too much (food) waste'

3.1.3 Procedure

The target sentences were recorded on DAT recorder using a high quality unidirectional head-mounted microphone (Shure SM58) in a sound proof booth. Speakers read the target sentences on a computer screen. They were instructed to read each sentence naturally, with no special emphasis on any part of the sentence. They were also asked to read the materials at a rate of speech they would consider normal. Also, the speakers were given some time to familiarize themselves with the task by practicing with a small number of randomly selected sentences from the data. The sentences were presented in random order. Speakers repeated the target sentences in different sets so the different repetitions of each sentence were separated. After recording, the sentences were examined to ensure that the test word was produced with a pre-nuclear rising accent. Eight utterances were discarded and replaced by instances from the extra reading materials (recall that six repetitions were recorded and only the first four were used); the problem with these utterances was that the speakers produced the test word with narrow focus. All the materials for this experiment as well as Experiment 2 were recorded in one session.

3.1.4 Measurements

The recorded sentences were digitized at a sampling rate of 16 kHz. They were analyzed using the acoustic speech analysis software Praat (version 4.3.01; Boersma & Weenink 2005). All the measurements were made on simultaneous visual displays of waveform, wideband spectrogram and f0 tracks. The following segmental and f0 landmarks were identified in each utterance:

---

[5] Arvaniti et al. (1998) tested for the effect of consonantal duration differences in Greek by designing materials in which the post-accentual syllables (there was a sequence of three unaccented syllables following the accented syllable of the test word) included stops, fricatives or nasals. Their results suggested that post-accentual segmental composition does affect the alignment of the H, with nasals resulting in earlier alignment of the H, fricatives resulting in later alignment, and stops grouping with either one or the other category, depending on the speaker.
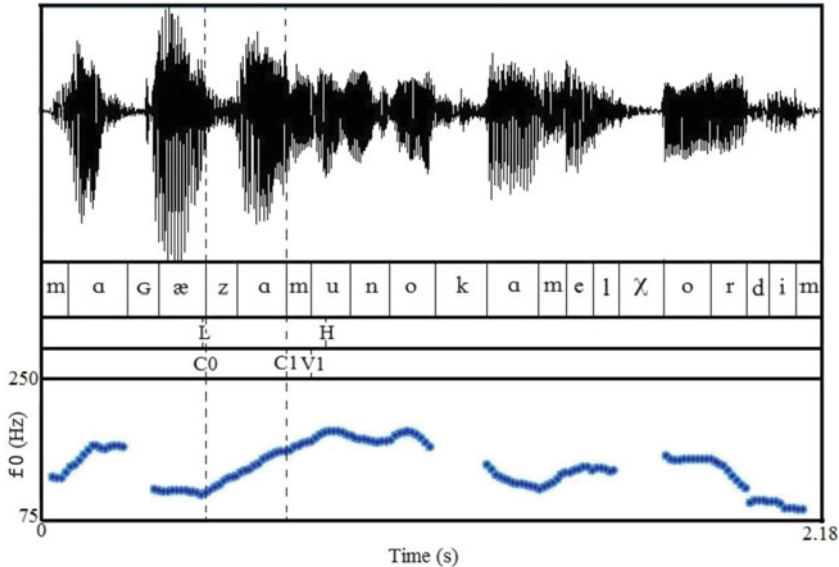
Figure 2 (Colour online) Waveform (upper panel) and f0 contour (lower panel) of the utterance /ma ɢæˈzamuno kamel χordim/ 'We fininshed our food' produced by a speaker of Experiment 1. The two vertical dashed lines mark the beginning and end of the accented syllable /za/.

C0: beginning of the consonant of the accented syllable
V0: beginning of the vowel of the accented syllable
V1: beginning of the vowel of the first post-accentual syllable
 L: location of the f0 minimum
 H: location of the f0 maximum

In general, the pre-nuclear rises showed their f0 minimum at the onset of the accented syllable. A measurement was made of this f0 minimum at the onset of the rise. The end of the rise was realized as an observable f0 peak around the beginning of the first post-accentual syllable (this can be seen in Figure 2). The rise was then followed by a plateau that ended at a higher f0 point at the right edge of the accentual phrase. As indicated in Figure 2, the measurement of f0 maximum taken for each rise was the first highest f0 point between the onset of the rise and the end of the accentual phrase. This always occurred within the first post-accentual unstressed syllable. In other words, the accentual H was located at the first observable f0 maximum at the beginning of the plateau. As discussed in Section 2 above, the plateau in Persian pre-nuclear APs is attributed to the additional presence of an H- phrase accent, for a more complete specification of the pre-nuclear rises as L+HH- (Sadat-Tehrani 2009). The separation between the accentual H of the bitonal pitch accent L+H and the following H- phrase accent is justified with the observation that the two H tones do not share a single tonal target, but are realized as discernable f0 peaks on different locations in pre-nuclear accentual phrases.

In some cases, the identification of peaks and valleys was not trivial.[6] One common problem was the occurrence of f0 discontinuities or dips at the closure or release of the nasals. The confounding effects of this phenomenon were compensated for by ignoring f0 points that were clearly perturbations and taking the adjacent highest or lowest point as the measurement

---

[6] Results of all measurements of scaling and alignment were double-checked by the author on different days.

f0 target. In six tokens, the end of the accentual rise was not marked by a local salient f0 peak but only a plateau where no clear f0 value emerged as the highest. These tokens were discarded and not used for further analysis. Segmental boundaries across vowels and sonorants were identified following standard criteria of segmentation (Peterson & Lehiste 1960).

Choices of segmental reference points for expressing peak alignment were based on previous research and the following hypotheses about the alignment behavior of Persian nuclear peaks. Preliminary visual examination of the data showed that the beginning of the accented syllable and the beginning of the following unstressed vowel function as possible anchor points for the accentual L valleys and H peaks, respectively. Following Atterer & Ladd (2004) and Schepman et al. (2006) the alignment of f0 targets was defined with respect to nearby acoustic landmarks. Thus, the alignment of accentual L is reported relative to C0 and the alignment of accentual H relative to V1 (the onset of the following unstressed vowel), although H is also reported relative to V0, as a distant landmark, to examine the effects of the factors (e.g. syllable structure and vowel type of the accented syllable) that are predicted to affect H alignment. In order to account for possible inter-speaker variation in speech rate, all measures of alignment were normalized using 'average syllable duration' as a measure of articulation rate. Specifically, each alignment measure was divided by the average syllable duration for the four syllable structure × vowel type conditions for each speaker. The resulting ratios were then multiplied by 100 to express the values as whole numbers.[7] Also, all f0 data were converted to ERB[8] units before conducting statistical analyses (Glasberg & Moore 1990) as these units are more appropriate for comparison between male and female f0 data, and they give a better approximation of perceptual distance between f0 levels (Arvaniti & Ladd 2009).

3.2 Results

3.2.1 L alignment

Figures 3 and 4 plot the normalized mean distance from the start of the accentual L target to the onset of the accented syllable (normalized C0toL) as a function of syllable structure (open vs. closed) and vowel type (short vs. long) for individual speakers (F = female, M = male) and for all speakers. The data reveal no consistent effects of syllable structure or vowel type on accentual L location: Taking the 0 as the beginning of the accented syllable, L is generally aligned slightly after the initial consonant of the accented syllable irrespective of the syllable structure or durations of the vowels in the accented syllable.

A two-way repeated-measures ANOVA on the speaker means was conducted taking normalized C0toL as the dependent variable, syllable structure and vowel type as the independent variables and speaker as the random factor. As expected, the result did not reveal an effect of syllable structure ($F(1,11) < 1$) or vowel type ($F(1,11) < 1$) on accentual L location.

3.2.2 H alignment

Earlier studies on H alignment in pre-nuclear rising accents (Silverman & Pierrehumbert 1990 for English, Prieto et al. 1995 for Spanish, Arvaniti et al. 1998 for Greek) have shown a high correlation between the duration of the accented syllable and the location of the accentual H

---

[7] For example, the mean syllable duration of the accented syllable across all the four syllable structure × vowel type conditions for Speaker M1 was 193.7 ms. One value of V0toH for each of the four phonological conditions mentioned above for this speaker was as follows: open short = 116 ms; open long = 162 ms; closed short = 201 ms; closed long = 248 ms. To normalize these values, they were divided by the mean syllable duration, e.g. 193.7 and multiplied by 100. The results were 59.88, 83.63, 103.76 and 128.03, respectively.

