

Statistical Evaluation of the Influence of Stress on Pitch Frequency and Phoneme Durations in Farsi Language

D. Gharavian^{1,2} & S.M. Ahadi¹

¹Electrical Engineering Department, Amirkabir University of Technology, Tehran, Iran

²Shahid Abbaspour University, Tehran, Iran
gharavian@pwit.ac.ir, sma@cic.aut.ac.ir

Abstract

Stress is known to be an important prosodic feature of speech. The recognition of stressed speech has always been an important issue for speech researchers. On the other hand, providing a large corpus with the coverage of all different stressed conditions in a certain language is a difficult task. Farsi (Persian) has been no exception to this. In this research, our aim has been to evaluate the effect of stress on prosodic features of Farsi language, such as phoneme duration, pitch frequency and the pitch contour slope. These might be valuable in further research in speech recognition. As the main influence of stress is on vowels, the effect of stress on such parameters as duration and pitch frequency and its slope on the phoneme level and for vowels has been evaluated.

1. Introduction

Speech recognition is finding an important role in providing human-machine interaction. However, once the speech gets informal, the problems with speech recognition increase. One of the important prosodic parameters of speech is stress, which may cause changes in the duration and pitch frequency and its slope. Stress plays an important role in helping human beings to understand speech [1,2], hence, ignoring it in speech recognition will be irrational and may make recognition more difficult [3]. Examples of the application of prosodic features to HMM-based speech recognition systems have been their use as the extensions to the speech feature vectors [4], in a post-processing phase [5], integrated in the speech recognition system, e.g. to remove some search paths in the Viterbi search [6] and in the hybrid systems and multi-stage recognition [7]. In tonal languages, the most important prosodic feature is the pitch frequency, while in stress accent languages, other parameters such as duration and energy are also important and stress can cause increase in both [8]. In Farsi, the parameter that gets the largest impact from stress is the pitch frequency, where the most important effect is the change in the slope of the pitch. Furthermore, in such languages, stress can also cause increase in phoneme duration and energy [9]. Observations have also shown that the effect of stress has been mostly on the vowels, in comparison to the other phonemes [10].

Our previous work was concerned with the duration of different phonemes, and especially vowels, in normal Farsi speech. The results have shown the high dependency of vowel durations to the syllable types [11]. As there has not been any statistical evaluation of the effect of stress in Farsi language before, in this paper, the amount of influence of stress on

phoneme duration, pitch frequency and the pitch contour slope for each of the vowels in different syllables will be evaluated. We will try to extract some rules for their changes from the results and will show that the amount of influence of the stress depends on the syllable type.

2. Corpus

The only available corpus of Farsi continuous speech is *Farsdat* [12]. This corpus consists of 6000 sentences from 300 speakers (both male and female), who have randomly uttered some of the 392 pre-defined available sentences. The sentences are normally uttered and do not contain any stressed parts. We have used about 1800 of the above utterances (from a prominent dialect) to build an HMM-based speech recognition system.

A male speaker has been used in this work to create a stressed corpus. This speaker has uttered all the above-mentioned 392 sentences three times in a normal manner. These utterances have been used to adapt the HMM system to the new speaker. We call this set of models M1. Furthermore, this speaker has also uttered 154 of the above sentences in an overall 468 different stress conditions. Using M1 models, the stressed M2 models have been created.

Time alignments for both normal and stressed sentences with their transcriptions were required to give the vowel boundaries in these sentences. These were used to extract the vowel duration and pitch contours. The two sets of models, M1 and M2, were used in finding the vowel boundaries in normal and stressed speech. The *hidden Markov model toolkit (HTK)* [13] has been used for all training, adaptation and time-alignment experiments.

3. Vowels and Syllables

In Farsi, there exist about 30 different phonemes, of which six are vowels. These are shown in Table 1. We call /æ/, /ɛ/ and /o/ weak vowels and /a/, /i/ and /u/ tense vowels. There is another phoneme, namely /ow/, which is a diphthong and has specifications similar to a vowel. The changes in the parameters of /ow/ according to stress will also be presented here together with the six mentioned vowels.

