The role of F0 and duration in the identification of wh-in-situ questions in Persian

Zohreh Shiamizadeh a,b,*, Johanneke Caspers a,b, Niels O. Schiller a,b

a Leiden University Centre for Linguistics, Leiden University, 2311 BX, Leiden, The Netherlands
b Leiden Institute for Brain and Cognition, LUMC, 2300 RC Leiden, Leiden, The Netherlands

ARTICLE INFO
Article history:
Received 20 January 2017
Revised 26 June 2017
Accepted 28 July 2017
Available online 31 July 2017

Keywords:
Persian
Wh-in-situ questions
Declaratives
Prosodic correlates
Manipulation
Sentence type identification

ABSTRACT
Prosody plays an influential role in the recognition of Persian wh-in-situ questions (Shiamizadeh et al., in press). Perception of speech intonation is supported by several prosodic correlates (Lehiste, 1970). For instance, perception studies demonstrate that prosodic correlates do not contribute similarly to the identification of contrasts (oppositions) in speech intonation, and F0 contributes more prominently to the categorization of contrast in speech intonation (Lehiste, 1976; Peng et al., 2012). The current study manipulates a number of different prosodic correlates to investigate their relative contribution to sentence type categorization in Persian. This study also investigates whether identification is improved in congruent cue conditions in comparison to conflicting cue conditions. For this purpose, a perception experiment was designed in which native Persian speakers were required to listen to manipulated stimuli and decide if they perceived a wh-question or a declarative sentence. Based on the literature, we expect a primary role for F0 and a secondary role for duration in the perception of sentence modality contrasts. We also predict improved identification of sentence type in matching cue conditions in comparison to mismatching cue conditions. The results of the study confirm our predictions, i.e. a primary role of F0, and an increased identification in the matching cue condition is applicable to the perception of the contrast between Persian wh-in-situ questions and declaratives.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

One linguistic function that is frequently used in daily conversation is interrogativity. Wh-questions are interrogative expressions that use a wh-phrase to enquire about desired information. Wh-questions can be divided into two groups: fronted and wh-in-situ questions. In fronted wh-questions, the wh-phrase moves to the beginning of the sentence to form a wh-question (see Example 1), whereas in wh-in-situ questions the wh-phrase does not move to the sentence-initial position (Carnie, 2007; Chomsky, 1977). One of the languages that is characterized by wh-in-situ is Persian (Abedi et al., 2012; Adli, 2007; Gorjian et al., 2012; Kahnemuyipour, 2009; Karimi, 2005; Karimi and Taleghani, 2007; Lotfi, 2003; Megerdoumian and Ganjavi, 2000; Mirsaedi, 2006; Toosarvandani, 2008). In Persian, wh-questions are in-situ by default. The wh-phrase does not move to the beginning of the sentence, rather it occurs at the same site where its declarative counterpart is expected to occur (see Example 2b).1

\[ \text{(1) a. Mary carries a book.} \]
\[ \text{b. What does Mary carry?} \]
\[ \text{(2) a. Maryam diruz ketab xærid.} \]
\[ \text{Maryam yesterday book buy.PAST.3SG.} \]
\[ \text{“Maryam bought a book yesterday.”} \]
\[ \text{b. Maryam diruz \( \bar{t} \)i xærid?} \]
\[ \text{Maryam yesterday what buy.PAST.3SG.} \]
\[ \text{“What did Maryam buy yesterday?”} \]

1 The wh-phrase can optionally move to the earlier parts, including the beginning of the sentence for non-syntactic reasons (Abedi et al., 2012; Adli, 2007; Gorjian et al., 2012; Kahnemuyipour, 2009; Karimi, 2005; Karimi and Taleghani, 2007; Lotfi, 2003; Megerdoumian & Ganjavi, 2000; Mirsaedi, 2006; Toosarvandani, 2007). These authors claim that the movement of the wh-phrase to earlier parts of the sentence is not triggered by the syntactic (+wh) feature. Therefore, Persian cannot be categorized as a wh-movement language. Adli (2007), Kahnemuyipour (2001), Karimi (2005), Karimi and Taleghani (2007), Lotfi (2003) and Toosarvandani (2008) claim that the wh-phrase moves to earlier parts of the sentence to receive contrastive focus. (1) is an example of a sentence in which the wh-phrase “chi” (what) moves to the beginning of the sentence to receive contrastive focus. The declarative and wh-in-situ question counterpart of this example is given in (2a) and (2b) within the text.

