EVALUATION OF THE EFFECT OF STRESS ON FORMANTS IN FARSI VOWELS

D. Gharavian[1,2] & S.M. Ahadi[1]

[1]Electrical Engineering Department, Amirkabir University of Technology, Tehran, Iran
[2] Power & Water Institute of Technology, Tehran, Iran
gharavian@pwit.ac.ir, sma@cic.aut.ac.ir

ABSTRACT

Stress is known as an important prosodic property of the speech. Recognition of stressed speech, due to the effects of stress on acoustic features of speech, has always been considered more difficult, in comparison to unstressed speech. On the other hand, preparing a large speech corpus to include all different possible occasions of stressed speech, for the purpose of training in speech recognition, is very difficult. This is true for any language and Farsi (Persian) is no exception. Due to the importance of formant frequencies, they have long been considered in the recognition task. In this paper, we report on our efforts to find out the potential effects of stress on the formant frequencies. The results of our statistical evaluations demonstrate changes in the formant frequencies in certain directions, due to stress, that may be found helpful in improving automatic recognition of stressed speech.

1. INTRODUCTION

One of the effects known as a source of problem, for a speech recognition system, is the prosodic feature applied to speech by the speaker. A wide variety of such features can be counted such as the speaker gender, his/her speaking style or type of speech, stress, etc. Although some of these features may help the human listener to get a better understanding of the speaker's speech, they usually lead to a decrease in the performance of a speech recognition system trained under normal conditions [1,2,3].

Apparently, to be able to improve the recognition results, we need some information regarding the prosodic features of the speech. Such information may be utilized in several different ways in a speech recognition system [4,5,6]. In order to acquire such information, we need to investigate the effects of such prosodic changes, e.g. stress, on the speech features and the way they affect the recognition process. In previous works, the vowel durations, pitch frequency and its slope were investigated [8,9]. In this work, we try to extend these results to cover the effects of stress on formants and vowel energies. The use of formants in speech recognition has already been considered in several different ways. If appropriate relationships between the stress and the formant frequencies are found, then these relationships may be used in a formant-based speech recognition system (or a system that uses formant parameters for improving recognition) to further improve the recognition of vowels.

2. THE SPEECH CORPUS

The Farsi continuous speech corpus, FARSDAT [10] was used in this work. This corpus is the only continuous speech corpus of Farsi, which is available to public. It consists of 6000 sentences from 300 speakers, each uttering 20 sentences selected from a set of 392 available sentences. The sentences have been uttered in a normal condition with a high SNR. No particular stress has been used by these speakers. 1800 native speaker sentences from this corpus are used for building a 15-Gaussian phoneme-based HMM system. Training, adaptation and forced alignments have been carried out using HTK [11].

An extension to the speech corpus was made using the speech of a male speaker. This speaker has uttered the 392 sentences of the corpus normally 3 times. The HMMs have been adapted to the unstressed speech of this speaker. 154 sentences (from the total of 392) have been selected and uttered several times by the same speaker with stress applied to different parts (words) of them. This has led to a total of 468 stressed sentences. Using these stressed sentences, a set of stressed HMMs has been created. These HMM systems have been used for forced-alignment of the stressed and unstressed speech with their phonemic transcriptions. These time alignments have been used for statistical evaluations of formants and other parameters of interest.

3. VOWELS AND SYLLABLES IN FARSI

Farsi includes a total of about 30 phonemes, among which 6 vowels exist. These vowels are shown in Table 1. These are

Table 1. Vowels of Farsi

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

divided into two groups of weak (/æ/,/ɛ/,/o/) and tense (/ɑ/,/i/,/u/) vowels. Another phoneme, /ow/, although a diphthong, behaves similar to vowels. Hence, in this work, in cases where there exist enough data, will be used among the vowels for the evaluation of the effect of stress. According to the most popular categorization, the number of syllables in Farsi is three, namely CV, CVC and CVCC.

