The Analysis of the FCM and WKNN Algorithms Performance for the Emotional Corpus SROL

Marius ZBANCIOC[1,2], Silvia Monica FERARU[1]
[1]*Institute of Computer Science" Romanian Academy of Iasi, 700054, Romania*
[2]*Technical University "Gheorghe Asachi" of Iasi, 700050, Romania*
zmarius@etti.tuiasi.ro; monica.feraru@iit.academiaromana-iasi.ro

***Abstract*—The purpose of this research is to find a set of relevant parameters for the emotion recognition. In this study we used the recordings from the emotion database SROL which is part of the project "Voiced Sounds of Romanian Language". The database was validated by human listeners. The recognition accuracy of the correct expressed emotion (neutral tone, joy, fury and sadness) for the entire database was 63.97%. We used for the classification of input data the Recurrent Fuzzy C-Means (FCM) and WKNN algorithms. We compared the cluster position with the statistical parameters extracted from vowels in order to establish the relevance of each parameter in the recognition of the emotions. For the extracted parameters for each vowel (mean, median and standard deviation of fundamental frequency - F0 and F1-F4 formants, jitter, and shimmer) the FCM algorithm gave satisfactory results in the phonemes recognition, but not to the emotions. For this reason we used WKNN algorithm in classification, which provided the errors around 20-30% comparing with FCM algorithm when the classification errors are around 40-50%.**

***Index Terms*—emotional speech database, FCM and WKNN algorithm, recurrent coefficient, statistical parameters.**

I. INTRODUCTION

Many researchers want to develop and to improve automatic classification techniques. In the literature there are many studies regarding the vocal parameters in order to recognize the emotional states [1]. One of these parameters is fundamental frequency (F0), and even if there are lots of automatically extracting methods for this prosodic information, we can not say yet that it was found a complete algorithm for this problem [2]. From this parameter can be constructed features vectors which include mean of F0, max and min of F0, variance of F0, intensity distributions, and patterns for F0 rising segments [3].

Other important parameters used by the researchers [17] are: the energy, the duration, the formants (F1-F2) and their bandwidths (BW1, BW2). Other spectral features [18] used in the emotions classification are the mean value and standard deviation of 13 Mel Frequency Cepstral Coefficients (MFCC) (mfcc01_mean to mfcc13_mean and mfcc01_std to mfcc13_std).

Ververidis and his colleagues [19] used Bayers classifier and different parameters for emotions recognition using DES database. They used spectral features, pitch features, and intensity (energy) features (a total number of 87 parameters). The recognition rates was: for neutral 56% , for surprise 65%, for happiness 39%, for sadness 72% and for anger 40%.

In [20], the fundamental frequency F0, energy, speaking rate, first three formants (F1, F2, and F3), and their bandwidths (BW1, BW2, and BW3) were used. As a classification methods it were used KNN with a accuracy rate of 55%, neural networks KNN with a accuracy rate of 65% (normal state 55-65%, happiness 60-70%, anger 60-80%, sadness 60-70%, and fear 25-50%.), etc.

New methods or combinations of methods are welcome to improve the current results. The paper [21] proposed different methods for emotions classification as GMM, HMM, weighted Bayesian Classifier, MLP. 80% from the database was for training and 20% was for testing. The results obtained using GMM likelihoods were: for anger 76.7%, for fury 70.7%, for happiness 33.3%, for sad 95.2%, for surprise 59.1% and for neutral 76.4%. Using HMM likelihoods the results were: for anger 79.1%, for fury 85.4%, for happiness 50.0%, for sad 88.1%, for surprise 54.6% and for neutral 76.4%. Combining GMM and HMM likelihoods they obtained: for anger 81.4%, for fury 100%, for happiness 83.3%, for sad 92.9%, for surprise 50.0% and for neutral 90.9%.

The pitch, jitter, the first three formants (FI, F2, F3), the speech duration, the speech energy, the zero crossing and 14 LP coefficients were used as the parameters in the paper [22]. The results obtained were: for happy 91.53%, for sad 66.67%, for anger 18.33%, for disgust 80.00%, for fear 70.00% and for the surprise 85%.

Other parameters used by [23] are: pitch, energy, speed of speech, Mel-Frequency Cepstrum Coefficients (MFCCs) and gender. The recognition rates obtained are for neutral 60.8%, for surprise 59.1%, for happiness 56.4%, for sadness 85.2%, and for anger 75.1%.

In order to improve the results, [24] proposed an algorithm "to combine the decisions from eight hidden Markov model (HMM) classifiers". They used "Beihang University mandarin emotion speech database - BHUES" and analyzed six emotions. The recognition rates for 8 speakers without gender information were: 77.54% for sad, 48.19% for anger, 56.78% for surprise, 54.48% for fear, 53.88% for joy, and 61.11% for disgust.