1. \( \bar{t} \)i maryam diruz xærid? what Maryam yesterday buy.PAST.3SG. “What did Maryam buy yesterday?”

* Corresponding author at: Leiden University Centre for Linguistics, Leiden University, 2311 BX, Leiden, The Netherlands.

E-mail addresses: zshiamizadeh@gmail.com, z.shiamizadeh@hum.leidenuniv.nl (Z. Shiamizadeh).

http://dx.doi.org/10.1016/j.specom.2017.07.005
0167-6393/© 2017 Elsevier B.V. All rights reserved.
Prosody plays an influential role in marking interrogativity (e.g., D’Imperio, 2000; Haan, 2001; Vion and Colas, 2006). According to Van Heuven and Haan (2000), hyper- and hypo-theory of speech production (Lindblom, 1990) suggests prosodic interrogativity cues will be stronger when there are none or fewer lexi-co-syntactic interrogativity markers in the sentence. Shiamizadeh et al. (minor revision) delineated that this proposal is applicable to Persian wh-in-situ questions. According to Shiamizadeh et al. (minor revision), which is a production study, prosodic correlates mark the pre-wh part in wh-in-situ questions as opposed to declaratives in Persian.

The production of speech intonation is mediated through a combination of cues, namely F0, intensity and duration (Lehiste, 1970; Peng et al., 2012). Similarly, the perception of speech intonation is assumed to be conveyed through several different cues (Hazan and Rosen, 1991; Lehiste, 1976; Peng et al., 2012). In line with a number of studies (e.g., Hazan and Rosen, 1991; Lehiste, 1970, 1976; Peng et al., 2012; Van Heuven and Van Zanten, 2005), our previous studies demonstrate that in the production (Shiamizadeh et al., minor revision) and perception (Shiamizadeh et al., in press) of the pre-wh part in wh-in-situ questions vs. statements in Persian a combination of several acoustic cues creates the prosodic marking of sentence modality contrast. The result of the studies by Shiamizadeh et al., (minor revision; in press) brings the focus to the relative contribution of prosodic correlates in the perception of wh-in-situ questions as opposed to statements in Persian. A perception experiment in which relevant prosodic correlates are manipulated can investigate the relative contribution of prosodic correlates to sentence modality contrast identification in Persian.

1.1. Background

1.1.1. Production and perception of prosodic correlates of Persian wh-in-situ questions

To our knowledge, Shiamizadeh et al. (minor revision; in press) is the first systematic and experimental study investigating the role of acoustic correlates of prosody in the production and perception of wh-in-situ questions vs. declaratives in Persian. In a production study, Shiamizadeh et al. (minor revision) compared the prosodic correlates of wh-in-situ questions and their declarative counterparts to investigate whether acoustic correlates of the pre-wh part mark wh-in-situ questions as opposed to declaratives in the absence of the wh-phrase at the beginning of wh-questions. They found that a higher level of pitch register, a higher F0 onset and a shorter duration contributed to the prosodic distinction of the pre-wh part in wh-questions as opposed to declaratives. A steeper inclination of the F0 contour and a greater excursion size of the pre-wh words were two additional features that give rise to the prosodic markedness of the pre-wh part in wh-questions.

The production study (Shiamizadeh et al., minor revision) was followed by a perception study (Shiamizadeh et al., in press). The purpose of the perception study was to investigate whether the prosody of the pre-wh part can cue the identification of wh-in-situ questions in the absence of the wh-phrase at the sentence-initial position. The results revealed that prosody guides perception of wh-in-situ questions before the wh-phrase is uttered by the speaker.

1.1.2. Empirical background

Speech intonation is mainly conveyed through variations in F0, but this variation is often accompanied by variation in intensity and duration patterns (Cooper and Sorensen, 1981; Ladd, 1996; Lehiste, 1970, 1976; Peng et al., 2012; Van Heuven and Van Zanten, 2005). Similarly, recognition of speech intonation is mainly achieved by relying on F0 variation as the primary cue, while other cues such as duration and intensity are considered secondary cues (Lehiste, 1976; Peng et al., 2012). The primary role of F0 and the secondary role of duration in the perception of speech intonation can be attributed to the recognition of questions where the syntactic interrogativity markers are absent or occur later in the sentence.