4. EFFECT OF STRESS ON FORMANTS

In this paper, we try to give some statistical evidence on the effect of stress on the first 3 formants. The formants have been found using McCandless formant tracker [12]. After extracting formants, several discontinuities were found in the frequency contours, which could easily be spotted visually. In order to compensate for these, within the voiced regions, the values of the neighboring frames were used. After calculating the required statistics, some outlier points were observed. A variance analysis approach was taken to overcome this problem. Here, the data points, which were more than 1.5 times standard deviation away from the distribution mean, were ignored. The value of 1.5 was chosen empirically.

As the main motive for this research has been to find a way to improve the recognition rate in a speech recognition system, and since during the recognition the type of the syllable is not known, the results that are independent of the syllable type are more important. Table 2 depicts the frequency of each Farsi vowel in the CV and CVC syllables of our corpus. About 10% of these are stressed. The frequencies for CVCC syllables are relatively small. According to this table, the distribution of vowels in syllables is not consistent and depends very much on the types of vowel and syllable. Hence, an analysis, which does not explicitly take the syllable type into account, can be considered approximate. In this paper, for the sake of briefness, we only report the results, regardless of the syllable type. However, throughout our discussions, some important behaviors that are related to the syllable type will be pointed out.

In the following analysis, a number of points have been considered. As mentioned earlier, due to the small number of CVCC syllables, they have not been taken into account. Furthermore, since only little data has been available for the vowel /ow/, the results for this vowel are not very reliable. Therefore, at some points, these results have been omitted.

Only the first 3 formants and their slopes have been used. In order to find a formant slope, in the region of a vowel, the formant contour has been approximated with a line. The slope of this line has been used as the slope of that formant's contour.

Table 2. The frequency of CV and CVC syllables for each of the Farsi vowels.

Vowel	CV	CVC
/æ/	1432	2376
/ɑ/	2379	1176
/ɛ/	2615	713
/i/	1257	595
/o/	628	484
/u/	123	46

The results of the change in the formant means are shown in Tables 3, 4 and 5 and Figure 1. Due to the space limitations, the results for the standard deviations are not shown in the figures. The formant slopes are in Hz/frame and S/U means the value of the stressed parameter to the unstressed one.

4.1. Formant means and variances

The region of change in stressed and unstressed cases for F1 is between 280Hz and 500Hz, for F2 between 1000Hz and 2000Hz and for F3 between 2400Hz and 2700Hz. From Tables 3 to 5, it can be seen that F3 has the lowest variance and F1 the

Table 3. F1 statistics in stressed and unstressed conditions. The syllable type is not taken into account.

Vowel	Unstressed		Stressed		S/U	
	Mean	SD/ Mean	Mean	SD/ Mean	Mean	SD/ Mean
/æ/	442.86	0.36	452.53	0.36	1.006	1.02
/ɛ/	349.83	0.29	389.16	0.32	1.080	1.11
/o/	347.19	0.23	349.83	0.23	1.007	1.01
/ɑ/	418.95	0.29	426.43	0.30	1.055	1.02
/i/	307.89	0.26	291.05	0.24	0.905	0.95
/u/	303.37	0.19	299.90	0.21	1.135	0.99
/ow/	365.45	0.22	321.19	0.17	0.773	0.88

Table 4. F2 statistics in stressed and unstressed conditions. The syllable type is not taken into account.

Vowel	Unstressed		Stressed		S/U	
	Mean	SD/ Mean	Mean	SD/ Mean	Mean	SD/ Mean
/æ/	1610.4	0.12	1639.6	0.16	1.30	1.02
/ɛ/	1703.6	0.16	1640.7	0.22	1.39	0.96
/o/	1252.4	0.24	1151.1	0.25	1.03	0.92
/ɑ/	1295.4	0.19	1355.1	0.23	1.18	1.05
/i/	1867.1	0.17	1796.7	0.21	1.24	0.96
/u/	1198.2	0.28	1349.1	0.31	1.11	1.13
/ow/	1116.0	0.29	1236.9	0.34	1.18	1.11

Table 5. F3 statistics in stressed and unstressed conditions. The syllable type is not taken into account.