The extracted parameters for emotion recognition in [25] are: the energy, pitch, linear prediction cepstrum coefficient (LPCC), Mel frequency cepstrum coefficient (MFCC). In

This work was supported by the Romanian Academy in the research priority program "Cognitive Systems".

Digital Object Identifier 10.4316/AECE.2012.03005

the literature, the best accuracy using GMM is 78.77%, for speaker independent recognition 75% and for speaker dependent recognition 89.12%. The accuracy using HMM is 76.12% for the speaker dependent and 64.77% for the speaker independent [26]. Using ANN the results are 51.19% for speaker dependent and 52.87% for speaker independent. The accuracy is 64% in case where it is using the information given by pitch and by the energy, using KNN for the four emotions classification [27]. Using SVM, the accuracy is 75% for the speaker independent and 80% for speaker dependent [26-27].

There is a need now to bring the knowledge about the emotions recognition in the voice, for Romanian language [3-5]. This need is felt in the linguistic domain, as well as science and computer technology, psychology and medicine. A few studies have been made on the Romanian language [11, 12].

Most studies were performed on happiness, anger, sadness, fear, disgust, panic, anxiety states. The recognition rates for a few emotions as 4-6 emotions "are considered successful around 70% and pure chance guess" when there is 16 emotions studied [10]. The focus on this study is on 3 emotions (happiness, sadness, and fury states) and neutral tone. The sentences are from the SRoL database and they contain all the vowels from the Romanian language. The speakers are feminine and masculine persons with ages between 20-30 years, without pathological manifestation and without professional voices.

Among the difficulties encountered in the studies regarding the emotion recognition/classification are the speaker variability, the emotional speech databases which most of them are private and can not be accessed and others.

The emotion recognition can be made using different methods and algorithms. In [6], the researchers "proposed a new entropy-based measure which makes possible a comparison between human labelers and machine classifiers". They obtained a recognition rate around 60% for 4 emotions and with 5 persons which validated the corpus. They said that first the emotions must be classified by humans. It is good to know also that the more similar emotions we have, the results will be more confusing.

K-Nearest Neighbor classifier (kNN) and Fisher's linear classifier [7] were chosen as they are commonly used in the classification [4].

Frank Dellaert and his colleagues explored some standard pattern recognition techniques and compared their performance. In their study they used maximum likelihood Bayes classifier (MLB), Kernel regression (KR) and K-nearest neighbors (KNN). The best results are obtained using a K-nearest neighbor's classifier [8].

In our study we used the algorithm Fuzzy C-Means (FCM) in order to make a classification of the vocal signal according with the emotional states for Romanian language. In addition to previous research, we have included in the set of features the values of jitter and shimmer.

After [9], "the jitter is a measure of period-to-period fluctuations in fundamental frequency and number of voiced frames in the utterance and the shimmer is a measure of the period-to-period variability of the amplitude value".

After [10] "the jitter was obtained by counting the number of changes in sign of the pitch derivative in a window and the shimmer was obtained by counting the number of changes in sign of the intensity derivative in a window and".

II. Emotion Database Validation

The emotional database is a part of the SRoL annotated corpus and contains at this moment 396 sound files (199 with male voices and 197 with female voices) pronounced for neutral tone and three simulated emotions states: joy, sadness and fury. It contains the recordings of 7 phrases pronounced by 25 speakers, 11 of them female and 14 male.

The pronounced sentences are: "Vine mama / Mother is coming", "Aseară/ yesterday evening", "Cine a făcut asta? / Who done that?", "Ai venit iar la mine/ You came back to me", "Omul meu îl lucră / My man done it", "Îţi vei câştiga locul dorit / You will get the desired place", "Oricum îţi poţi câştiga locul dorit / Anyway, you can get the desired place".

Each phrase is pronounced for several times (on average about 4-5 times in the recording file). Thus we can estimate that the total number of sentences is about two thousands.

After the corpus validation made by three people we obtained two type of database: one (named DB100) in which all valuators has indicated the same emotional state and another (DB_validated) in which two of the three valuators (the majority of listeners) have confirmed the same emotion.