Tapping into the relative contribution of melodic cues and durational cues for the identification of interrogatives vs. declaratives is a relatively untouched territory. To our knowledge, Cangemi and D’Imperio (2013) is the only study investigating the relative contribution of duration in the perception of declaratives as opposed to interrogatives. According to the production experiment by Cangemi and D’Imperio (2013), the contrast between statements and questions in the Neapolitan regional variety of Italian is intonationally specified by different tunes and different durational patterns. Durational differences between these two sentence modalities are localized at specific portions of the sentence: the first sound segment is longer in statements, while the last sound segment is longer in questions. In the same study, Cangemi and D’Imperio (2013) investigated the relevance of the durational differences between statements and questions in the perception of the correct sentence type. In their perception study, Cangemi and D’Imperio (2013) used resynthesized sentences with three different F0 contours: question F0 contour, ambiguous F0 contour and statement F0 contour. Then, they implemented three different durational patterns (question duration, ambiguous duration and statement duration) on stimuli in each group of F0 contour. Therefore, three different durational patterns were transplanted on a stimulus with the same F0 contour. For example, there were three versions of a stimulus with a question F0 contour: a) with question duration, b) with ambiguous duration, c) with statement duration. They predicted that if duration influences sentence type identification, items with different durational patterns but the same F0 contour would receive different responses. The result of the perception experiment showed that a) F0 cues are relevant in the perception of sentence modality contrasts, whereas durational cues do not affect identification of the correct sentence type, and b) the reaction time (RT) to the stimuli with ambiguous F0 is longer than the RT to other conditions.

This result implies that, in comparison to durational cues, F0 cues contribute more to the identification of sentence modality contrast. Furthermore, the cooperation of F0 and durational cues have a stronger effect on the perception of correct sentence type than conflicting F0 and duration. Augmented identification of speech contrast in the cooperating cue condition in comparison to the conflicting cue condition has been reported in other studies which investigated the role of acoustic cues to the identification of minimal pairs (e.g., Hazan and Rosen, 1991).

2. Research questions, approach and hypothesis

2.1. Research questions and approach

To investigate the relative contribution of F0 and durational cues, F0 contour and durational pattern must be manipulated. Therefore, this study addresses the following questions:

1) Does F0 duration or their cooperation contribute more to the distinction of wh-in-situ questions from declaratives in Persian?
(2) Is there a difference between RT to the congruent and incongruent cue conditions?

The answer to the first research question potentially sheds light on the applicability of two proposals to sentence type identification in Persian: a) the primary role of F0 and the secondary role of duration in the recognition of speech intonation contrasts (Lehiste, 1976; Peng et al., 2012), b) increased categorization of contrasts in the congruent cue condition in comparison to the conflicting cue condition (Hazen and Rosen, 1991).

To answer the two research questions above, F0 and durational patterns of the pre-wh part of declaratives and their matching wh-questions were manipulated. Then, a forced-choice identification task was designed in which Persian native speakers listened to the manipulated pre-wh part of wh-in-situ questions and their matching declaratives. They were required to decide whether what they heard was going to be a wh-question or a declarative. The participants were asked to opt for either a declarative or a question by pressing a key as quickly as possible. Their reaction times were recorded.

2.2. Hypothesis

The primary role of F0 variation in the perception of speech intonation is reported in the literature (e.g., Cangemi and D’Imperio, 2013; Lehiste, 1976; Peng et al., 2012). Other studies demonstrated an increased effect of congruent cue conditions on the identification of speech contrasts (Hazen and Rosen, 1991) in comparison to the conflicting cue conditions. The above-mentioned findings can be logically extended to the perception of wh-in-situ question in Persian. Therefore, we hypothesize that:

(a) F0 is the dominant contributor to the categorization of sentence modality (in-situ wh-questions versus statements) in Persian. In other words, listeners will draw on F0 cues even in duration manipulated items to decide on the sentence type. This means F0 manipulated items will receive more correct (expected) responses³ than duration manipulated items.

(b) Identification of the sentence type in congruent cue conditions (stimuli where both F0 and duration are manipulated) is augmented in comparison to conflicting cue conditions (stimuli in which either F0 or duration is manipulated). Therefore, more identification responses will be given to the items with congruent F0 and duration in comparison to the items with conflicting F0 and duration.

