

Persian word accent is deletable

Hamed Rahmani¹, Toni Rietveld¹, Carlos Gussenhoven¹

¹ Radboud University Nijmegen

h.rahmani@let.ru.nl, a.rietveld@let.ru.nl, c.gussenhoven@let.ru.nl

Abstract

Two experiments were conducted to determine whether the Persian word accent disappears in two putative 'deaccenting' contexts, post-focal regions and 'presupposed' embedded clauses, to the extent that accentual minimal pairs become homophonous. A production experiment showed low F0 plateaus on the post-focal and presupposed words, while a perception experiment showed that such words are not recognized above a just-noticeable-difference (JND) baseline. The results confirm that accents are deleted and that accent location contrasts are neutralized.

Keywords: accent, deaccentuation, post-focus compression, perception, Persian

1. Introduction

Word prosodic marking in Persian involves the presence of accent, meaning that it does not have any phonetic cues other than F0 [1]. Accentual minimal pairs of phonological words are highly frequent, mainly due to a rule which assigns an accent to the final syllable of morphological words, skipping right-edge clitics [2]. An example is /vpnet/ [vp.nét] 'pickup' vs. /vpn-et/ [vp.net] 'your bathtub'.

One major question about Persian word accent is whether it is deletable. A frequently made claim is that accent is removed after the focus constituent, such that accentual minimal pairs become homophonous [3, 4, 5, 6, 7]. Recent experimental evidence, however, has shown that while the pitch range of the words is considerably reduced after focus, the tonal structure remains intact, as suggested by above-chance identification of the post-focal words by the listeners [1]. The aim of the present study is to reconsider this issue. We believe that the validity of the focus-related data used in [1] is questionable. Specifically, the status of 'focus' is unclear in their design. They used a reading task in which the focused words were printed in bold letters. Since the researchers did not use any question/answer paradigm, it is unclear how each subject interpreted the focus. Possibly, some subjects pronounced the target words merely with more emphasis rather than with focus prosody. This consideration motivated us to replicate [1]'s study with a more realistic elicitation task.

Two experiments were conducted to investigate whether accents are deleted. To place the study in a wider perspective, we decided to include different putative deaccenting contexts in our corpus. In addition to a contrastive focus structure (here referred to as Focus A), there is a way of expressing information structure distinctions by means of prosody in Persian, which is triggered by certain complex sentences with

clausal complements (here referred to as Focus B). In such constructions, there are two possible intonational forms for the embedded clause, neither of which can be easily seen as being more 'neutral' than the other. If the complement clause is accented, it is interpreted as 'asserted' or can be treated as being new to the addressee. If the complement is unaccented, it is interpreted as 'presupposed', or can be treated as being familiar to the addressee.

2. Experiment I

2.1. Methodology

We conducted a production experiment in which speakers had to pronounce a set of pre-designed sentences in different focus contexts. We built up a corpus of sentences featuring one minimal pair contrasting a simple noun ([tro.béʃ] 'proper name') and a cliticized noun ([tro.beʃ] 'his swing') as target words. The target words were to be examined across two focus types: Focus A and Focus B. The corpus was divided in two blocks as given in Table 1. There were three conditions for Focus A: 1-neutral, 2-focal (the target word was contrastively focused), post-focal (the first word of the sentence was contrastively focused). For Focus B, there were two conditions: 1- asserted (the embedded clause is asserted), 2-presupposed (the embedded clause is presupposed).

Table 1. Corpus structure

	Focus type	Carrier sentence
Block 1	Focus A	[un tobeʃ-e]
		'That is Tabesh/his swing.'
Block 2	Focus B	[midunam tobes-e]
		'I know that is Tabesh/his swing.'

The carrier sentences were created in such a way that they provide the minimal context for all focus conditions in each block. The target words were cliticized by the copular clitic [-e] 'is', which makes them part of trisyllabic phonological words that contrast in having the accent on the antepenultimate syllable (*initial accent* condition) or on the penultimate syllable (*final accent* condition). By having a word-final unaccented syllable in all cases, we avoided local phrase-final effects on the realization of the target syllables. To elicit the intended focus conditions, carrier sentences were preceded by context sentences, with which they formed mini-dialogues (see Table 2 for examples). Within each block, some practice dialogues were provided, which were different from the test materials.

Table 2. Mini-dialogues to elicit the asserted (top panel) and the presupposed (bottom panel) conditions of Focus B for the word [tvbé/].