[8] Equivalent Rectangular Bandwidth or ERB is a measure used in psychoacoustics, which gives an approximation to the bandwidths of the filters in human hearing.
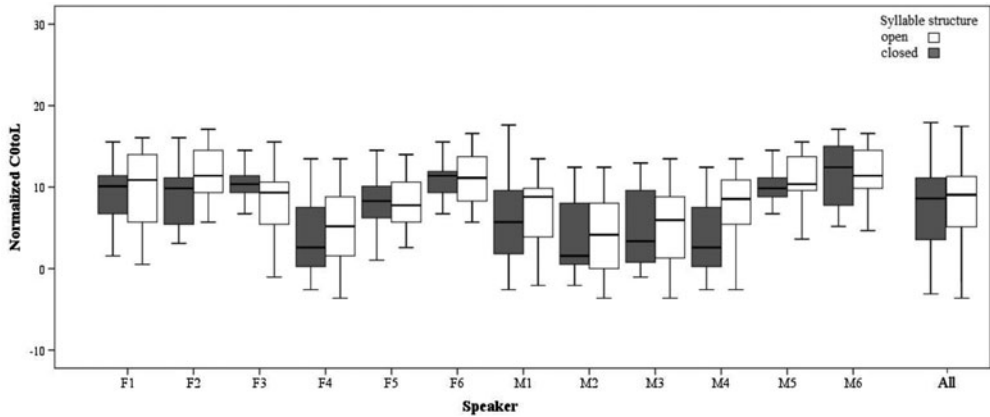
Figure 3 Normalized mean distance from accentual L to the onset of the accented syllable (normalized COtoL) as a function of syllable structure (open vs. closed) for individual speakers and for all speakers.
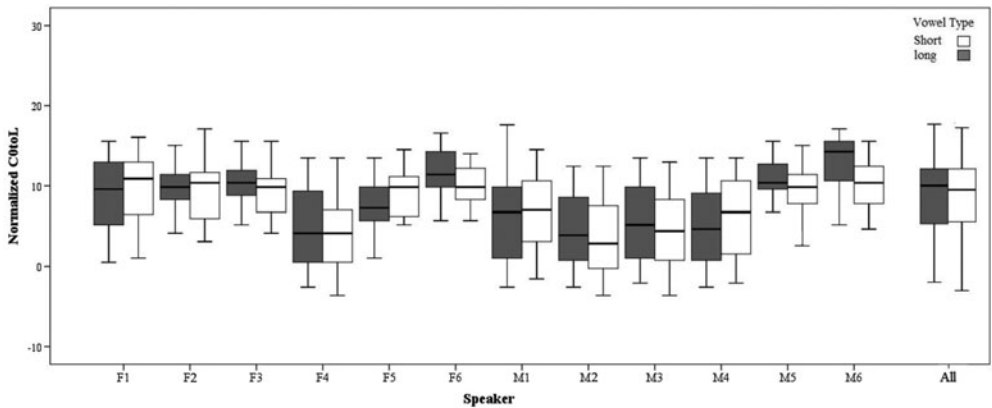

Figure 4 Normalized mean distance from accentual L to the onset of the accented syllable (normalized COtoL) as a function of vowel type (short vs. long) for individual speakers and for all speakers.

target, i.e. the longer the accented syllable, the further the H target is from the onset of the accented syllable. The graphs in Figure 5 plot the mean normalized V0toH as a function of the duration of the accented syllable in the two syllable structures (open syllables and closed syllables) and vowel types (short vowels and long vowels). As can be seen, the regression lines in the graphs reveal a strong positive correlation between the two variables (normalized V0toH and syllable duration) for both syllable structure conditions (correlation coefficients are $R^2 = .71$ for open and .77 for closed syllables, both significant at $p < .0001$) and vowel types (correlation coefficients are $R^2 = .75$ for short and $R^2 = .84$ for long vowels, both significant at $p < .0001$). The graphs also reveal a consistent difference in normalized H delay as a function of syllable structure and vowel type, as the data are clearly separated into two regions along the y-axis in the two graphs: Peaks in open syllables (black circles) are less delayed than those in closed syllables (gray circles). Also, peaks in short vowels (black circles) are less delayed than those in long vowels (gray squares). This difference in normalized peak delay seems to be highly correlated with the difference in duration as values of syllable duration along the x-axis are different across the two syllable structure conditions and vowel types.
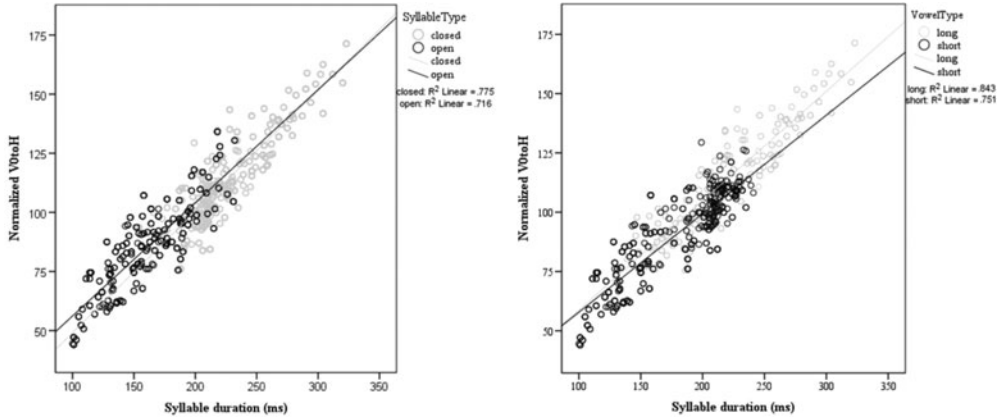
Figure 5 Mean normalized V0toH as a function of the duration of the accented syllable in the two syllable structures (left panel) and vowel types (right panel) for all speakers.

These results revealed a potential effect of syllable structure and vowel type on accentual peak alignment, namely that peaks in closed syllables and long vowels are further from the onset of the accented syllable than those in open syllables and short vowels, due to differences of duration across the two syllable structures and vowel types.

Therefore, the effects of syllable structure and vowel type were investigated in more detail. First, vowel duration was compared across the two vowel types, and also the two syllable structures since it might be suspected that closed syllables cause the vowel to be shortened (Klatt 1973). The results of a two-way repeated-measures ANOVA revealed that the short vowels [e] and [æ] were indeed significantly shorter than the long vowels [uː] and [aː] ($F(1,11) = 56.28$; $p < .001$); however, the vowel duration for open syllables was not significantly different from closed ones ($F(1,11) < 1$).

Second, patterns of accentual H alignment were compared across the two syllable structure conditions and vowel types to test the predictions of the segmental anchoring hypothesis (Hypothesis B). Accentual Peak location was analyzed quantitatively in two different ways, namely relative to the onset of the accented vowel (normalized V0toH) and relative to the beginning of the first post-accentual vowel (normalized V1toH). Figures 6 and 7 plot the mean H peak delay (normalized V0toH) as a function of syllable structure and vowel type for individual speakers and for all speakers. The data reveal consistent effects of both syllable structure and vowel type on H delay: for all speakers, peaks are further from the onset of the accented vowel for closed syllables and long vowels than for open syllables and short vowels.

