Economy Transdisciplinarity Cognition www.ugb.ro/etc	Vol. 15, Issue 1/2012	267-272

Speech Emotion Analysis in the Romanian Language

Silvia Monica FERARU
Romanian Academy, Iasi, Romania
monica.feraru@gmail.com

Abstract: *In the communication in the international negotiation, the people must adopt certain ways to deliver the message in order to support and enforce in a most effective strategies and tactics adopted by the negotiating team. The analysis of the emotion expressiveness in the voice, setting new theoretical methodologies and algorithms to identify and classify the emotions are based on several practical and scientific applications in the inter-disciplinary field which touches: computers sciences, psychology and cognitive sciences. Several scientists are focused on the improvement of the performances of classification and emotional recognition methods, in order to find new optimum parameters to solve the constraints (classification errors, complexity of the algorithms etc.) met in several applications.*
The aim of this paper consists of developing and improving the analysis for the emotional expressiveness of the vocal signals in Romanian language. We use the only emotional corpus annotated for the Romanian language, which is part of the Romanian Language Sounds project (SRoL) which contains more than 300 folders with emotional stales (i.e., 900 speeches) as well as some folders with standard tonalities. The studied emotions are: happiness, sadness, fury and neutral tone.
This inter-disciplinary theme has been chosen in order to contribute to the development of several principles, methodologies and possible concepts concerning the emotional classification and recognition.
Keywords: *Emotional analysis, speech database, inter-disciplinary field*

Introduction

The negotiation plays an important role in international commercial transactions. It now covers a wide range of areas, such as the politico-diplomatic, social, cultural and, especially, economic. The negotiation is a social process, a form of communication between partners in an intercultural context. To succeed in a negotiation you must have self-control on your emotions and to know where to insist and to lead a discussion by a desired end.

In the economic field, the largest and the most important negotiations are commercial, and among them those related to international economic affairs. The communication can be oral, written, in an international context, or extra-verbal. From the last category we mention the speech emphasis, the intonation, the pauses, the rhythm, the gesture, etc. We must take into account also the body language, the time, the look, the facial expressions, the body posture, etc..

Speech emotion plays an important role and it is an inter-disciplinary field of research with contributions coming from psychology, acoustics, speech science, linguistics, medicine, engineering, and computer science. The analysis of the emotional expressiveness [1, 254-262; 2, 27-32] in order to establish theoretical methodologies and algorithms to classify the emotional states [3, 1-10; 6, 207-212] have many practical and scientific applications [7; 8, 54-61]. These analyses are laborious and require considerable resources, however, when it comes from large corpora of speech recordings.

Speech emotion analysis refers to the use of various methods to analyze the vocal behavior. There is a set of objectively measurable voice parameters that reflect the affective state a person. The most affective states involve physiological reactions which in turn modify different aspects of the voice production process. In the fury state we can observe the changes in respiration and the increase in muscle tension, which influence the vibration of the vocal folds and vocal tract shape, affecting the acoustic characteristics of the speech [9, 143-165]. Speech emotion analysis is complicated by the fact that vocal expression is an evolutionarily old nonverbal affect signaling. Voice researchers still debate the extent to which verbal and nonverbal aspects can be neatly separated. However, that there is some degree of

independence is illustrated by the fact that people can perceive mixed messages in speech utterances – that is, that the words convey one thing, but that the nonverbal cues convey something quite different. Voice cues are divided into those related to: (a) fundamental frequency (F0), (b) vocal perturbation (short-term variability in sound production), (c) voice quality, (d) intensity and (e) temporal aspects of speech (speech rate), as well as various combinations of these aspects (prosodic features).

Current research directions include cross-cultural studies, comparisons of emotion portrayals with natural expressions, comparisons of different theoretical approaches, attempts to develop tools for automatic decoding of emotions, and multimodal approaches in emotional expression [10, 42-40].

The aim of the paper was to organize and to process data for the classification of emotional states in Romanian language. The work is a continuation of the research concerning the analysis and emotional expressiveness in voice [11, 61-65;13]. The paper is structured as follows: the second section is dedicated to the SRoL corpus, the third section describes the methodological aspects of the analysis, and in the next section we present the obtained results and in the end the conclusion are drawn.

