**Advances in Intelligent Systems and Technologies**
Proceedings ECIT2010 – 6th European Conference on Intelligent Systems and Technologies
Iasi, Romania, October 07-09, 2010

# Methodology and Instrument for Statistic Analysis of the Romanian Language for a Population Selection

Ioan Păvăloi, Horia-Nicolai Teodorescu, Monica Feraru

*Romanian Academy, Iaşi Branch, Institute of Computer Science, Romania*

**Abstract**. *We describe a methodology and a program created for statistic analysis of formantic characteristics of vowels, using files sub-classes selected according to certain characteristics given by user. The program may by used to analyze the differences which occur to the vowel level of the Romanian language for the formants values, for a population selection compared with the whole population. First are described the methodology and the tool used. We briefly present data about variability coefficient (asymmetry) of formants for several words and short sentences. This application allows a refinement in the analysis of voice emotivity, at the level of vocalic sounds.*

## 1. Introduction

Emotional expressiveness analysis [1], [6] of theoretical methods and algorithms for identification of emotional states [3], [4], [8] have many practical [2], [5] and scientific uses in the field of computer science, in the linguistic field, in psychology, etc. But these analyses are laborious and require vast resources when used for large corpuses of voice recordings.

Corpus SRoL [7] that we used in the analysis is a free corpus which can be found at the address (http://www.etc.tuiasi.ro/sibm/romanian_spoken_language/index.html). The recordings are based on two types of protocols:

- recording protocol, containing information about the noise, the microphone used, the soundboard and the corresponded drivers;

- documentation protocol, which contains a questionnaire regarding the healthy state of the speaker. Speaker profile includes linguistic, ethnic, medical, educational, professional information about the speaker.

Recorded sentences are: *Mother is coming* (*Vine mama*, in Romanian), *Who did that*? (*Cine a făcut asta*?, in Romanian), *Last night* (*Aseară*, in Romanian), and *You came back to me* (*Ai venit iar la mine*, in Romanian). Were analyzed emotional states of happiness, sadness, fury and the neutral tone.

The research aim is to present the methodology for organizing and processing data in order to perform statistical analyses and data mining. A good organization of the data allows the user more flexibility in accessing information, validation of existing data, high speed in data processing, moreover development of new applications in the field of spoken language processing. Unitary approach of a large data amount offers an efficient way for achieving complex statistic processing. In sections two and three are presented methodology for organizing / processing of data and the program WinCollection. In section four we'll show preliminary results. In the last section we derive conclusions and some requirements for future work.

## 2. The methodology for input data organization and validation

The aim of the methodology is to allow an efficient processing, error detection and fusion of raw data files, consisting in wav and text files. Intermediary text files that contain the values of F0 and F1-F3 formant are produced from .wav files. The output files are generated after data validation and error correction. Using this methodology, we created a flexible program for statistic analysis. The program named WinCollection can use a large class of data.

### 2.1. Input data

At this moment, SRoL corpus, that refers to database with emotional records, contains 25 speakers and 78 .wav files. Each file has a number of 3 to 7 pronunciations. All records are .wav files. TextGrid files are generated using Praat program [9]. These files contain annotations at sentences, words, syllables and phonemes levels. During annotation the following types of pauses were established and annotated correspondingly:

i) intra-speaking pauses (occurring inside syllables and words) are marked by character „$";

ii) inter-speaking pauses (occurring between sentences) are marked by character „ " (blank);

iii) brief pauses that are not noticed by the listeners, but are detected at the level of instrumental analysis; these pauses are marked with the character „%".

From the files .wav and .TextGrid, we obtained two other text files, containing values of the F0-F3 formant. All these files, along with text files Codes.txt, Selections.txt,

Phrases.txt and Emotions.txt (files created by users, and described later), will provide input data for program WinCollection. To facilitate the identification and the data manipulation, the name of each input file contains speaker's code and the sentence. For example, for the speaker with code 11861, for joy state, the sentence „vine mama", we'll find the following files in the sub-directory joy: vine_mama11861_[16b].wav, vine_mama11861_[16b].TextGrid, F0_vine_mama_11861, F1-F3_vine_mama_11861.