(c) The RT to congruent cue conditions is shorter than the RT to conflicting cue conditions.

3. Methodology

3.1. Participants

Twenty-four native speakers of Persian, twelve males and twelve females, took part in this experiment. All of the participants were brought up in Tehran. Fifteen participants lived in Delft. They came to the Netherlands in 2014 or 2015⁴ to continue their education at the Technical University of Delft. Nine participants were students at the University of Qom, Iran. Their age range was between 26 and 40. None of them reported any hearing impairment.

3.2. Materials

3.2.1. Speakers

Part of the sentences elicited from the native Persian speakers in the production experiment on the prosodic correlates of Persian wh-in-situ questions (Shiamizadeh et al., minor revision) were used as the basic materials for the current experiment. To control for the effect of gender on the listeners’ performance in the perception experiment, we chose a male and a female speaker.

Selecting the speakers who keep the two sentence types most distinct in their speech would limit the generalizability of the results to only these speakers. To broaden the generalizability of the results of the current experiment, we picked out those speakers who are most representative of all participants of the production experiment. Speakers whose mean value of the acoustic measurements were closest to the mean value of the acoustic measurements in the production of all speakers (cf. Section 1.1.1) were selected.

3.2.2. Selection of the stimuli

Twenty pairs of sentences elicited from a male and a female speaker in the production experiment by Shiamizadeh et al. (minor revision) comprise the stimuli of this experiment. The structure of the wh-questions and declaratives is illustrated in (3) and (4), respectively.

(3) Subj Adv Wh-phrase Verb
(4) Subj Adv AdP/ IDO/ AdjT/ AdjM/ AdjP Verb

Subject is abbreviated as Subj, adverb as Adv, direct object as DO, adjunct of time as AdjT, adjunct of manner as AdjM and adjunct of place as AdjP. As (4) shows ADP, IDO, AdjT, AdjM and AdjP replace the wh-phrase in declaratives. Therefore, they will be referred to as declarative wh-phrase counterparts (DWC) in the remainder of the paper. An example of a declarative and a matching wh-question is given in (5a) and (5b).

(5) a. mohammadæmin pariruz kærd? kej æsr æmin
Mohammadamin two days ago afternoon swim kej do. PAST. 3SG.

b. mohammadæmin pariruz kærd?
Mohammadamin two days ago when swim kærd?

c. "When did Mohammadamin swim two days ago?"

Five different wh-phrases, two different nouns as the subject, two words as the adverb, two words in each category of DWC and five verbs were used as sentence constituents of the original stimuli in the production experiment. The word constituents of the declaratives and wh-questions are presented in Appendix A. As Appendix A shows, the verb varies along the wh-phrase.

A repeated measures multivariate analysis of variance showed that the interaction effect between the nouns used as the subject and the sentence type (F (5,65) = .397, p > .05; Wilk’s A = .970, η² = .030) and between the words used as the adverb and the sentence type (F (6,12) = .432, p > .05; Wilk’s A = .968, η² = .032) on the acoustic features described in Section 1.1.1 (dependent variables) were not significant. Therefore, we decided to include just one noun as the subject and one word as the adverb in the stimuli of this experiment. Variation in other sentence constituents is kept intact.

The pre-wh part of sentences were separated from the remaining part of the sentence in Praat version 6.0.04 (Boersma and Weenink, 2014) and were used as the stimuli of the current experiment. To prepare the stimuli for this experiment, acoustic correlates of the prosody of the pre-wh part must be manipulated. The manipulation conditions are F0 manipulation (F0M), duration

---

3 Expected and correct responses are defined in Section 3.4.
4 The data were collected in October, 2015.
manipulation (DM) and F0 and duration manipulation (FODM) and resynthesized version of original sentences (RSO).

The total number of stimuli equals 160 (1 subject x 1 adverb x 2 DWCS\(^5\) x 5 wh-phrase and matching verbs x 2 sentence types x 4 manipulation conditions x 2 speakers). Although only the pre-wh parts of the sentences comprise the stimuli of the current experiment, variation in DWCS, wh-phrases and their matching verbs are included in the formula to clarify how we arrived at 160 stimuli.

