ISSN 1799-2591
Theory and Practice in Language Studies, Vol. 5, No. 2, pp. 426-434, February 2015
DOI: http://dx.doi.org/10.17507/tpls.0502.26

Vowel Classification and Vowel Space in Persian

Nasim Esfandiari
University of Isfahan, Isfahan, Iran

Batool Alinezhad
University of Isfahan, Isfahan, Iran

Adel Rafiei
University of Isfahan, Isfahan, Iran

***Abstract*—This article aims to develop an acoustic vowel space in Persian speech. There are several aspects in this survey which make it different from what has been done before. The first is related to the issue of speech material. The need for more natural choice in voice qualities in recent years exhort us not relying on citation form or artificial sound produced in laboratories. Furthermore, the formant frequencies were not extracted from specified vowels, in specified context. In contrast, we are interested in the shape of vowel space determined by extremely large collections of vowel tokens, with whatever distribution of categories and context they may have in the read text. Thirdly, the vowels selected in the database for calculation of the area of vowel space are being stratified for locating in stressed or unstressed syllables, or being uttered by male or female speakers. So, we are simultaneously dealing with four groups. But the most important aspect is related to the methodology used for better plotting of vowels. Either F_1*F_2 or F_1 *F_2-F_1 is leaded to better vowel classification is a matter being evaluated by two parameters: (a) linear discriminant analysis and (b) scatter reduction.**

***Index Terms*—vowel space, outlier, linear discriminant analysis, scatter reduction**

I. Introduction

In acoustic and accordingly other related practical sciences such as signal processing, the study of the characteristics of vowel sounds has been taken a vast amount of attention. The significance of this consideration is particularly justified with the essentiality of vowels in intelligibility of speech. So, this article aims to add to vowel literature by presenting Persian an acoustic vowel space.

In articulatory phonetics, vowels are defined as a voiced sound produced by the absence of any occluding, diverting, or obstructing in the vocal tract, allowing the breath stream free passage through the larynx and oral cavity. The articulatory features that distinguish different vowel sounds are said to determine the vowel`s quality. Daniel Jones (cited in Ladefoged & Johnson, 2001) developed the cardinal vowel system to describe vowels in terms of common features height (vertical dimension), backness (horizontal dimension) and roundedness (lip position).

In Acoustic phonetics, it has been established that in the production of vowels, vocal resonances are altered by the articulator to form distinguishable vowel sounds. That is, the repetitive closure of vocal folds set the different volumes of air in throat and mouth into vibration and as such a sound wave is produced. The resonances of vocal tract which are called formants are decisive means at determining the qualities of vowels (Ladefoged & Johnson, 2001; Ladefoged, 2006). There is extensive evidence going back to the nineteenth and early part of twentieth century that vowel-quality distinction depends on the first two resonances of vocal tract (Traunmuller & Lacerda, 1987, Ladefoged & Johnson, 2001, Ladefoged & Maddieson, 1999, Livonen, 1996; Harrington, 2010; Sundberg, 1977). The jaw opening, which constricts the vocal tract toward the glottal end and expands it toward the lip end, and the shape of the body of the tongue are respectively the deciding factors for the first and second formants. It should be mentioned that some researchers (Fant, 1973) consider F_3 for vowels and followed F_1 vs F'_2 plane for plotting vowel space in which F'_2 is a weighted average of F_2 and F_3. But we did not pursue F_3 and relying upon first two formants in plotting Persian vowel space in this survey for following reasons. Firstly, third formant, according to Ladefoged & Johnson(2001), has very little function in distinguishing the vowel and can be predicted fairly accurately from the frequencies of the first two formants. Secondly, the first two formants are used to contrast one vowel with another in nearly every language but others are used less frequently. Additionally, the frequency of the third formant is very much affected by the position of the lips. It so happens that Persian has no vowel with the same tongue positions but different lip position (Haghshenas, 1997; Alinezhad & hosseininibalam, 2012; Bijankhan, 2013). So, considering the first two frequencies seems sufficient in plotting Persian vowel space.

In vowels, as it is implied, the frequency of lower formants is mainly used to categorize vowel. The higher the tongue in the mouth when producing the vowel, the lower F_1. The further forward the tongue in the mouth when producing the vowel the higher F_2. The acoustic theory of speech production (Harrington, 2010) shows that there is a relationship between phonetic height and F_1 and phonetic backness and F_2 from which it follows that if vowels are plotted in the plane of the first two formant frequencies with decreasing F_1 on the y-axis and decreasing F_2 on the x-axis, a shape resembling the vowel quadrilateral emerges. In this way, the relationship between the first and the second formant is being summarized in a vowel space plot. The vowel space illustration provides a graphical method of showing where a speech sound, such as a vowel, is located in both acoustic and articulatory space. This was first demonstrated by Essner (1947) & Joos (1948) and since then the F_1*F_2 plane has become one of the standard ways of comparing vowel quality in phonetics.