- A: Have you heard about the new teacher?
- B: I know he is called Tabesh.
- A: He says his name is Tabesh but I don't think so.
- B: I know he is called Tabesh.

The order of blocks and of test dialogues within each block was randomized per speaker. Materials were presented in the form of a booklet with one dialogue per page in Standard Persian orthography. A training session preceded the recording, during which speakers practiced the example dialogues. In the recording session, the experimenter pronounced the context sentences and the speakers read aloud the responses from the booklet. Speakers were free to repeat themselves if they thought they had not pronounced a sentence correctly.

Eight native speakers, all with university education, participated in the experiment, four male and four female, aged from 27 to 37. The recording was done using a Shure SM58 vocal cardioids microphone (44.1 kHz, mono channel, 16-bit). All 80 response sentences (10 sentences × 8 speakers) were recorded and digitally saved for analysis.

2.2. Results

A Praat script was used to carry out the fundamental frequency (F0) measurements [8]. Time-normalized F0 curves pooled over all eight speakers are reported in Figures 1 and 2 for Block 1 and Block 2, respectively. Time-normalized F0 was collected at 12 equally spaced points in each syllable. We had no missing data.

Accented syllables are substantially higher than their unaccented counterparts in *neutral* and *focal* pronunciations of Focus A and in *asserted* pronunciation of Focus B. There is evident pitch compression in *post-focal* condition of Focus A and in *presupposed* condition of Focus B, realized as low plateaus. In Figure 1, a comparison of *neutral* and *focal* contrasts (top and middle panels) with *post-focal* contrasts (bottom panel) suggests that post-focal words are deaccented. For Focus B, a similar pattern of deaccentuation is observed in *presupposed* condition (see Figure 2).

To statistically evaluate the effect of focus and accent on the targeted syllables, we used four separate repeated measures ANOVAs with F0 values as dependent variable and FOCUS (neutral, focal, post-focal for Focus A; asserted, presupposed for Focus B) and LOCATION (initial accent, final accent) as independent variables. To minimize the carryover effects of the preceding syllables, F0 values were obtained from the second half of the syllables. This resulted in 48 data points per condition (6 measuring points × 8 speakers). Each analysis included data from one focus type (Focus A, Focus B) and one syllable position (first syllable, second syllable). The analyses are reported in Tables 3 and 4. A highly significant interaction of FOCUS and LOCATION was found in each ANOVA, due to the fact that post-focal condition of Focus A and presupposed condition of Focus B markedly reduced the F0 of the accented syllables.

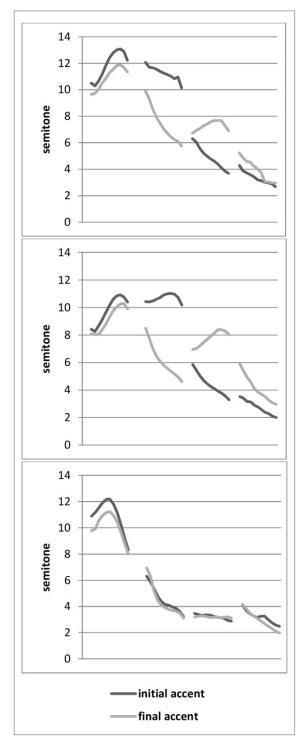


Figure 1. Mean F0 contours in semitones (reference: 100 Hz) for [un tobefe] on normalized time scale with target words in neutral (top), focal (middle) and post-focal (bottom) pronunciations, pooled over 8 speakers.

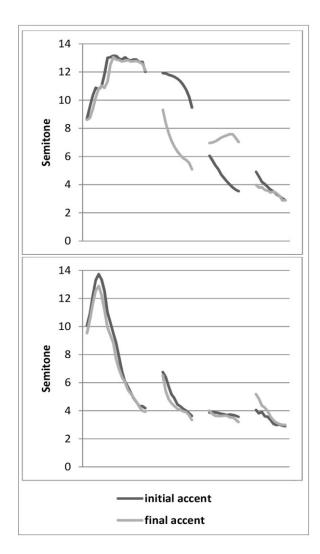


Figure 2. Mean F0 contours in semitones (reference: 100 Hz) for [midunam tubefe] on normalized time scale with target words in asserted pronunciation of the embedded clause (top), and presupposed pronunciation of the embedded clause (bottom), pooled over 8 speakers.