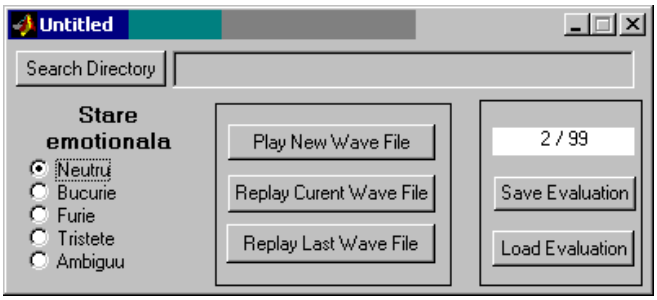

Figure 1 A screenshot of the emotion validation interface

The valuators have the possibility to listen the wav. files how many time they want, the option to return to previous records and the possibility to make the validation in sections. In fig. 1 is exemplify the interface between the valuators and the computer. When a human listener is not sure about the expressed emotion then he can use the "ambiguous" button.

TABLE I. CONFUSION MATRIX FOR FEMALE VOICES VALIDATION (IN PERCENTS)

Female	Neutral	Joy	Fury	Sadness	Ambiguous
Neutral	60.99%	2.13%	4.96%	30.50%	1.42%
Joy	3.85%	80.13%	9.62%	2.56%	3.85%
Fury	4.58%	10.46%	73.20%	8.50%	3.27%
Sadness	12.77%	2.84%	2.13%	78.72%	3.55%

TABLE II. CONFUSION MATRIX FOR MALE VOICES VALIDATION

Male	Neutral	Joy	Fury	Sadness	Ambiguous
Neutral	59.03%	3.47%	6.25%	20.83%	10.42%
Joy	22.00%	54.67%	10.00%	3.33%	10.00%
Fury	24.53%	5.66%	61.01%	3.77%	5.03%
Sadness	39.58%	5.56%	4.17%	43.06%	7.64%

After the validation of the sound recordings from the emotional database SRoL, the recognition accuracy of the emotions (neutral tone, joy, fury and sadness) was 73.43% for female speakers and only 54.6% for male speakers. The average accuracy recognition percent for the entire database

was 63.97%. Most confusion was found between the three emotional states and the neutral tone for the masculine voices (see Table II). For the female voices the biggest confusion was between neutral tone and sadness.

After validation, it was preserved a number of 263 files (66.41%) in DB_validated and 145 files in DB100 (36,61%).

TABLE III. NUMBER OF VOWEL OCCURRENCES FOR FEMALE VOICES (NUMBER OF FILES / NUMBER OF VOWELS)

Vowel	Neutral	Joy	Fury	Sadness
e	4 / 25	10 / 97	8 / 62	8 / 76
i	5 / 32	10 / 104	6 / 72	9 / 72
a	6 / 34	10 / 191	8 / 132	9 / 141
a+	7 / 27	8 / 67	8 / 62	7 / 46
o	5 / 37	6 / 74	8 / 73	7 / 40
u	5 / 50	4 / 36	6 / 39	2 / 21
a-	5 / 29	4 / 44	6 / 53	2 / 29

TABLE IV. NUMBER OF VOWEL OCCURRENCES FOR MALE VOICES (NUMBER OF FILES / NUMBER OF VOWELS)

Vowel	Neutral	Joy	Fury	Sadness
e	6 / 23	4 / 19	4 / 41	3 / 10
i	7 / 26	7 / 55	4 / 44	5 / 25
a	7 / 40	3 / 10	7 / 78	5 / 19
a+	3 / 6	5 / 27	7 / 27	2 / 5
o	2 / 10	4 / 29	6 / 25	4 / 12
u	1 / 8	4 / 37	3 / 20	3 / 15
a-	1 / 12	7 / 40	3 / 23	3 / 21

In Table III and IV is specified the number of occurrences for every vowel depending on the expressed emotion for the DB100 validated database.

III. Feature Vectors And Statistical Parameters

In this study we used only the parameters extracted from the vowels of the Romanian language with Praat software. Each sound file (wav) was annotated manually generating a file with the *TextGrid* extension. In the statistical analysis was used only the "phoneme" annotation level, which specify the duration and the position of each vowel phoneme. Praat is free software for acoustic analysis which can extract the pitch F0, the F1-F4 formants, the pulses and the intensity of the voiced signals.

The pitch is extracted using an analysis window of 40ms and a time step of 10ms in the frequency band [75, 500] Hz with the autocorrelation method, because compared with other methods is more robust, accurate and noise-resistant.

The formants F1-F4 are computed using a window length of 25ms and a time step of 6.25 ms using the LPC coefficients obtained with the Burg algorithm.

The statistical parameters are computed for each vowel (their boundaries were drawn in the annotation files). In the feature vectors are included the mean, the standard deviation and the median of F0 and of the formants F1-F4. The median is the central value of the sorted input vector.

In addition to previous research, we have included in the set of features the values of jitter and shimmer.