WinCollection program reads .TextGrid files and shows a range of annotation errors, the most common one being an incompatibility between two annotation levels, for example „ra+" at syllable level and „r" and „a+" at phoneme level. Note: both in Praat files and in program's input files the codes for /â/ is a-, for /ă/ is a+, for /ş/ is sh and for /ţ/ is tz.

Values of F0 and F1-F3 were automatically computed by Praat program and saved accordingly in text files. In some situations, F0 appears as undefined and these situations are:

- between pronunciations – there is no formant, neither F0 nor F1-F3. During processing the program ignores areas with values undefined by the Praat program.
- areas with no vocalic sound;
- areas with existing vocalic sound, but where Praat program cannot detect F0; this is considered an F0 detection error;
- when Praat program generates aberrant values for a vocalic sound – much lower or much higher than a normal value (for example, outside the range 70-500Hz for men).

Text files with instant values of formants were obtained with moving windows of 0.025s, step of the window being 0.01s. The files processed by the program have a maximum length of 10 seconds that corresponds to an average of 3 pronunciations for recorded sentence. Program WinCollectin runs only with files from Praat program.

The files are grouped in directories and each directory is dedicated to a certain emotional state. Strings that describe emotional states are lines of the text file Emotions.txt. They are used to find input data of a certain emotional state. If an emotional state has no directory assigned, the program signals an error. An error is also reported when the input data contains a directory whose name is not an emotional state. So this input file is used to check automatically if names are correct.

The codes.txt file contains data about all speakers whose records are in that directory. Data in the Selection.txt file allow the user to select a certain sub-population in order to obtain statistic values of interest about it. Subsequently these statistic values for selected group can be compared with same statistic values for another group or for entire population. Information about the speaker is available and described in speaker's profile file.

The file selections.txt contains 12 fields, each having meanings identical to those in the file codes.txt. Each field is a selection criterion and contains the string ALL, which indicates that the criterion will be taken into account all speakers, and a string whose value indicates which speakers will be selected from the database. Finally we select a sub-population that satisfies all criteria of Selections.txt file. Selection criteria are listed in Table 4.

Phrases.txt file is meant to store sentences used to create pronunciations records. This file is used to find errors in annotation sequence. Strings from Phrases.txt file are compared with corresponding annotation strings and are reported errors such as missing letters, inverted letters and other errors of this type that occurs during annotation. This file also contains encoded information about the vowels, namely:

| | | | |
|---|---|---|---|
| C1 | Isolated, sustained | C2 | Beginning of a word |
| C3 | End of a word | C4 | Interior of the word |
| C5 | Context CVC | C6 | Context CV_ |
| C7 | Context _VC | C8 | Accented syllable |
| C9 | Non-accented syllable | | |

## 2.2 Input data validation and errors reporting

As input data may contain errors, their validation is required, followed by manual correction. Data validation is performed by checking all input data in TextGrid annotation files created by Praat. Several criteria are used to eliminate incorrect values supplied by this program. Each annotation file TextGrid assigned to a record has two corresponding files: a file containing values of F0 and a file containing values of the formants F1-F3. Assignment is made automatically – missing one of them is reported as an error. These criteria are decided by the user and could be changed taking into consideration the variation ranges of fundamental and the formants. Existing segments from annotated files with their time markers allow validation or invalidation of formants existence. For example, if a pause corresponds to a segment between t1 and t2 (time markers), in the formants file, for the same segment, F0 and F1-F3 will be invalidated.

After processing the input data, the program reports the detected errors. The most common errors found at the input data processing level are: missing annotation for an existing record from the analyzed directory; annotation errors (appearance at a certain level of a phoneme that do not exists in recorded sentence); absence of the file that contains information about formant F0 for a certain record; absence of the file that contains information about formants F1-F3 for a certain record.

