THE PUBLISHING HOUSE
OF THE ROMANIAN ACADEMY

PROCEEDINGS OF THE ROMANIAN ACADEMY, Series A,
Volume 12, Number 3/2011, pp. 257-265

MICRO-ACOUSTICAL ANALYSIS AND CLASSIFICATION OF THE FRICATIVE /*V*/

Alina UNTU (HULEA)[1], Horia-Nicolai TEODORESCU[1,2]

[1] "Gheorghe Asachi" Technical University of Iasi, Iasi, Romania
[2]Institute for Computer Science Iasi Branch of the Romanian Academy, Iasi, Romania
E-mail: hteodor@etti.tuiasi.ro

A thorough analysis of the statistical acoustical features of the pronunciations of the /*v*/ consonant in the Romanian language, in the _/*v*/V contexts, that is in words starting with /*v*/ followed by a vowel reveal several different pronunciations. We determine that beyond the voiced / non-voiced/ fluctuating (voiced – unvoiced) classes, the /*v*/ in the _/*v*/V contexts fall into distinct subclasses. The results show that, while at the perceptive level Romanian speakers do not distinguish different /*v*/ pronunciations, at the acoustical level, the /*v*/ in Romanian spoken language has several forms.

Key words: Phonetics, Formants, K-means, Prototypes, Hierarchical clustering, Cluster validity.

1. INTRODUCTION

The fricative /*v*/ has variable pronunciations across various languages. In Greek and Japanese, /*v*/ is not present, but the consonant /*b*/ is pronounced close to Romanian /*v*/. The Romanian /*v*/ is often slipping toward the /*f*/ and sometimes toward /*b*/, similarly to the 'be corta' in Spanish. Context-dependent and slightly erroneous pronunciations of /*v*/ in Romanian as induced by various deficiencies of the dental state may produce pronunciation patterns that make the /*v*/ be recognized as /*v*/, as /*f*/, or as /*b*/. The Romanian /*v*/ has various instantiations that partly resemble to Spanish /*v*/ - 'be corta' and to German /*f*/. The rationale of the research is as follows: Because /*v*/ is a voiced, dento-labial fricative, produced by forcing air through the lower lip against the upper teeth, pathologies of one of these articulators may induce modifications in the pronunciation of /*v*/. It is known that the /*v*/ consonant may be produced with a pitch (vibrations of the vocal folds), but the pitch may be unstable. This instability may favor several clusters of /*v*/ pronunciations, leading to variants of /*v*/. Therefore, we investigate the pronunciations patterns of the /*v*/ consonant which are present at the formantic level, taking into account the presence/absence of the pitch (denoted by F0) and statistical parameters of the first formant, F1. We focus on the words containing /*v*/ in the _/*v*/V context, where V denotes a vowel, and _ the beginning of the word. Three main /*v*/ types occur, namely with F0 present, absent, or fluctuating; each of these classes may have one or two subclasses. The mixture of the various types of /*v*/ pronunciations may differ from speaker to speaker, from one pronunciation to another and from context to context. The proportions of the various types of /*v*/ in speech may be sensitive to the dental pathologies, emotional state, and fatigue; hence the relevance of this research in several application fields.

In section two, we describe the speech database and signal processing. The sections three and four present the /*v*/ clustering task and techniques, and the cluster validity indices. Section five consists in the analysis of the F1 distributions. The last two sections include results, their discussion, and conclusions.

2. DATABASE AND DATA PREPROCESSING

Our study is based on a gnatophonic sounds corpus which is composed of recordings belonging to 29 speakers, 12 female speakers and 17 male speakers, all native Romanians, the majority from the NE Romania (Moldavian region), with normal denture and with various dental pathologies. A number of 19 recordings are available on the "Romanian spoken language" site [1,2], while the others are protected (reserved). The corpus was introduced in [3] and its developments were reported in [4,5]. It contains recordings with words including /*v*/, /*f*/, /*s*/, /*sh*/, /*z*/ and /*j*/ consonants in CV (consonant-vowel) and VCV (vowel-consonant-consonant) contexts (V = /*a*/, /*e*/, /*i*/, /*o*/, /*u*/, /*â*/). Speakers spelled between 8 and 21 words containing the /*v*/ consonant, one, two or three times; there are 299 /*v*/ pronunciations for female speakers and 338 /*v*/ pronunciations for male speakers. The recording and documentation methodologies, as well as the database organization are extensively described in [4] and [5]. All the recorded .wav files were filtered for noise removing with a band-pass 70–7 000 Hz filter, set with GoldWave™ tool. The original .wav files were separated into sub-files containing words with /*v*/, /*f*/, /*s*/, /*ş*/, /*j*/, /*z*/ consonants, in CV or VCV context. The recordings were manually annotated and segmented with the Praat™ application, taking into account the auditory perception, the presence/absence of the fundamental frequency (F0), and the formantic patterns [6]. We used a Praat™ script for formants (F0-F4) and time extraction of the recorded words.