It is a widely accepted fact that there exist only three types of syllables in Farsi language, i.e. CV, CVC and CVCC, where, V represents a vowel and can be any of the six above-mentioned vowels plus /ow/ C can be any consonant (the remaining 23 phonemes).

Table 1: Vowels of Farsi

Vowel	Example	Meaning
/æ/	sæbr	Patience
/ɑ/	xɑb	sleep
/ɛ/	tʃɛrɑɟ	lamp
/i/	diruz	yesterday
/o/	gozæft	mercy
/u/	ruz	day

4. Effect of Stress on Prosodic Parameters

4.1. Pitch frequency

As mentioned earlier, the pitch frequency is the most important prosodic parameter in Farsi. In this work, pitch extraction has been carried out using the technique devised by Medan *et al.* [14] utilizing the *Edinburgh speech tools library* [15]. Using the time alignment results, the vowel area of the speech signal was found and the average value of pitch frequency was calculated, according to the pitch frequency contour. In this section, the changes in this average value due to stress are evaluated.

Table 2 is based on the changes in the average value of the pitch frequency without taking into account the syllable type. In each part of the table, the parameters of distributions based on the samples seen in the experiments are reported, where “Std” represents standard deviation and “Std/Mean” is a measure of closeness of the pitch frequency values to the mean. The “S/U” column represents the ratio of these changes in the stressed case to that of the unstressed case. In Figure 1, the changes in the mean value of the pitch frequency in different syllables are displayed.

Table 2: Changes in the pitch frequency of Farsi vowels disregarding the syllable type.

Vowel	Unstressed			Stressed			S/U
	Mean	Std	Std/Mean	Mean	Std	Std/Mean	
/æ/	125.9	20.57	0.163	145.3	19.14	0.131	1.15
/ɑ/	131.4	20.62	0.157	146.6	17.82	0.122	1.16
/ɛ/	131.5	20.56	0.156	147.5	20.91	0.141	1.12
/i/	131.2	20.68	0.158	138.9	18.22	0.131	1.06
/o/	129.3	20.07	0.152	140.8	18.47	0.131	1.09
/u/	136.1	20.27	0.149	145.2	20.85	0.144	1.07
/ow/	129.1	20.03	0.155	150.6	17.98	0.119	1.17

The following conclusions can be made using the results presented in Table 2 and Figure 1:

- The pitch frequency for all vowels is increased due to stress.
- The Std/Mean ratio is reduced due to stress. This means that the pitch frequency in stressed case is less susceptible to changes.
- The increase in the pitch frequency is about 10 percent.
- In CV and CVC syllables the increases are very close, while in CVCC syllables, they are higher.

- For weak vowels, this increase is slightly more.
- The results for /ow/ display larger increases in different syllables, in comparison to vowels.

4.2. Pitch Contour Slope

The role of the slope of the pitch contour in Farsi language is more important, compared to the pitch frequency average [9]. In order to calculate this parameter, we approximated the pitch frequency with a line and used the slope of it as the prosodic parameter of interest. The results of this analysis obviated the need to omit some sparse data. The histogram analysis of the pitch frequency slope resulted in the fact that most of the values constitute a Gaussian histogram around zero (The mean value of the pitch frequency slope is negative and small) and only a few dispersed values remain out of the distribution which can be omitted. In order to omit the sparse data, a variance analysis was carried out. In this technique, the data, which are more than half the standard deviation away from the distribution mean, are discarded. This has been determined experimentally. In Table 3, the results of pitch frequency analysis, disregarding the syllable type and in Tables 4, 5 and 6, these results for the three syllable types are reported.