Detected errors are recorded in a special file named Dc.log. The same file is used to store other types of errors detected in subsequent processing steps. Then the file will be used to correct the corpus. Program (block Errors Analysis shown in fig. 1) also analyzes other error types that may occur in the corpus, for example an incomplete corpus where an emotion record is missing, or a situation when the file with F0 value exists, but the file with F1-F3 values is missing, etc. Detected errors are included in the Dc.log file. After error correction, all input data must be reprocessed for obtaining the binary files. Also, when a new record is inserted in an existing collection, all input data are reprocessed and a new, updated binary file is created.

F0 from Figure1 contains all files generated by Praat and stores F0 values for .wav files of that directory. Similarly, files F1-F3 in Figure1 contains all files generated by Praat storing these formants values.

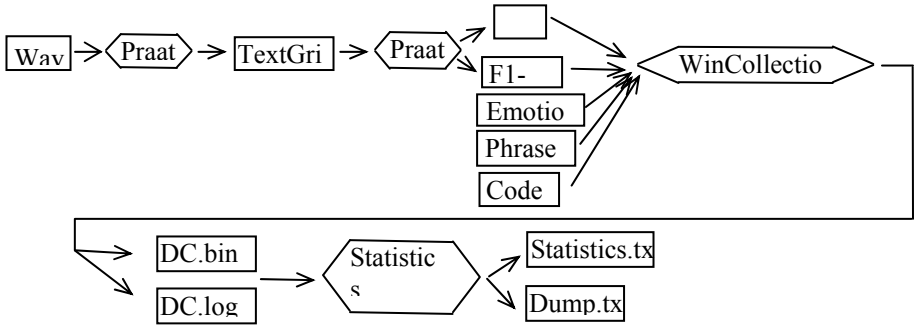


Fig. 1. Data processing diagram

### 2.3. Output data

Results of processing program are stored in two text files, Dump.txt and Statistics.txt (see Figure2). The Dump.txt text file contains:
- general information about collection (birth date, number of version for that collection, number of version of WinCollection program, etc);
- a summary of database (number of speakers, sentences, states, number of records for each state and sentence);
- input data, F0 and F1-F3 values;
- a set of basic statistic results for each vowel, including the number of occurrences for each emotional state (results are about minimum value, the maximum value, the average value and the dispersion calculated for F0 and F1-F3 values for each vowel, for each appearance of the vowel in a record).

Statistics.txt file created by WinCollection program contains following information for each vowel: number of occurrences for each emotional state; average value for subpopulation, for F0 and F1-F3 formants. These data are presented sorted by: emotion–speaker-phoneme, emotion-phoneme-speaker, and phoneme-speaker-emotion.

### 2.4. Statistical features – Variability and asymmetry indices

The program computes a variability coefficient inter-state and intra-state for formants values, and also a general statistic of formants for vowels in Romanian language („a”, „e”, „i”, „u”, „ă”) depending on emotional state.

In statistic analysis, we introduced a variability coefficient (asimmetry) defined by following formula (see [10] for full explanations):

$$\eta_j = \frac{1}{4}\sum_{k=0}^{3}\frac{1}{N}\sum_{\varphi=1}^{N}\frac{1}{M}\sum_{\Psi=1}^{M}\frac{F_k[j;\varphi,\psi]-\overline{F}_k[\psi]}{\overline{F}_k[\psi]},\quad \overline{F}_k[\psi]=\sum_{j=1}^{J}\sum_{\varphi=1}^{N}\frac{F_k[j,\varphi]}{NJ}$$

Above, ψ denotes the number of phonemes occurring in the sentence, φ denotes the sentence, *k* stands for the number of the current formant, and j specifies the current speaker index. Another „asimmetry coefficient" in speaker's pronounciation $v_j$ compared with the general average at the same state and the same vowel, for the formant *k* is defined with the following formula:

$$\alpha_{\varphi,\psi,k}[j]=\frac{\overline{F_k}[j,\varphi,\psi]-\overline{F_k}[\varphi,\psi]}{\overline{F_k}[\varphi,\psi]},\qquad \overline{F_k}[\varphi,\psi]=\frac{\sum_{j=1}^{J}\overline{F_k}[\varphi,\psi,j]}{J}$$

Statistic processing (the estimation of emotional state variability) is made on entire population, and on a certain sub-population. The selections.txt file is used to select a sub-population in order to make a statistic processing. The structure of the Selections.txt file is described in 2.1. Because there is a large amount of input data and of computed statistic data, the interpretation of this data is a quite difficult problem.