3. THE COMPONENTS OF THE /V/ CLUSTERING

For the /*v*/ classification, we performed the clustering according to the steps proposed in [7]: pattern representation, which includes feature extraction and selection; definition of a similarity measure between patterns; clustering or grouping using clustering techniques; data abstraction (if needed); assessment of output (if needed). A pattern, according to [8], is a feature vector *p* of *d* measurements, in our case $p = (F0, F1, F2, F3, F4, t)$, where F0–F4 are the /*v*/ consonant formants, *t* is the time and *d* is the dimensionality of the pattern. The components of the *v* vector are named features. For the /*v*/ classification we used a pattern set $P = \{v_1, v_2, ...v_n\}$, which in the clustering process is viewed as a $n \times d$ pattern matrix.

While the available features are the values of the formants F0-F4 and the consonant duration, we focused on the F0 and F1 formants only. Following [9], we used in the initial step two types of features: a quantitative feature represented by continuous values (F1) and a qualitative feature represented by ordinal values obtained after the evaluation of the presence / absence of the F0. The first features we selected are the first and the third quartiles values of the F1 formant (Q1, Q3) of all /*v*/ consonants. The first and third quartiles are the values corresponding to 25% and respectively 75% in the F1 formant distribution. We chose these quartiles because they frame the median value and because they provide some rough information about the standard deviation of F1. The second feature we selected is the percentage of time the pitch is present, according to the method already described in [5]. Three qualitative values are assigned to F0: present, absent and fluctuating; we represented these qualitative degrees by the numbers 10, 0, and -10. The feature extraction process produced three quantitative features represented by continuous variables: F0, Q1 and Q3. We applied the Euclidian distance as proximity measure [8,10], as we done in [5] and [11].

To simplify the classification, we *a priori* separated the /*v*/ pronunciations into three main classes: $\mathrm{F0} = 10$, $\mathrm{F0} = 0$ and $\mathrm{F0} = -10$, for both male and female speakers. Then we applied the clustering methods for each of these main classes. The feature vector used for clustering inside the main classes is $v = \{Q1, Q3\}$. We created .txt files which contain the Q1 and Q3 values, separately for each F0 class and for each gender, necessary for /*v*/ consonant classification. Three .txt files resulted for each gender: one for $\mathrm{F0} = -10$, one for $\mathrm{F0} = 10$ and another for $\mathrm{F0} = 0$. For each file, the set of patterns was organized in a $n \times d$ matrix, where *n* is the number of /*v*/ pronunciations with F0 present / absent or fluctuating and $d = 2$ is the dimension of the *v* vector. The last three steps of the clustering task are described in the subsequent sections, namely the clustering techniques in Section 4 and the data abstraction and the output assessment in Section 6.

4. CLUSTERING

Clustering of the /*v*/ pronunciations was performed by both partitional and hierarchical methods. The reason for applying different clustering types was to obtain a valid classification of the /*v*/ pronunciations. The input of the clustering task was composed by three .txt files for each gender, which contain the patterns sets for the F0 main classes. These files were processed with Matlab software. The code for clustering algorithms was written by us, as well as extra code lines for the subclasses representation. For hierarchical clustering we used the Matlab Statistic Toolbox for dendrograms representation.

4.1. Partitional algorithms

Prototypes clustering (A1). This classification technique is based on randomly choosing of one prototype for each supposed subclass. A prototype is an element which by visual inspection is assumed to belong to a class and represents the centroid of that class. For example, for $\mathrm{F0}=-10$, for the female speakers, having the unclassified /v/ types represented in Fig. 1 in Q1-Q3 plane, we randomly chose the prototypes $p1=(313;789)$ for the first subclass and $p2=(332;1688)$ for the second subclass. For the female speakers, for $\mathrm{F0}=10$ we randomly chose the prototypes $p3=(271;582)$ and $p4=(581;1726)$ and for $\mathrm{F0}=0$ the prototypes $p5=(265;\ 317)$ and $p6=(1458;1823)$. For the male speakers we randomly chose the following prototypes: (i) for $\mathrm{F0}=-10$, the prototypes $p1=(340;749)$ and $p2=(840;1367)$; (ii) for $\mathrm{F0}=10$, the prototypes $p3=(250;800)$ and $p4=(800;1200)$; (iii) for $\mathrm{F0}=0$, the prototypes $p5=(300;500)$ and $p6=(1078;1259)$.