Jitter is the average variation of the F0, expressed as the difference between two consecutive pitch period T0=1/F0.

$$Jitter(phoneme) = \frac{1}{N-2}\sum_{k=3}^{N}\left|T_k - T_{k-1}\right|$$

$$= \frac{1}{N-2}\sum_{k=3}^{N}\left|(P_k - P_{k-1}) - (P_k - P_{k-2})\right|$$

$$= \frac{1}{N-2}\sum_{k=3}^{N}\left|P_k - 2P_{k-1} + P_{k-2})\right| \quad (1)$$

where N is the number of pulses corresponding to the duration of a phoneme/vowel. The pulses are used by Praat to compute the pitch time period and we stored these values in a file. The pulse values P_{k-1}, P_k mark the boundaries of each fundamental period extracted from the voice signal s.

Shimmer represents the average variability of the peak-to-peak amplitude between two consecutive pitch periods.

$$Shimmer = \frac{1}{N-2}\sum_{k=3}^{N}\left|20\cdot\log_{10}\frac{A_k}{A_{k-1}}\right| \quad (2)$$

$$= \frac{1}{N-2}\sum_{k=3}^{N}\left|20\log_{10}\frac{\max(s(P_k : P_{k-1})) - \min(s(P_k : P_{k-1}))}{\max(s(P_{k-1} : P_{k-2})) - \min(s(P_{k-1} : P_{k-2}))}\right|$$

where N is the number of pulses Pk and s is the input signal.

The software application computes the classification accuracy of the emotions / phonemes performing the following steps:

1. For each recording from DB100 is determined the boundaries of the phonemes based on the annotation files.

2. For each vowel, the prosody parameters are computed. The correction of the F0 and the formants values which not satisfy certain restrictions [12] are made. The short phonemes are not taken into account.

3. The parameters associated to the phonemes (vowels) are stored in a tree data structure according to the gender of the speaker and to the emotion expressed. It is possible to make an analyze for all speakers or only male/female speakers.

4. The classification algorithms are applied (with unsupervised learning - Fuzzy C-means FCMR recurrent, respectively supervised learning - WKNN)

5. The classification error is determined. For WKNN algorithm, the variation error for different values of k is analyzing.

IV. Recurrent Fuzzy K-Means Algorithm

The clustering algorithms divide the input sets based on their similarity. The most known are the hierarchical algorithms (agglomerative and divisive) and the partitioning algorithms. Among the partitioning algorithms, we mention EM (Expectation Maximization) based on the probabilistic models, QT (quality threshold), algorithms based on graph theory and K-means based on the calculation of Euclidean distance and squared distance error. From K-means algorithm was developed a series of other algorithms including the FGKA genetic algorithms (Fast Genetic K-means Algorithm) and fuzzy techniques such as Fuzzy C-means (FCM).

For the data clustering, we used an improved version of the FCM algorithm which gives a confidence coefficient to each input pattern associated with every center of a cluster/ fuzzy partition. These coefficients are computed with the recurrent function in each new iteration of the algorithm. The FCMR algorithm (fuzzy c-means recurrent) eliminates

one of the major disadvantages of the classical FCM algorithm, related to the fact that the learning is unsupervised and does not know the real number of clusters from the input data set. In this situation, the cluster centers found by FCM not converge to the real centers, which are compensated in the new algorithm FCMR by introducing the recurrent confidence coefficients.

In [13-15] Teodorescu proposed a model of the recurrent fuzzy systems with the justification that people tend to adjust their judgments based on the preliminary results. Thus, the knowledge (rules) that lead to achieve worst results should be associated with the lower coefficients compared with those that achieve better results. Because the recurrent fuzzy systems - Teodorescu model is a generalization of the existing classical systems (Mamdani, Sugeno, etc.), any classical fuzzy system (fuzzy algorithm) in any area of the analysis, can be transformed into a fuzzy model with recurrence to the confidence degree in meta-knowledge / rules [16].

The input data set $\hat{v} = \{x_1, x_2, ..., x_n\}$ are vectors of n features, and the distance between two vectors is computed based on the Euclidean distance, but we can use any kind of distances, as Mahalanobis, Pearson, Manhattan, Chebyshev, Lee, Hamming, etc.. We note with N the size of the input data set $X = \{\hat{v}_k\}$, $k = \overline{1, N}$. The FCM algorithm determines the distance from the each cluster center c_j to each feature vector $\hat{v}_k$, and it associated a membership degree $\mu_{j,k} : R \to [0,1]$. We also note with C the number of clusters/partitions and with U the matrix (of size CxN) which keeps all the membership grades.

