JOURNAL OF SOCIAL SCIENCES & HUMANITIES OF SHIRAZ UNIVERSITY
VOL. 23, NO. 1, Spring 2006 (SER. 46)
(Special Issue in English Language and Linguistics)

Speech Segmentation in L2: Stress VS. Position

Dr. Z. Fotovatnia*
Islamic Azad University, Najafabad Branch

ABSTRACT

This study aimed to investigate the process of speech segmentation of Farsi speaking EFL learners by manipulating position (initial/medial) and stress (stressed/unstressed). One hundred BA students of English performed a phoneme localization task in which they were asked to locate some target phonemes. Both the participants' reaction times (RTs) as well as their accuracy in localizing target phonemes were measured and analyzed. The results showed that speech segmentation was mainly affected by position, while stress did not produce any significant effect.

Keywords: 1. Speech segmentation 2. Speech segmentation strategies 3. Listening comprehension 4. Reaction time 5. Accuracy.

1. Introduction

Speech segmentation has been the focus of many research studies, as it is a prerequisite to listening comprehension (Mattys, 2000). The main purpose of these studies has been finding out the units used by native speakers of a language in the process of speech segmentation. These units can be compared with the ones used by learners of that language to detect the problems they have in breaking down the speech into meaningful units. Most scholars believe that the reason for this problem is the students' first language (Best, 1995; and Flege, 1995). Best (1995) and Flege (1995) attributed the failure of second language learners in segmentation to L1-based strategies that are used to segment L2 speech. They believed that phonemes, clusters, or other phonological units are perceived incorrectly in L2 because second language learners map them onto L1 categories and distinctions. Therefore, such learners fail to make appropriate distinctions in L2.

Regarding the English language, some models of speech processing postulate certain intermediate representations such as segments (Barry, 1984), and syllables (Norris, 1994) between the acoustic signal and the stored lexical representation in the segmentation process. That is, listeners utilize these units to segment speech into meaningful constituents. In English, however, suprasegmental features like stress are believed to play an important role because stress possesses physical saliency (Lehiste, 1970), phonemic stability (Altmann and Carter, 1989), and perceptual distinctiveness (Bond and Garnes, 1980). Most experiments show that stress interacts with the position of the syllable in English. In other words, it has been shown that stressed syllables are more important in speech segmentation because they are considered as word initials (Field 2003). On the contrary, unstressed syllables are not considered as word initials, or if they are, they are considered to initiate function words which are not as important as content words in the process of speech perception and comprehension (Cutler and

* Assistant Professor of TEFL

Butterfield, 1992; Gow and Gordon, 1993; Mattys and Samuel, 2000).

Word and sentence stress, as Roach (1995) maintained, are realized through loudness, pitch, quantity and quality of the vowels involved, and are different across languages. In English phonology, for example, stress depends strongly on intensity while in Farsi it is the pitch that gives prominence to the stressed syllable (Sepanta, 1976; Yarmohammadi, 1996). Furthermore, Farsi and English are languages with different systems of rhythm and stress; English being a stress-timed while Farsi being a syllable-timed language (Zawadzki, n.d). The difference lies in that it takes different amounts of time to pronounce either a stressed or an unstressed syllable in English, while there is a given amount time (on average) between two consecutive stressed syllables, and that time is roughly a constant. On the other hand, in Farsi every syllable takes up roughly the same amount of time when pronounced and there is no difference between a stressed and unstressed syllable in this respect.

Despite the difference mentioned, the position of stress seems not to be very distinctive in the process of speech segmentation in Farsi probably because it is more predictable than in English. Many scholars believe that stress falls on the final syllable of nouns, adjectives, and simple verbs in Farsi (Ferguson, 1957; Same'I, 1996; Yarmohammadi, 1996). This generalization fails when derived verbs are considered because certain affixes attract stress (for example, *miravam*, I go). However, as Kahnemuyipour (2001) showed, the position of stress can be explained without considering the category of the word concerned. This is possible by making a distinction between word-level, phrasal-level, and intonational-level stress. He argued that stress can be assigned on more than one level. At the word-level, the stress rule is "End Rule Right" while at the phrase level it is "End Rule Left". The position of stress moves to the utmost right when phrases in combination are involved i.e., at the intonatinal level. That is, at the word level, the final syllable of the word attracts the stress while at the phonological level, the phonological unit attracts the stress. Prefixed verbs fail to follow the word-final rule in Farsi, as they are the combination of a prefix which is a phonological word and the verb itself which is another phonological word. Based on the phrase-level stress rule, stress is put on the prefix because it is the initial unit in the phrase.