Vowel	Unstressed		Stressed		S/U	
	Mean	SD/ Mean	Mean	SD/ Mean	Mean	SD/ Mean
/æ/	2502.6	0.058	2548.2	0.056	0.969	1.018
/ɛ/	2540.0	0.064	2585.4	0.062	0.078	1.018
/o/	2450.6	0.067	2496.4	0.064	0.967	1.019
/ɑ/	2501.0	0.061	2543.8	0.072	1.189	1.017
/i/	2593.6	0.065	2605.2	0.070	1.078	1.004
/u/	2469.7	0.057	2464.8	0.070	1.231	0.998
/ow/	2506.9	0.061	2620.7	0.086	1.422	1.045

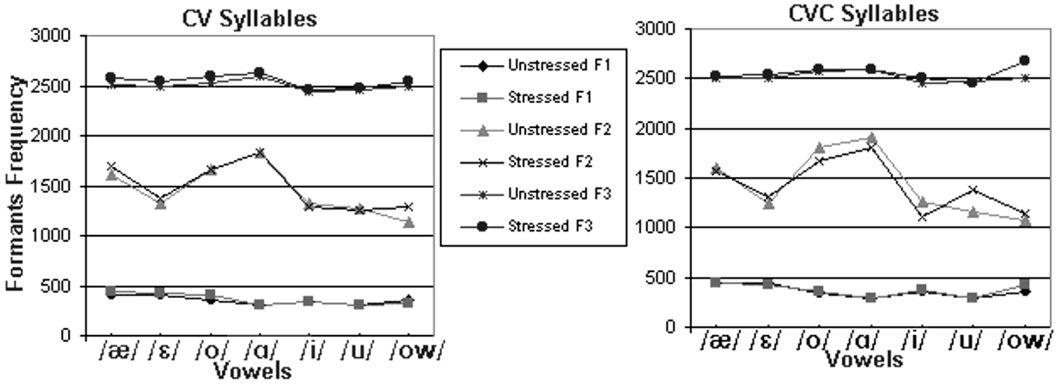

Figure 1. Changes in the formant means in CV and CVC syllables.

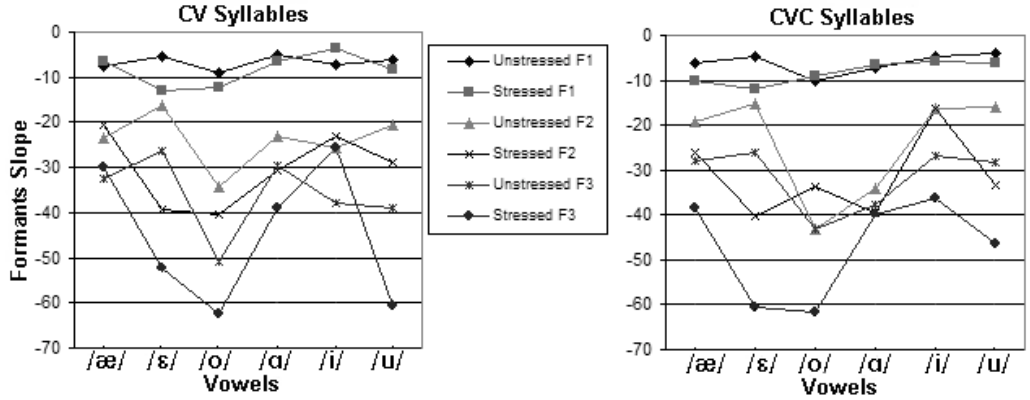

Figure 2. Changes in the formant slope means in CV and CVC syllables.

highest. In fact, as the formant frequency increases, the change in its value, due to either stress or context, gets smaller. It can be concluded that higher formants may be less effective in the recognition of the stressed speech as their change due to stress is smaller [13].