Another difficult problem is comparison between statistic data calculated for two sub-populations and their interpretation.

The selection of a sub-population in order to accomplish a statistic processing is made as follows:

- selections.txt file offers information about data required to be processed at a certain moment in time; this is made by an appropriate selection of fields described in 2.1.
- DC.bin file is read step by step and selected data are loaded in memory;
- selected data are used for statistic processing;
- processing results are recorded in Dump.txt and statistics.txt text files, described in 2.2.

### 2.5 Data collection

WinCollection Program creates data collection as a binary file named DC.bin, located in data collection's directory. This file is obtained by processing all input text file and will be subsequently used in statistic analysis of emotional voices. The only input files that remain unprocessed are .wav files that contain records. If during processing, process errors are reported, necessary corrections will be made, than processing will continue. This structure allows statistic processing of information. Data validation and error reporting were discussed in detail in section 2.2. Moreover, data are stored in only one file and performing their validation in the same time. If it is necessary to add new information at input data, for example new records, all data will have to be reprocessed, as it was shown

before, and a new data collection will be created. Data reprocessing will be done until no error will appear.

The structure of the binary file is:

| Header | Emotions | Sentences | Speakers | Speakers' data |
|---|---|---|---|---|

where:

- *Header* - contains information about date and hour of creation, number of collection's version, number of version of program used to create the collection, data about person who created the collection. These are general information about data collection.
- *Emotions* - contains emotional states from collection and their names. (Emotions names are: joy, sadness, anger and neutral tone.).

| n - number of emotions | emotion 1 | emotion 2 | ……. | emotion n |
|---|---|---|---|---|

- *Sentences* - contains number of sentences that were used to produce records, and also the offset of each sentence. For each sentence exist the sentence itself and description of characteristics for each vowel, as it was shown before in 2.1.

| No of sentences | sentence 1 offset | sentence 2 offset | … | sentence n offset | sentence 1 description | sentence 2 description | … | sentence n description |
|---|---|---|---|---|---|---|---|---|

where *sentence i description* is :

| string (sentence) | N – number of vowels | vowel 1 description | vowel 2 description | … | vowel n description |
|---|---|---|---|---|---|

- *Speakers* - contains a number of speakers from collection and an offset at information about each speaker. This information is available in Codes.txt file. These are information stored in speaker' profile which could be founded at address http://www.etc.tuiasi.ro/sibm/romanian_spoken_language/pdf/Speaker_Profile.pdf. Based on this information a selection could be made, in order to create statistic processing for a sub-population.
- *Speakers' data* - contains data of a record for a certain emotional state and sentence:

- string that indicates the place where is stored the .wav file that contains the record;
- F0 values and F1-F3 values obtained as a result of annotation and processing using Praat program;
- minimum value, maximum value, average value and dispersion value for F0 and F1-F3 for each phoneme.

Due to the fact that input data shown in 2.1 are located in numerous files and this number of files depends upon number of speakers, number of sentences and number of emotional states, the main idea was to collect all data in a singular file in order to accomplish several purposes:

- input data validation and correction of reported errors;

- selection of a sub-population of interest and statistic processing of selected data;
- data portability.

## 3. WinCollection Program

The WinCollection Program is written in Visual C++ and consists of three separate modules:

- pre-processing module;
- processing module;
- module for generation of data collection.

Pre-processing module read input data, validates it and reports founded errors. This operation is accomplished in five steps:

1. Processing information about speakers from Codes.txt file; all information will be stored in an array; every element array stores all data about a speaker;

2. Reading Phrases.txt file and processing information about sentences used to produce existing records; all sentences will be stored in an array of strings; for every sentence, there is an array of elements that describes every vowel in sentence;

3. Processing information about emotional states for which exist records, from Emotions.txt file; loading information from Emotions.txt file about emotional states for which exist records in a string array;

4. Processing information about selections from Selection.txt file; table 4 show information about selection criteria used;

5. Processing data input from annotation file and from files containing values of F0 and F1-F3 formants.

The pre-processing module generates the Dc.log file. In this text file is recorded information about data processing steps and about errors occurred. For every step, it indicates the processed emotion, the sentence, the speaker, and the .txt file. For each error it records the emotion, the sentence, the speaker, and the name of the .txt file, the number of line that contains the error and the error description. Some errors stop the preprocessing of the file where it occurs.

