STUDIES ON THE PROSODY OF THE ROMANIAN LANGUAGE: THE EMOTIONAL PROSODY AND THE PROSODY OF DOUBLE-SUBJECT SENTENCES

Horia-Nicolai TEODORESCU [*], [**], Monica FERARU [**], Diana TRANDABĂȚ [*]

[*] Institute for Computer Science, Romanian Academy, Bd. Carol I nr 8, Iasi, Romania
[**] Technical University of Iasi, Iasi, Romania

Corresponding author: Horia-Nicolai Teodorescu, hteodor@etc.tuiasi.ro

We present a study of the prosody – seen in a broader sense – that supports the theory of the interrelationship function of speech. Both "pure emotions" and specific syntactical constructions, like the double-subject sentences are meant to show a relationship of the speaker with the general context. The analysis goes beyond the basic prosody, as related to pitch values, pitch trajectory, sound duration and pauses; the analysis also aims to determine the change in higher formants. The refinement in the analysis asks for finer tools. Methodological aspects are discussed, including limitations of the currently available tools. Some conclusions are drawn.

Key words: Romanian spoken language, prosody, emotions, double subject, formant analysis.

1. INTRODUCTION

The study of natural languages has seen a tremendous progress during the years 1960-1990, while the spoken language analysis and interpretation lagged behind. Recent years, a dramatic shift of interest has been produced, fuelled by applications like virtual reality, video-games, human-computer speech interaction, medical applications, and security, along with the more traditional domains like telephony. The European Union needs, especially the preservation of the linguistic and cultural heritages of the constituent countries also boosted the interest in spoken languages. It is the belief of the first author that the next big revolution in communications, after the completion of the current version of Internet will be the development of a voice-based, multimedia (MM) enhanced version of the WWW, with spoken databases and MM-full capabilities – the "MM-WWW". In view of preparing that future step, moreover to take full advantage of the national language heritage, countries have to develop their own versions of spoken language databases and knowledge bases. In an effort to provide such tools for the Romanian language, a high-priority project of the Romanian Academy has been promoted by the first author and a multi-disciplinary research team has been formed in the Institute for Computer Science of the Romanian Academy. The team, comprising linguists, computer scientists, electronic engineers, and biomedical engineers has collected, based on a rigorous methodology, a spoken language data- and knowledge-base which, presently, includes more than 1000 recordings. The database, named "The Sounds of the Romanian Language" is made publicly available through Internet, at the address [11].

One of the components of the "language-base" is the analysis and knowledge extraction of the language-based communication processes, including emotion communication and specific language constructions that are related to emotions. Presently, a challenge in artificial intelligence development is the emotional interpretation of the interaction between human and computer. Speech is a subtle and rich communication; it transfers not only the linguistic information, but also information about the personality and the emotional state of the speaker. The emotion is a motivation-related answer adapted to the social environment. The prosody is defined like "the rhythm and intonation aspects of a language", while the intonation represents "the expression manner; raising and lowering of the pitch in voice" (Merriam-Webster). Containing information about the speaker and about the environment, the prosody is a communication manner which includes the attitude, the emotions etc. [12].

According to Fakotakis [5], emotions are classified in "basic" emotions, with different intensity levels, and in "non-basic" emotions (the "mixed" emotions). In 1992, Johnson-Laird and Oatley stated that there are five basic emotions; in 1998, after other researches and studies, they concluded that there are four basic emotions: happiness, anger, sadness and fear. In this paper, we focus only on two emotions.

In order to create an emotional database, it is necessary to have a number of speakers who try to simulate the emotions in different contexts [4]. A different set of subjects listen to the recordings and try to identify the emotion that the speaker has tried to simulate. The experimental analysis of Buluti, Narayanan and Syrdal [2] showed that the emotion's recognition is not perfect for emotions like sorrow, sadness, happiness, and the neutral tone. Their recognition rate was 92.1% for the neutral tone, 89.1% for sorrow, 89.7% for sadness, and 67.3% for happiness. It is important to notice that the recordings have been made by professionals (actors in most cases) with professional voices [13]. In our research, the minimal recognition score for emotions has been 67%.

2. BASES OF OUR APPROACH

The research, as envisioned by the first author, has been aimed to verify several speculative hypotheses, namely:

- The emotions are represented by a complex characteristic of the voice; that complex goes beyond the accepted set of characteristics: duration of phonemes, duration of pauses, and pitch trajectory, including the first formants, moreover the higher formants and the subtle mixture of linear and nonlinear processes of speech generation.
- The same emotion is represented differently in presence of different interlocutors, depending on the relationship the speaker has with the interlocutor.
- If the inter-relationship theory of the prosody proposed in [12] were to be true, some of the syntactical constructions of the language should be specifically aimed to the inter-relationship promotion; therefore, such constructions have to be looked for, determined and analyzed. One of these constructions, identified by the first author as possibly related in a specific way to the emotion and inter-relationship representation, is the double-subject construction.