In interpreting the details of vowel spaces, the traditional articulatory descriptions of vowels are related to formant frequencies, although there is no consenus in this issue. It is claimed that the correlation between the second formant frequency and the degree of backness of a vowel is not as good as that between the first frequency and the vowel height since the second formant is considerably affected by both backness and lip rounding. So to eliminate some of the effect of lip rounding, the second formant is considered in relation to the first. In other words, the degree of backness is best calculated to the differences between first and second formant frequencies (Ladefoged & Johnson,2001; Bijankhan, 2013). That is, the close they are together, the more "back" a vowel sounds. In this way, representing a vowel space in F_1* F_2-F_1 plane became one of the traditions in acoustic analysis of vowels.

In Persian, either F_1*F_2 or F_1*F_2-F_1 is more appropriate in representing and classifying vowels is a question to be brought forth for discussion in this paper for the purpose of providing Persian a vowel space plane. In fact, accepting plotting vowels on F_1*F_2-F_1 plane theoretically, we are prompt to investigate practically whether F_1*F_2-F_1 plotting is leaded to a better vowel classification than F_1*F_2 plane or not. The deliberation of this matter is performed considering two parameters through which vowel classification is evaluated: (a) linear discriminant analysis and (b) scatter reduction. Then, by applying the result achieved in the first stage, a vowel space for standard Persian in continuous speech is developed.

There are three aspects in this survey which make it different from what has been done in Persian before in providing vowel spaces: (1) Since in recent years, more emphasis has been placed on the research and production of more natural sounding male and female voices, we do not rely on citation form or artificial sound produced in laboratories. That is, the need for more natural choice in voice qualities is one at the major issues that has been addressed in speech synthesis in recent years, especially when considering voice output communication aids (VOCAs) and increasing needs of users of such devices. (2) Since Peterson and Barney's classic(1952)article on vowel formant patterns, the acoustic space of vowels has been studied for many languages. In most, if not all of these, the formant frequencies were extracted from specified point in specified phonetic contexts. In contrast, we are interested in the shape of vowel space determined by extremely large collections of vowel tokens, with whatever distribution of categories and context they may have in the read text. Of course it should be mentioned that we have considered prosodic context to see whether the vowel is placed in stresses syllable or unstressed one. (3) The vowels selected in the database for calculation of the area of vowel space are being stratified for locating in stressed or unstressed syllables, or being uttered by male or female speakers. So, we are simultaneously dealing with four groups: (a) vowels located in stressed syllable uttered by female, (b) vowels located in stressed syllable uttered by male, (3) vowels located in unstressed syllable uttered by female, (4) vowels located in stressed syllable uttered by male.

To be briefly familiar with Persian vowel system, it can be stated that linguistics (Alinezhad & Hosseinibalam, 2012; Bijankhan, 2013; Haghshenas, 1997) are unanimously agreed upon six-vowel /i/, /e/, /æ/, /u/, /o/, /a/ existed in Persian vowel system in which /i/, /e/, /æ/ is considered as being front and /u/, /o/, /a/ as back vowels. In Persian, the division of vowels based on rounded-unrounded feature is the same as the division based on front-backness. So, this feature is not contrastive in Persian.