The pre-processing module reads all sub-directories containing records about a certain emotional state, each emotional state having a correspondent sub-directory with the same name the state has. In each sub-directory, program identifies all groups of four files (.wav file, corresponding annotation together with files containing values of formants F0, F1-F3). Identification is made based on speaker's code and on sentences for which records were made.

The processing module computes minimum and maximum value, average value and dispersion value for F0 and F1-F3, for each phoneme. In the future it will be developed in order to make also other primary processing. Also it creates Dump.txt and Statistics.txt text files that were described above.

The module for generation of data collection creates DC.bin binary file that contains entire data collection.

## 4. Example of application for statistic analysis

The examples presented are not final results. We exemplify the results obtained with the data from SRoL corpus.

Table 1 shows the values of the variability coefficient for the „a", „e", „i", „u", „ă" phoneme, for eleven speakers depending on the emotional state. The biggest value and the smallest value of the variability coefficient are in sadness state respectively neutral tone. We observe that the speaker 05392 has the smaller values of the variability coefficient for all four emotional states comparing with the other speakers. By the variability coefficient, we present the difference between speakers.

TABLE 1. The variability coefficient for the vowels, for eleven speakers according to the emotional state

| Speaker | Happiness state | Fury state | Sadness state | Neutral tone |
|---|---|---|---|---|
| 20048 | 0.28 | 0.27 | 0.17 | 0.25 |
| 30197 | 0.15 | 0.14 | 0.24 | 0.20 |
| 263315 | 0.14 | 0.12 | 0.16 | 0.16 |
| 12312 | 0.13 | 0.12 | 0.13 | 0.11 |
| 10031 | 0.11 | 0.15 | 0.14 | 0.12 |
| 11861 | 0.16 | 0.17 | 0.22 | 0.17 |
| 12121 | 0.17 | 0.20 | 0.19 | 0.18 |
| 32167 | 0.11 | 0.09 | 0.11 | 0.07 |
| 83714 | 0.25 | 0.21 | 0.23 | 0.27 |
| 26653 | 0.26 | 0.15 | **0.34** | 0.13 |
| 05392 | 0.07 | 0.07 | 0.09 | **0.05** |

Table 2 presents the asymmetry coefficient values obtained for the speaker #20048, phoneme "a". We observe that the asymmetry coefficient values for F0 and F1 helps us to identify the emotional state for a speaker. The asymmetry coefficient values for F2 and F3 are smaller and are of no use in determining the emotional state for this specific case.

TABLE 2. The asymmetry coefficient values for F0-F3 formants according to the emotional state, for speaker 20048f, vowel „a"

| 20048/ phoneme "a" | The asymmetry coefficient values for F0 | The asymmetry coefficient values for F1 | The asymmetry coefficient values for F2 | The asymmetry coefficient values for F3 |
|---|---|---|---|---|
| Happiness state | 0.70 | 0.22 | 0.13 | 0.12 |
| Sadness state | 0.56 | 0.07 | 0.11 | 0.06 |
| Fury state | 0.78 | 0.21 | 0.04 | 0.14 |
| Neutral tone | 0.61 | 0.20 | 0.08 | 0.13 |

Comparing the asymmetry coefficient values for F0-F3 between the speakers (table 3), we can make some observations about speakers and emotional state. For example, speaker 20048 has the asymmetry coefficient value for F0 triple than the speaker 30197, 26653 and 263315, and double for speaker 12312, in the happiness state. Thus, we are able to distinguish one speaker from another, for example the speaker 20048 from speakers 30197, 26653, and 263315.