Next, the alignment of the accentual H target was analyzed relative to the beginning of the post-accentual vowel so as to find the exact anchoring point of the H target. Figures 8 and 9 plot the mean normalized V1toH as a function of syllable structure and vowel type for each speaker and for all speakers. The dashed horizontal line (at value 0 on the y-axis) indicates the beginning of the post-accentual vowel. The data show that the accentual H target is reached at the beginning of the post-accentual vowel in both syllable structure conditions and vowel types. Although there is some between-speaker variability in the data (as some speakers, for instance, tend to place the H targets before the onset of the post-accentual vowel, while others tended to place them slightly after the beginning of this vowel), all speakers generally tend to place the H target in the vicinity of the onset of the post-accentual vowel. Thus, when the position of accentual H is measured with reference to the onset of the post-accentual vowel, closed syllables or long vowels exhibit no later alignment than open syllables or short vowels.
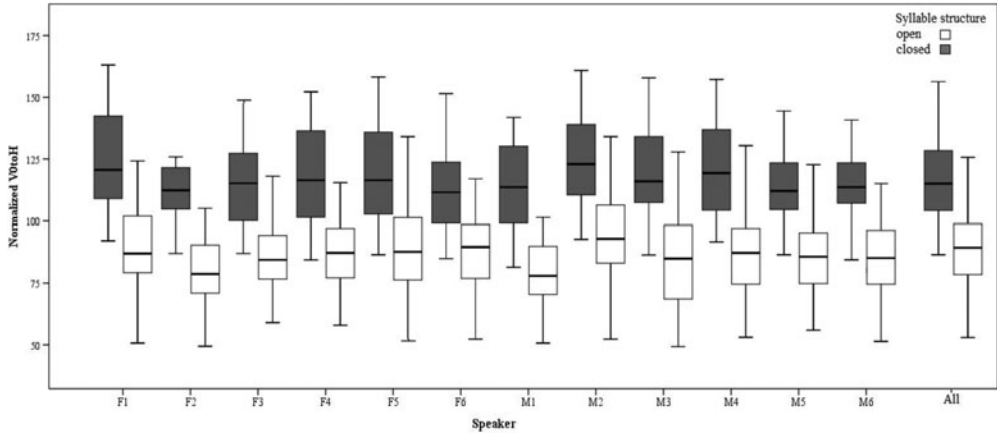
Figure 6 Normalized mean distance from the onset of the accented vowel to accentual H (normalized VOtoH) as a function of syllable structure for individual speakers and for all speakers.
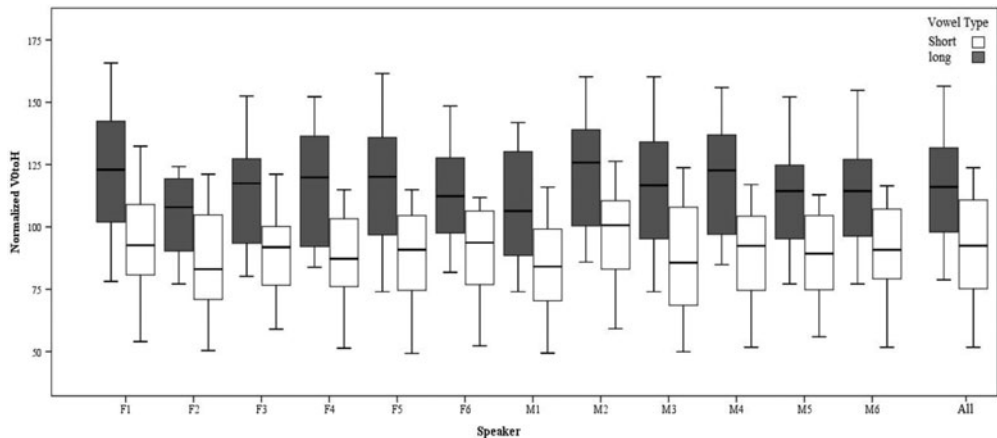

Figure 7 Normalized mean distance from the onset of the accented vowel to accentual H (normalized VOtoH) as a function of vowel type for individual speakers and for all speakers.

Table 1 ANOVA summaries of the effects of syllable structure and vowel type on two measures of accentual H location, namely normalized VOtoH and normalized V1toH.

	Normalized VOtoH	Normalized V1toH
Syllable structure	$F(1,11) = 428.39; p < .001$	$F(1,11) < 1$
Vowel type	$F(1,11) = 354.18; p < .001$	$F(1,11) < 1$

Two-way repeated-measures ANOVAs were performed for the two measures of accentual H alignment, i.e. normalized V0toH and normalized V1toH, with syllable structure and vowel type as the independent variables and speaker as the random factor. The results are presented in Table 1. As can be seen, the ANOVAs for normalized V0toH reveal a main significant effect of both syllable structure and vowel type. The interaction between vowel type and syllable structure was not statistically significant for normalized V0toH ($F(1,11) < 1$). However, no
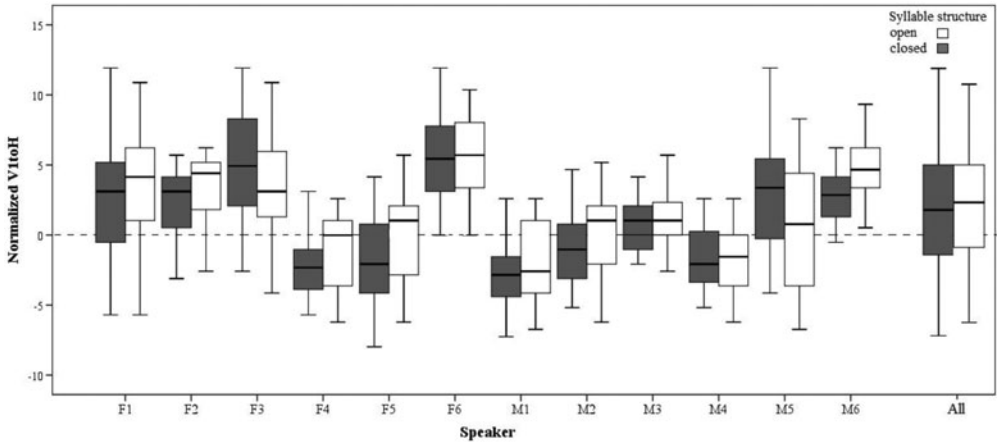
Figure 8 Normalized mean distance from the onset of the first post-accentual vowel to accentual H (normalized V1toH) as a function of syllable structure for individual speakers and for all speakers.
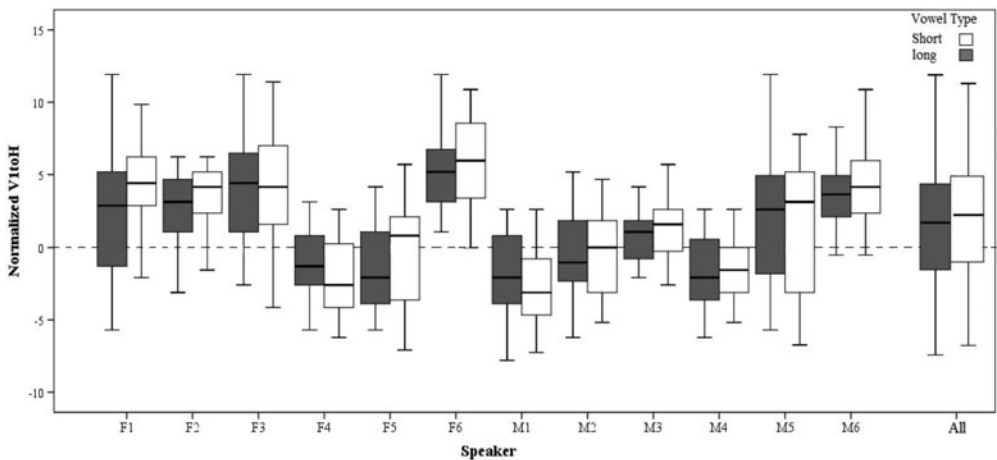

Figure 9 Normalized mean distance from the onset of the first post-accentual vowel to accentual H (normalized V1toH) as a function of vowel type for individual speakers and for all speakers.

effect of syllable structure or vowel type was found on accentual H alignment relative to the onset of the first post-accentual vowel (normalized V1toH).

These data suggest strongly that Persian pre-nuclear rising accents remain anchored to the onset of the post-accentual vowel, regardless of the syllable structure or vowel type of the accented syllable. In other words, the duration of pre-nuclear rising accents in Persian is adjusted to keep the end of the rise aligned with its segmental anchoring point as segment durations decrease or increase with syllable structure or vowel type. The result is most likely indicating that the two accentual tones of Persian pre-nuclear rising accents are aligned with the segmental string independently of one another (segmental anchoring hypothesis).