The FCMR algorithm pseudocode is the following:

1. The matrix initialization with random numbers $U^{(0)}$ and its normalization, setting the number of clusters $2 \le C \le N$, the error convergence of the algorithm ε and the exponent m applied to the membership degrees. In addition, a matrix of the recurrent confidence coefficients is introduced: $\{CF_{jk}^{(0)} = 1\}$, t=0;

2. A set C of partitions' centers $\{\hat{c}_j{}^*\}$ is computed using a relationship with the recurrent confidence coefficients. The distances $\{d_{jk}{}^*\}$ from each element of the X data set to each center are computed;

3. Based on the distances, the objective function J* and the value of the membership grades for next step $U^{(t+1)}$ are computed;

4. With the new values of the matrix $U^{(t+1)}$ the recurrent coefficients for the t +1 time is updated;

5. If the differences in the objective function $\left|J^{(t+1)} - J^{(t)}\right|$ are above the imposed convergence limit (ε imposed error) and the maximum number of the iterations is not reached, then return to step 2.

By simulation, it can be concluded that the performance of the FCMR proposed algorithm may be improved if the confidence coefficients are not applied from the beginning but with a certain delay after a consistent movement of the cluster's centers has been observed.

Figure 2 shows the situation when we have a real number M=3 clusters and C=2 specified clusters of the partitioning algorithm. The classical FCM algorithm places a center between two clusters while the FCMR algorithm achieve all three clusters' centers and the FCMR algorithm applied with "delay" is more robust and frequently finds only two clusters. The simulations have been made by running the algorithms for 50 times.

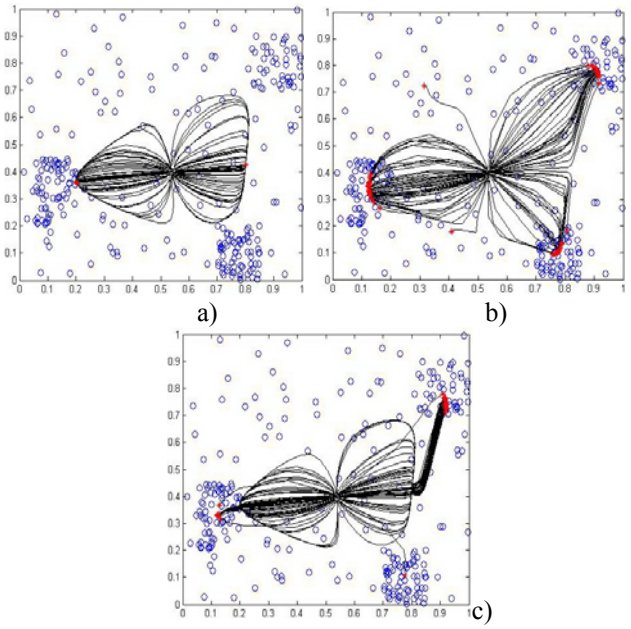

Figure 2.a) FCM classical; b)FCM recurrent; c)FCMR applied with "delay"

The FCMR algorithm partially solves another disadvantage of the classical K-means and FCM algorithms, which can converge only to a local, not global maximum and that the final values of the cluster centers depends on the selected values of the partitions / clusters' centers.

V. Weighted K-Nn Algorithm

The K-means and FCM clustering algorithms have provided satisfactory results for the vocal phonemes recognition, because the distribution of the data in the feature space allows finding the clusters. The statistical parameters associated with the F1 and F2 formants and with F0 - the fundamental frequency was the most important in the vowel recognition. The remaining parameters help to improve the percentage of the classification.

Because of the data overlapping from the feature vectors and of the difficulty of establishing the clusters it was chosen for the emotion recognition, the supervised classification KNN algorithm.

As in the case of the traditional FCM algorithm where were introduced the recurrent coefficients, for the KNN algorithm it was introduced a vector of weights which express for each set of features the "strength" of belonging to a class. If an instance of training set gives good results in classification then it will be assign with a higher weight, respectively if it not helps in classification, it will be assign with a lower weight.

The implemented KNN algorithm made the classification according to the following pseudo-code:

for i = 1 : *number of patterns* from the test data set

- computes the Euclidean distance between the current vector v_i and other vectors from the training data set

$$d(v_i, v_j) = \sqrt{(x_{i1} - x_{j1})^2 + ... + (x_{in} - x_{jn})^2} = \sqrt{\sum_{h=1}^{n} (x_{ih} - x_{jh})^2} \quad (3)$$

- the vectors are sorted by the distance and it is obtained $v^* = \{v_1^*, \ldots, v_N^*\}$
- the nearest k instances $v^*[1:k]$ are stored
- the current vector is classified in the class that contains the most of the k nearest neighbors $v^*[1{:}k]$
- if we have more classes which contain the same number of neighbors, the final decision is taken according to the mean distance to the current vector. Thus the nearest instance will have more weight in the classification process.

end for

For the WKNN algorithm, in the above algorithm has introduced a first stage for computing the weights associated to each feature vector based on the classification performance for the number of k neighbors chosen. In addition to the determining of the associated class to a vector, the final decision is made by computing a weighted sum of the distance between the current vector and each class.