Previous studies show the importance of stress and position in the process of speech segmentation for native speakers of English. Investigating the influence of the same factors on the speech segmentation of English learners may allow some insights into the strategies that Farsi learners of English use when listening to English input.

Given the view that L1 strategies are used to map L2 speech, it seems likely that Farsi learners do not take the stressed syllable as word initial because the majority of Farsi words are not initially stressed at either the word level or the intonational level. An experiment by Shiri (cited in Yarmohammadi 1996) seems to support this view, as Farsi learners who were asked to repeat English words and expressions significantly put the stress on final syllables. In the present study, however, there was an attempt to use a phoneme localization task which was used to tackle this issue for native speakers of English. The speech localization task which measures the participants' accuracy and speed in localizing target phonemes is useful for two reasons (Sanders and Neville, 2000). First, speech segmentation is accomplished by the recognition of word onsets. Therefore, the use of phoneme localization which shows the way subjects detect target phonemes and the way they perceive them as word initial or word medial can be revealing in this respect. Second, this task does not intrude natural speech so that it can provide reliable information on online segmentation processes, especially when sentences are constructed to include target words.

The purpose of the present study was to determine whether Farsi speaking learners of English used the same strategies that native speakers of English use in the process of speech segmentation. In other words, the focus of the study was to determine if Farsi speaking learners of English depended on stress and position in the process of speech segmentation.

2. Method

2.1. Participants

The participants included one hundred freshman students studying for a B.A. degree in English at the University of Najafabad in the academic year of 2001-2002 selected through a general proficiency test. The proficiency test was the OPT (Oxford Placement Test) and the participants who were selected were those between one SD above and one SD below the mean.

2.2. Materials and instruments

Sixteen single phonemes and eight phoneme combinations were selected to be used in this experiment. Vowels were discarded from the experiment because of the following reasons. First, it was necessary to find words which included the phonemes in the four different stress positions: stressed-initial, stressed-medial, unstressed-initial, and unstressed-medial, but vowels tended to be pronounced differently in each of these positions (Matteys, 2000). Second, there is usually much more agreement on the onset of consonants than vowels. That is, it is easier to measure phonemic onset for consonants than vowels (Sanders and Neville 2000).

The target words were mainly taken from Sanders and Neville (2000), but a few more were added by the researcher in an effort to include the phonemes these researchers disregarded. The words were divided into the following five categories: (a) initially stressed words with the phoneme or phoneme combination at the beginning, (b) medially stressed words with the phoneme or phoneme combination at the beginning, (c) initially stressed words with the phoneme or phoneme combination in the middle, (d) medially stressed words with the phoneme or phoneme combination in the middle, and (e) words without the target phoneme. As mentioned, the two initial and medial positions were allocated to the target phonemes because phonemes in initial positions were expected to be localized faster and more accurately than phonemes in the medial position (Pitt and Samuel 1990; Sanders and Neville 2000). Furthermore, considering the participants' level of proficiency (pre intermediate) and their previous experience with the computer, adding to the number of phoneme positions could have made the task of responding very difficult.

The words varied in the number of letters they were made of and in their parts of speech. However, no significant differences resulted by assigning words to the four aforementioned categories (a, b ,c ,d) as far as the number of letters was concerned, $F(3, 151) = 1.93, P=.12$. Also, no significant differences resulted by assigning words to the four aforementioned categories as far as their part of speech was concerned. That is, none of the following factors were found to be significantly different in each group; the number of nouns, $x2$ (3, N=85) = 0.67, $p < .05$, the number of verbs $x2$ (3, N= 38) =1.67 $p < .05$), and the number of adjectives, $x2$ (3, N=26) =2.36 $p < .05$.