The results of correlation analyses on the speaker means revealed a strong relationship between normalized L-to-H interval and normalized LtoV1 ($R^2 = .76$; $p < .001$), suggesting that rise time is not fixed and is highly correlated with the temporal distance stretching from the onset of the accented syllable to the onset of the first post-accentual syllable. On the

contrary, there was not a strong correlation between normalized L-to-H interval and pitch excursion ($R^2 = .041$; $p = .27$). Thus, while the duration of a pre-nuclear rise in Persian is strongly correlated with the duration of the associated segmental material, the amount of f0 change (pitch excursion) is unaffected by such differences.

These results support the conclusion that the accentual L and H tones of Persian pre-nuclear rising accents are independently aligned with respect to the segmental string, and not with respect to each other.

4 Experiment 2

The results of Experiment 1 showed very clearly that the accentual H tone in proparoxytones is consistently aligned with the vowel in the post-accentual syllable. Experiment 2 was designed to examine the effect of within-word position of stress (lexical stress pattern) on the alignment of the accentual H target. In particular, the purpose of this experiment was to explore to what extent the pattern of H alignment emerged from the proparoxytones could be extended to paroxytones (words with penultimate lexical stress) and oxytones (words with final lexical stress). Previous research has generally emphasized the effect of prosodic pressure from the upcoming word boundary or accent on the alignment of accent peaks (Bruce 1977 for Swedish; Silverman & Pierrehumbert 1990 for English; Prieto & Shih 1995, Prieto et al. 1995 and Face 2002 for Spanish; Arvaniti et al. 1998 for Greek). Specifically, Arvaniti et al. (1998) found that the presence of an adjacent word boundary or accent triggers a relatively earlier alignment of H target, while a later word boundary or accent leads to a later alignment of H. They observed that the phonological condition that results in strong stability effects in Greek involves at least two unaccented syllables following the accented one, preferably within the same word as the accent.

This experiment tested whether the pressure from the right-hand prosodic context would result in regular temporal adjustments as reported in Silverman & Pierrehumbert (1990), Prieto et al. (1995) and Arvaniti et al. (1998). It was expected that in such cases, the position of the accent relative to both the word boundary and the upcoming accent would affect the position of the accentual H tone with respect to the segmental content. More specifically, the prediction was that the pressure of an adjacent word boundary produces time pressure environments that may result in earlier alignment of the H tone.

In addition, the present experiment examines whether the accentual H target exhibits consistent differences in scaling in clash and non-clash conditions. We find contradictory results in the literature concerning the scaling of the H peaks in clash environments. While Prieto & Shih (1995, for Spanish) and Arvaniti et al. (1998, for Greek) report clear stability effects on H, Face (2002, for Spanish) observes an effect of distance between accents on H1 (the first H target) scaling; that is, H1 tends to be lower as the distance to the following accent decreases. He suggests that this effect is due to the fact that the time allowed for producing the rising gesture is reduced in clash situations.

4.1 Method

4.1.1 Speakers
The same 12 speakers as in Experiment 1 were used in another recording session one day after Experiment 1. The target sentences were read six times by the speakers, and four repetitions for each speaker were used for analysis.

4.1.2 Materials
The data consisted of three sets of sentences with five sentences each. The target words in each set of sentences were oxytones, paroxytones or proparoxytones, and the distance from the target words' test syllables to the next accented syllables varied from one to two (given the

two-syllable distance between the accented syllable and the word boundary in proparoxytones, only one condition, namely the condition involving two intervening unaccented syllables, could be obtained for this stress pattern). Clearly, the location of the word boundary in the sequence of unaccented syllable(s) intervening between the test accent and the following accent varied. For example, when there were two intervening syllables, both syllables belonged to the target word in proparoxytones, while they belonged to the following word in oxytones. All test syllables and the following unaccented syllables were exclusively short open CV syllables. This enhanced time pressure between accents. The consonants were nasal /m n/ or liquid /l/, or /r/, to avoid differences in alignment due to vowel type. Examples of the sentences included in the corpus are given in (3)–(5), with the test words underlined (a full list of the test sentences used in Experiment 2 can be found in Appendix B).

(3) *Proparoxytones*
/na:ˈmæ-muːn-o   ˈbaːzeʃ   kærd-iːm/   (two intervening unaccented syllables)
 letter-our-RA   open.its   do-PAST.1PL
 'We opened our letter.'

(4) *Paroxytones*
 a. /naˈme-æm-o   ˈbaːzeʃ   kærd-æm/   (one intervening unaccented syllables)
  letter-my-RA   open.its   do-PAST.1SG
  'I opened my letter.'

 b. /naˈme- æm-o   bærˈgæʃt   daːd-ænd/ (two intervening unaccented syllables)
  letter-my-RA   return   give-PAST.3PL
  'They returned my letter.'

(5) *Oxytones*
 a. /na ːˈme   bærˈgæʃt   χorde/   (one intervening unaccented syllables)
  letter   return   hit.PRES PERF.3SG
  'The letter was returned.'

 b. /naːˈme   baːzχaːˈni   χaːhæd   ʃod/ (two intervening unaccented syllables)
  letter   read.out   will   become.PAST.3SG
  'The letter will be read out.'

4.1.3 Procedure and measurements

The analysis procedure and segmentation of the materials were the same as used in Experiment 1. In particular, a measurement was made of the distance in time between the accentual L and H targets, as well as the f0 difference of the accentual L and H tones in ERB, using the same procedure as in the previous experiment.

 In order to obtain a pattern with two adjacent pitch accents of the pre-nuclear type, as we had expected, the participants were asked to read sentences naturally at a relatively slow speech rate.[9] Despite of this, some instances of accent deletion were found in the data (about six tokens for each speaker), which were discarded and replaced by instances from the extra reading materials.

---

[9] Asking the participants to speak relatively slowly could be a problem as it could minimize the tonal crowding effects the materials were meant to create. To address this issue, the speaking rates were statistically compared between Experiment 1 and Experiment 2 using the duration of the accented syllable as a measure of speech rate. Mean syllable duration was calculated for each speaker across the two experiments. The effect was not found to be significant for any speaker (all $ps > .1$), suggesting that speakers' productions across the two experiments did not reflect a significant difference in speaking rate, and that variability (if found) in the alignment measures for the materials in Experiment 2 could not be attributed to differences in speaking rate.
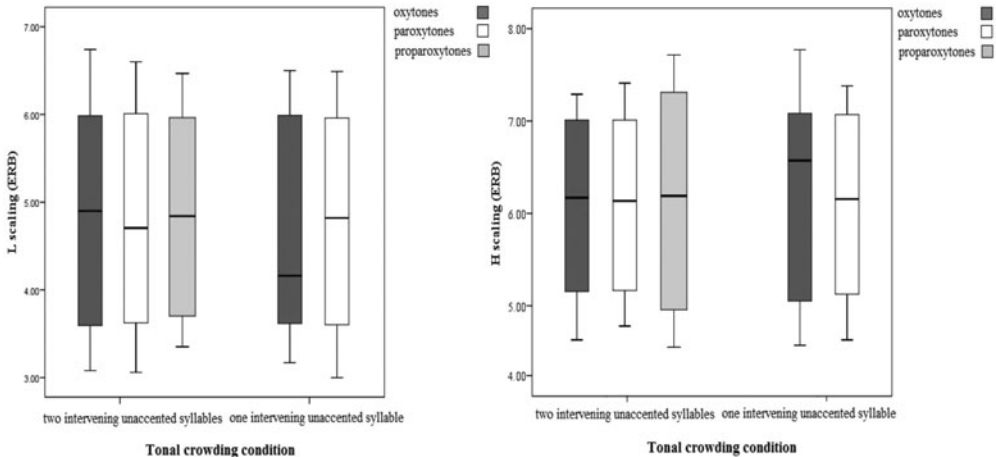
Figure 10 Mean accentual L (left panel) and H (right panel) values in ERB across different stress patterns (oxytones, paroxytones and proparoxytones) and tonal crowding conditions (one and two intervening unaccented syllables) for all speakers.

4.2 Results

4.2.1 L and H scaling

The graphs in Figure 10 plot mean L and H values (in ERB) across different stress patterns (oxytones, paroxytones and proparoxytones) and tonal crowding conditions (one and two intervening unaccented syllables). The results show that both accentual L and H tones are rather constant across different stress and tonal crowding conditions.