We intend to implement a mechanism to recurrent change of the weights analyzing the classification error obtained to each iteration with the current vector of weights and finally to select an optimal weight vector.

VI. Results And Conclusion

Combining the emotional states ('Joy', Sadness', ,Fury' and ,Neutral') with the gender of the speakers (,Female' and ,Male') are obtained Ns = 8 situations notated with 'JF', JM', 'SF', 'SM', 'FF', 'FM', 'NF', 'NM'. The structure DB where are stored the feature vectors has the dimension $[Ns \times Nv \times Np]$. *Np*=23 is the number of statistical parameters and Nv=7 the number of the vowels (analyzed phonemes).

We compared in the bi-dimensional space of characteristics F1 and F2 the cluster centers (in Fig.4) with the mean values of each vowel (in Fig.3). We have observed that formant space is properly partitioned; the FCMR is robust and places the partition centers in the same areas (Fig. 5). For the n-dimensional space of features {F0, F1-F4, jitter, shimmer} we computed the accuracy of classification of emotions, but results have been around a percentage of 41% (for all speakers), for the database validated by human listeners with an accuracy recognition percent of 63.97%.

From simulations it was observed that higher formants F3-F4 not help to differentiate emotions. By eliminating the superior formants from the features vectors were obtained the same results in classification.

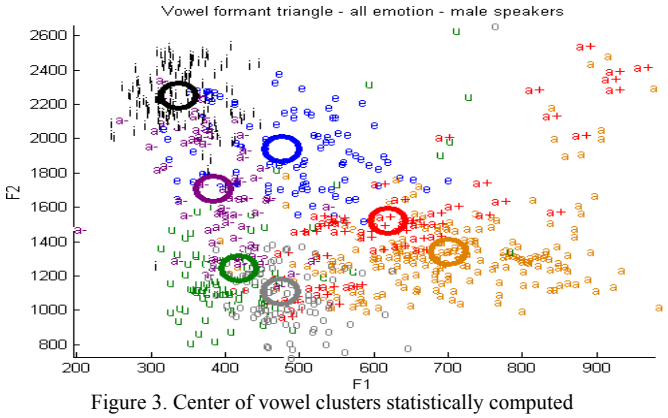

Figure 3. Center of vowel clusters statistically computed

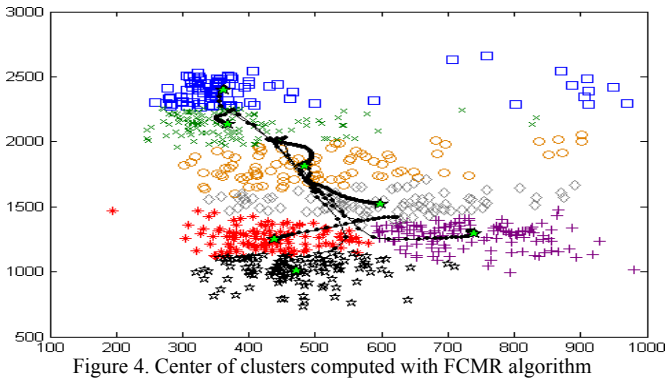
Figure 4. Center of clusters computed with FCMR algorithm

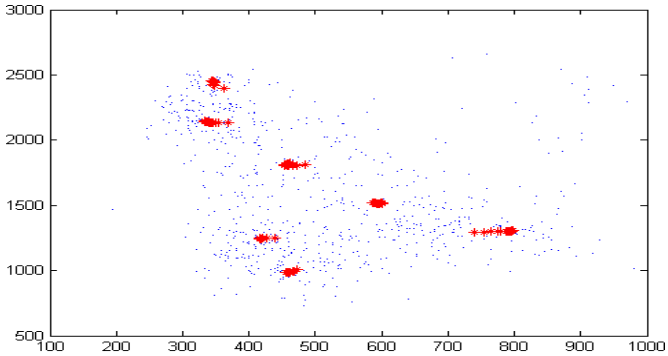
Figure 5. Center of clusters computed with FCMR algorithm over 50 simulations