II. Methodology

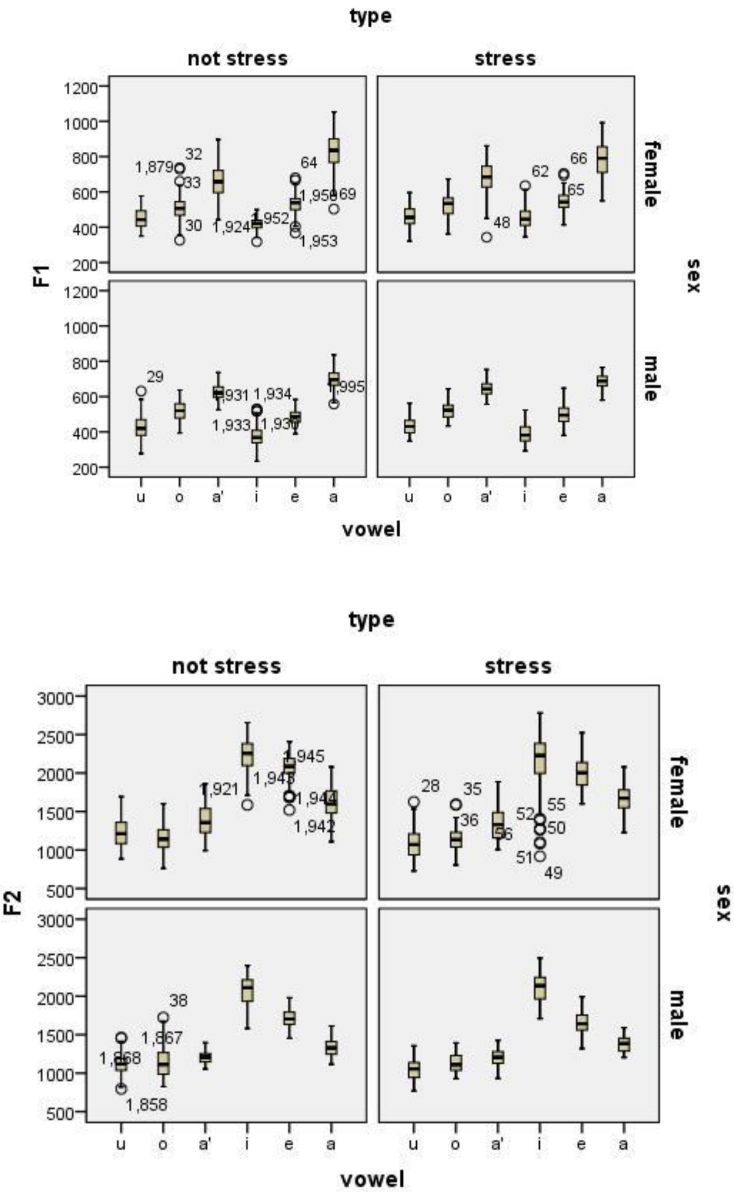
A. Data Representation

The speech material consists of recordings of IRIB Broadcasts of 10 news reporter of Persian who were stratified for their gender (5 male, 5 female). The broadcaster, aged 35-50 years, were born and raised in Tehran, Iran. They can be regarded as professional language users in standard version of Persian as they have all passed successfully many courses and examination in being expertized to speak well to be understood by Iranian population who are interested in following news. The news in question were broadcast and recorded in August and September, 2013.

B. Formant Extraction and Outliers

Tokens of the vowels were identified from simultaneous inspection of three displays (raw wave-form, spectrum, and spectrogram). Formant values calculated by the program's LPC algorithm, using a window of 20ms and a band width of 300Hz, were read off the spectrum display at a point which was judged as indicating the main tendency of the vowel without consonantal interference, following a procedure described by Harington, et al (2000).

In spite of considering all the related issues carefully, there occur some errors especially in large speech corpora due to formant tracking which is inevitable. So, when such errors occur, it is likely that they will show up as outliers in ellipse plot and will be deleted. Statistically, an outlier is an observation point that is distant from other observation. Outliers, if not deleted, can have dramatic effect on the ellipse orientation since they are usually far from the ellipse's center (Harrington, 2010). The technique we follow to detect outliers is box plot pattern. In this technique, each box shows the median, quartiles and extreme values within a category. The center of the distribution can be approximated by the median or second quartile. So, half of the data values fall between the first and third quartiles. In this paper, since the same data are drawn into two different plane, three box plots (the first for the first formant, the second for the second formant and the third for the gap between formants) is needed for detecting outliers as follow (Fig. 1):



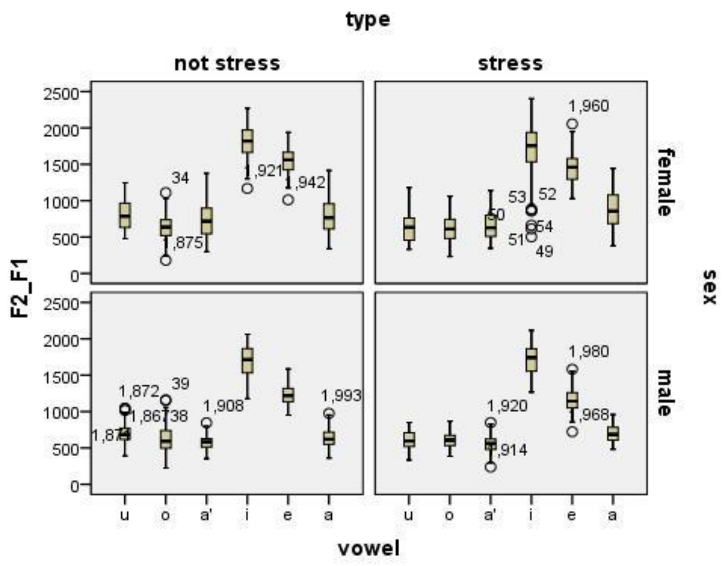


Figure 1. Three box plots for detecting outliers in F_1, F_2, & F_2-F_1

Outliers are shown in the Fig. 1 with circles. They included the cases with values between 1.5 to 3 box lengths from the upper or lower edge of the box. It should be mentioned that the box length is the inter quartile range. As it is shown in the table I below, the vowel tokens decrease from 1970 to 1910 when outliers omitted.