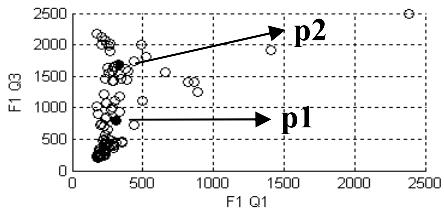

Fig. 1 – The graphical representation of the /v/ types in Q1-Q3 plane, for $\mathrm{F0}=-10$, for 12 female speakers, before prototype-based classification.

The Q1 and Q3 values for each F0 type are stored in three .txt files for each gender: F1.txt $(\mathrm{F0}=-10)$, F2.txt $(\mathrm{F0}=0)$ and F3.txt $(\mathrm{F0}=10)$ for the female speakers, respectively M1.txt $(\mathrm{F0}=-10)$, M2.txt $(\mathrm{F0}=0)$ and M3.txt $(\mathrm{F0}=10)$ for the male speakers. They are organized in n by m arrays, where n is the number of Q1 and Q3 values and $m=2$ represents the number of the F1 quartiles took into account for the /v/ classification. The names of the arrays are the same with the names of the .txt files. For describing the algorithm we explain only the case with **F1** array.

Having the two prototypes $p1$ and $p2$ for $\mathrm{F0}=-10$, for the female speakers, we computed the Euclidian distances between each of the two prototypes and all the **F1** array elements, $\mathrm{F1}=[\mathrm{f}_i]$, $f_i=(\mathrm{Q1}_i,\mathrm{Q3}_i)$, $i=\overline{1{:}n}$. The Euclidian distance is a similarity measure used to evaluate the proximity of objects in two or three-dimensional space, and it is known that works well when a data set has compact or isolated clusters. These distances were stored in two vectors: $\mathbf{B}[\mathrm{d}(f_i,p1)]$ and $\mathbf{D}[\mathrm{d}(f_i,p2)]$. The vector **B** contains the distances between all **F1** array elements and $p1$ prototype, while the vector **D** contains the distances between all **F1** array elements and $p2$ prototype. We separated the **F1** array elements as follows: if $d(f_i,p1)<d(f_i,p2)$, then f_i belongs to the subclass C1; else, if $d(f_i,p2)<d(f_i,p1)$, then f_i belongs to the subclass C2. Next, we defined two vectors, **C11** and **C12**, as follows. The vector **C11** contains the positions of the elements (line indices) from **F1** array for which the corresponding distances from **B** vector are lower than the corresponding distances from **D** vector. The vector **C12** contains the positions of the elements (line indices) from **F1** array for which the corresponding distances from **D** vector are lower than the corresponding distances from **B** vector. Having the **C11** and **C12** vectors we returned two arrays which contain the elements of the two subclasses. The classification results after performing this type of algorithm, for both male and female speakers are presented in Section 6.

K-means algorithm (A2). The second classification technique applied for the /v/ subclasses separation is the k-means algorithm. For performing this classification, a priori knowledge on the number of clusters is needed. By visual inspection of the Q1-Q3 graphics for all F0 classes, we assumed that we have two clusters for each F0 class ($k=2$, where k is the number of clusters) [12]. We exemplify the case for $\mathrm{F0}=-10$, for female speakers. The algorithm main steps are:

- Randomly choose a centroid for each F0 subclass. For the subclass C1, denote by $c1^{[0]}$ the prototype and for the subclass C2, denote by $c2^{[0]}$ the prototype.
- Compute the distances between the chosen centroids and all the **F1** array elements.
- Split the **F1** array elements in two subclasses according to: if $d(c1^{[0]},f_i)<d(c2^{[0]},f_i)$, then f_i belongs to the subclass C1; else if $d(c2^{[0]},f_i)<d(c1^{[0]},f_i)$, then f_i belongs to the subclass C2.

- Compute the new values of the centroids ($c1^{[1]}$, $c2^{[1]}$) of the C1 and C2 subclasses. Compute the new distances between the new centroids and all the **F1** array elements.
- Repeat until the centroids stop moving [8, 13].

We applied the *k*-means algorithm only for $F0=-10$ and $F0=10$ classes, because for $F0=0$ we have too few elements. The number of iterations needed for centroids converge were four respectively two, for female and male speakers for $F0=-10$, and six respectively twelve for female and male speakers for $F0=10$. The classification results obtained with the k-means algorithm will be presented in Section 6.