Sections 3.2.3–3.2.7 explain the manipulation process. In these sections, the stimuli with manipulated acoustic features are called target sound (TS) and the stimuli from which the acoustic features are transplanted to the target sound are called the source sound (SS). For example, if the F0 contour of a declarative is replaced by the F0 contour of a question, the declarative is the TS and the question is the SS. It must be noted that all steps of the manipulation process were automatically accomplished by running scripts in Praat.

3.2.3. Stylizing the pitch contour

The preliminary step in manipulating F0 is stylizing the pitch contour. Before stylizing the pitch contour, the erroneous F0 values (octave errors) were corrected in the original F0 contour using Praat functions. Then, the original pitch contour of the utterance was stylized using a 1 semitone frequency resolution in Praat. Thus, two F0 contours were available, the original or non-stylized pitch contour (panel 2 in Figs. 1 and 2) and the stylized pitch contour (panel 3 in Figs. 1 and 2). The stylized contour produces points which can be manipulated. All pitch points in the stylized pitch contour of the pre-wh part of the sentence were removed except for 6 points: the F0 onset, the valley (L) and peak (H\(^\dagger\)) of the subject pitch accent, the L and H\(^\dagger\) of the adverb pitch accent and the end of the fall of the subject pitch accent (ESPA) (see Figs. 1 and 2). The F0 onset and the L and H\(^\dagger\) of the subject and adverb pitch accents are kept because these are the F0 points that make the prosody of wh-in-situ questions distinct from that of declaratives (Shiamizadeh et al., minor revision). ESPA was also kept so that no audible difference existed between the original and the stylized pitch curves. The end of the fall of the adverb pitch accent is not considered here because the relevant pitch point is not produced by the stylized pitch. In addition, adding a pitch point manually as the end of the fall of the adverb pitch accent does not affect the audible similarity between the original and the stylized pitch curves. The pitch points were interconnected with straight lines.

After stylizing the pitch contour, two steps were taken to accurately determine the location and the F0 value of the valleys and the peaks of subject and adverb pitch accents. At the first step, L and H\(^\dagger\) were indicated in the stylized pitch. At the second step, the valley and the peak points (L and H\(^\dagger\)) produced in the stylized pitch contour were automatically moved to the maximum and minimum F0 of the original F0 contour using Praat functions (this method was adopted from Sadat Tehrani, 2009). The place and the value of the minimum and maximum F0 of the original F0 contour determined by Praat functions (the second step) either coincide with or closely correspond to the L and H\(^\dagger\) produced by the stylized pitch (the first step). There were no F0 plateaus.

3.2.4. F0 manipulation

To investigate the role of F0 in the identification of wh-questions, the F0 contour was manipulated. After stylizing the pitch contour, F0 manipulation was performed. The pitch points which were kept during the stylization process were manipulated: the F0 onset, the L and H\(^\dagger\) of the subject and adverb pitch accents, and the end of the fall of the subject pitch accent. These pitch points in the stylized pitch of the TS were manipulated to create the intonational characteristics of their corresponding pitch points in the stylized pitch of the SS.

In the F0 manipulation process, F0 values were manipulated, but the time coordinates and duration were left unaltered. Therefore, in a question where F0 was manipulated, the question keeps its original durational pattern but its F0 contour is changed to the F0 pattern of its declarative counterpart. The correct (expected) response to this item is “declarative”. Similarly, in a declarative where F0 was manipulated, the declarative keeps its original durational pattern but its F0 contour is changed to the F0 pattern of its corresponding question. The correct (expected) response to this item is “question”.

3.2.5. Duration manipulation

Duration was manipulated to inspect the role of duration in the identification of declaratives as opposed to wh-questions. Since the pitch contour was stylized in the F0 manipulated stimuli, the pitch contour in duration manipulated items has to be stylized as well. The pitch was stylized in the same manner described in Section 3.2.3.

To manipulate duration, the duration tier of the SS must be transplanted to the TS. Therefore, duration of the duration tier of the TS was multiplied by the quotient of the duration of the SS divided by the duration of the TS.

In duration manipulated items, only the duration was manipulated and the F0 was left unaltered. Therefore, in a question where the duration was manipulated, the question keeps its original F0 contour but its duration is switched to the durational pattern of its declarative counterpart. The correct (expected) response to this item is “declarative”. Similarly, in a declarative where duration was manipulated, the declarative keeps its original F0 contour, but its duration is changed to the duration of its corresponding question. The correct (expected) response to this item is “question”.