TABLE I.
DEMONSTRATING THE NUMBER OF TOKENS BEFORE AND AFTER DELETION

before deleting

Vowel	Number
/u/	216
/o/	328
/a/	344
/i/	341
/e/	383
/æ/	358
Total	1970

After deleting

vowel	Number
/u/	208
/o/	316
/a/	340
/i/	323
/e/	368
/æ/	355
Total	1910

The distribution of vowel tokens in different groups and the effect of deleting the outliers can be examined in the scatter diagram. In scatter diagram, the vowel tokens are displayed as a collection of points, each having the value of the first formant determining the position on the horizontal axis and the value of the second formant determining the position on the vertical axis. In Fig. 2, the diagram plots the scattering before implementing outlier detecting and Fig.3 displays scattering after outlier deleting.

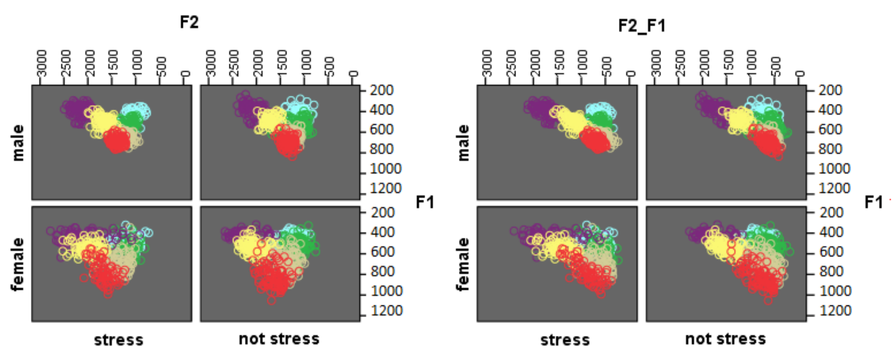


Figure 2. Scatter plot before deleting outliers

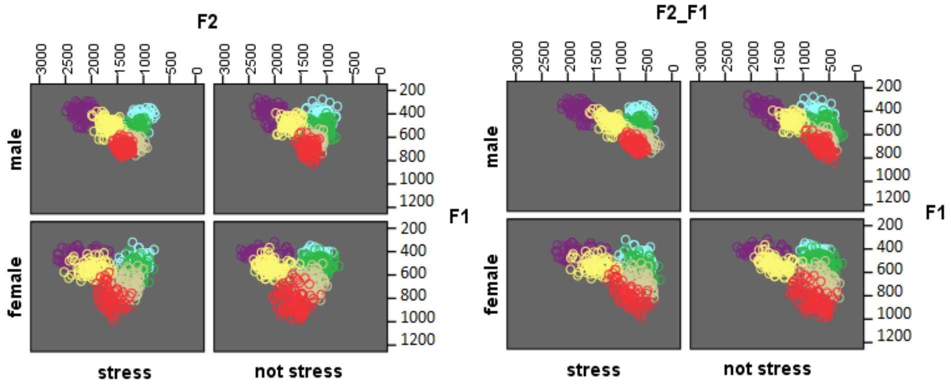


Figure 3. Scatter plot after deleting outliers

C. F_1*F_2 or F_1*F_2-F_1 Plane

It has been argued that representing vowel space based on the frequency gap between the first two formants (F_2-F_1) on horizontal axis can be considered as a better correlator in tracing tongue position. Assenting to this opinion, we are prompt to see how effectively plotting vowel space based on F_1*F_2-F_1 separate the Persian vowels into distinct groups corresponding to the phonetic categories of Persian in comparison with plotting on F_1*F_2 plane. Two quantifiable method is proposed here for this evaluation: (1) Linear discriminant analysis and (2) scatter reduction.