Table 3. ANOVA summary of the effects of Focus A and accent location on F0 measures of the syllable [tv] (top panel) and the syllable [be] (bottom panel) of [tvbefe]

Effects	ANOVA
Focus	$F_{2,94} = 694.07 *, p < .001, \eta^2 = .937$
Location	$F_{1,47} = 289.65 *, p < .001, \eta^2 = .860$
Focus × Location	$F_{2,94} = 192.46 *, p < .001, \eta^2 = .804$
Focus	$F_{2,94} = 150.41 *, p < .001, \eta^2 = .762$
Location	$F_{1,47} = 551.17 *, p < .001, \eta^2 = .921$
Focus × Location	$F_{2,94} = 157.48 *, p < .001, \eta^2 = .770$

Table 4. ANOVA summary of the effects of Focus B and accent location on F0 measures of the syllable [tv] (top panel) and the syllable [be] (bottom panel) of [tvbe[e].

Effects	ANOVA
Focus	$F_{1,47} = 524.26 *, p < .001, \eta^2 = .918$
Location	$F_{1,47} = 358.22 *, p < .001, \eta^2 = .884$
Focus × Location	$F_{1,47} = 155.92 *, p < .001, \eta^2 = .768$
Focus	$F_{1,47} = 205.19 *, p < .001, \eta^2 = .814$
Location	$F_{1,47} = 324.58 *, p < .001, \eta^2 = .874$
Focus × Location	$F_{1,47} = 278.27 *, p < .001, \eta^2 = .856$

3. Experiment II

3.1. Methodology

The result of Experiment I suggested the tonal distinction between the members of the minimal pair is lost in *post-focal* condition of Focus A and in *presupposed* condition of Focus B. If the accent in fact disappears in these conditions, the members of the minimal pair should become homophonous. We designed a word identification task to verify this prediction.

As stimuli, we used all 80 sentences recorded in Experiment I. Two experimental blocks were constructed, one containing Focus A sentences (Block 1), and the other containing Focus B sentences (Block 2). The experiment was presented using a Multiple Forced Choice task in Praat [8]. Listeners were instructed that they would hear several sentences, each containing one of the two words [tóbeʃ] 'his swing' and [tobéʃ] 'proper name', and that they should choose which of these words each sentence contained. Target words were displayed on the screen in standard Persian orthography in two clickable boxes.

The order of the stimuli in each block was randomized per listener. The order of blocks was counterbalanced across the listeners. Subjects were first given four practice trials to gain familiarity with the task, under the supervision of the experimenter. During the test proper, subjects could listen to each stimulus as often as desired, but they were reminded that once they made their choice, it could not be changed. A total of 22 native listeners, different from the speakers in Experiment I, were recruited, 12 male and 10 female, aged from 17 to 45.

3.2. Results

We opted for an analysis based on *d*-prime (*d*'), a sensitivity index used in Signal Detection Theory [9], given by (1):

$$d' = z$$
 score (hit rate) – z score (false alarm rate) (1)

where *hit* rate is the proportion of correct responses to one stimulus type and *false alarm* rate is the proportion of incorrect responses to the other stimulus type. The advantage of a *d*' analysis over an analysis of mean percentages correct is

that the measure d' eliminates any biases in the response rates that may have crept in due to the decision rules a subject uses. We calculated hits and false alarms for the minimal pair in each experimental block. d' values were obtained for each listener. In general, higher d' scores suggest better performance. For purposes of our study, we took d' = 1.35 to be the baseline performance. This value corresponds with correct performance on 75% of trials [9], which is often seen as the correct rate for a just-noticeable-difference (JND) [10]. Where d' would otherwise be undefined (a hit or false alarm rate of zero or 1), rates of 0 were replaced with [0.5 / N] (N: number of equal stimuli), and rates of 1 were replaced with [(n - 0.5) / N] (N: number of unequal stimuli) [9]. In our data, these values were 0.062 and 0.937, respectively. The maximum d' score we obtained was 3.06 and the lowest was -1.15. Figure 3 presents averaged d' values with corresponding standard errors pooled over 22 listeners for each experimental block. We had no missing data.