4.2. Hierarchical clustering (A3)

According to the hierarchical clustering classification technique, we do not need to *a priori* know the number of clusters. The hierarchical clustering algorithms, single linkage (**A3a**), average linkage (**A3b**), complete linkage (**A3c**), and weighted linkage (**A3d**) create clusters recursively and differ with respect to how the distance between clusters is computed. They merge smaller cluster into larger ones or split larger clusters into smaller ones [8,14]. In the single linkage algorithm or the Nearest Neighbor Method the distance between two clusters is computed as the minimum distance between all element pairs from one cluster to another. In the other three algorithms the computation of the distance between two clusters is based on the maximum distance (complete linkage), average between minimum and maximum distance (average linkage) and weighted average between minimum and maximum distance respectively (weighted linkage). It is known that single linkage method is susceptible to chaining effects more than the other hierarchical clustering methods, which means that single clusters are added one at a time. The results show that this method is not useful for the /*v*/ class separation.

In the hierarchical clustering, each element is considered to constitute a cluster. The algorithm for the single and complete linkage methods follows the steps: (i) Compute the *n* by *n* array which contains the Euclidian distances between all elements pairs from **F1**, **F2**, **F3**, **M1**, **M2** and **M3** arrays ; (ii) Search the nearest two elements (single linkage) or the farthest two elements (complete linkage) and group them in a cluster; (iii) Compute the distances between the new cluster and the other elements; (iv) Repeat the above two steps until obtaining a single cluster.

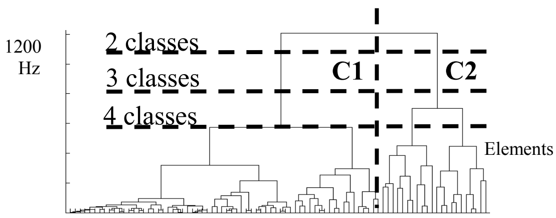

Fig. 2 – The dendrogram for $F0=-10$, average linkage method, male speakers.

For the other two methods the algorithm is the similar, differing only in the way the distances between clusters are computed. A hierarchical clustering algorithm yields a dendrogram – a pictorial description of the hierarchical linkages between elements and the levels where they are grouped. An example of dendrogram for $F0=-10$, average linkage method, for the male speakers is given in Fig. 2, where on the *Ox* and *Oy* axis are the elements respectively the distance in Hz between the dendrogram levels.

As Fig. 2 shows, considering three levels of the dendrogram separation, the average linkage method of the hierarchical clustering may separate the /*v*/ types in two, three, respectively four subclasses. For the classification based on hierarchical clustering we obtained the dendrograms using Matlab Statistics Toolbox. The dendrogram elements shown in Fig. 2 represent the line indices of the **M1** array which contain the Q1 and Q3 values for $F0=-10$, for the male speakers. We separated the **M1** array elements in two subclasses taking into account the first partition level of the related dendrogram; then we plot them in the Q1-Q3 plane. The results of the /*v*/ classification based on hierarchical clustering are given in Section 6.

4.3. Cluster validity measurement indices

To justify that there are two /*v*/ subclasses for each F0 type, in case of *k*-means clustering we used the centroid distance which measures the heterogeneity of the clusters merged at any given step to form a new cluster, and is computed as the Euclidian distance between the centroids of the two clusters. If the two clusters are less heterogeneous, then this distance will be small; otherwise it will be large [10,12]. The iteration process stops when the centroids of the two subclasses stop significantly moving. This occurs in our experiments after two to six iterations.

The results obtained by *k*-means clustering show that there are two subclasses for each F0 type, for both male and female speakers. To validate this hypothesis we computed the asymmetry coefficient SI which gives information about the existence of the two subclasses taking into account the condition that if the asymmetry coefficient SI is greater than 10% then we have a clear separation of the subclasses, $SI = |d(c1,c2)| / \min(|d1|,|d2|)$. In this formula, $d(c1,c2)$ is the Euclidian distance between the two centroids $c1 = (Q1_1, Q3_1)$ and $c2 = (Q1_2, Q3_2)$ after they stop moving, $d(c1,c2) = \sqrt{(Q1_1 - Q1_2)^2 + (Q3_1 - Q3_2)^2}$. The distances $d1$ and $d2$ are the modules of the position vectors of the two centroids, $d1 = \sqrt{Q1_1^2 + Q3_1^2}$, $d2 = \sqrt{Q1_2^2 + Q3_2^2}$. The SI values determined for $\text{F0} = 10$ and $\text{F0} = -10$, for male and female speakers are 1.28 / 2.09 respectively 1.79 / 2.33. These values show that for both male and female speakers the SI coefficients have values greater than 10%, indicating that there are two different subclasses, one for $\text{F0} = 10$ and the other for $\text{F0} = -10$.