D. Linear Discriminate Analysis

Linear Discriminate analysis (LDA), proposed by reference (Weenink, 1999) and has been implemented in the Praat program, builds a predictive model for group membership. The model is a standard pattern recognition technique that uses the pooled within-groups covariance matrix of the acoustic variables to classify cases [Adank, 2004]. LDA assumes that the within-groups covariance matrices are equal across categories. It composed of a discriminate function based on linear combination of the predictor variable (In this paper, F_1*F_2 and F_1*F2- F_1) that provide the best discrimination between the group (in this paper vowel categories). See table II & III below.

TABLE II.
DISCRIMINANT ANALYSES (LDA)

Two vowel spaces	Percentages correctly classified vowel tokens	Percentages correctly classified gender	Percentages correctly Classified Stress
HZ(F2-F1 * F1)	1404/1910=73.5%	1208/1910=63.2%	989/1910=51.8%
HZ(F2 * F1)	1404/1910=73.5%	1208/1910=63.2%	989/1910=51.8%

TABLE III.
PERCENTAGES CORRECTLY CLASSIFIED VOWEL TOKENS FOR EACH GROUP OF GENDER AND STRESS

Two vowel spaces	Female-unstress	Female-stress	male-unstress	male-stress	Average
HZ(F2-F1 * F1)	423/519=81.5%	309/404=76.5%	397/489=81.2%	418/498=83.9	80.775%
HZ(F2 * F1)	423/519=81.5%	309/404=76.5%	397/489=81.2%	418/498=83.9	80.775%

In tables II & III , the percentages of correct classification of vowel tokens based on two variables (F_1*F_2 and F_1* F_2-F_1) for gender and stress classification is shown. The results are surprising since all percentages are predicted to be the same. That is, there is no discrepancy between the percentages correctly classified vowel tokens for LDA F_1*F_2 and the percentages correctly classified vowel tokens for LDA F_1* F_2- F_1 in all groups and in this way, no improvement is achieved. So, it can be concluded that no one does have preference over other in classifying and discriminating vowel tokens into vowel categories.

E. Scatter Reduction

Scatter reduction parameter is selected for getting an indication of the distribution of the vowels in two mentioned planes (F_1*F_2 VS F_1*F_2-F_1). Disner (1980) claims that scattering is precisely quantifiable in terms of the reduction of the size of the ellipses needed to encompass the data points for each vowel. In other words, the smaller the resulting ellipses are, the more successful the plane is in classification of vowels. That is, reducing within-category variance has the greatest impact on classification (Weenink, 2001).

As it is clearly observed in Fig. 3, there are enormous spreads within each vowel class within each group for these two kinds of plane. So, it is not possible to decide which one is optimal in discriminating vowels visually. Statistically we appealed to Euclidean or straight-line distances between all vowel tokens and the centroid of the same vowel category. In two dimensional space, at first, two vowel spaces (F_1*F_2 and F_1*F_2-F_1) for the same data are developed and thus two centroid for each vowel category, one in F_1*F_2 and other F_1*F_2-F_1, are calculated. Then, the Euclidean distance is calculated by summing the square of horizontal and vertical distances between the points and taking the square root [Harrington, 2010]. The results are shown in table IV for more convenience.

TABLE IV.
THE EUCLIDEAN DISTANCE FOR EACH GROUP IN TWO PLANES

Sex	Type	Vowel	dis(F_1* F_2- F_1)	dis(F_1*F_2)
Female	not stress	/u/	.2128	.2195
		/o/	.2799	.2983
		/a/	.4113	.4051
		/i/	.3049	.3792
		/e/	.3249	.3315
		/æ/	.5290	.5221
	Stress	/u/	.3800	.3865
		/o/	.3025	.2963
		/a/	.3867	.4030
		/i/	.4311	.4899
		/e/	.3185	.3452
		/æ/	.5873	.5482
Male	not stress	/u/	.3960	.4306
		/o/	.4218	.5049
		/a/	.1663	.1656
		/i/	.5397	.5688
		/e/	.2273	.2583
		/æ/	.2869	.2874
	Stress	/u/	.2660	.3105
		/o/	.2517	.2871
		/a/	.2049	.2097
		/i/	.4259	.4614
		/e/	.3872	.3906
		/æ/	.2135	.2299

It seems F_1*F_2- F_1 plane to have smaller Euclidean distance to the centroid of vowel categories which is proved scientifically by means of Wilcoxon test. In the Wilcoxon test, ranks are based on the absolute value of the difference between the two test variables. The sign of the difference is used to classify cases into one of three groups: differences below 0 (negative ranks), above 0 (positive rank), or equal to 0 (ties). Tied cases are ignored. In these data, 5 cases have negative differences and the sum of their ranks equals 43. The other cases have positive differences, whose ranks sum to 257 (table V).