3.2.6. F0 and duration manipulation

Manipulating the duration tier of F0 manipulated items provides us with items in which both F0 and duration are manipulated. The duration was manipulated as explained in the Section 3.2.5. For instance, a statement gets both the F0 contour and durational pattern of its corresponding question and a question receives the F0 contour and durational pattern of its corresponding statement. The correct (expected) answer to a question which gets the F0 and durational properties of its corresponding declarative is “declarative”. And the correct (expected) answer to a declarative to which the F0 and duration of its corresponding question is transplanted is “question”.

3.2.7. Resynthesized original sounds

To manipulate the stimuli of this experiment, the pitch contour of each sound was stylized as a first step and the resynthesized version of the sound was extracted from the manipulation object of each sound as the last step. Therefore, the resynthesized version of each sound with a stylized pitch contour is required as the baseline condition of the current experiment. Pitch stylization procedure is explained in Section 3.2.3. After stylizing the pitch contour of each sound, the resynthesized version of the sound was automatically extracted from manipulation object of the sound file in

\(^5\) As mentioned earlier, there are five categories of DWCS which correspond with the number of wh-phrases. DWCS replace the wh-phrases in declaratives. Therefore, the use of the categories of the DWCS in declaratives varies along with the use of the wh-phrases in wh-questions, i.e. if a wh-phrase asking about ADO is used in the wh-question, a category of the DWC referring to ADO is used in the declarative. This means that the coefficient of wh-phrases in the formula accounts for the number of categories of DWCS as well. However, since two words are used in each category of DWCS, the coefficient of 2 is defined for DWC in the formula.
3.3. Procedure

A forced-choice sentence categorization task was designed in E-prime 2.0.10 (Psychology Software Tools, 2012). The subjects were seated in front of a computer in a quiet room. The experiment started with the presentation of written instructions on the computer screen. Participants could take as much time as they wanted to read the instructions. They were allowed to ask questions about the instructions if they had any. Then, they were introduced to the task by performing a practice session. The practice session included 16 non-experimental items (1 item x 2 sentence types x 4 manipulation conditions x 2 speakers). The pre-wh part of one declarative and one wh-question from each manipulation condition read by two different speakers were played to them over headphones (Sennheiser PC 141 Headset). Participants were instructed to decide whether what they heard is going to be a wh-question or a declarative. After hearing each stimulus, they had four seconds to opt for either a wh-question or a declarative by pressing either the “M” key or “Z” key on the keyboard. If participants did not give a response within four seconds, the experiment automatically proceeded to the next trial. To help participants not confuse which key they needed to press for which option (declarative or wh-question), a full stop and a question mark along with the letters M and Z appeared on two opposite sides of the screen, at the same time as a stimulus was played to them. Three seconds after pressing the key, the next stimulus was played. Three seconds were entered as the inter-item interval (ITI). During the ITI the Persian equivalent of the word “next item” was shown on the screen as an indication that the next stimulus was coming. The presentation order of the items of the practice session was the same for all participants. They were allowed to do the practice session two times if they wanted. Having accomplished the practice session, partic-
Participants embarked on the main part of the experiment when they felt ready.

The stimuli were divided into four blocks. Each block contained 40 items. The participants should have at least a three-minutes break between blocks. After the break, they were asked to press the space bar to continue to the next block when they felt ready to start again. Every block started with a warming-up which consisted of two items other than the main items of the block. The purpose of including warming-up items was to make participants ready for the new block after the break. The sequence in which the blocks and the sentences within each block were presented was randomized for each participant. The order in which the full stop and the question mark and the corresponding letters (M or Z) was displayed on the screen was fixed for individual participants, whereas it was counterbalanced for different subjects. The procedure of the main session was identical to that of the practice session. The experiment lasted about 35 min for each participant.