Table 3: Changes in the pitch frequency slope of Farsi vowels disregarding the syllable type

Vowel	Unstressed			Stressed			S/U
	Mean	Std	Std/Mean	Mean	Std	Std/Mean	
/æ/	-0.30	0.490	-1.62	-0.22	0.380	-1.74	0.72
/ɑ/	-0.11	0.175	-1.56	-0.02	0.017	-0.93	0.16
/ɛ/	-0.71	0.745	-1.06	-0.46	0.577	-1.25	0.65
/i/	-0.22	0.332	-1.48	-0.11	0.209	-1.91	0.49
/o/	-0.39	0.525	-1.35	-0.15	0.053	-0.35	0.38
/u/	-0.09	0.131	-1.50	-0.02	0.010	-0.52	0.21
/ow/	-0.02	0.007	-0.28	-0.08	0.001	-0.17	0.29

Table 4: Changes in the pitch contour slope of Farsi vowels in CV syllables after the variance analysis

Vowel	Unstressed			Stressed			S/U
	Mean	Std	Std/Mean	Mean	Std	Std/Mean	
/æ/	-0.50	0.581	-1.17	-0.24	0.213	-0.87	0.49
/ɑ/	-0.12	0.168	-1.39	-0.02	0.012	-0.54	0.18
/ɛ/	-0.78	0.741	-0.94	-0.54	0.604	-1.11	0.69
/i/	-0.19	0.266	-1.38	-0.12	0.257	-2.22	0.60
/o/	-0.64	0.615	-0.96	-0.29	0.069	-0.24	0.46
/u/	-0.15	0.207	-1.36	-0.02	0.008	-0.40	0.12
/ow/	-0.03	0.006	-0.18	-0.01	0.001	-0.10	0.26

The results reported in Tables 3 to 6, lead to the following conclusions:

- Stress causes the pitch contour slope to decrease.
- The pitch contour slope is less steep for weak vowels in comparison to tense vowels.
- The Std/Mean ratio decreases after applying the stress. This means that less slope diversities are seen.

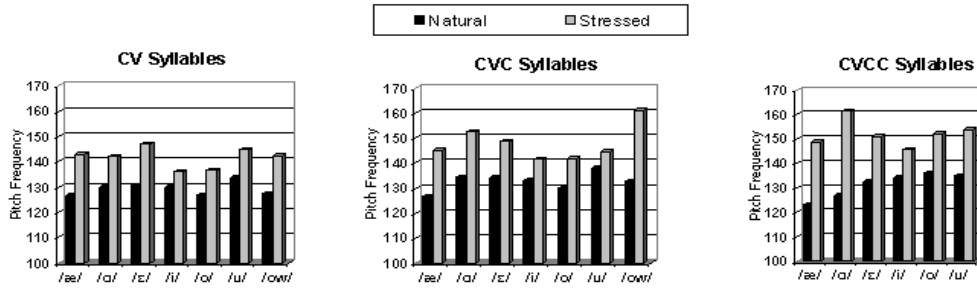


Figure 1: Changes of the pitch frequency in different syllable types, due to stress.

Table 5: Changes in the pitch contour slope of Farsi vowels in CVC syllables after the variance analysis

Vowel	Unstressed			Stressed			S/U
	Mean	Std	Std/Mean	Mean	Std	Std/Mean	
/æ/	-0.28	0.040	-1.44	-0.27	0.464	-1.71	0.97
/ɑ/	-0.09	0.151	-1.72	-0.02	0.019	-1.05	0.21
/ɛ/	-0.40	0.428	-1.06	-0.23	0.080	-0.35	0.57
/i/	-0.21	0.233	-1.12	-0.11	0.027	-0.24	0.53
/o/	-0.22	0.343	-1.53	-0.10	0.037	-0.36	0.46
/u/	-0.05	0.077	-1.44	-0.02	0.011	-0.56	0.38
/ow/	-0.01	0.003	-0.26	-0.01	0.001	-0.24	0.55

Table 6: Changes in the pitch contour slope of Farsi vowels in CVCC syllables after the variance analysis