TABLE 3. The asymmetry coefficient values for F0-F3 formants in happiness state, for five speakers, vowel "a"

| Happiness state/ Phoneme "a"/ Speakers | The asymmetry coefficient values for F0 | The asymmetry coefficient values for F1 | The asymmetry coefficient values for F2 | The asymmetry coefficient values for F3 |
|---|---|---|---|---|
| 20048 | **0.70** | 0.22 | 0.13 | **0.12** |
| 12312 | 0.31 | 0.19 | 0.13 | **0.03** |
| 26653 | 0.22 | **0.27** | **0.03** | 0.10 |
| 30197 | 0.23 | 0.14 | **0.14** | 0.10 |
| 263315 | **0.20** | **0.07** | 0.06 | 0.06 |

In the table 4 we exemplify the speaker's selection made according to sex, age, place of childhood, maximum education level, habits, voice education level, etc.

The results presented in table 5 show the asymmetry coefficient values for F0-F3 formants according to the emotional state, for speaker 20048f, vowel „a" from ten speakers, before the selection. Table 6 presents the asymmetry coefficient values for F0-F3 formants in happiness state, for two feminine speakers born after 1979, for the vowel "a", with data automatically selected by the program based on the criterion "born after year [1979].

TABLE 4. The meanings of the notation in the speakers selection

| | |
|---|---|
| 01 | The speaker code |
| 02 | The sex (M or F) |
| 03 | The age (year of birth) |
| 04 | The birth place and childhood - the region: NE - Moldova, Arges, Maramures etc. |
| 05 | The region of the primary education - idem, the region |
| 06 | The maximum education level: B(basic) = 8 classes or less; L= secondary school; C= university degree, G(graduate) = bachelor's or doctor's degree |
| 07 | The profession (university degree): Phl = philology, Th = technical; Med= doctor, AB= technical medical; H = humanist; Sci -science |
| 08 | The habits / work environment: S - smoking, A – intensive alcohol, T – toxic environment |
| 09 | The education voice: N = no; Y = yes, teaching, public speaking |
| 10 | pathology phonatory = V; respiratory R; neural N; dentition D; maxillo-facial MF |
| 11 | The day of the registration |
| 12 | Where to find on the site SRoL |

We present the results obtained (table 5 and 6) by the asymmetry coefficient for one type of selection: feminine speakers and born after 1979, from all 11 speakers.

In the table 5, we exemplify for speaker 20048, the differences of the asymmetry coefficient values for F0-F3, after the selection between the emotional states, for "a" phoneme. For fury state, we have the largest values, while for sadness, the asymmetry coefficient values of F0 is the smallest.

TABLE 5. The asymmetry coefficient values for F0-F3 formants according to the emotional state, for speaker 20048f, vowel „a", after the selection

| 20048/ phoneme "a" | The asymmetry coefficient values for F0 | The asymmetry coefficient values for F1 | The asymmetry coefficient values for F2 | The asymmetry coefficient values for F3 |
|---|---|---|---|---|
| Happiness state | 0.03 | 0.11 | 0.06 | 0.01 |
| Sadness state | 0.01 | 0.05 | 0.14 | 0.12 |
| Fury state | 0.13 | 0.12 | 0.05 | 0.12 |
| Neutral tone | 0.07 | 0.10 | 0.02 | 0.10 |

In table 6, the asymmetry coefficient values for F0, F1 and F2 formants in happiness state, for vowel "a" after the selection are smaller comparing with the values for F3, when the differences between speakers are meaningful.

TABLE 6. The asymmetry coefficient values for F0-F3 formants in happiness state, for two feminine speakers, vowel "a" after the selection

| Happiness state/ Phoneme "a"/ speakers | The asymmetry coefficient values for F0 | The asymmetry coefficient values for F1 | The asymmetry coefficient values for F2 | The asymmetry coefficient values for F3 |
|---|---|---|---|---|
| 20048 | 0.03 | 0.11 | 0.06 | 0.01 |
| 83714 | 0.02 | 0.08 | 0.04 | 0.07 |

Comparing the asymmetry coefficient values before and after selection (table 3 and 6), we observe that there are significant differences for F0.