1. Romanian Database

The Romanian speech database – SRoL, currently located at the address http://www.etc.tuiasi.ro/sibm/romanian_spoken_language/en/arhiva_tot_en.htm is conceived as an Internet-based "dictionary of sounds and words" for the Romanian language. The content of the site can be used for educational purposes such as analysis of sounds, analysis of specificities of the Romanian language pronunciation compared to other languages, Romanian language learning aided by computer, as well as for research purposes. The corpus [14, 280-290] includes files with vowels, consonants, diphthongs, sentences with emotional states, linguistic particularities for the Romanian language, dialectal voices, and gnathosonic and gnatophonic sounds. The database contains female and male voices; the speakers are aged between 25-35 years; they are from the middle area of Moldova and have no manifested pathologies. The section of SRoL devoted to emotional speech contains files with emotional states of mind and comparison between them. The analyzed emotional states are happiness, sadness, fury, and the neutral tone. The recorded sentences are: Mother is coming (Vine mama, in Romanian), Who did that? (Cine a făcut asta?, in Romanian), Last night (Aseară, in Romanian), You came to me again (Ai venit iar la mine, in Romanian), My man done sapped him (Omul meu îl lucră, in Romanian), You will win / get the desired place (Îţi vei câştiga locul dorit in Romanian), Anyway, you can win / get the desired place, anyway (Oricum îţi poţi câştiga locul dorit in Romanian). Note that changing the topic of the sentences, when it is grammatically possible, usually the emotional state is modified.

2. Methodological Aspects and Analysis Tools

The recordings were made with a sampling frequency of 22050 Hz, 24 bits, using Goldwave (www.goldawave.com). Every speaker pronounced the sentence for three times, following the recording protocol. The persons were previously informed about the objective of the project and they signed an informed consent in accordance with to the Protection of Human Subjects Protocol to the U.S. Food and Drug Administration and with Ethical Principles of the Acoustical Society of America.

Today, there is no standard model for the emotional annotation process. For the annotation we have used Praat software (www.praat.org.). We automatic computed the formants values (F0, F1, F2, F3) of the all vowels from the Romanian language on the entire duration of the vowel. Based on the files of the formants values and on the vowels (/a/, /e/, /i/, /o/, /u/, /ă/) we try to classifier the emotional states using Matlab software.

Unsupervised learning allows the identification of the completely new concepts based on the known data, for example the algorithm K-means. In statistics and data mining, k-means clustering is a method of cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean.

Supervised learning is a type of inductive learning which is based on a set of examples of the problem and form an evaluation function in order to allow the classification of new data sets, for example the algorithm k-NN (k-nearest neighbor). The k-nearest neighbor algorithm (k-NN) is a method for

classifying objects based on closest training examples in the feature space. The training examples are vectors in a multidimensional feature space, each with a class label. The training phase of the algorithm consists only of storing the feature vectors and class labels of the training samples. In the classification phase, k is a user-defined constant, and an unlabeled vector is classified by assigning the label which is most frequent among the k training samples nearest to that query point. We used the Euclidean distance as the distance metric. The accuracy of the k-NN algorithm can be severely degraded by the presence of noisy or irrelevant features, or if the feature scales are not consistent with their importance. The best choice of k depends upon the data; generally, larger values of k reduce the effect of noise on the classification, but make boundaries between classes less distinct. A good k can be selected by various heuristic techniques, for example, cross-validation. The special case where the class is predicted to be the class of the closest training sample (i.e. when k = 1) is called the nearest neighbor algorithm.

The training phase for kNN consists of simply storing all known instances and their class labels. If we want to tune the value of 'k' and/or perform feature selection, n-fold cross-validation can be used on the training dataset. The testing phase for a new instance 't', given a known set 'I' is as follows:

- compute the distance between 't' and each instance in 'I';
- sort the distances in increasing numerical order and pick the first 'k' elements;
- compute and return the most frequent class in the 'k' nearest neighbors, optionally weighting each instance's class by the inverse of its distance to 't'.

Some of the advantages of kNN for classification are:
- a very simple implementation;
- a robust with regard to the search space; for instance, classes don't have to be linearly separable;
- the classifier can be updated at little cost as new instances with known classes are presented;
- a few parameters to tune: distance metric and k.