Vowel	Unstressed			Stressed			S/U
	Mean	Std	Std/Mean	Mean	Std	Std/Mean	
/æ/	-0.05	0.083	-1.77	-0.02	0.011	-0.747	0.31
/ɑ/	-0.27	0.031	-1.15	---	---	---	---
/ɛ/	-0.04	0.045	-1.08	---	---	---	---
/i/	-0.04	0.054	-1.42	---	---	---	---
/o/	-0.03	0.012	-0.45	---	---	---	---
/u/	-0.02	0.005	-0.21	---	---	---	---

- Enlargement of the syllable increases the “S/U” ratio. This indicates smaller changes in the pitch slope. This is observed for CV and CVC syllables. However, as the number of tokens available for the case of CVCC is reduced due to the application of variance analysis, the results are not statistically valid anymore and not reported.
- As the syllable gets larger, the “S/U” ratios for weak and tense vowels get closer.
- /ow/ has the least steep slope, but demonstrates more slope change in comparison to tense vowels.

The above results show that the slope parameter is more disperse in comparison to the average pitch frequency. Furthermore, there exist a number of cases in the results that do not comply with the above conclusions. In a few cases, such differences are seen for /æ/, /ɛ/ and /i/ vowels. Further investigations on the pitch slope histograms, in two stressed and unstressed situations, have brought into light several issues. After the application of stress, the pitch slope histogram has become wider so that most of the data are not around zero, but they are in fact divided into two groups. This results in the omission of plenty of data due to the reduction of the variance analysis range. Further detailed investigations of the data in the distribution results in further conclusions. As the most important one, we can conclude that the presence of a plosive consonant, such as /b/, /d/ and /p/, at the start of the syllable results, in most cases, in an extraordinary decrease in the pitch slope.

4.3. Duration

This section is dedicated to the evaluation of the effect of the stress on the vowel durations. The results reported in Table 7 and Figure 2 display the noticeably large and orderly influence of stress on duration. Table 7 includes the results for different vowels, disregarding the syllable type, while in Figure 2, the results are given for different syllable types.

Table 7: Changes of the vowel durations due to stress

Vowel	Unstressed			Stressed			S/U
	Mean	Std	Std/Mean	Mean	Std	Std/Mean	
/æ/	95.4	45.4	0.48	117.8	59.3	0.50	1.24
/ɑ/	110.5	46.0	0.42	172.3	57.8	0.33	1.56
/ɛ/	59.9	29.8	0.50	73.6	33.9	0.46	1.23
/i/	97.9	46.8	0.48	140.2	62.6	0.45	1.43
/o/	86.3	42.3	0.49	98.5	43.0	0.44	1.14
/u/	124.3	52.4	0.42	189.2	48.3	0.26	1.52
/ow/	165.1	50.4	0.31	242.0	49.5	0.20	1.47

From the results presented in Table 7:

- Stress increases the vowel duration.

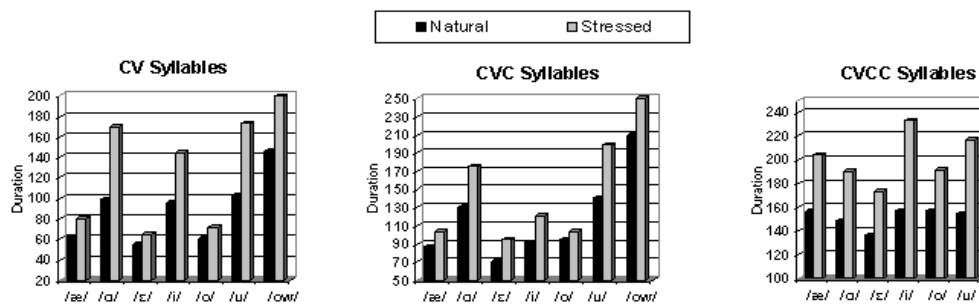


Figure 2: Changes of the vowel durations in different syllable types, due to stress.