TABLE V.
DEMONSTRATING RANKS BASED ON WILCOXON TEST

Ranks

		N	Mean Rank	Sum of Ranks
dis(F_1*F_2) and- dis(F_1* F_2- F_1)	Negative Ranks	5[a]	8.60	43.00
	Positive Ranks	19[b]	13.53	257.00
	Ties	0[c]		
	Total	24		

a. dis(F_1*F_2) < dis(F_1* F_2- F_1)
b. dis(F_1*F_2) > dis(F_1*F_2-F_1)
c. dis(F_1*F_2) = dis(F_1*F_2-F_1)

Therefore, the null hypothesis referring to the equality of two variable is not only rejected by nonparametric Wilcoxon test but the preference of F_1* F_2- F_1 plane over F_1* F_2 is also confirmed, considering the greater amount of positive ranks in table V.

III. CONCLUSION

This paper aims to develop an acoustic vowel space in Persian natural speech. The vowel spaces are commonly represented in formant chart in which the lower formant at each vowel (F_1) is plotted on y-axis and the upper formant (F_2) on the x-axis. But it is argued that the difference between the first and second formant frequencies (F_2-F_1) on the abscissa can be regarded as a better correlator in tracing tongue position. This plotting (F_1* F_2- F_1) is somewhat in agreement with the notion that backness corresponds to the distance between formant two and formant one since second formant is affected by both backness and lip rounding.

Accepting plotting vowels on F_1*F_2-F_1 plane theoretically, we are prompt to investigate whether F_1*F_2-F_1 plotting is leaded to a better vowel classification than F_1*F_2 plane or not. As it is said before, the delibration of this matter is performed considering two parameters through which vowel classification is evaluated: (a) linear discriminant analysis and (b) scatter reduction. It has been shown that LDA not only counted no differences in the degree of classification provided by these two kinds, but it also reports equality. Considering scatter reduction parameter, although it appears to be not considerable difference between them at the first glance, statistically the differences are proved to be meaningful. Thus plotting vowel space, based on F_1*F_2-F_1 plane, in addition to represent tongue position much more accurately based on mentioned literature, it is accompanied with less scattering and as a result more classification in vowel space. So, we decide presenting Persian vowel space in F_1*F_2-F_1 plane as follows (Fig. 4):