Sentences were constructed around target words in a way that each target phoneme or phoneme cluster occurred in the selected position in the target word and nowhere else in the sentence. For each phoneme or phoneme cluster, one filler was also added to the list such that the target phoneme could not be found anywhere in the filler, as the participants were asked to first identify the existence of the target phoneme and then localize them. The fillers were discarded from data analysis In addition, the sentences were given to a sample of the same population who did not take part in the experiment to see whether they could predict the target words before hearing them. It was reasoned that if the participants could predict the target words based on their previous context, they would not respond to the target phoneme based on its phonological properties. The sentences were presented in a cloze test for the sample to complete. It was decided that if the sentences were filled in with the correct words by only 25% of the students, the sentences would be excluded and replaced by other sentences. However, no such cases were observed. Moreover, to provide context for the target words, the sentences were constructed such that the target words did not occur within three words to the beginning or to the end of the sentences. An ANOVA test was run to compare the length of the sentences in the four conditions. No significant differences were observed, $F(3, 151) = 0.81$, $P=.97$. Furthermore, to rule out the possibility of any differences across conditions with respect to the position of the target words in the constructed sentences, an ANOVA test was run. No significant differences were observed, $F(3,151) = 0.17$, $P=0.91$. As the first sentence shows, the target word *daughter,* which contains the target phone *d* in the initial position (SI), occurs as the seventh word to the beginning and the fifth word to the end of the sentence.

A native English speaker read the sentences and recording was made. Then, using the *Goldwave Software*, the phoneme onset time from the beginning of the sentence, and the time it took for each sentence to complete were computed to provide a basis for the measurement of the reaction time and accuracy in the following phases of data analysis. The reason that motivated the measurement of both the reaction time and accuracy in this study was that according to Birdsong (1989), providing converging evidence resulting from two measures could be more useful in presenting a more complete picture of the phenomenon. This is because each measure assesses a different ability. Accuracy is concerned with the learner's knowledge representations of L2 (Bley-Voman, 1990) while reaction time manifests online processing tasks (White and Genesee, 1996).

To meet the particular requirement of the study, a special software was designed to be used on a Pentium III in Macromedia Authorware version 4.00 environment.

2.3. Procedure

Each participant was told to put on a headset and sit at the computer close to an assistant who gave instructions on how to use the computer and how to proceed through the program. The participants were informed that both their accuracy and their speed in the localization task were measured. To make the participants more familiar with the task and to give them more opportunity to do the task well, they were asked to go through a practice trial for as many times as they thought they needed to make themselves ready for the main task. The practice trial was especially designed to be similar to the main task; however, it included different sentences. The participants received feedback on how well they performed in the practice trial. The program started as a phoneme was displayed on the screen of the computer while the same phoneme was played through the headset. The participant was told to look at the phoneme on the screen, listen to it on the headset, and repeat the phoneme. After that, s/he was required to listen carefully to a sentence which was played through the headset, and to locate the target phoneme. The participant was then asked to react to the presence of the target phoneme in the sentence by pressing one of the buttons on the mouse as soon as s/he heard the target phoneme i.e., to left-click if the target phoneme was located initially, and right-click if the target phoneme was located medially. Afterwards, the program initiated a new cycle of displaying a target phoneme and playing the related sentence by pressing one of the buttons on the keyboard. As a result, the participant could decide on the time s/he was ready to proceed through the rest of the program.

3. Data analysis

Two factors, the position of the target syllable (initial, medial) and stress (stressed, unstressed) were considered in relation to both reaction times and accuracy. The files were made ready to be processed by the SPSS Program (version 10).

3.1. Reaction Times

Reaction times (RTs) were measured from the onset of the target phonemes. It was decided to use a cutoff for RTs to prevent aberrant latencies from having too much influence on the means. However, the selection of the cutoff is arbitrary. For example Mattys (2000) used a 100-1500ms cut-off, but as the participants in this study were not native speakers of English, it was preferred to use slower RTs. To this aim only RTs between 200-2000 milliseconds were used in data analysis. As a result, some of the data were discarded from the analysis. The result was then classified into two categories: one showed the time latency of the responses and the other their accuracy. Mean RTs were calculated for each of the four conditions. Results of the analysis are shown in Table 1.

Table 1: Mean RTs (in milliseconds) on phoneme localization as a function of stress and position of the target-carrying syllables.

		Position	
Stress		**Initial**	**Medial**
Stressed	1	008.28	1153.77
Unstressed		1006.85	1181.14

Table 1 shows that the participants were faster in localizing the phonemes that occurred initially and that their decision was not affected by stress.

A Repeated Measures ANOVA was performed on the RTs, examining stress and position and a main effect of position, $F(1, 64) = 37.13, p = 0.0001$ was found. The results showed that stress could not influence speech segmentation. The mean RTs on position are illustrated in Figure 1.