![Waveform and Pitch](image)

**Fig. 2.** The wave form, original (non-stylized) pitch contour and stylized pitch contour of the pre-wh part of a wh-question. The first panel presents the wave form, the second panel displays the non-stylized pitch contour and the last panel contains the stylized pitch contour. In the stylized contour, only the points designating F0 onset, F0 onset, L and H* and the end of the fall of the subject pitch accent (ESPA) are kept. In the first tier of the third panel, the “L” and “H*” which correspond with the word “mohamedzamin” in the second tier represent the valley and the peak of the subject pitch accent, ESPA is the end of the fall of the subject pitch accent and the “L” and “H*” which correspond with the word “pariruz” in the second tier demonstrate the valley and the peak of the adverb pitch accent. The second tier in the third panel represents word boundaries.

<table>
<thead>
<tr>
<th>Manipulation condition</th>
<th>Stimulus type</th>
<th>Expected response</th>
</tr>
</thead>
<tbody>
<tr>
<td>F0 manipulation</td>
<td>F0qDs</td>
<td>Question</td>
</tr>
<tr>
<td>Duration manipulation</td>
<td>F0sDq</td>
<td>Question</td>
</tr>
<tr>
<td>F0 and duration manipulation</td>
<td>F0qDq</td>
<td>Question</td>
</tr>
<tr>
<td>Resynthesized original</td>
<td>Declarative+</td>
<td>Declarative</td>
</tr>
</tbody>
</table>

* F0q means that a stimulus has the F0 contour of a question and Ds means that a stimulus has the durational pattern of a statement.

Sentence type can be defined for resynthesized original items since they are not manipulated. However, sentence type of the manipulated stimuli cannot be determined because the prosodic correlates of the stimuli are manipulated in manipulation conditions. Hence, abbreviations for prosodic correlates were used to define the stimulus type in manipulation conditions.
3.4. Data analysis

As mentioned previously, the F0 contour and durational pattern were manipulated to investigate their relative contribution to sentence type categorization. Therefore, an expected response was defined for each stimulus based on the manipulated prosodic feature in that stimulus. To be more specific, F0 contour of an FOM stimulus, durational pattern of a DM stimulus and F0 and durational pattern of a FODM stimulus determines the expected response to that stimulus. Table 1 presents the expected response to the stimuli in different manipulation conditions.

The responses given by participants were transferred from E-prime to SPSS version 22 (IBM SPSS, 2012). If the given response is compatible with the expected response, it is coded as correct and if it is not compatible with the expected response, it is coded as incorrect. Thus, response is reported in terms of percentage of correct answers. To correct for possible response bias, responses were transformed to A’ scores (Stanislaw and Todorov, 1999). In addition to percentage of correct answers, A’ scores of the responses to each manipulation condition are reported as well. A repeated measures ANOVA (RM-ANOVA) was run to investigate the effect of manipulation condition on response accuracy.

RT data were also transferred from E-prime to SPSS. RT was calculated as the time between response and stimulus offset. A separate RM-ANOVA was performed to inspect the influence of manipulation condition on RT.

4. Results

4.1. Accuracy

Table 2 presents the percentage and number of declarative and question responses given by participants across manipulation conditions.

Table 3 presents the response accuracy across manipulation conditions. As Table 3 illustrates, the response accuracy increases in the following order: DM < FOM < FODM < RSO.

As mentioned in Section 3.4, the responses were transformed to A’ to correct for the possible response bias (Stanislaw and Todorov, 1999). A’ values range from .5 to 1; .5 shows inability of the listeners to discriminate the two sentence types and 1 presents perfect performance. Mean A’ score for each manipulation condition is presented in Fig. 3.

To inspect the effect of the manipulation condition on response accuracy, a RM-ANOVA was run with aggregated responses as the dependent variable and the manipulation condition as the independent variable. The multivariate test demonstrated a main effect of the manipulation condition on the response accuracy (F(3,21) = 227.161, p < .01; Wilks’ Lambda = .030, η² = .970). Pairwise comparison tests using the Bonferroni correction delineated that the difference between all manipulation conditions is significant p < .01 except for the difference between the FOM and DM items p > .05.

4.2. Reaction time analysis

Reaction times were calculated from the stimulus offset. Table 4 provides mean and standard deviation of RT across manipulation conditions. All RT data is reported in seconds.