Two-way repeated-measures ANOVAs on the speaker means were conducted separately for accentual L and H scaling values with stress pattern (having three levels: oxytones, paroxytones and proparoxytones) and number of unaccented syllables (having two levels: one and two) as the independent variables, and speaker as the random factor. The results showed that neither stress pattern nor the number of intervening unaccented syllables exerted a significant main effect on L and H scaling (all Fs < 1). In general, both accentual L and H tones showed stability effects in scaling across speakers in stress and tonal crowding conditions. This seems to indicate that both accentual L and H function as production real targets: speakers make the effort to maintain target values of L and H across different stress or tonal crowding conditions.

4.2.2 L and H alignment

Two one-way repeated-measures ANOVAs were performed across speakers for normalized C0toL and V0toH with stress pattern as the independent variable and speaker as the random factor. The results revealed no effect of stress condition on the alignment of the accentual L ($F(2,22) < 1$). As can be seen in Figure 11, L is generally placed at the onset of the accented syllable in all stress patterns. However, the results for the measure of accentual H alignment produced a significant main effect of stress ($F(2,22) = 332.73$; $p < .001$). As shown in Figure 11, peaks are less delayed in oxytones than paroxytones and also delayed in paroxytones than proparoxytones. Post-hoc Scheffe tests revealed that H distance to the onset of the accented vowel was significantly larger in proparoxytones than in both paroxytones and oxytones (all pairwise comparisons, ps $< .001$), and also significantly larger in paroxytones than oxytones (ps $< .001$).
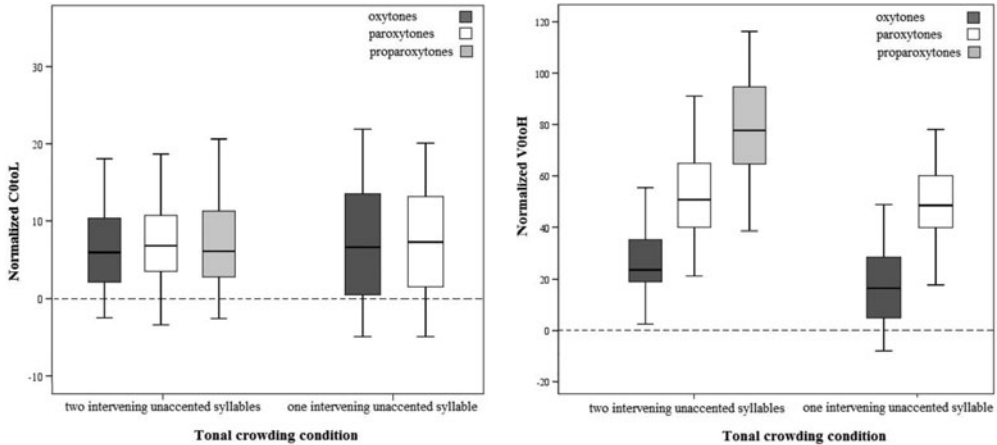
Figure 11 Mean values of normalized C0toL (left panel) and V0toH (right panel) across different stress patterns and tonal crowding conditions for all speakers. The dashed horizontal line in each figure represents the temporal position of the respective segmental landmark.

Also, the data including oxytones and paroxytones with one and two syllables intervening the test accent and the following accent were submitted to two separate two-way repeated-measures ANOVAs on the speaker means for normalized C0toL and V0toH with stress pattern (having two levels: oxytones and paroxytones) and number of unaccented syllables (also having two levels: one and two) as the independent variables and speaker as the random factor. There was no effect of stress pattern ($F(1,11) = 1.89$; $p = .29$) or number of unaccented syllables on accentual L location ($F(1,11) < 1$). As expected, the ANOVA for the accentual H alignment, on the other hand, revealed a significant main effect of stress pattern ($F(1,11) = 292.68$; $p < .001$). However, as illustrated in Figure 11, no effect of number of unaccented syllables was found on accentual H alignment ($F(1,11) = 2.65$; $p = .14$). Thus, the results suggest that unlike within-word position of stress, the number of unaccented syllables does not affect the alignment of the accentual H target.

Given the clear effects of within-word position of stress on the location of the accentual H target and consistent differences in accentual H alignment across stress conditions, the alignment of accentual H was analyzed relative to the onset of the first post-accentual vowel so as to find out the exact anchoring of the H across different stress conditions. Figure 12 plots the normalized mean distance between the accentual H peak relative to the onset of the first post-accentual vowel (normalized V1toH) in proparaxytones and paraxytones for each speaker and for all speakers (given that peaks in oxytones are located in the vicinity of the syllable boundary, the data for this stress condition were excluded from the analysis of this measure). Negative numbers indicate alignment before the relevant segmental landmark, while positive numbers indicate alignment after this frontier.

The data for proparoxytones replicate the pattern of H alignment from Experiment 1: all peaks are realized in the vicinity of the onset of the first post-accentual vowel. Peak delay measures for paroxytones, however, have mainly negative values, suggesting that peaks for such words are aligned earlier with respect to the first post-accentual vowel and located well within the preceding consonant.

A one-way repeated-measures ANOVA was conducted on the speaker means taking V1toH (peak distance to onset of the first post-accentual vowel) as the dependent variable and the number of post-accentual syllables within the word as the independent variable. The result showed a significant main effect of stress pattern on accentual H location ($F(1,11) = 380.56$; $p < .001$).
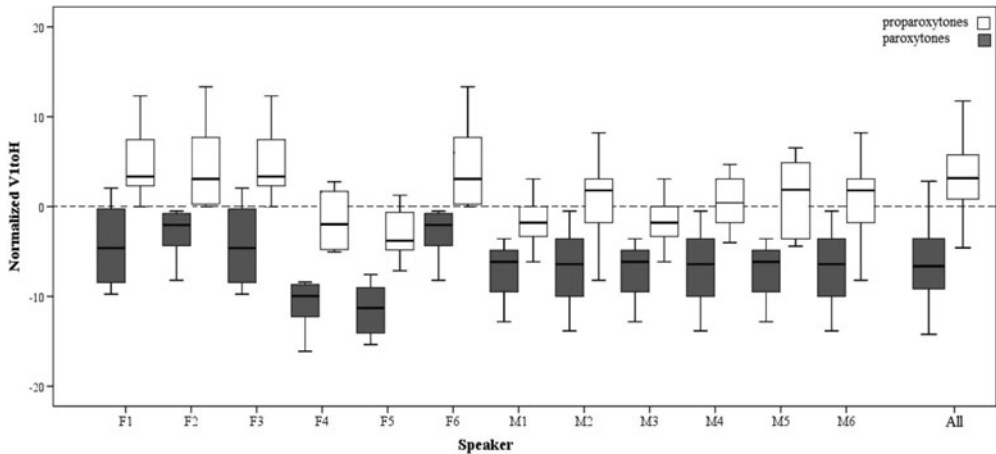
Figure 12 Normalized mean distance from the onset of the first post-accentual vowel to accentual H (normalized V1toH) in proparaxytones and paroxytones for each speaker and for all speakers. The dashed horizontal line indicates the beginning of the first post-accentual vowel.

In sum, the results show that the alignment of the accentual H pre-nuclear peak is consistently affected by within-word position of stress: as the distance between the pitch accented syllable and the word boundary decreases, the pitch accent peak occurs progressively earlier in the accented word.

5 Discussion and conclusions

This study examined the phonetic realization of pre-nuclear rising accents in Persian in order to provide a more accurate phonological representation of such rising movements. The first experiment explored the variability in the timing of the tonal targets as a function of syllable structure and vowel type of the target accented syllable. Results revealed that in proparaxytones, the alignment of the accentual H tones was affected by variations in the duration of the accented syllable in consistent ways: accentual H pre-nuclear peaks were earlier in open and short syllables than closed and long syllables. However, when the alignment of the accentual H was measured with reference to the onset of the first post-accentual vowel, syllable structure or vowel type failed to produce significant effects: The accentual H target was aligned on average within 10 ms after the onset of the first post-accentual vowel. The results support the segmental anchoring hypothesis (Hypothesis B), whereby the duration of pitch movements in speech is finely adjusted to the duration of the accompanying segmental material. Thus, the results clearly show that the duration of pre-nuclear rising accents in Persian is not constant, as claimed by the fixed rise time hypothesis (Fujisaki 1983, 't Hart et al. 1990 and others), but is instead governed by the coordination of the movement with the segmental string (Ladd et al. 1999 and Dilley, Ladd & Schepman 2005 for English, Ladd et al. 2000 for Dutch, Atterer & Ladd 2004 for German).