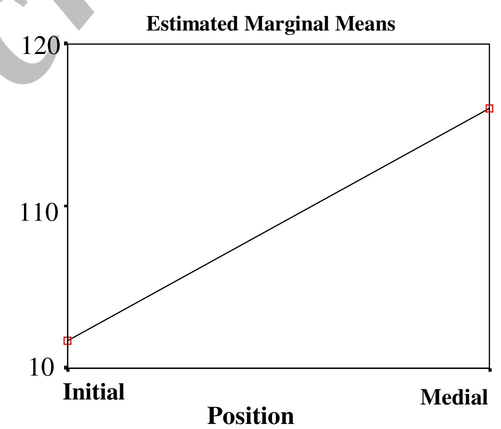

Figure 1: Performance as a function of position.

As Figure 1 shows, phonemes were located faster when they occurred initially than medially. Stress seemed to have no effect on the performance.

3.2. Accuracy of Responses

The next part of the analysis focused on the accuracy of responses. To this end, it was decided to study these issues at different time bands, as it was observed that the performance was different when the reaction times increased. Since it was reasoned that the participants might have used different strategies in making automatic and delayed responses to speech signals (Cutler, cited in Field 2003), the data were categorized into three time bands (referred to as SAT): (a) from 200 to 592 msec, (b) from 593 to 899 msec, and (c) from 900 to 1399 msec,

A Repeated Measures ANOVA was performed on the means of responses in different conditions and the following results were obtained. First, there were significant main effects of position, $F1\ (1, 64) = 49.14,\ p < .000$; and stress, $F2\ (1, 64) = 5.82,\ p < .02$. Table 2 Shows mean performances on both the initial and medial positions.

Table 2. Means of performance on accuracy as a function of position.

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Initial	45.8387	31	7.33074	1.31664
	Medial	30.8387	31	8.69903	1.56239

Figure 2 illustrates the means of performance in both the initial and medial positions.

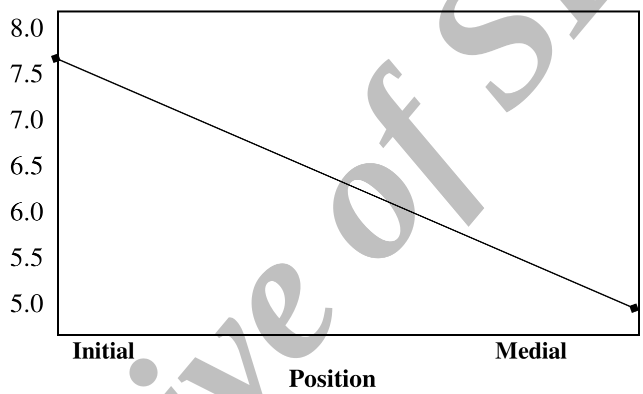

Figure 2: Accuracy performance in terms of position.

Careful observation of figure 2 shows that the participants were more accurate to locate target phonemes when the phonemes occurred initially.

Table 3 shows the means of performance on stressed and unstressed syllables. Comparison of the means shows that the participants were more accurate to locate target phonemes when they were included in unstressed syllables.

Table 3. Means of performance on accuracy as a function of stress.

stress	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Stressed	6.204	.171	5.855	6.554
Unstressed	6.575	.187	6.194	6.957

Figure 3 is used to illustrate the means in both conditions.

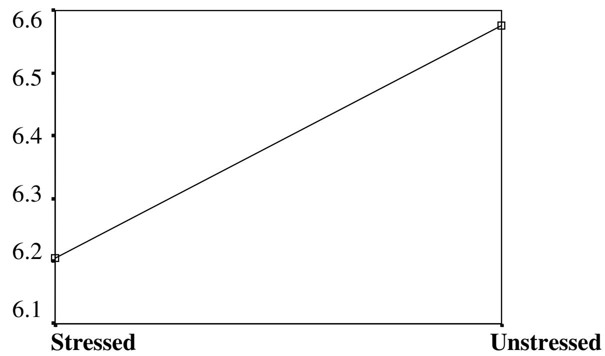

Figure 3: Accuracy performance in terms of stress.