For hierarchical clustering, to justify that the cluster solution is valid we applied two validity indices R-Squared and RMSSDT (Root mean squared standard deviation). RMSSDT is a validity index measuring the homogeneity of the clusters. If its value is low then we have a good separation of the clusters. The R-Squared index measures the heterogeneity between clusters. Its value ranges between 0 and 1. The zero value means that the clusters are the same and the 1 value means that the clusters are different [14]. The obtained values for these indices for each method of the hierarchical clustering, for $\text{F0} = 10$ for the male speakers, and for $\text{F0} = -10$ and $\text{F0} = 10$ for the female speakers, are given in Table 1. For $\text{F0} = -10$, for the male speakers all classification methods gave the same separation of the /v/ subclasses. For this reason, in this case we did not compute the validity indices. The values given in Table 1 are rounded values.

Table 1

The validity indices RMSSDT and R-Squared (RS) for **A3b**, **A3c**, **A3d** methods, for the male and female speakers

	Female speakers				**Male speakers**	
	F0 = -10		**F0 = 10**		**F0 = 10**	
Method	*RMS-SDT*	*RS*	*RMS-SDT*	*RS*	*RMS-SDT*	*RS*
A3b	31	0,98	39	0,89	42	0,89
A3c	27	0,95	31	0,98	42	0,96
A3d	43	1,00	42	0,97	45	0,96

The values of the RS index in Table 1 show that for both male and female speakers the results are closest to 1, meaning that we have two heterogeneous subclasses. The lowest RS index is for **A3b** and $\text{F0} = 10$, for both male and female speakers. The method A3b produces different separations of the /v/ subclasses. To decide which method is the best we have to consider simultaneously both the RS and RMSSDT indices. For $\text{F0} = -10$, for female speakers, for **A3d** method, both the RS and RMSSDT indices are the highest, meaning that in this case the two subclasses are less homogeneous. **A3b** and **A3c** methods have close indices values and give the same separation. For both male and female speakers, for $\text{F0} = 10$, the methods **A3c** and **A3d** have close RS index values. The RMSSDT index is different only for the female speakers. These two methods give the same subclass separation for both genders.

5. ANALYSIS OF THE F1 FORMANT DISTRIBUTION FOR THE /V/ SUBCLASSES

Our analysis is based on the quartiles of the distributions of the formants. Namely, for determining the /v/ classes, we computed the second quartile of the F1 (Q2) which represents the median value situated at the 50% of the F1 distribution. Computing the asymmetry coefficient η, we characterized the symmetry of the F1 distribution taking into account the following two rules: (i) If $\eta \approx 1$, then the F1 distribution is symmetrical; (ii) If $\eta \neq 1$, then the F1 distribution is asymmetrical and $\eta = \left(\overline{\eta_k}\right)$, $\eta_k = (Q1_k + Q2_k)/(2Q2_k)$, where *k* is the number of elements from a subclass. For each /v/ subclass we stored the values for Q1, Q2 and

Q3 in an Excel file. First we computed the η_k coefficient for each element of a subclass and then we computed the average of the η_k coefficients of that subclass. The asymmetry coefficients obtained for each /*v*/ subclass, for both genders are given in Table 2.

Table 2

The η coefficients for the /*v*/ subclasses

Subclass	C1 F	C1 M	C2 F	C2 M	C3 F	C3 M	C4 F	C4 M	C5 F	C5 M	C6 F	C6 M
η	1.27	1.14	**1.69**	0.98	1.16	1.14	**1.68**	1.25	1.03	1.14	0.95	0.98

Table 2 shows that the F1 distribution of the /*v*/ subclasses is relatively symmetrical for all /*v*/ classes for the male speakers, while for the female speakers it tends to be asymmetrical for the C2 and C4 subclasses.

6. CLASSIFICATION RESULTS AND DISCUSSION

The aim of the paper was to validate the /*v*/ classes previously obtained for 19 speakers by increasing the number of speakers from 19 to 29. All classification methods gave similar results excepting the case for the male speakers, for $F0 = 10$ where complete and weighted linkage methods of the hierarchical clustering and prototypes classification gave a separation of the /*v*/ classes and average linkage and *k*-means algorithm produced a different class separation. We denote the classification methods as: **A1** prototypes classification; **A2** *k*-means classification; **A3** hierarchical clustering; **A3a** single linkage method; **A3b** average linkage method; **A3c** for complete linkage method; **A3d** for weighted linkage.