## 5. Conclusions and future work

This paper presents the methodology to organize data in order to process them and shows the advantage of this organization; it also presents the application and its capabilities. We described the steps to obtain the data collection and importance to use it in various statistic processing. Brief analysis realized shows differences between speakers and emotional states.

The analysis shows the differences which occur at the vowel level of the Romanian language for the f0 and the formants values, for a population selection comparing with whole population from corpus SRoL, regarding the emotional state. Specific to this methodology is data validation, error correction and the adaptability of the WinCollection program to various statistics.

In the future we intend to achieve automatic comparison of computed statistical values for a subpopulation of the entire population. The program will automatically notify the higher values than a set threshold option, and automatically compare the statistical values computed for two subpopulations.

**Acknowledgments**. We acknowledge the partial support of the Romanian Academy "priority research" project "Sisteme cognitive"

**Authors contributions**: The first author implemented in C++ the program WinCollections. H.N Teodorescu has initiated the research and established its aims, largely determined the requirements for the program package, proposed some of the solutions for data organization, contributed to writing the paper. He also introduced he indices described in sub-section 2.4. The third author made the annotation files and obtained with Praat the values of the F0 and F1-F3 formants. All authors contributed to the analysis results, identify solutions to improve and validate the program and the results.

## References

[1]. Feraru M., Teodorescu H.N., Classification of the Emotional States in Speech using the SRoL Database–Preliminary Results, *Proc. 3rd Int. Conf. Electronics, Computers and Artificial Intelligence*, Pitesti, România, (2009), ISBN 1843-2115, 27-32.

[2]. Kienast M., Sendlmeier W. F., Acoustical Analysis of Spectral and Temporal Changes in Emotional Speech. An Acoustic Framework for Detecting Fatigue in Speech Based Human-Computer-Interaction, *Lecture Notes in Computer Science*, (2008), ISBN 978-3-540-70539-0, 5105/2008, 54-61

[3]. Mcgilloway S., Cowie R., Douglas-Cowie E. et al., Approaching Automatic Recognition of Emotion from Voice: A rough Benchmark. *Proc. ISCA Workshop Speech and Emotion, Newcastle*, (2000), 207-212

[4]. Nakatsu R., Solomides A., Tosa N., Emotion Recognition and its Application to Computer Agents with Spontaneous Interactive Capabilities. *Proc. IEEE Int. Conf. Multimedia Computing and Systems*, Florence, Italy, (1999), 2, 804-808.

[5]. Scherer K., A., Cross-Cultural Investigation of Emotion Inferences from Voice and Speech: Implications for Speech Technology. *Proc. Conf. Spoken Language Processing (ICSLP)*, China, (2000), http://www.unige.ch/fapse/emotion/publications/pdf/icspl00_crosscul.pdf.

[6]. Teodorescu H.N., Feraru M., A study on Speech with Manifest Emotions, 10th International Conference on Text, Speech and Dialogue, *Lecture Notes in Computer Science, Springer Verlag*, ISBN 978-3-540-74627-0, 4629/2007, 254-262.

[7]. H.N. Teodorescu, L. Pistol, M. Feraru, M. Zbancioc, D. Trandabăț, Sounds of the Romanian Language Corpus, SRoL, © (2010) (http://iit.iit.tuiasi.ro/romanain_spoken_language/ index.htm)
[8]. Ververidis D., Kotropoulos C., Emotional Speech Recognition: Resources, Features, and Methods, *Speech Communications*, 48: 9, (2006), 1162-1181.
[9]. Boersma, P., Weenink, D., Praat: doing phonetics by computer, Institute of Phonetic Science, University of Amsterdam, *www.praat.org*.
[10]. Teodorescu, H.N., Păvăloi, I., Feraru, M., Aspecte metodologice de organizare a datelor şi de analiză statistică a vocilor emoționale, Conferința Resurse Lingvistice și Instrumente pentru Prelucrarea Limbii Române - ConsILR2010, Mai 6-7, Bucureşti, România Editura Universității „Al.I. Cuza", (2010), Iasi