The second experiment examined the variability in the scaling and timing of pitch accents as a function of the proximity of the word boundary and of the following accent. With regard to scaling, the evidence was that the pitch levels of both accentual L and H points are not consistently affected by tonal crowding. No strong tendency was observed among speakers to undershoot pitch peaks or valleys in the tonal crowding conditions.

Regarding the alignment behavior of tonal targets, the data revealed clear effects of stress conditions on accentual H location. Peaks were located earlier as the distance of the target accented syllable and the word boundary decreased. The accentual H target was realized

early in oxytones, and progressively later in paroxytones and proparoxytones. Accentual L targets, on the other hand, were highly stable across different stress conditions. These prosodic word effects seem to suggest the possibility that either the end of the word or the word (AP) H- phrase accent is acting as a prosodic context that exerts prosodic pressure on accentual H tonal alignment. The results are consistent with findings from several other languages. For example, the findings of Silverman & Pierrehumbert (1990) for English, Prieto et al. (1995) for Spanish and Prieto (2005) for Catalan, report clear effects of prosodic word boundaries on peak alignment. In all these languages, peaks tend to shift backwards as their associated syllables approach the end of the word. Also, as in Persian, in languages like Spanish (Prieto et al. 1995) and Greek (Arvaniti et al. 1998) it has been shown that f0 movement toward a tonal target starts consistently at the onset of the accented syllable irrespective of the degree of proximity of the word boundary.

Unlike stress pattern, distance between accents did not strongly affect the position of accent peaks in Persian. Accentual H alignment patterns were not quantitatively different between tonal crowding and non-crowding conditions for most speakers. Similarly, no significant differences were found on accentual L alignment in crowding vs. non-crowding environments. The data thus reveal that accentual L and H alignment is largely unaffected by the number of unaccented syllables if there is at least one unaccented syllable intervening between accents. The lack of peak alignment effect in tonal crowding context can be accounted for if we assume that a pre-nuclear pitch accent in Persian is separated from the following accent by a right H- phrase accent aligned with the AP end. Thus, it may be argued that the H-phrase accent acts as the most immediate prosodic context to trigger peak retraction while the upcoming pitch accent as the remote tonal context fails to drastically affect surface accentual H alignment patterns.

In general, our results replicate and extend earlier findings of Arvaniti et al. (1998) for Greek, Xu (1998) for Chinese, Ladd et al. (1999) for English, Atterer & Ladd (2004) for German, and Schepman et al. (2006) for Dutch, showing that in the absence of prosodic pressure from the upcoming material, i.e. when the accented syllable is not in the vicinity of the word boundary or the next accent, the two tones of pre-nuclear pitch accents in Persian are consistently aligned with respect to the segmental material, and the stability effects are pervasive under changes of segmental or syllable structure composition. Thus, the results demonstrate that a successful quantitative model of alignment for Persian must contain at least two factors, namely the distance in time between the onset of the accented syllable and the onset of the first post-accentual vowel and the distance in syllables to an upcoming word boundary.

The findings just summarized readily lend themselves to an auto-segmental interpretation, in which the accentual rise is treated as a sequence of a low (L) tone and a high (H) tone associated with specific points in phonological structure, rather than a description based on distinctive pitch movements ('t Hart et al. 1990).

Within the auto-segmental metrical framework of intonational phonology, there are two contrastive possibilities for the alignment of tonal targets in a bitonal pitch accent (see e.g. Frota 2014): targets may align with respect to one another or they may align with respect to the segmental string. Accents of the first type are rises and falls, i.e. the f0 peak and valley are tightly timed with respect to each other, regardless of factors that may cause variability in the timing of the two tonal targets. Accents of the second type, on the other hand, are simply pairs of two tones, i.e. the L and H tones are timed independently with the segmental string, and the rise (or fall) between them is epiphenomenal (simply resulting from interpolation between the two targets). Findings of segmental anchoring as presented in the current study suggest that Persian pre-nuclear rising accents fall into the second category. One possible phonological account of segmental anchoring is in terms of the idea of secondary association proposed by Pierrehumbert & Beckman (1988). Thus, it may be assumed that a given pitch accent as a whole would be associated with the accented syllable as the head of the prosodic domain, but the individual tones making up the pitch accent could have secondary associations to
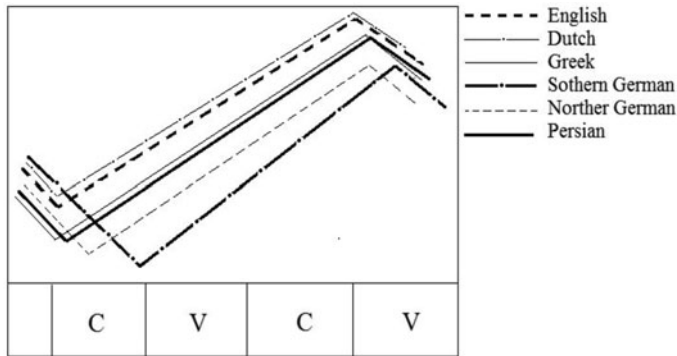
Figure 13 Schematic representation of the alignment of L and H tones relative to a stressed syllable in Persian (based on Experiment 1), Northern and Southern German (based on Atterer & Ladd 2004), English (Ladd et al. 1999), Greek (Arvaniti et al. 1998) and Dutch (Schepman et al. 2006).

specific heads or edges within that domain. This approach allows cross-linguistic differences of segmental anchoring to be treated as detailed differences of secondary association. For example, on the basis of the Greek, Dutch and Persian findings, we might assume that the pre-nuclear accentual rise is a starless tonal sequence L+H, underlyingly associated with the accented syllable, but in Dutch, the H tone is secondarily associated with the right edge of the accented syllable, while in Greek and Persian, the H is associated with the left edge of the following unaccented vowel.

However, as pointed out by Atterer & Ladd (2004) in their cross-linguistic comparison of alignment of tonal targets, the problem with an association-based account of alignment differences is that it assumes that one pattern of alignment is categorically distinct from another, while some differences in alignment among languages are so small that it is rather difficult to analyze them in terms of distinct patterns phonological association. Specifically, Atterer & Ladd (2004) studied patterns of alignment of L and H tones in Southern and Northern German and compared their data with those from three other languages, namely Dutch, Greek and British English. Their principal finding was that the patterns of alignment of L and H in these languages are so subtly different that cannot be accounted for in terms of finely differentiated phonological specifications of tonal association. A solution adopted by Atterer & Ladd (2004), and also advocated by many others, including Beckman, Hirschberg & Shattuck-Hufnagel (2005) and Arvaniti (2016), is to describe differences of segmental anchoring in terms of quantitative phonetic realization rules, which is in keeping with the notion of language-specific phonetic rules developed by Pierrehumbert (1980). Within this approach, different languages could share a given pattern of tonal targets, namely an L followed by an H, but realize it in different ways by choosing different values in a continuum of phonetic alignment.

A cross-linguistic comparison between the Persian data and data from other languages reveals subtle differences in alignment (as shown in Figure 13) which might best be interpreted in terms of continuous phonetic alignment rules as proposed by Atterer & Ladd (2004). Phonetically the pattern of L alignment we find in Persian is slightly later than those of English (Ladd et al. 1999) and Greek (Arvaniti et al. 1998) and rather earlier than Southern German (Atterer & Ladd 2004). For H, the Persian pattern of alignment is quite comparable to what Arvaniti et al. (1998) found for Greek and what Atterer & Ladd (2004) found for Northern German, and rather later than the findings of Ladd et al. (1999) for English and Dutch. One interpretation of this comparison would be that language-specific differences in the alignment of pitch movements may be a matter of what Ladd (2006) calls phasing: the same f0 change can be aligned earlier or later. This in turn suggests that the two targets of a

bitonal pitch movement are not independently aligned at specific places in structure; rather the whole movements are aligned relative to whole syllables. Thus, Southern German aligns both L and H later than Northern German, Greek and Persian, which in turn align both L and H later than Dutch and English. This, in general, may provide some evidence for Xu's (1998) idea that the rise is, at some level of analysis, a unitary phonological event, the alignment of which is specified as a whole.