The hierarchical clustering based on single linkage method (**A3a**) is useful only for the separation of the /*v*/ subclasses for $F0 = -10$, male speakers and $F0 = 0$ for both male and female speakers. For the other cases, using this method, the classification is affected by a chaining effect which does not allow us to separate the /*v*/ subclasses. For female speakers, for $F0 = -10$ the **A1**, **A3b** and **A3c** classification methods gave the same results of the main class separation, which are presented in the Fig. 3a, which shows that the objects in C1 subclass have Q3 values lower than the objects in C2 subclass, moreover never have Q1 larger than 497 Hz. We also performed the classification without the two outliers with Q1 larger than 1000 Hz and we obtained the same classification results, which mean that the supposed outliers do not have any influence on the classification. The **A2** and **A3d** classification methods gave similar results with the other algorithms excepting 2/66 (2 out of 66) respectively 10/66 elements from C1 subclass, which are incorporated in the C2 subclass. Fig. 3c shows the distributions for F1, for two classes plot in Fig. 4.

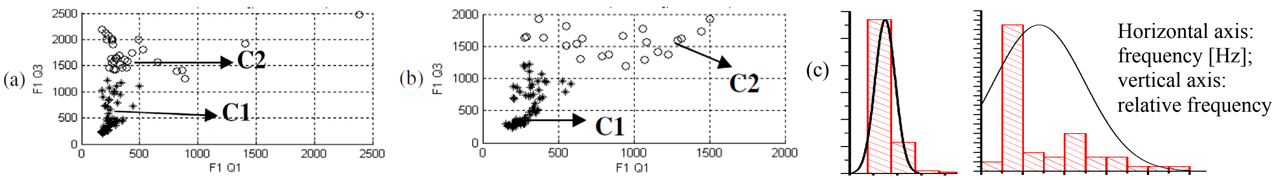

Fig. 3 – The resulted /*v*/ subclasses C1 and C2 for $F0 = -10$ obtained with **A1**, **A3b** and **A3c** classification methods, female speakers (a) and with **A1**, **A2**, **A3a**, **A3b**, **A3c** and **A3d** classification methods, male speakers (b). Right (c) – qualitative representation of the histograms for F1, classes determined by complete linkage, Fem., F0=+10, C3, and for Fem., F0=+10, C4 (see Fig. 4).

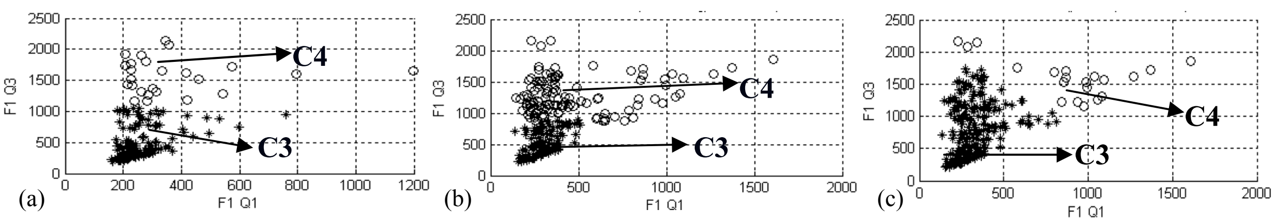

Fig. 4 – The resulted /*v*/ subclasses C3 and C4, for F0 =10, obtained with **A3c** and **A3d** classification methods, female speakers (a), male speakers (b) and with **A1** classification method for male speakers (c).

For male speakers and for $F0 = -10$, all the classification methods gave the same results. The obtained subclasses are presented in Fig. 3b which shows that the objects in C1 subclass have Q3 values lower than the objects in C2 subclass, moreover never have Q1 larger than 587 Hz. For the female speakers, for $F0 = 10$ the **A3c** and **A3d** classification methods gave the same results, while the **A1**, **A2** and **A3b** methods gave similar results excepting 2/171, 23/171 respectively 9/171 elements from C1 subclass that are incorporated in the C2 subclass. The classification /v/ types obtained with the methods that gave the same results are shown in Fig. 4a, which shows that the objects in the subclass C3 have Q3 values lower than the objects in the C4 subclass and Q1 has similar ranges. We also performed the classification without the three outliers with Q1 larger than 700 Hz and we obtained the same classification results, which mean that the supposed outliers do not have any influence on the classification.

For male speakers, for $F0 = 10$, the **A3c** and **A3d** classification methods gave the same results, while the **A2** provide similar results excepting 2 out of 116 elements from C4 subclass which are incorporated in C3 subclass, while the methods **A1** and **A3b** gave a different /v/ class separation – 96 out of 116 respectively 101 out of 116 elements from C4 subclass are incorporated in C3 subclass. The results for the methods having similar outcome, respectively for the methods that differ in outcome are plot in Fig. 4b and c.