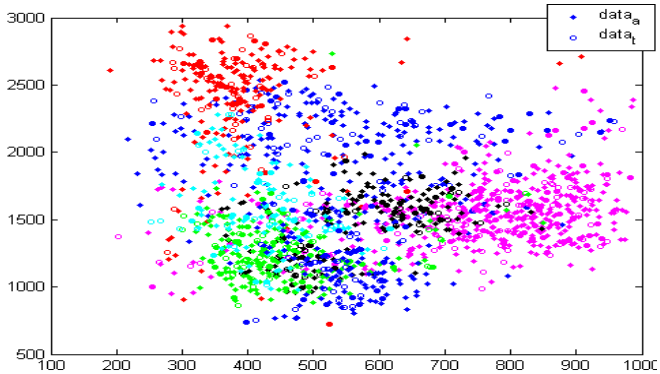

Figure 6. The vowels classification with the KNN algorithm in the F1,F2 features space (classification error - 26%)

For each vowel was extracted the features vectors with the following statistical parameters: 1.name of the phoneme, 2. number of occurrences, 3. mean_F0, 4. std_F0, 5. median_F0, 6. mean_F1, 7. std_F1, 8. median_F1,..., 15. mean_F4, 16.std_F4, 17.median_F4, 18.jitter, 19.shimmer.

For the FCM algorithm, the classification error of the emotion was 48.38% for the parameters (3, 6, 9), 50.58% for (6,9) and 60.10% for all parameters. The difficulties are for

the "ă" vowel, where the error classification of FCM is around 65%. Comparing with FCM algorithm, the KNN algorithm gives better results if the features vectors are complete (they contain all the parameters).

TABLE V. RECOGNITION ERROR OF THE VOWELS FOR FCM AND KNN ALGORITHMS, ALL EMOTIONS-MALE SPEAKERS

Parameters	**6, 9**	**3,6,9**	**all**
FCM	50.58%	48.38%	60.10%
KNN	33.94% (k=37)	28.75% (k=6)	30.28% (k=4)

The KNN algorithm gives better results for a small number of neighbors, k<5 and the error is situated around 20% for female voice and 30% for male voice.

TABLE VI. EMOTION RECOGNITION ERRORS (%) BASED ON THE ANALYZED VOWELS USING FUZZY K-NN AND WKNN ALGORITHMS

Vowels	**e**	**i**	**a**	**a+**	**u**	**a-**	**o**
Fuzzy KNN	34.48	29.79	38.79	25.71	33.77	29.63	24.19
WKNN	32.27	29.95	35.14	22.67	27.19	28.30	29.58

The SRoL corpus contain a sets of software instruments for pitch extraction (using four methods: auto-correlation, cepstral, HPS - Harmonic Product Spectrum and AMDF - Average Magnitude Difference Function) and for the detection of formants using concatenated parts of "smoothed" spectrum. In this paper we used only features extracted from Praat files, but we intend to expand this research using our instruments for pitch detection.

For future research will take into account only the parameters coming from phonemes located in fixed positions, preferably in the center of sentences where the emotion is expressed better. Another approach which will be analyzed consists in extracting parameters not at phoneme level, but for an entire word or phrase.

ACKNOWLEDGMENT

The authors thank all those who contributed with recordings and colleagues that contributed to the building of the website SRoL. The authors acknowledge the support of the Romanian Academy in the framework of the program "Cognitive systems".

REFERENCES

[1] K.R. Scherer, "Vocal communication of emotion: A review of research paradigms", *Speech Communication*, vol. 40, pp. 227-256, 2003.

[2] W. Hess, "Pitch determination of speech signals: algorithms and devices", *Springer-Verlag*, Berlin, Germany 1983.

[3] S. McGilloway, R. Cowie, E. Douglas-Cowie, S. Gielen, M. Westerdijk, S. Stroeve, "Approaching automatic recognition of emotion from voice: a rough enchmark", in *Proc. of the ISCA Workshop on Speech and Emotion*, Belfast, Northern Ireland, pp. 200-205, 2000.

[4] G. Klasmeyer, "An automatic description tool for timecontours and long-term average voice features in large emotional speech databases", in *Proc. of ISCA Workshop on Speech and Emotion*, Belfast, Northern Ireland, pp. 66-71, 2000.

[5] M. Slaney, G. McRoberts, "Baby ears: a recognition system for affective vocalization", in *Proc. of ICASSP*, 1998.

[6] S. Steidl, M. Levit, A. Batliner, E. Noth, H. Niemann, "Of all things the measure is man" automatic classification of emotions and inter-labeler consistency, in *Proc. of ICASSP*, pp. 317-320, 2005.

[7] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Recognition*, 2nd edition. New York, John Wiley & Sons Inc., 2001.