- Effects are stronger on the durations of tense vowels.
- The change for weak vowels is about 20 percent and for tense vowels about 50 percent.
- The Std/Mean ratio reduces after the application of stress. This ratio indicates the sparsity of different samples of each vowel. In fact, the increase in the stress results in less diversity in vowel durations.
- The larger the syllable, the closer get the “S/U” ratios in the tense and weak vowels.
- The increase in duration for CV syllables is about 20 percent for weak vowels and about 60 percent for tense vowel, for CVC syllables is about 10 percent for weak vowels and about 30 percent for tense vowels and for CVCC syllables is about 30 percent for both weak and tense vowels.
- The change in /o:/ duration is less than that of tense vowels, but more than that of weak vowels.

Previous investigations indicate that even for a single vowel, when considered in multi-syllable words, the average duration of the vowel in the ending syllables of the word is more than that in the starting syllables, i.e. moving toward the end of the word, the vowel duration increases. The effect of stress on the starting syllables of a word is more than that of the ending syllables. Pause has also an important role on the duration of the vowel of the previous syllable.

5. Conclusions

In this paper, through a series of statistical analyses, we have shown that stress has an influence on all three prosodic features discussed. The obtained values demonstrate that the highest influence is on the pitch contour slope. The vowel duration stands second, whilst the least influence is measured on the average pitch frequency. However, with these conclusions, it is not possible to foresee either the amount of influence of each of these parameters on the performance of a stressed speech recognizer, or the most influential one in recognition. The answer to these questions will be available once these results have been utilized in the design of a speech recognition system. Furthermore, in order to be able to get clearer results from the implementation of prosodic parameters in a speech recognition system, several levels of

stress will be needed as speech tags, since not all different parts of one utterance receive the same amount of stress.

6. References

- [1] X. Huang, et.al, *Spoken language processing*, Prentice-Hall, 2001.
- [2] E. Shriberg, et.al, “Prosody-based automatic segmentation of speech into sentences and topics”, *Speech Communication*, 32(1-2), September 2000.
- [3] B. Shneierman, “The limits of speech recognition”, *Communication of the ACM*, vol. 43, no. 9, September 2000.
- [4] Y. Cheng and H.C. Leung, “Speaker verification using fundamental frequency”, *In Proc ICSLP’98*, Sydney.
- [5] M. Kemal Sonmez, et.al, “A lognormal tied mixture model of pitch for prosody-based speaker recognition”, *In Proc. Eurospeech’97*, Rhodes, Greece.
- [6] C. Wang and S. Seneff, “Lexical stress modeling for improved speech recognition of spontaneous telephone speech in the JUPITER domain”, *In Proc Eurospeech 2001*, Aalborg, Denmark.
- [7] K. Hirose and K. Iwano, “Detection of prosodic word boundaries by statistical modeling of MORA transitions of fundamental frequency contours and its use for continuous speech recognition”, *In Proc. ICASSP 2000*, Istanbul, III-1763.
- [8] R. Silipo and S. Greenberg, “Automatic transcription of prosodic stress for spontaneous English discourse”, *The 14th International Congress of Phonetic Sciences*, San Francisco, August, 1999.
- [9] F. Almasganj, “Structural analysis of the Farsi utterance using prosodic features”, *PhD Thesis*, Tarbiyat Modarres University, Tehran, 1998.
- [10] L. Hitchcock and S. Greenberg, “Vowel height in intimately associated with stress accent in spontaneous American English discourse”, *In Proc Eurospeech 2001*, Aalborg, Denmark.
- [11] D. Gharavian, H. Sheikhzadeh and S.M. Ahadi, “An experimental multi-speaker study on Farsi phoneme duration rules using automatic alignment”, *In Proc SST2000*, Canberra, Australia.
- [12] M. Bijankhan et al., “The speech database of Farsi spoken language”, *In Proc. SST’94*, Perth, Australia.
- [13] S.J. Young et al., *The HTK Book*, Cambridge University Eng. Dept., 2001.
- [14] Y. Medan, E. Yair and D. Chazan, “Super resolution pitch determination of speech signals”, *IEEE Trans. Sig. Proc.*, Vol. 39, No.1, January 1991.
- [15] Edinburgh Speech Tools Library, available at http://festvox.org/docs/speech_tools-1.2.0/x2152.htm.