Fig. 5 – The resulted /v/ subclasses C5 and C6 obtained with **A1, A3a, A3b, A3c, A3d** classification methods, for $F0 = 0$, female speakers (a) and male speakers (b).

Fig. 4b shows that the objects in C3 have values of Q3 lower than the objects in C4, moreover never have Q1 larger than 511. Fig. 4c shows the classification results obtained with **A1** classification method which is similar with **A3b** method excepting 5/20 elements from C4 subclass from **A1** method which in **A3b** method are included in C3 subclass. Notice in Fig. 3c that the distributions of F1 may significantly differ from Gauss distribution, moreover may significantly differ between classes – which adds to the rationale of using the quartiles Q1 and Q3 as features. For $F0 = 0$, for female and male speakers, the classification results obtained only with **A1**, **A3a**, **A3b**, **A3c** and **A3d** methods are presented in Fig. 5.

We analyzed the Q1 and Q3 distributions of the obtained /v/ classes using the graphical representation named box chart or box plot. This type of graphic displays differences between populations using five-number summaries: minimum value, lower quartile (Q1), median (Q2), upper quartile (Q3), and the maximum value, and gives information about the degree of dispersion and skewness in the data. Examples of the Q1 and Q3 distributions of the obtained /v/ classes for the female speakers are given in Fig. 6.

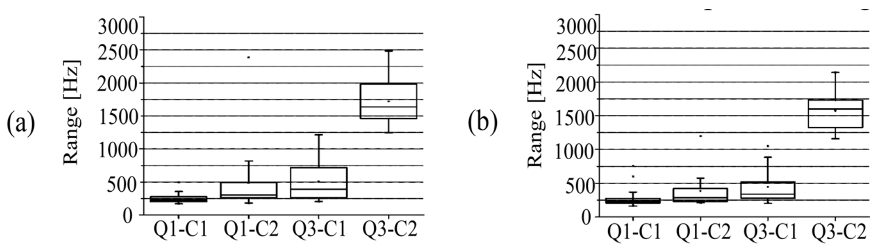

Fig. 6 – The box charts for Q1 and Q3, for female speakers, for $F0 = -10$, C1 and C2 classes (a), and for $F0 = 10$, C3 and C4 classes (b).

Fig. 6 shows that the Q1 and Q3 of the C1 and C3 classes have lower values than the Q1 and Q3 of the C2 and C4 classes and the distributions of the C2 and C4 classes are larger than those for C1 and C3 classes. All the Q1 and Q3 distributions are asymmetrical and the right tail is longer than the left tail meaning that the mass of the distribution is concentrated on the left, excepting the Q3 distribution of the C4 class. The presented box charts show that the obtained /*v*/ classes are distinct.

7. CONCLUSIONS AND FUTURE WORK

Continuing the work presented in [5] on a larger set of speakers and with several refinements, we analyzed and classified the /*v*/ types from the Romanian language taking into account the statistical acoustical features of the /*v*/ pronunciations. We applied three classification techniques; all produced similar results excepting the case of $F0 = 10$, for male speakers, where two methods of the hierarchical clustering algorithm and **A2** classification technique, on one side, and **A1** and **A3b** method on the other side produce different /*v*/ subclasses. The results obtained for 19 speakers in [5] and those obtained for 29 speakers reported in this paper are similar. We can confidently conclude that there are three main /*v*/ types: with F0, without F0 and with F0 fluctuating. For F0 fluctuating and present there are two subclasses and for F0 absent we do not have sufficient elements to decide how many /*v*/ subclasses are. We have too few elements for the main class, $F0 = 0$, for both male and female speakers, to derive statistically valid conclusions. While performing the classification based on prototypes and on hierarchical clustering the results suggested that there may be two subclasses, this conclusion can not be supported at this stage.

While this study has potential applications in phonology, linguistics and dentistry, more work is needed for the validation of the /*v*/ classes by phonetician experts and by persons with high hearing acuity. Also, future work is required for developing a speech database with dentistry pathological cases and with recordings belonging to persons from different regions of the country. We should analyze in the future the relation between the /*v*/ classes and both the dialects and the dental pathologies.

Notice: Part of this paper was presented in a preliminary version as [15].

Authors' contributions: A.U.(H.) has recorded 15 out of 29 subjects, pre-processed the recordings, made the computations, wrote Matlab code and used Matlab toolboxes for determining the /v/ classes. She also made a visual inspection of the Praat™ files and came out with the hypothesis that more than two classes of /*v*/ may exist. **H.N.T.** established the structure of the research and of the paper as well as the recording and pre-processing methodologies, proposed the dendogram interpretation, proposed analyzing the F1 formant distribution, the method of analyzing the clusters validity for the k-means classification etc. He interpreted and checked the results and coordinated all the research steps. Both authors contributed to writing the paper and agreed the final form.