[8] F. Dellaert, Th. Polzin, A. Waibel, "Recognizing emotion in speech", in *Proc. of ICSLP*, vol. 3, pp. 1970 – 1973, 1996.

[9] Xi Li, Jidong Tao, Michael T. Johnson, J. Soltis, A. Savage, Kirsten M. Leong, John D. Newman, "Stress and emotion classification using jitter and shimmer features", *In Proc. of ICASSP*, pp. 1081-1084, 2007.

[10] A. Noam, "Classifying emotions in speech: a comparison of methods", in *Proc. of 7th European Conference on Speech Communication and Technology*, Aalborg, Denmark, pp. 127-130, 2001.

[11] H.N. Teodorescu, M. Zbancioc, M. Feraru, "The analysis of the vowel triangle variation for Romanian language depending on emotional states", in *Proc. of ISSCS Conference*, Romania, ISBN 978-1-4577-0201-3, pp. 331-334, 2011

[12] H.N. Teodorescu, M. Zbancioc, M. Feraru, "Statistical characteristics of the formants of the Romanian vowels in emotional states", in *Proc. of the Int. Conf. on Speech Technology and Human-Computer Dialogue*, Romania, ISBN 978-1-4577-0439-0, pp. 13-22, 2011,http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5940725

[13] H.N Teodorescu, "Recurrent Rules-Based Fuzzy Decision-Making and Control", in *Proc. of WSAS Conference*, Udine, Italy, 2004.

[14] H.N. Teodorescu, "Fuzzy systems with recurrent rules in population and medical models", in *Proc. of the American Conference on Applied Mathematics World Scientific and Engineering Academy and Society* Stevens Point, Wisconsin, USA, ISBN: 978-960-6766-47-3, pp. 343-349, 2008.

[15] H.N Teodorescu, "Fuzzy Systems with Recurrent Rules. A new type of fuzzy systems and applications", *Intelligent Systems*, pag 157-166, Editors: H.N.Teodorescu, Iaşi, România, Ed. Performantica, ISBN 973-7994-85-X, 2004.

[16] M. Zbancioc, "Recurrent fuzzy rules (Teodorescu's fuzzy systems) in economic process modeling", in *Proc. of 15th International Conference on Control Systems and Computer Science*, Bucureşti, România, 2005.

[17] C.M. Lee, S. Narayanan, "Emotion recognition using a data-driven fuzzy inference system", *in Proc. of Eurospeech*, Geneva, , pp. 157-160, 2003.

[18] M. Grimm, K. Kroschel, "Rule-based emotion classification using acoustic features", in *Proc. Int. Conf. on Telemedicine and Multimedia Communication*, 2005.

[19] D. Ververidis, C. Kotropoulos, I. Pitas, "Automatic emotional speech classification", in *Proc. of Internat. Conf. on Acoustics, Speech and Signal Processing*, Montreal, vol. 1, pp. 593–596, 2004.

[20] Valery A. Petrushin, "Emotion recognition in speech signal: experimental study, development, and application", in *Proc. of the Sixth International Conference on Spoken Language Processing* ICSLP 2000.

[21] Dan-Nmg Jiang, LiaHong Cai, "Speech emotion classification with the combination of statistic features and temporal features", *IEEE International Conference on Multimedia and Expo (ICME)*, pp.1967-1970, 2004.

[22] Aishah AM. Razak, Mohd Hafizuddin Mohd Yusof, Ryoichi Komiya, "Towards automatic recognition of emotion in speech", pp.548-551

[23] Kuan-Chieh Huang, Yau-Hwang Kuo, "A novel objective function to optimize neural networks for emotion recognition from speech patterns", in *Proc. of the second World Congress on Nature and Biologically Inspired Computing*, Kitakyushu, Fukuoka, Japan, pp. 413-417, 2010

[24] Liqin Fu, Changjiang Wang, Yongmei Zhang, "A study on influence of gender on speech emotion classification", in *Proc. of 2nd Int. Conference on Signal Processing Systems*, pp. 534-537, 2010.

[25] Ashish B. Ingale, D. S. Chaudhari, "Speech Emotion Recognition", *International Journal of Soft Computing and Engineering* (IJSCE) ISSN: 2231-2307, Volume-2, Issue-1, 2012.

[26] M. E. Ayadi, M. S. Kamel, F. Karray, "Survey On Speech Emotion Recognition: Features, Classification Schemes, And Databases", *Pattern Recognition* vol. 44, pp. 572-587, 2011.

[27] D. Ververidis, C. Kotropoulos, "Emotional speech recognition: resources, features and methods", *Elsevier Speech Communication*, vol. 48, no. 9, pp. 1162-1181, 2006.