We have also observed that for F1, F2 and F3, generally, as the syllable gets larger, the changes due to stress get smaller. In other words, the effect of stress on formants is inversely related to syllable size. In comparison to the effect of stress on such parameters as duration and pitch frequency [9], one can conclude that formants are inversely affected by the stress.

In contrast to most other vowels, /ow/ shows a smaller F1 mean under stress. Furthermore, the F1 mean for the cases of /i/ and /u/, the F2 mean for /ɛ/, /i/ and /o/ and the F3 mean for /u/, whether syllable type is taken into account or not, are usually either reduced or do not change considerably under the stressed conditions. Moreover, the change in the mean values of F1 and F3 for weak vowels is more than tense vowels, while for F2, more changes are seen in tense vowels in comparison to the weak ones.

Generally, an increase in the standard deviation of formants due to stress is observed. The only exceptions are F1 in the cases of /i/ and /ow/ and F3 in the cases of /æ/ and /ɛ/ for CV

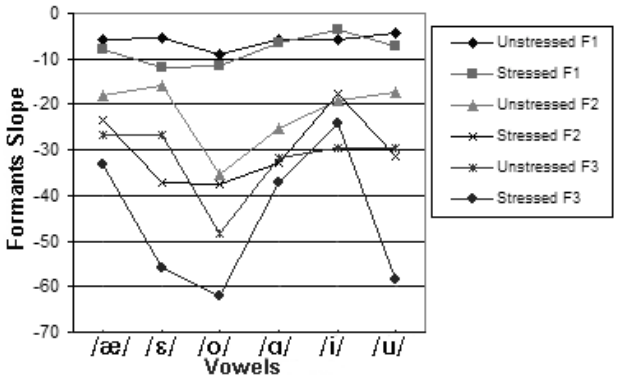

Figure 3. Changes in formant slope means, regardless of the syllable type.

syllables and /o/ in both syllable types. In the case of F3, a larger change in the standard deviation is observed for CVC syllables, compared to CV ones. For F1 and F2, the changes in their standard deviations in the case of /æ/ is larger in CVC syllables, while for /o/ and /u/, CV syllables show a larger change. Altogether, it seems that the change in the standard deviations of different syllables is irregular and needs further investigations.

4.2. Formant slopes

The changes in the formant slopes, however, are more promising. The overall results for the formant slopes are shown in Figures 2 and 3. Interestingly, almost for all formants and in all syllables, the formant slopes are negative. According to the statistics concerning the formant slopes, in a minimum of 90% of all frames (case of /æ/) for F1, a minimum of 85% of all frames (case of /ow/) for F2 and a minimum of 99.7% (case of /æ/) for F3, negative slopes are observed. The changes in the means of the formant slopes, due to stress, are more regular in comparison to the formant mean changes. This is exactly in contrast to the slope of the pitch, which, in spite of being highly influenced by the stress, does not show much regularity [9].

Generally, stress leads to a steeper slope. However, for F1, in the cases of vowels /o/ and /æ/ in CV syllables and /ɛ/ and /i/ in CVC syllables the slope (negative) mean values increase. The same effect is seen for F2 in the cases of /o/ and /æ/ in CV syllables and /ɛ/ and /o/ in CVC syllables and for F3 in the cases of /o/ and /æ/ in CV syllables. The changes in the slope means of the 3 formants of tense vowels are usually more than weak vowels. The only exceptions are /ɛ/ and /i/ vowels' F1 in CV and F3 in CVC syllables.

Among all the vowels, /ow/ demonstrates the highest change in the slope means for all 3 formants, due to stress. The range of change of slope means is generally between 1 and 2. The exact value depends on the formant and vowel type. In higher formants the value of slope and the amount of its change due to stress are increased. For F1 and F3, the stress usually leads to a decrease in the standard deviation of the formant slopes, i.e. the dispersions of slopes within the vowel are decreased. The only exceptions are F1 of /o/ and /æ/ in CVC syllables and F3 of /æ/ in CV syllables. For F2, however, less consistency is observed.