Some of the disadvantages of the algorithm are:
- the expensive testing of each instance, as we need to compute its distance to all known instances;
- the sensitiveness to noisy or irrelevant attributes, which can result in less meaningful distance numbers;
- the sensitiveness to very unbalanced datasets, where most entities belong to one or a few classes, and infrequent classes are therefore often dominated in most neighborhoods.

3. Analysis and Results

We used the k-NN algorithm (simple and with normalization) on the vowels of Romanian language in order to recognize the emotional state (happiness, fury and sadness) from the recordings of SRoL corpus. Depending on the vowels and on the emotions, the percentage was different. We used two types of the files: one which use the F0, F1, F2 formants values, and other which use only the F1 and F2 formants values.

In the next table we presents the results obtained for the vowels /a/, /e/ and /i/ (on seven sentences), from 19 speakers (11 males and 8 females), for two emotional states (sadness state and neutral tone).

The classification rates for the vowels /a/, /e/, /i/ depending on two emotions

Table 1.

Emotional states	**Sadness state**			**Neutral tone**		
F0/F1/F2	*/a/*	*/e/*	*/i/*	*/a/*	*/e/*	*/i/*
	87%	80%	89%	90%	85%	89%
F1/F2	*/a/*	*/e/*	*/i/*	*/a/*	*/e/*	*/i/*
	86%	100%	73%	100%	100%	80%

The figure 1 exemplifies the spatial representation of the /a/, /e/, /i/ vowels, for the sadness and neutral tone, in the Romanian language. We note with red points the formats values of the /a/ vowel, with pink the values of the /e/ vowel, and with blue the values of the /i/ vowel.

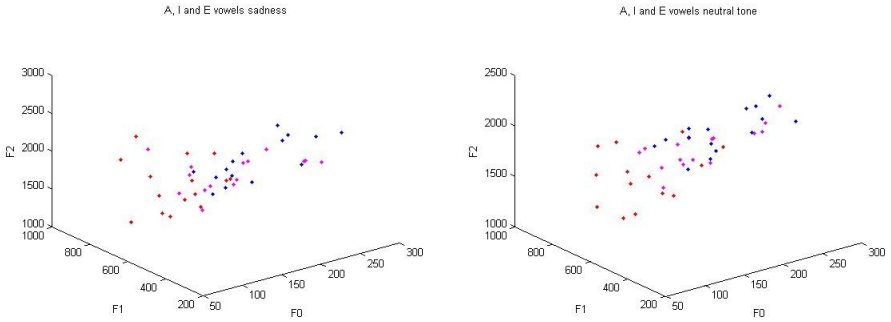

Figure 1. The spatial representation of the formants values of the /a/, /e/, and /i/ in sadness and neutral tone, in Romanian language

The recognition systems of emotional states must be trained by speaker in order to distinguish the fury. The emotional intra-speaker states can be clearly distinguished, but we cannot specify the emotional inter-speaker states.

The recognition rates for the /u/ vowel based only on the average values of the F1 and F2 formants (18 speakers) are: happiness – 83%, for fury – 83% , for sadness – 88%,and neutral tone - 83% ; for the /o/ vowel (10 speakers) are happiness – 80%, for fury – 80%, for sadness – 80%, and neutral tone 90%.

The classification rates based on the average values of the F0, F1 and F2 formants using the KNN algorithm with normalization are the follows: for k=5, in neutral tone we obtained 100% for the /ă/ vowel, 83% for the /i/ vowel, 42% for the /o/ vowel, 60% for the /a/ and /e/ vowels and 33% for the /u/ vowel. For the sadness state, k=4 the classification rates are: 33% for the /ă/ vowel, 50% for the /i/ vowel, 75% for the /o/ vowel, 66% for the /a/, 80% for the /e/ vowels and 40% for the /u/ vowel. For the fury state, k=5, the results are: 16% for the /ă/ vowel, 80% for the /i/ vowel, 100% for the /o/ and /a/ vowels, 71% for the /e/ vowel and 66% for the /u/ vowel. For joy, k=5, the classification rates are: 60% for the /ă/ vowel, 66% for the /i/ vowel, 75% for the /o/ vowel, 50% for the /a/ vowel, 42% for the /e/ vowel and 14% for the /u/ vowel. We notice that after the neutral tone, the sadness state is better recognizing that the joy state and the fury state is less recognized. In the fig. 2 and 3 we represented the average values of the F0, F1, F2 for the /ă/, /i/, /o/, /a/, /e/ and /u/ vowels in the emotional states analyzed. We note with red points the formats values of the /ă (a+)/ vowel, with blue the values of the /i/ vowel, with magenta the values of the /o/ vowel, with cyan the values of the /a/ vowel, with green the values of the /e/ vowel and with black the values of the /u/ vowel.