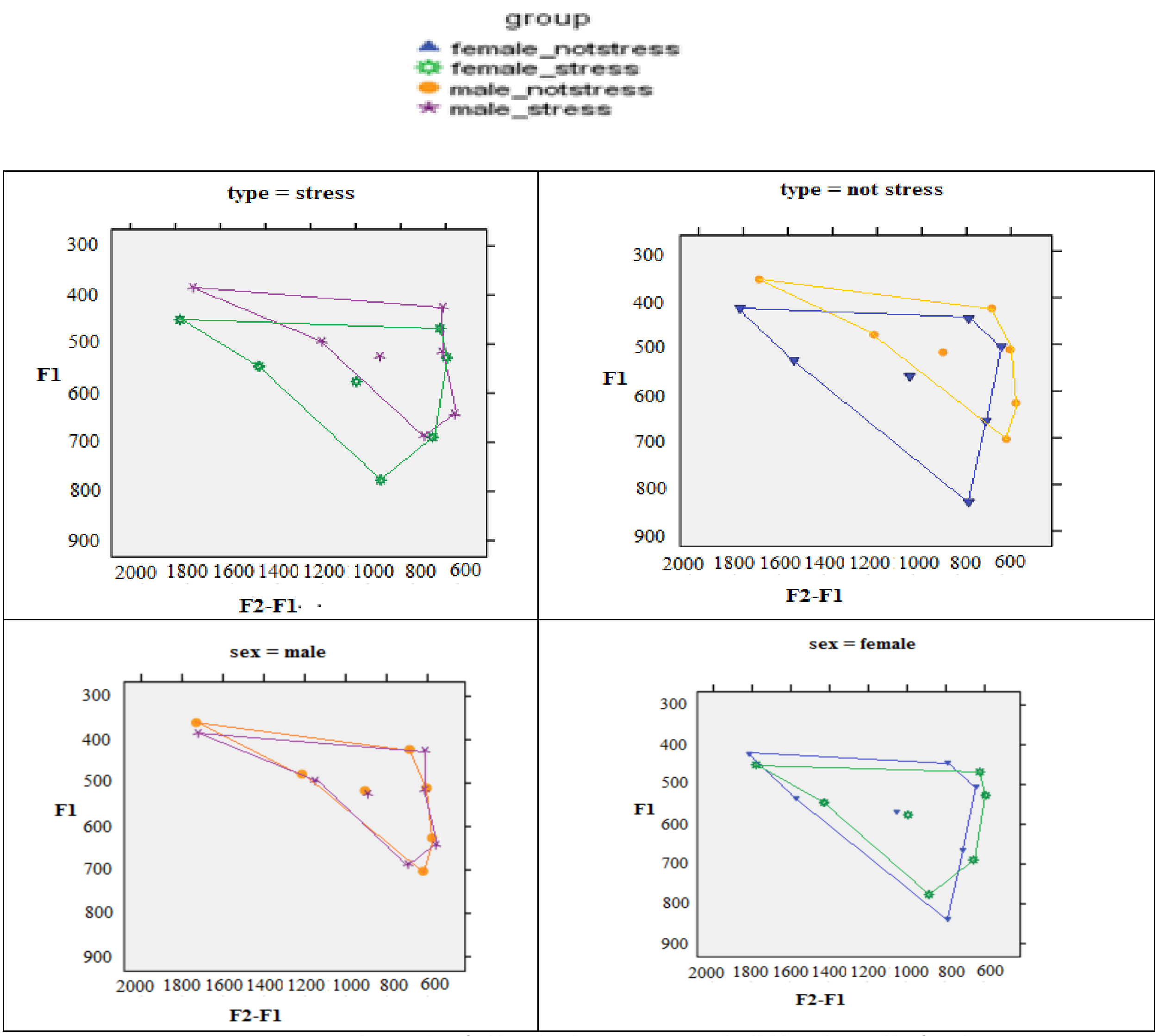


Figure 4. Persian vowel space classified based on being in un/stressed syllable and uttered by fe/male

For more direct comparison, all vowel spaces are being overlapped in Fig. 5:

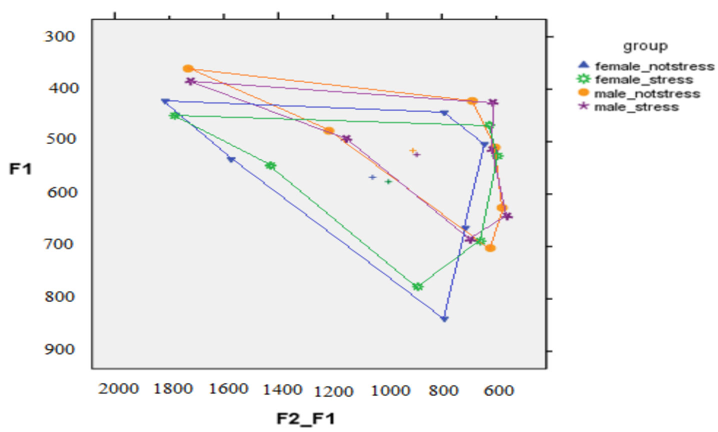


Figure 5. Overlapping of all vowel spaces

The characteristics of the drawn vowel spaces including the amount of F_1, F_2-F_1, N which shows the number of vowel tokens, and the center of the vowel spaces for each group is summarized in table VI:

TABLE VI.
THE CHARACTERISTICS OF VOWEL SPACES

Sex	Type	Vowel	F2_F1	F1	N	Center F2_F1	Center F1
Female	not stress	/u/	790.3636	444.4364	55	1054.07	567.85
		/o/	643.3750	504.7188	96		
		/a/	712.8542	664.3333	96		
		/i/	1814.8429	422.2000	70		
		/e/	1572.8404	533.6277	94		
		/æ/	790.1546	837.7835	97		
	Stress	/u/	625.3250	468.3250	40	995.62	576.31
		/o/	595.1636	527.3091	55		
		/a/	658.4156	689.7532	77		
		/i/	1779.1860	450.3256	86		
		/e/	1428.0946	545.4459	74		
		/æ/	887.5342	776.7123	73		
Male	not stress	/u/	687.7143	422.4464	56	906.28	517.07
		/o/	604.4342	511.0263	76		
		/a/	578.5526	626.0526	76		
		/i/	1730.7581	360.6129	62		
		/e/	1214.2500	479.2500	80		
		/æ/	621.9655	703.0460	87		
	Stress	/u/	612.4348	425.4130	46	892.16	524.91
		/o/	614.0143	515.8429	70		
		/a/	560.0588	642.3059	85		
		/i/	1721.5227	384.6364	88		
		/e/	1149.8795	494.3855	83		
		/æ/	695.0795	686.8750	88		

Finally, it should be mentioned that although it has been demonstrated using $F_1{*}F_2$-F_1 is more effective in reducing scattering, it is not intended the above as a criticism of plotting the vowel space on $F_1{*}F_2$ in any other respect. Furthermore, the result is only limited to natural version of Persian language and can not be generalized into other languages.