Figure 3 shows that the most accurate responses were given to the target phonemes that were carried by the unstressed syllables. What is inferred from Table 2 and Table 3 is that the degree of stress affects performance, but not as much as position.

Second, there were significant interactions between position and stress, *F3 (1, 64) = 45.22, p < .000* as well as those between SAT and position, *F4 (2, 64) = 14.28, p < .000*.

Table 4 shows the means of performance in the four conditions of IS (initially stressed), IU (initially unstressed), MS (medially stressed), and MU (medially unstressed). These conditions can be arranged from the most accurate to the least accurate as follows IS, IU, MU, and MS.

Table 4. Means of performance on accuracy as a function of stress and position.

Position	Stress	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Initial	stressed	8.140	.217	7.697	8.583
	Unstressed	7.140	.275	6.578	7.702
Medial	stressed	4.269	.253	3.753	4.785
	Unstressed	6.011	.332	5.332	6.690

Careful analysis of the means revealed that stressed syllables were more facilitative than unstressed syllables in initial position. An opposite alternative was the case regarding unstressed syllables. That is, performance on unstressed syllables was better than performance on stressed syllables in medial position. Figure 4 shows the result of the interaction between position and stress.

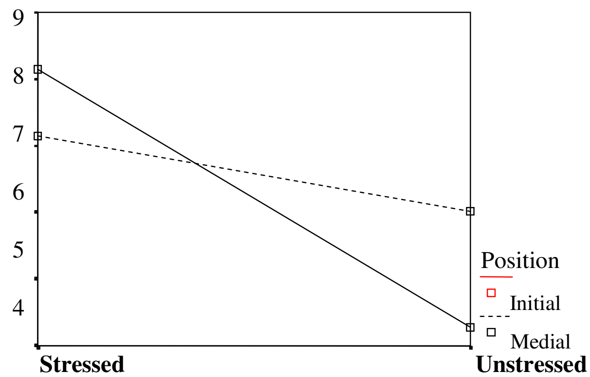

Figure 4: Interaction of stress and position based on accuracy of responses.

The next significant effect was the interaction between the reaction time bands (SAT) and position. Table 6 shows the means of performance as a function of SAT and position.

Table 6: Means of performance on accuracy as a function of reaction time bands and position.

SAT	Position	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1	Initial	9.194	1.014	7.123	11.265
	Medial	3.274	.293	2.676	3.873
2	Initial	7.984	.563	6.834	9.134
	Medial	6.532	.562	5.385	7.679
3	Initial	5.742	.493	4.735	6.749
	Medial	5.613	.410	4.776	6.450

***Note.* 1: The first time band, 2: The second time band, 3: The third time band**

Further analysis of the results showed that the effect of position diminished as reaction times increased, *t (88) = 5.73, p = .000* in the first time band; *t (77) = 1.24, p = .27* in the second time band; and *t(30) = 0.30, p= .76* in the third time band. The effect of position was found above the chance level only in the first time band. Figure 5 shows these results.

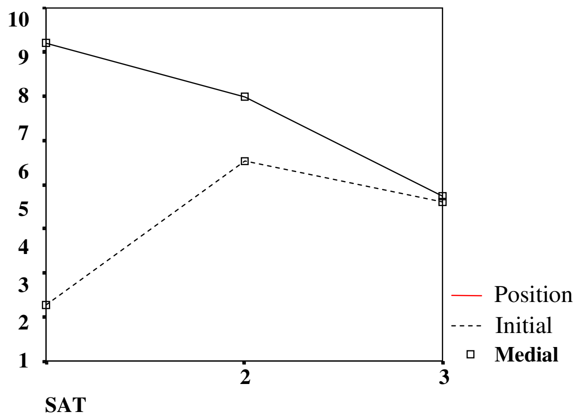

Figure 5: Interaction of reaction time bands (1: the first time band, 2: the second time band, 3: the third time band) and position based on accuracy of responses.

The last interaction which reached a near significant level was the one between position, stress, and SAT, *F (2, 64) = 3.17, P < .05*. Further analysis of the variables involved revealed that performance was different in each band. Figure 6, figure 7, and figure 8 show the accuracy of performance in each time band.

As Figure 6 shows, position interacts with stress in all bands. However, increasing reaction times brought about subtle changes in the pattern of stress-position interaction. Paired sample *t*-tests were computed to show whether the differences were significant.