According to Table 2, the RT to different manipulation conditions increases as follows RSO < FODM < DM < FOM. RT data was submitted to a RM-ANOVA with manipulation conditions as independent variable. A multivariate test showed a main effect
of the manipulation condition \( F (3,21) = 16.231, p < .001; \) Wilks' Lambda = .301, \( \eta^2 = .699 \). Pairwise comparison tests revealed that the difference between all manipulation conditions is significant \( (p < .03) \), except for the difference between DM and FOM and between DM and FODM. The p-value was adjusted for multiple comparisons using the Bonferroni correction test.

5. Discussion and conclusion

The present study was designed to weight prosodic cues to the identification of wh-in-situ questions as opposed to declaratives in Persian. The result is consistent with the prediction that F0 variation is the more important contributor to the sentence modality contrast in Persian and that sentence type identification is improved in the congruent cue condition in comparison to the incongruent cue condition. Sentence type categorization across manipulation conditions is improved in the following order: DM < FOM < FODM < RSO. Sentence type identification in the FOM items is higher in comparison to the sentence type identification in the DM items. A possible justification is that listeners draw on F0 cues even in the duration manipulated condition to decide on the sentence type. This finding corroborates what has been proposed in the literature: F0 performs a prominent role and duration a secondary role in the recognition of speech intonation \( (\text{Lehiste, } 1976; \text{ Peng et al., } 2012; \text{ Cangemi and D’Imperio, } 2013) \). Categorization of sentence type in the RSO and FODM items (congruent cue condition) was augmented in comparison to the FOM and DM items (incongruent cue condition). This result suggests that the proposal of the increased identification of speech contrast in the congruent cue condition in comparison to the conflicting cue condition \( (\text{e.g. Hazan and Rosen, } 1991) \) is applicable to wh-question vs. statement identification in Persian.

RT to different manipulation conditions increases as follows: RSO < FODM < DM < FOM. In line with our predictions, the RT to congruent cue conditions is shorter than the RT to the incongruent cue conditions.

To our knowledge, no study has investigated whether or not dialectal differences affect the role of prosody in the production and perception of interrogatives, including wh-in-situ questions vs. statements in Persian. Therefore, we suggest that the results of this study might be limited to standard Persian.

The general conclusion of this study sets the ground for further research: whether the dominant role of F0 and the improved perception of the sentence type in congruent cue conditions in comparison to conflicting cue conditions is extendable to other Persian structures that lack syntactic markers and are prosodically marked.
Appendix A

Structure and word constituents of the sentences from which the stimuli of this experiment is extracted. Only the pre-wh part (subject and the adverb) comprise the stimuli of this experiment.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Adverb</th>
<th>3rd constituent of the sentence</th>
<th>Verb</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) mohamadزارمین</td>
<td>1) ۲پارسیز</td>
<td>Inanimate DO&lt;sup&gt;2&lt;/sup&gt;</td>
<td>1) کارف (shoe)</td>
</tr>
<tr>
<td>(Mohamadamin)</td>
<td>(two days ago)</td>
<td>2) کیف (bag)</td>
<td>2) دزی شد (repair)</td>
</tr>
<tr>
<td>2) mohamadزارلی</td>
<td>1) ۲کودش</td>
<td>Animate DO</td>
<td>1) ژنک (Yas)</td>
</tr>
<tr>
<td>(Mohamadali)</td>
<td>(where)</td>
<td>2) نوزه (Naz)</td>
<td></td>
</tr>
<tr>
<td>3) کودش</td>
<td>1) ۲دیوژل (jungle)</td>
<td>Adjust of Place&lt;sup&gt;2&lt;/sup&gt;</td>
<td>1) ژوهر (noon)</td>
</tr>
<tr>
<td>4) کیف</td>
<td>2) کودک (street)</td>
<td>Adjust of Time</td>
<td>2) ژارز (afternoon)</td>
</tr>
<tr>
<td>5) تژتوی (how)</td>
<td>1) ۲ژند (sadly)</td>
<td>Adjust of Manner</td>
<td>1) ژندر (swim)</td>
</tr>
<tr>
<td>2) ژند (sadly)</td>
<td>2) ژندر (carefully)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>2</sup> DWC refers (declarative wh-phrase counterpart) to the categories which replace the wh-phrase in declaratives.

<sup>3</sup> DO refers to direct object. The object marker “ra” occurs after direct object in declaratives and wh-in-situ questions.

<sup>4</sup> The preposition “tu” which means “at” precedes the adjunct of place in declaratives but not in wh-in-situ questions.

References