In general, our findings of segmental anchoring provide little evidence for the starred tone interpretation assumed in early auto-segmental theory, according to which one of the two tones of a bitonal accent must be aligned with the accented syllable, while the other tone merely leads (or trails) the starred tone by a fixed temporal interval. In addition, starring one of the two tones in the Persian LH pitch accents would pose problems for Persian intonational system in which there is no contrast of alignment between L+H* and L*+H (Sadat-Tehrani 2007, 2009). Rather, following both language-specific and cross-linguistic evidence presented in this study, we may suggest that the most appropriate representation of pre-nuclear rising accent in Persian is the starless sequence of a low (L) and a high (H) tone, namely LH, where the L is realized at the beginning of the stressed syllable and the H early in the vowel of the post-tonic syllable according to phonetic implementation rules of alignment for Persian. What is still not quite clear is the phonological association of tones in intonational systems, as well as the definition of the phonetic properties of association. These are issues that will need to be addressed in future investigation so that the phenomenon of segmental anchoring will be better understood.

Acknowledgements

Appendix A. Experiment 1: Speech materials

Test words are underlined.

(A1) *Open short syllables*

    a. /uːnhaː   <u>sæˈnæd-eʃuːn</u>   gom   ʃod/
       they   <u>deed-their</u>   lost   become.PAST.3SG
       'our deed was lost'

    b. /maː   <u>koˈmod-emuːn</u>   χæraːb   ʃod/
       we   closet-our   damaged   become.PAST.3SG
       'our closet tore apart'

    c. /maːʤeraː-e   <u>ʤa ːˈ leb-etuːn</u>   hæme ʤaː   pæχʃ   ʃode/
       story-GEN   interesting-your   everywhere   spread   become.PRES PERF.3SG
       'your interesting story has spread everywhere'

    d. /maː   <u>tæˈlæb-emuːn</u>   vuːsuːl   ʃod/
       we   claim-our   paid   be.become.PAST.3SG
       'our claim was paid'

(A2) *Open long syllables*

    a. /uːnhaː   <u>ʔoˈmiːd-eʃuːn</u>   ʔæz   dæst   ræft/
       they   <u>hope-their</u>   from   hand   go.PAST.3SG
       'their hope was gone'

b. /ʃoma: <u>beh'nu:d-etu:n</u> ku:tʃek-tær ʔæz reza:st/
  you Behnud-your small-er from Reza.be.PSES.3SG
  'your son, Behnud, is smaller than Reza'

c. /ma: <u>po'ma:d-emu:n</u> pejda: ne-mi:-ʃe/
  we lotion-our found not-DUR-become.PRES.3SG
  'our lotion is not available'

d. /pæχʃ-e <u>su:' ru:d-emu:n</u> ɢætʔ ʃod/
  playing-GEN song-our stopped become.PAST.3SG
  'our song was stopped to be played'

(A3) *Closed short syllables*

a. /ma:ʃi:n-e <u>sæ'mænde-mu:n</u> χu:b ka:r ne-mi:-kone/
  car-GEN Samand-our well work not-DUR-do.PRES.3SG
  'our Samand does not work well'

b. /ɢa:mæt-e <u>bo'lænde-tu:n</u> be dʒæzza:bi:jæt-etu:n ʔæfzu:de/
  height-GEN tall-your to attraction-your add.PRES PERF.3SG
  'your height has added to your attraction'

c. /ka:rt-e <u>huʃ'mænd-emu:n</u> sa:der næ-ʃode/
  cart-GEN intelligent-our issued not-become.PRES PERF.3SG
  'our intelligent card has not been issued'

d. /mu:ha:-e <u>be'lond-eʃu:n</u> ri:zeʃ da:re/
  hair-GEN blond-their loss have.PRES.3SG
  'their blond hair is being lost'

(A4) *Closed long syllables*

a. /ʃæhr-e <u>bodʒ'nu:rd-emu:n</u> χejli: bozorg ni:st/
  city-GEN Bojnurd-our very large not.be.PRES.3SG
  'Bojnurd, our city, is not very large'

b. /ma: <u>pæs'ma:nd-emu:n</u> zi:ja:d ni:st/
  we waste-our too much not.be.PRES.3SG
  'we do not have too much (food) waste'

c. /ʃæhr-e <u>mæ'la:rd-emu:n</u> ɢærb-e kærædʒ ʔæst/
  city-GEN Malard-our west-genetive Karaj be.PRES.3SG
  'Malard, our city, is in the west of Karaj'

d. /ru:zna:me-e <u>le'mu:nd-eʃu:n</u> di:ru:z montæʃer næ-ʃod/
  newspaper-GEN Lemund-their yesterday published not-become.PAST.3SG
  'their newspaper Lemund was not published yesterday'

Appendix B. Experiment 2: Speech materials

Test words are underlined.

(B1) *Proparoxytones*

a. /<u>na:'mæ-mu:n-o</u> 'ba:z-eʃ kærd-i:m/
  letter-our-RA open-its do-PAST.1PL
  'we opened our letter'

b. /<u>ma:'læ-mu:n-o</u> 'gom-eʃ kærd-i:m/
  trowel-our-RA lost-its do-PAST.1PL
  'we lost our trowel'

c. /ʃuː'n æ-muːn-o   'gom-eʃ   kærd-iːm/
    <u>comb-our-RA</u>   lost-its   do-PAST.1PL
    'we lost our comb'

(B2) *Paroxytones*

  a. (i) /na'me-æm-o   'baːz-eʃ   kærd-æm/
          <u>letter-my-RA</u>   open-its   do-PAST.1SG
          'I opened my letter.'

      (ii) /na'me-æm-o   bær'gæʃt   daːd-ænd/
          <u>letter-my-RA</u>   return   give-PAST.3PL
          'they returned my letter'

  b. (i) /ma'le-æm-o   'gom-eʃ   kærd-æm/
          <u>trowel-my-RA</u>   lost-its   do-PAST.1SG
          'I lost my trowel'

      (ii) /ma'le-æm-o   bo'lænd   kærd-æm/
          <u>trowel-my-RA</u>   up   do-PAST.1SG
          'I raised my trowel'

  c. (i) /ʃuː'ne-æm-o   'ʔaːb-eʃ   keʃiːd-æm/
          <u>comb-my-RA</u>   water-its   draw-PAST.1SG
          'I rinsed my comb with water'

      (ii) /ʃuː'ne-æm-o   baː   'ʔaːb   ʃost-æm/
          <u>comb-my-RA</u>   with   water   clean-PAST.1SG
          'I washed my comb'

(B3) *Oxytones*

  a. (i) /na'me   bær'gæʃt   χorde/
          <u>letter</u>   return   hit-PRES PERF.3SG
          'the letter was returned'

      (ii) /na'me   baːzχa'ni   χaːhæd   ʃod/
          <u>letter</u>   read out   will   become.PAST.3SG
          'the letter will be read out'

  b. (i) /ma'le   nːi'jaːz   næ-daːr-ænd/
          <u>trowel</u>   need   not-have-PRES.3PL
          'they do not need a trowel'

      (ii) /ma'le   læbe-'haː-æʃ   χæraːb   ʃode/
          <u>trowel</u>   edges-PL-its   damaged   become.PRES PERF.3SG
          'the edges of the trowel are spoiled'

  c. (i) /ʃuː'ne   ʔæz   'dʒiːb-eʃ   ʔoftaːd/
          <u>comb</u>   from   pocket-his   fall.PAST.3SG
          'the comb fell down from his pocket'

      (ii) /ʃuː'ne   laːje   'muː-æʃ   giːr   kærd/
          <u>comb</u>   through   hair-his   caught   do.PAST.3SG
          'the comb was caught through his hair'

References

Abolhasani Zadeh, Vahideh, Carlos Gussenhoven & Mahmood Bijankhan. 2010. The position of clitics in Persian intonational structure. *Proceedings of the Speech Prosody 2010 Conference*, Chicago, IL, 104–108.