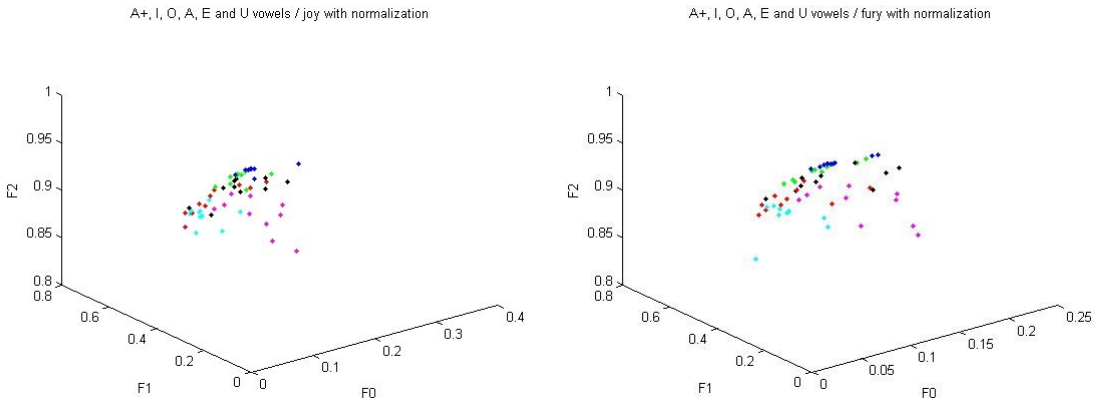

Figure 2. The representation of the average values of the F0, F1, F2 for the /ă/, /i/, /o/, /a/, /e/ and /u/ vowels in joy and fury states

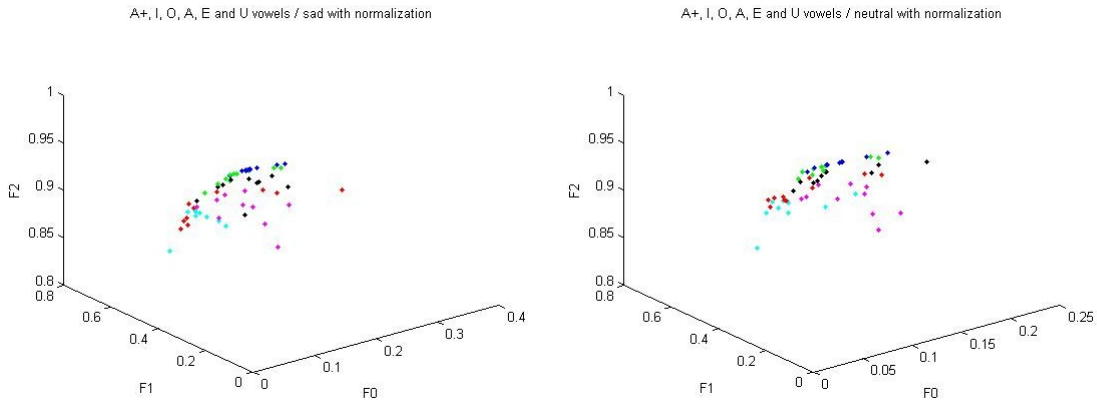

Figure 3. The representation of the average values of the F0, F1, F2 for the /ă/, /i/, /o/, /a/, /e/ and /u/ vowels in sadness state and neutral tone

The results obtained by comparing the Romanian language with Swedish, English, German and Spanish languages are offered in the next table [15; 16, 1517-1520].

The classification rates of the three emotional states depending on the languages

Table 2.

Languages	Happiness state	Fury state	Sadness state	Mean
Swedish	92%	83%	69%	81%
English	66%	75%	100%	80%
German	83%	96%	80%	86%
Spanish	79%	96%	91%	89%
Romanian	83%	82%	85%	83%

Following the results we observe that the Spanish language is more expressive, and the less expressive is the English language. We think that the emotions in speech might have a connection to the syntax, to the intonation, to the mood of the speaker and to the own self control, and also if the speakers is a native language or not.