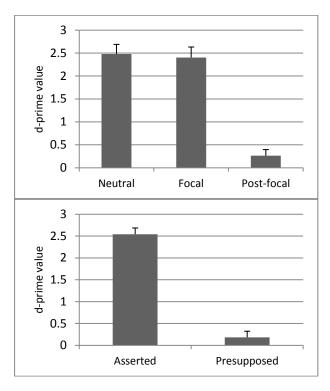


Figure 3. Mean d'scores in a word identification task for the minimal pair [tibbe]/tvbéf], as obtained in neutral, focal and post-focal conditions of Focus A (top panel) and in asserted and presupposed conditions of Focus B (bottom panel), pooled over 22 listeners.

d' values for each block were submitted to a separate repeated-measures ANOVA with FOCUS (*neutral*, *focal*, *post-focal*) as factor for Block 1, and FOCUS (*asserted*, *presupposed*) as factor for Block 2. The analyses revealed highly significant effect of FOCUS for both Block 1 [F(2,40)=71.82, p<.001, η_p^2 =.782] and Block 2 [F(1,20)=215.17, p<.001, η_p^2 =.915]. For Block 1, the Sidak post-hoc test showed that *post-focal* was significantly different from *neutral* (p<.001) and *focal* (p<.001). Taken together, listeners performed substantially

worse in *post-focal* condition of Focus A and in *presupposed* condition of Focus B.

Finally, one-tailed t-tests were used to compare different performance conditions against the JND baseline (i.e. *d*'=1.35). The results showed that the identification was significantly above the baseline in *neutral* (t=5.41, *p*<.000) and *focal* conditions (t=4.63, *p*<.000) of Focus A, as well as in *asserted* condition of Focus B (t=8.39, *p*<.000). The identification of *post-focal* words in Focus A and of *presupposed* words in Focus B did not reach the baseline level, suggesting that accent location contrasts are neutralized in these conditions.

4. Conclusions

The purpose of the current study was to determine whether Persian word accents are deleted in post-focal regions and 'presupposed' embedded clauses, such that accentual minimal pairs become homophonous. The results of two experiments confirmed that accents are deleted. A production experiment showed low F0 plateaus on the post-focal and presupposed words, while perception experiment showed that deaccented members of minimal pairs are not recognized above the just-noticeable-difference (JND) baseline.

The finding that Persian deaccents after the focus is in agreement with Xu's suggestion that post-focal compression of the pitch range (PFC) may be an areal characteristic covering Europe as well as parts of Asia [11]. Persian, having no phonetic stress [1], represents a clear case of a language that neutralizes the accent contrast. English and Dutch equally deaccent words in the post-focal region, but because distinctions between primary and secondary stress may be preserved [12], they are not uncontroversially neutralizing in the way Persian is.

5. Acknowledgements

We thank our participants at the University of Tehran and Sajjad Peyvasteh, Mehran Ghajar and Joop Kerkhoff for practical and technical support.

6. References

- [1] Abolhasanizadeh, V., Bijankhan, M., & Gussenhoven, C. (2012). The Persian pitch accent and its retention after the focus. *Lingua*, *122*(13), 1380-1394.
- [2] Ferguson, C. A. (1957). Word stress in Persian. *Language*, 123-135.
- [3] Vahidian-Kamyar, T. (2001). Næva-ye goftar dær farsi [Melody of speech in Persian]. Mashhad: Ferdowsi University Press.
- [4] Eslami, M. (2000). Senaxt-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 computeraided speech recognition]. Tehran: Tehran University PhD thesis.
- [5] Sadat-Tehrani, N. (2007). The intonational grammar of Persian (Doctoral dissertation, University of Manitoba).
- [6] Hosseini Seyed, A. (2014). The Phonology and Phonetics of Prosodic Prominence in Persian (Doctoral dissertation).
- [7] Scarborough, R. (2007). The intonation of focus in Farsi. UCLA working papers in phonetics. No. 105, pp. 19-34.
- [8] Boersma, P., & Weenink, D. (2014). Praat, a system for doing phonetics by computer.
- [9] Macmillan, N. A., & Creelman, C. D. (1991). Detection theory: A user's guideCambridge University Press. *New York*.

- [10] Durrant, J. D., Dunnick, N. R., Lovrinic, J. H., McCallum, R. W., & Sandler, C. M. (1977). *Bases of hearing science*. Williams
- [11] Xu, Y., Chen, S. W., & Wang, B. (2012). Prosodic focus with and without post-focus compression: A typological divide within the same language family?.
 [12] Gussenhoven, C. (1991). The English rhythm rule as an accent deletion rule. *Phonology*, 8(01), 1-35.