The changes in the standard deviation of formant slopes of tense vowels are usually more than that of the weak vowels. An exception is the case of the weak vowel /ɛ/ and the tense vowel /i/ for F1 and F2 in CVC syllables. The highest rate of change in the standard deviation of the formant slope is seen for /ow/.

From the above discussion we can conclude that the formant slopes, similar to the pitch slope, are more important than their means. Due to their stronger and more consistent changes in the stressed conditions, in comparison to the mean values, a better role can be expected from them in helping to improve recognition. The standard deviations of the formant slopes, similar to that of the pitch slope, tend to decrease after the stress is applied. Furthermore, we can draw a rough conclusion that, here, F1 and F3 behave similarly. The exceptions counted for above are usually related to certain groups of vowels in certain conditions (formants and syllable types). This points out that these vowels have certain properties that cause this behavior. Although relatively small, such groups should be paid more attention in future research works. Among these, the vowel groupings of (/ɛ/ and /i/), (/o/ and /æ/) and (/ow/) can be counted.

5. CONCLUSIONS

The effect of stress on the formant frequencies has been addressed in this paper. We have shown that stress affects the formants and their slopes. The effect on the formant slopes, due to its consistency, seems to be more important, compared to that of the formant values. Formant slopes have shown consistent and regular changes due to stress. The most regular changes belong to the third formant. The implementation of the relationships defining such effects can be useful in improving speech recognition results under stressed conditions.

6. REFERENCES

[1] X.Huang, A.Acero and H-W.Hon, *Spoken language processing, A guide to theory, Algorithm, and system development*, Prentice Hall, 2001.

[2] Elizabeth Shriberg, Andreas Stolcke, Deilek Hakkani-Tur, Gokhan Tur, "Prosody-based automatic segmentation of speech into sentences and topics", *Speech communication*, 32(1-2), September 2000.

[3] Ben Shneilerman, "The limits of speech recognition", *Communication of the ACM*, vol. 43, no. 9, September 2000.

[4] York Cheng, and Hong C.Leung, "Speaker verification using fundamental frequency", in *Proc. ICSLP,* Sydney, Australia, 1998.

[5] M.Kemal Sonmez, Larry Heck, Mitchel Weintraub and Elizabeth Shriberg, "A lognormal tied mixture model of pitch for prosody-based speaker recognition", in *Proc. Eurospeech*, Rhodos, Greece, vol. 3, pp. 1391-1394, September 1997.

[6] Chao Wang and Stephanie Seneff, "Lexical stress modeling for improved speech recognition of spontaneous telephone speech in the JUPITER domain", in *Proc. Eurospeech*, Aalborg, September 2001.

[7] Keikichi Hirose and Koji 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,* Istanbul, III-1763, June 5-9, 2000.

[8] 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.

[9] D.Gharavian and S.M.Ahadi, "Statistical Evaluation of the Influence of Stress on Pitch Frequency and Phoneme Durations in Farsi Language", in *Proc. Eurospeech*, Geneva, September 2003.

[10] M. Bijankhan *et al.*, "The speech database of Farsi spoken language", in *Proc. SST'94,* Perth, Australia.

[11] S.J. Young *et al*., *The HTK Book*, Cambridge University Eng. Dept., 2001.

[12] S.S.McCandless, "An algorithm formant extraction using linear prediction spectra", *IEEE Transactions on acoustics, speech and signal processing*, ASSP-22, n.2, pp.135-141, April 1974.

[13] Lizao, Wei Lu, Ye Jeang and Zhen Yang Wu, "A study on emotional feature recognition in speech", in *Proc. ICSLP*, Beijing, 2000.