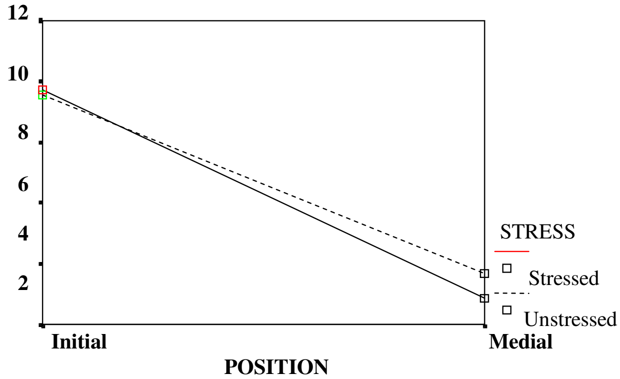

Figure 6: Interaction of stress, position, and sat in the first band on accuracy of responses.

As figure 6 shows, in the first band no significant effect of stress on initially located phonemes was observed; the accurate responses were the same for the unstressed as well as the stressed syllables in initial position. In other words, position superseded stress in the fastest time band. Figure 7 shows the interaction of stress and position in the second time band.

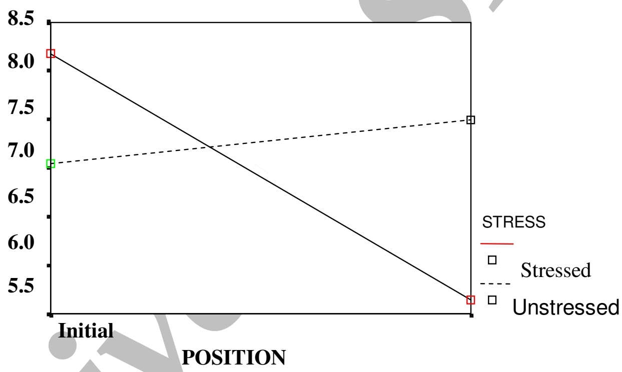

Figure 7: Interaction of stress, position, and sat in the second time band on accuracy of responses.

In the second band, the interaction of position and stress became significant in all conditions, IS/IW $t\ (87) = 3.71\ p = .0001$; MS/MW $t(77)= 4.8,\ p = .0001$; IS/MS $t(79)= 4.47,\ p = .0001$; an IW/MW $t(30)= 2.58,\ p < 0.01$. The conditions can be arranged from the most accurate to the least accurate in the following order: IS, MW, IW, and MS. The participants were more accurate in locating stressed target phonemes when they occurred initially. On the contrary, unstressed phonemes were located more accurately when they occurred medially. Furthermore, performance on the unstressed medial phonemes improved more than performance on the same phonemes in the first band.

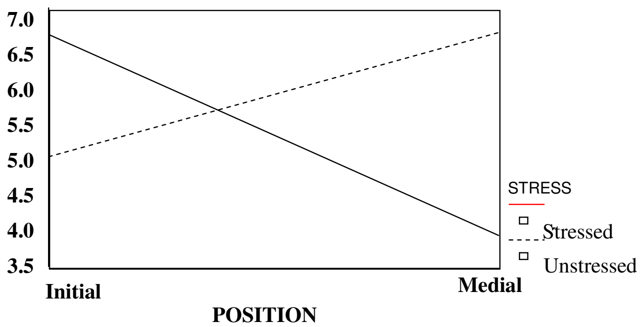

Figure 8: Interaction of stress, position, and sat in the third time band on accuracy of responses.

The improvement on the unstressed medial phonemes continued throughout the third time band: MS/MW *t (30) = 4.48, p = .0001*; IW/MW *t (30) = 3.04, p = .005.* That is, the participants tended to be more accurate in localizing unstressed syllables in the medial position.

Table 7 shows a summary of the statistical figures obtained in different phases of data analysis.