ACKNOWLEDGMENTS

We thank all the subjects who accepted to be recorded and agreed with the use of the recordings in our research. We thank the SRoL team and especially M. Feraru and L. Pistol for the contributions to the "Archive of gnathosonic and gnathophonic sounds". We also thank L. Pistol for providing us the Praat™ scripts used for this study. The work of HNT was performed in relation with and was partly supported by the grant Cognitive Systems and Applications–Priority Research of the Romanian Academy. A. Untu (Hulea)'s work was supported by EURODOC project.

REFERENCES

1. H.N. TEODORESCU, S.M. FERARU, D. TRANDABĂT, M. ZBANCIOC, R. LUCA, A. VERBUTĂ, M. HNATIUC, R. GANEA, O. VORONEANU, L. PISTOL, D. ŞCHEIANU, *Situl web Sunetele limbii române*, 2005–2007; http://www.etc.tuiasi.ro/sibm/romanian_spoken_language/ index.html.
2. H.N. TEODORESCU, D. TRANDABĂT, S.M. FERARU, M. ZBANCIOC, R. LUCA, *A corpus of the sounds in the Romanian spoken language for language-related education*, International Conference on Human and Material Resources in Foreign Language Learning – HMRFLL 2006, Murcia, Spain; C. Perinan Pasqual (Ed.), *Revisiting Language Learning Resources*, Cambridge Scholars Publishing (CSP), UK, Chapter Six, 2007, pp. 73–89.
3. H.N. TEODORESCU, S.M. FERARU, *Micro-corpus de sunete gnatosonice şi gnatofonice*, Proc. Conf. Resurse lingvistice şi instrumente pentru prelucrarea limbii române, Iaşi, România, 2007, pp. 21–30.
4. H.N. TEODORESCU, A. UNTU, *Corpus pentru gnatofonie: protocol, metodologie, adnotare,* Proc. Conf. Resurse Lingvistice şi Instrumente pentru Prelucrarea Limbii Române, Iaşi, România, 2010, pp. 51–60.

5. A. UNTU, H.N. TEODORESCU, *Pattern analysis of /v/ pronunciations in _/v/V contexts*, Proc. of The Eighth IASTED Int. Conf. on Signal Processing, Pattern Recognition, and Applications SPPRA 2011, Innsbruck, Febr. 16–18, 2011, pp. 253–260.
6. P.P.G., BOERSMA, *Praat, A system for doing phonetics by computer*, Glot International, **5**, *9/10*, pp. 341–345, 2002; http://www.fon.hum.uva.nl/praat/, accessing date: 1.04.2010.
7. A.K. JAIN, R.C. DUBES, *Algorithms for clustering data*, Prentice-Hall advanced reference series, Inc., Upper Saddle River, NJ, 1988.
8. A.K. JAIN, M.N. MURTY, P.J. FLYNN, *Data clustering: A review*, ACM Computing Surveys, **31**, 1999.
9. K.C. GOWDA, E. DIDAY, *Symbolic clustering using a new dissimilarity measure*, IEEE Trans. Syst. Man Cybern., **22**, 1992, pp. 368–378.
10. S. SHARMA, *Applied multivariate techniques*, John Wiley & Sons, Inc., 1996.
11. H.N. TEODORESCU, S.M. FERARU, *Classification in gnathophonics - preliminary results*, Proc. Second Int. Symposium on Electrical and Electronics Engineering, ISEEE, Galați, România, pp. 525–530, 2008.
12. S. SHARMA, A. KUMAR, *Cluster analysis and factor analysis*, In *The Handbook of Market Research: Dos and Don'ts*, R. Grover and M. Vriens (Eds.), 2005 (Sage Publications).
13. A. MOORE, *K-means and hierarchical clustering – Tutorial slides*; http://www-2.cs.cmu.edu/~awm/tutorials/kmeans.html.
14. F. KOVACS, C. LEGANY, A. BABOS, *Cluster validity measurement techniques*, The 6th Int. Symposium of Hungarian Researchers on Computational Intelligence, Nov. 18–19, Budapest, Hungary, 2005.
15. A. UNTU, H.N. TEODORESCU, *Classification of the pronunciations of the /v/ consonant in the context _/v/V in the Romanian language*, Proc. IEEE Conf. SPED 2011 (CD), Brașov, România, 18–19 May 2011.

Received May 31, 2011