Conclusions

We reported the preliminary results of the study which try to classify the emotional states in the Romanian language. We notice that some vowels give better results than the other in the emotional classification. For example, the /o/, /e/, /a/ vowels give good results for the fury, sadness and joy states comparing with the /u/ vowel in the classification of the emotional states.
The recognition rates using the k-NN algorithm are better in the neutral tone compared with sadness state. In the same emotional state, the differences between the vowels are better than in the case where we have the same vowels, but different emotional states.

In the future, we will analyze other type of classification algorithm (as MLP). We will make the comparisons between them in order to obtain better results. The study will be made on a larger number of sentences and subjects.

ACKNOWLEDGMENT
This paper is supported by the Sectoral Operational Programme Human Resources Development (SOP HRD), financed from the European Social Fund and by the Romanian Government under the contract number POSDRU ID 56185.

References

[1] Teodorescu H.N., Feraru M. (2007). 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.

[2] Feraru M., Teodorescu H.N. (2009). Classification of the Emotional States in Speech using the SRoL Database – Preliminary Results, Proc. 3rd Int. Conf. Electronics, Computers and Artificial Intelligence, Pitesti, România, ISBN 1843-2115, 27-32.
[3] Teodorescu H.N., Zbancioc M., Feraru M. (2011) Statistical characteristics of the formants of the Romanian vowels in emotional states, International Conference on Speech Technology and Human Computer Dialogue ISBN: 978-1-4577-0440-6, pp.1-10, http://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=5929232
[4] Ververidis D., Kotropoulos C. (2006) Emotional speech recognition: Resources, features, and methods, Speech Communications, 48: 9, 1162-1181.
[5] Nakatsu R., Solomides A., Tosa N. (1999). Emotion Recognition and its Application to Computer Agents with Spontaneous Interactive Capabilities. Proc. IEEE Int. Conf. Multimedia Computing and Systems, Florence, Italy, 2, 804-808.
[6] Mcgilloway S., Cowie R., Douglas-Cowie E. et al. (2000). Approaching Automatic Recognition of Emotion from Voice: A rough Benchmark. Proc. ISCA Workshop Speech and Emotion, Newcastle, 207-212.
[7] Scherer K., A. (2000). Cross-Cultural Investigation of Emotion Inferences from Voice and Speech: Implications for Speech Technology. Proc. Conf. Spoken Language Processing (ICSLP), China, http://www.unige.ch/fapse/emotion/publications/pdf/icspl00_crosscul.pdf.
[8] Kienast M., Sendlmeier W. F. (2008). 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, ISBN 978-3-540-70539-0, 5105/2008, 54-61.
[9] Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. Psychological Bulletin, 99, 143-165.
[10] Patrik N. Juslin, Klaus R. Scherer (2008), Speech emotion analysis, Scholarpedia, 3(10):4240.
[11] Feraru S.M. (2011), Emotional expressiveness in the Romanian and German language, Gr. T. Popa University of Medicine and Pharmacy, Publishing House, ISBN: 978-606-544-078-4, pp. 61-65, http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=6150420
[12] Feraru S.M., A preliminary study of the intra-speaker variability in speech, Journal of Inventics, Vol. 14 / 2011, No. 72, ISSN: 1210-3084, pp. 12-16
[13] Feraru S.M., Emotional expressiveness in Romanian language, WSEAS/INEEE International Conferences, 2012, April 18-20, Rovaniemi, Finland – to be published
[14] Feraru S.M., Teodorescu H.N., Zbancioc M.D. (2010), SRoL - Web-based Resources for Languages and Language Technology e-Learning, International Journal of Computers, Communications & Control, ISSN 1841-9836, E-ISSN 1841-9844, Vol. V (2010), No. 3, pp. 280-290
[15] Abelin A., Allwood J., (2000). Cross Linguistic Interpretation of Emotional Prosody, Proc. ISCA Workshop (ITRW) on Speech and Emotion: A conceptual framework for research, Belfast, (http://www.ling.gu.se/~abelin/abelin.pdf)
[16] Burkhardt F. et al. (2005). A database of German emotional speech, Interspeech 2005, Lisbon, Portugal, ISCA, pp. 1517- 1520