Abolhasani Zadeh, Vahideh, Mahmood Bijankhan & Carlos Gussenhoven. 2012. The Persian pitch accent and its retention after the focus. *Lingua* 122(13), 1380–1394.

Arvaniti, Amalia. 2016. Analytical decisions in intonation research and the role of representations: Lessons from Romani. *Journal of the Association for Laboratory Phonology* 7(1), 1–43.

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

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

Atterer, Michaela & D. Robert Ladd. 2004. On the phonetics and phonology of 'segmental anchoring' of f0: Evidence from German. *Journal of Phonetics* 32, 177–197.

Beckman, Mary E., Julia Hirschberg & Stefanie Shattuck-Hufnagel. 2005. The original ToBI system and the evolution of the ToBI framework. In Sun-Ah Jun (ed.), *Prosodic models and transcription: Towards prosodic typology*, 1–37. Oxford: Oxford University Press.

Boersma, Paul & David Weenink. 2005. Praat: Doing phonetics by computer (Version 4.3.01) http://www.praat.org/. [Computer program]

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

Caspers, Johanneke & Vinent J. van Heuven. 1993. Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall. *Phonetica* 50, 161–171.

Dehé, Nicole. 2010. The nature and use of Icelandic pre-nuclear and nuclear pitch accents: Evidence from f0 alignment and syllable/segment duration. *Nordic Journal of Linguistics* 33(1), 31–65.

Dilley, Laura, D. Robert Ladd & Astrid Schepman. 2005. Alignment of L and H in bitonal pitch accents: Testing two hypotheses. *Journal of Phonetics* 33, 115–119.

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

Eslami, Moharram. 2000. *Šenaxt-e næva-ye goftar-e zæban-e farsi væ karbord-e an dær bazsazi væ bazšenasi-ye-rayane'i-ye goftar* [The prosody of the Persian language and its application in computer-aided speech recognition]. Ph.D. thesis, Tehran University.

Eslami, Moharram & Mahmood Bijankhan. 2002. Nezame ahænge zæbane farsi [Persian intonation system]. *Iranian Journal of Linguistics* 34, 36–61.

Estebas-Vilaplana, Eva. 2000. *The use and realisation of accentual focus in Central Catalan*. Ph.D. dissertation, University College, London.

Face, Timothy. 2002. When push comes to shove: Tonal crowding in Madrid Spanish. *The Linguistic Association of Korea Journal* 10, 77–100.

Ferguson, Charles. 1957. Word stress in Persian. *Language* 33, 123–135.

Féry, Caroline & Frank Kügler. 2008. Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics* 36, 680–703.

Frota, Sónia. 2014. The intonational phonology of European Portuguese. In Sun-Ah Jun (ed.), *Prosodic typology II: The phonology of intonation and phrasing*, 6–42. Oxford: Oxford University Press.

Fujisaki, Hiroshi. 1983. Dynamic characteristics of voice fundamental frequency in speech and singing. In Peter F. MacNeilage (ed.), *The production of speech*, 39–55. New York: Springer.

Gili-Fivela, Barbara & Michelina Savino. 2003. Segments, syllables and tonal alignment: A study on two varieties of Italian. In Maria-Josep Solé, Daniel Recasens & Juan Rojas Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (ICPhS XV), Barcelona, vol. 4, 2933–2936. Barcelona: Casual Productions.

Glasberg, Brian & Brian Moore. 1990. Derivation of auditory filter shapes from notched-noise data. *Hearing Research* 47, 103–138.

Hayes, Bruce. 1979. The rhythmic structure of Persian verse. *Edebiyat* 4, 193–242.

Hellmuth, Sam. 2005. Pitch accent alignment in Egyptian Arabic: Exploring the boundaries of cross-linguistic alignment variation. *II Phonetics and Phonology in Iberia, PaPI*, 20–21.

Hellmuth, Sam. 2006. *Intonational pitch accent distribution in Egyptian Arabic*. Ph.D. dissertation, School of Oriental and African Studies, University of London. Available at http://www.sfb632.uni-tsdam.de/homes/samhellmuth/.

Kahnemuyipour, Arsalan. 2003. Syntactic categories and Persian stress. *Natural Language & Linguistic Theory* 21(2), 333–379.

Klatt, Dennis. 1973. Interaction between two factors that influence vowel duration. *The Journal of the Acoustical Society of America* 54, 1102–1104.

Ladd, D. Robert. 2003. Phonological conditioning of f0 target alignment. In Maria-Josep Solé, Daniel Recasens & Juan Rojas Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (ICPhS XV), vol. 1, 249–252. Barcelona: Casual Productions.

Ladd, D. Robert. 2006. Segmental anchoring of pitch movements: Autosegmental association or gestural coordination? *Rivista di Linguistica* 18(1), 19–38.

Ladd, D. Robert, Dan Faulkner, Hanneke Faulkner & Astrid Schepman. 1999. Constant 'segmental anchoring' of f0 movements under changes in speech rate. *The Journal of the Acoustical Society of America* 106, 1543–1554.

Ladd, D. Robert, Ineke Mennen & Astrid Schepman. 2000. Phonological conditioning of peak alignment of rising pitch accents in Dutch. *The Journal of the Acoustical Society of America* 107, 2685–2696.

Mahjani, Behzad. 2003. An instrumental study of prosodic features and intonation in Modern Farsi (Persian). MS thesis, University of Edinburgh. Available at http://www.ling.ed.ac.uk/teaching/postgrad/mscslp/archive/dissertations/2002-3/behzad_mahjani.pdf (accessed 2 February 2011).

Peterson, Gordon E. & Ilse Lehiste. 1960. Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America* 32(6), 693–703.

Pierrehumbert, Janet. 1980. *The phonetics and phonology of English intonation*. Ph.D. dissertation, MIT.

Pierrehumbert, Janet & Mary E. Beckman. 1988. *Japanese tone structure*. Cambridge, MA: MIT Press.

Prieto, Pilar. 2005. Stability effects in tonal clash contexts in Catalan. *Journal of Phonetics* 33, 215–242.

Prieto, Pilar & Chilin Shih. 1995. Effects of tonal clash on down-stepped H* accents in Spanish. *Proceedings of EUROSPEECH'95: Fourth European Conference on Speech Communication and Technology*, vol. 2, 1307–1310.

Prieto, Pilar & Francisco Torreira. 2007. The segmental anchoring hypothesis revisited: Syllable duration and speech rate effects on peak timing in Spanish. *Journal of Phonetics* 35, 473–500.

Prieto, Pilar, Jan van Santen & Julia Hirschberg. 1995. Tonal alignment patterns in Spanish. *Journal of Phonetics* 23, 429–451.

Sadat-Tehrani, Nima. 2007. *The intonational grammar of Persian*. Ph.D. thesis, University of Manitoba.

Sadat-Tehrani, Nima. 2009. The alignment of L + H* pitch accents in Persian intonation. *Journal of the International Phonetic Association* 39, 205–230.

Samareh, Yadollah. 1977. *The arrangement of segmental phonemes in Farsi*. Tehran: Tehran.

Schepman, Astrid, Robin Lickley & D. Robert Ladd. 2006. Effects of vowel length and 'right context' on the alignment of Dutch nuclear accents. *Journal of Phonetics* 34, 1–28.

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

't Hart, Johan, René Collier & Antonie Cohen. 1990. *A perceptual study of intonation: An experimental-phonetic approach*. Cambridge: Cambridge University Press.

Taheri Ardali, Morteza. 2010. The intonation of focus in declarative sentences in Persian, MA thesis, Allameh Tabataba'I University, Tehran.

Taheri Ardali, Morteza & Yi Xu. 2012. Phonetic realization of prosodic focus in Persian. *Proceedings of the 6th International Conference on Speech Prosody*, Shanghai, 326–329.

Toosarvandani, Maziar. 2004. Vowel length in Modern Farsi. *Journal of the Royal Asiatic Society* 14, 241–251.

Windfuhr, Gernot L. 1979. *Persian grammar: History and state of its study* (Trends in Linguistics). Berlin: Walter de Gruyter.

Xu, Yi. 1998. Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica* 55, 179–203.

Xu, Yi. 1999. Effects of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics* 27, 55–105.