REFERENCES

[1] Adank, P., Van Hout, R. & Smiths, R. (2004). A comparison of vowel normalization procedures for language variation research. *Journal of the Acoustical Society of America* 116(5), 3099-3107.

[2] Alinezhad, batool & Hosseinibalam, Fahimeh. (2012). Fundamental of Acoustic Phonetics. Esfahan: Esfahan university publication.

[3] Bijankhan, Mahmood. (2013). Phonetic System of the Persian Language. Tehran: SAMT publication.

[4] Chiba, t., and Kajiyama, M. (1941). The Vowel: Its Nature and Structure. Tokyo: Tokyo Publishing Company.

[5] Disner, S. (1980). Evaluation of vowel normalization procedures. *Journal of the Acoustical Society of America* 67(1), 253-260.

[6] Essner, C. (1947). Recherche sur la structure des voyelles orales. *Archives Neerlandasises de Phonetique Experimentale*, 20,40-77.

[7] Fant, G. (1973). Speech sound and features. MIT Press: Cambridge, MA.

[8] Haghshenas, A. M. (1997). Phonetics. Tehran: agah

[9] Harrington, Jonathon. (2010). Phonetic analysis of Speech Corpora. UK: Wiley-Blackwell publication.

[10] Harrington, Jonathon. (2010). Acoustic phonetics. In W.J. Hardcastle and J. Laver (eds). *The Handbook of Phonetic Sciences*. Blackwell: Oxford.

[11] Harrington,Jonathon & Palethorpe, and C. I. Watson. (2000). Monophthongal vowel changes Received Pronunciation: an Acoustic Analysis of the Queen's Christmas Broadcasts. *Journal of the International Phonetic Association*, Vol 30, 63-78.

[12] Henton, G. (1995). Cross-language variation in the vowels of female and male speakers. In *Proc. XIIIth ICPhS*, Volume 4, Stockholm, 420–423.

[13] Joos, M. (1948). Acoustic Phonetics. *Language*, 24, 1-136.

[14] Ladefoged, P. (2006). A course in phonetics, UK; Blackwell Publishing

[15] Ladefoged, P., and Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America* 29, 88-104.

[16] Ladefoged, P. & Johnson, K. (2001). *A* course in phonetics. USA: Wadsworth, Cengage learning.

[17] Ladefoged, P. & I. Maddieson. (1999). The Sounds of the World s languages. Massachusetts: Blackwell Publishers Ltd.

[18] Livonen, Antti. (1996). Finnish Speech Data Base as a Research tool .*Academy of Finland* 1.1.–31.12.

[19] Neary, Terance. M. (1989). Static, dynamic and relational properties in speech perception. *Journal of the Acoustical Society of America* 85, 2,088-113.

[20] Nordstrom, P. E. (1967). Female and infant vocal tracts simulated from male area functions. *J. Phonetics* 5, 81-92.

[21] Peterson, G.E. & Barney, H.L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* (24),175-184.
[22] Pols, L. C. W., Tromp. H. R. C., and Plomp, R. (1973). Frequency analysis of Dutch vowels from 50 male speakers. *Journal of the Acoustical Society of America* 53, 1,093-101.
[23] Sundberg, Johan. (1977). Acoustics of the singing voice. *Journal of the Acoustical Society of America* 236(3), 104-116.
[24] Traunmuller, H., and Lacerda, F. (1987). Perceptual relativity in identification of two-formant vowels. *Speech Communication*, 6, 143-57.
[25] Thomas, Erik. (2002). Instrumental Phonetics. In *The Handbook of Language Variation and Change*. Oxford: Blackwell.
[26] Weenink, David. (1999). Accurate algorithms for performing principal component analysis and discriminant analysis. *Proceeding of the Institute of Phonetic sciences*. University of Amsterdam.
[27] Weenink, David. (2001). Vowel normalization with the timit acoustic phonetic speech corpus. *Proceeding of the Institute of Phonetic sciences.* University of Amsterdam.
[28] Weitzman, R. (1992).Vowel Categorization and the Critical Band. *Language & Speech*, 35, 115-125.

Nasim Esfandiari, PhD candidate of Linguistics, University of Isfahan, Iran.

Batool Alinezhad associate professor in Linguistics, University of Isfahan, Iran.

Adel Rafiei, assistant professor in Linguistics, University of Isfahan, Iran.