Table 7: ANOVA summary table.

	df	sum of squares	means of squares	F	P
Reaction Times					
Position	1	1356442.27	1356442.27	37.134	.0001
Accuracy of Responses					
Position	1	581.25	581.25	49.143	.000
Stress	1	12.798	12.798	5.822	.02
Position*Stress	1	174.798	174.798	45.225	.000
SAT*Position	1	570.790	285.395	14.283	.000
SAT	2	159.425	79.712	2.206	.119
SAT*Stress	1	0.694	0.347	0.076	.927
SAT*Position*Stress	2	30.597	15.298	3.178	.051

In summary, analysis of the data obtained from the phoneme localization task leads to the following findings. First, a main effect of position on reaction times was found – initial phonemes were located faster than medial phonemes and stress did not interact with position. Second, regarding the accuracy of responses, different patterns were found in the three time bands. The main effect of position was observed exclusively in the first i.e., the fastest reaction-time band, but not in the other two time bands. Stress did not interact with position in the first time band. It means that initial stressed syllables were located as accurately as the unstressed ones. Unstressed syllables, however, were located more accurately than stressed syllables in the medial position. The effect of position was moderated in the second time band such that performance on the medial unstressed syllables improved more than the performance on initial unstressed syllables. The same trend of improvement was observed in the third time band. In other words, medial unstressed syllables outperformed the syllables in the other conditions. Performance on the medial stressed syllables was the worst of all other conditions in the three time bands.

4. Discussion

The main objective of this study was to examine the units that are used by EFL learners in the segmentation of English input. To this aim, the speed and accuracy of responses in localizing certain phonemes were analyzed. Understanding differences between the units that are used by English speakers and the units that are used by English learners is important, and the findings might be used to trace the problems learners have in the segmentation of speech and listening comprehension.

Understanding how the flow of speech is broken down into meaningful units in English has been a challenging enterprise for a long time. The literature on the issue shows that the unit of processing and segmentation is the primary-stressed syllable considered as the word onset (Mattys and Samuel, 2000). This strategy is efficient for

English listeners, as it can account for the processing of the majority of English words. Nevertheless, the processing of the words beginning with a unstressed or secondary syllable needs additional right-to-left processing to incorporate contextual information into the course of action to undo the misleading segmentation procedure based on the medial-stressed syllable.

The results obtained show that contrary to the reliance of English speakers on primary stress, Farsi speakers rely on position. That is, Farsi speakers use initial phonemes to break down the flow of speech into meaningful units in English. These differences might explain why Farsi learners of English feel unable to break down the flow of speech in English. The failure of second language learners in speech segmentation is believed to result from the units these learners use for second language processing, which are the same ones used in the processing of speech in their first language. Accepting the view that L1 units and strategies are used to segment L2 speech, one might suggest that Farsi learners of English fail to segment speech appropriately because they do not use the stressed syllable as a clue to word boundary detection in the same way English speakers do. As the results show, the test of speech segmentation failed to show any effect for stress on speech segmentation. That is, neither speed nor the accuracy of responses was affected by stress. Initial phonemes were located faster and more accurately with stress having no role in the process of such location. From this perspective, it might be argued that Farsi learners of English do not rely on the stressed syllable as the main word-boundary clue. This effect is observed in the first time band which is more reliable in tapping segmentation mechanisms.

What is striking about the performance on medial phonemes is their interaction with stress. The better performance of participants on unstressed syllables in medial position is similar to that of English speakers (Sanders and Nevill, 2000). Native speakers of English take the unstressed syllable as word-medial if the word is a content word such as "effective" or word initial if the word is a function word such as "effect of" (Cutler and Butterfield, 1992). The significant interaction of stress with position in medial phonemes in the second and third time bands is in accordance with the idea that the second and third time bands are probably more apt to be influenced by metacognitive knowledge such as word familiarity in English and metalinguistic strategies. This is also in line with Cutler (1990), who maintained that responses to automatic speech signals are made based on strategies one uses in the L1, contrary to the ones used to handle slightly delayed decisions based on the strategies that native listeners utilize.

The findings add a further dimension to the claim made by models of speech processing which hold that there is competition between phonetic and phonological routes and that the default status is prelexical, but as the reaction time increases, the prelexical code is ignored and lexical knowledge overrides the first code (Eimas et al., 1990). The participants used their prelexical knowledge for the segmentation of speech in English, which was based on their first language. Nevertheless, as their reaction times increased, the postlexical information seemed to take over and supersede their first approach.

The findings imply some pedagogical objectives. First, as the results indicate, the processing of second language speech is different from that of the first language in terms of the categories and strategies involved. Therefore, it is important to consider the differences between these processes in L1 and L2. New approaches to teaching consider the practicality of a method of instruction which focuses on metacognition awareness (Halpern, 1996). Raising the listeners' awareness could make them sensitive to the similarities and differences each language displays regarding its segmental and suprasegmental features and could make them focus on the development of the strategies second language learners use. This issue shows the importance of courses in English and Farsi phonetics and phonology. In other words, contrastive courses on English and Farsi phonetics and phonology help learners to focus on the similarities and differences between both languages as far as segmental and suprasegmental features are concerned.

References

Altmann, G., and Carter, D. (1989). *Lexical stress and lexical discriminability: Stressed syllables are more informative, but why?* **Computer Speech and Language,** 3, 265-275.

Barry, W. J. (1984). *Segment or syllable? A reaction-time investigation of phonetic processing,* **Language and Speech,** 27, 1-15.

Best, C. (1995). *A direct realist view of cross-language speech perception research with adults*. In W. Strange (Ed.), **Speech perception and linguistic experience**, Timonium, MD: York Press, 171-204.

Birdsong, D. (1989). **Metalinguistic Performance and Interlinguistic Competence**, Berlin: Springer-Verlag.

Bond, Z. S., and Garnes, S. (1980). **Misperception of fluent speech. In R. Cole (Ed.), Perception and Production of Fluent Speech**, Hillsdale, NJ: Erlbaum, 115-132.

Cutler, A., and Butterfield, S. (1992). *Rhythmic cues to speech segmentation: Evidence from juncture misperception,* **Journal of Memory and Language**, 31, 218-236.

Eimas, P. D., Marcovitz Hornstein, S. B., and Payton, P. (1990). *Attention and the role of dual codes in phoneme monitoring*. **Journal of Memory and Language**, 29, 160-180.

Ferguson, C. A. (1957). *Word stress in Persian*. **Language**, 33,3, 123-136.

Field, J. (2003). *Lexical segmentation in L2 listening*. **ELT Journal**, 57, 4, 325-334.

Flege, J. E. (1995). *Second language speech learning: Theory, findings and problems*. In W. Strange (Ed.), **Speech Perception and Linguistic Experience,** Timonium, MD: York Press, 233-277.

Gow, D. W., and Gordon, P. C. (1993). *Coming to terms with stress: Effects of stress location in sentence processing*. **Journal of Psycholinguistic Research,** 22, 45-578.

Halpern, D. F. (1996). **Thought and Knowledge: An Introduction to Critical Thinking**, Mahwah, New Jersey: Lawrence Erlbaum Associates Publishers.

Kahnemuyipour, A. (2001). **A Unified Account of Persian Stress. Linguistic Workshop**, Islamic Azad University of Khorasgan, Iran, Esfahan.

Lehiste, I. (1970). *The timing of utterances and linguistic boundaries*. **Journal of the Acoustical Society of America,** 51, 2018-2024.

Mattys, S. L. (2000). *The perception of primary and secondary stress in English*. **Perception and Psycholinguistics**, 62, 2, 253-265.

Mattys, S., and Samuel, A.G. (2000). *Implications of stress-pattern differences in spoken-word recognition*. **Journal of Memory and Language,** 42, 571-596.

Norris, D. G. (1994). *Shortlist: A connection model of continuous speech recognition*. **Cognition,** 52,189-234.

Pitt, M. A., and Samuel, A. G. (1990). *The use of rhythm in attending to speech*. **Journal of Experimental Psychology: Human Perception and Performance**, 16, 564-573.

Roach, P. (1995). **English Phonetics and Phonology: A Practical Course,** Cambridge University Press.

Same'I, H. (1996). *Verb stress in Persian: A reexamination*. **Name-ye Farhangestan,**

1, 4, 6-21. (in Persian).
Sanders, L. D., and Neville, H. J. (2000). *Lexical, syntactic, and stress-pattern cues for speech segmentation*. **Journal of Speech, Language, and Hearing Research**, 43, 1301-1321.
Sepanta, S. (1976). **New Studies on Seventh Congress of Iranologists,** Tehran: Melli University Publications, 2.
White, L. and Genesee, F. (1996). *How native is near-native? The issue of ultimate attainment in adult second language acquisition*, **Second Language Research**, 12, 233-365.
Yarmohammadi, L. (1996). **A Contrastive Phonological Analysis of English and Persian**, Shiraz University Press.
Zawadzki, H. (2002). *In tempo*, Retrieved July, http://www.Nceltr.mq.edu.au/ publications/books/intempo.htm.