3. EXISTING APPROACHES

The emotional speech analysis is a challenging research field, as shown by the increasing number of databases dealing with the voice characteristics in different emotional contexts and for various languages. For sake of comparison with our approach, we briefly present in this section several databases of emotional voice: the Greek database [5], the German archive [8], the Danish [7] and the Spanish databases [6]. Numerous other databases, for other languages, exist, but we limit our presentation to the ones described below because they are quite different and illustrate well the variety of approaches in the literature.

The recordings from the Greek emotional database [5] were made in a professional studio, in Athens, and the speakers were actors. The recordings were made on 16 bits, mono, with a sampling frequency of 44,1 kHz. The goal of this research was to improve the naturalness of synthesized voice. The recordings were made in three different contexts:

- in order to reflect the reaction of the speaker to a concrete stimulus (authentic emotion);
- preparing the environment in order to help psychologically the speaker to simulate the indicated emotion;
- simulating the emotions only by imagining a context.

Several sentences were selected from the press by linguists. The corpus included 10 words, 20 short phrases, 25 long phrases, and 12 fragments of fluent speech. The study was oriented towards the evaluation of the simulated emotional states by free answers (86.9%) and false answers (89.6%).

The German emotional database [8] contains six basic emotions: anger, happiness, fear, sadness, disgust, boredom and neutral tone. The recordings in that study included five short sentences and five long sentences; the recordings have been realized by 10 professional actors, 5 women and 5 men in a special room. Over 800 recordings have been made (7 emotions × 10 sentence × 10 actors × 2 versions). The validation commission, composed by 20-30 persons, listen the phrases and recognized 80% of the simulated emotional states. The recordings were made on 16 bits, mono, with a sampling frequency of 16 kHz. The database contains files with sentences and words, files with syllables and phonemes, in wave format, the

information about the results of the perception tests (recognized emotion, evaluation of natural language, the power of emotion), and the results of measuring the fundamental frequency, the energy, the duration, the intensity and the rhythm.

The Danish database [7] contains recordings of two words, nine sentences and two fragments of fluent speech, simulating happiness, surprise, sadness, anger and neutral tone, spelled by four professional actors. The emotional states were recorded in a room of the Aarhus theatre – a room considered to have an excellent acoustic. The emotions were correctly recognized in a proportion of 67%. The happiness state was mostly confused with surprise and the sadness state was confused with the neutral tone. 75% of the people listening to the recordings said that it was difficult to identify the recorded emotions. The recordings were made on 16 bits, mono, with a sampling frequency of 16 kHz. Each recording has attached video information along the voice signal. The database also contains information about the profile of the speaker, like weight, height, sex, how long he/she worked like an actor etc. To increase the consistency of the assessment of the emotion conveyed by the speech, researchers at the University of Aolborg use a questionnaire for the assessing persons. The questionnaire includes questions such as "how the emotions identification seems like", "what are the factors which bring to the correct identification of the emotions", and "additional remarks regarding the recordings".

The Spanish database [6] contains recordings with the following emotional states: happiness, desire, fear, fury, surprise, sadness and disgust; the sentences have been pronounced by eight actors. Every speaker recorded the text three times, with various levels of intensity of the emotions. The validation of the recorded emotions was made by a test based on the questions: "mark the emotion which was recognized in each recording", "mark the credibility level of the speaker", and "specify if the emotional state was recognized and at what level". The speakers have made 336 recordings; only 34 were selected, analyzed, and validated through the tests made on 1054 persons. The goal of the study was to describe a useful methodology in the validation of the simulated emotional states. They obtained a set of rules which describe the behavior of the important parameters of the speech associated with emotions. The obtained results are useful in generating synthesized speech with emotion. The analyzed parameters were the fundamental frequency trajectory, time and rhythm. The researchers represented graphically the wave form, the pitch contour and the energy. They obtained the following characteristics of the emotional modulation [6]: i) for happiness state: "increase of the average tone, increase of the variability of the tone, quick modulations of the tone, […] stable intensity, decrease [of] the silence time; […]"; ii) for fury state: "variation of the emotional intonation structure, short number of pauses, increase of the intensity from beginning till the end, variation of timber, increase of the energy; […]"; iii) sadness state: "decrease of the average tone, decrease of the variability of the tone, no inflexions of the intonation, decrease of the average intensity."

The above-quoted analysis is somewhat subjective and leaves many unanswered questions on the variation of objective parameters, like formants, from one emotion to another. In our research reported here, we specifically address the characterization of emotions using the objective parameters for the states reported in [6]. We also contrast the characteristics of the voice for the above emotions with the normal (i.e., no emotion) speech. The comparison of our results with the results reported in [6] may help identify inter-language variations for the emotional speech.

4. THE METHODOLOGY

4.1. Recording protocol

The database contains short sentences or phrases fragments, with different emotional states. The emotions are: sadness, happiness, anger and detaching state. The files are classified in A class (feminine voice) and in B class (masculine voice). The speakers are persons with age between 25-35 years, born and educated in the middle area of Moldova (Iasi, Vaslui, and Bacau counties), with higher education, and without manifested pathologies.

The recordings were made using the GoldWave™ application, with a sampling frequency of 22050 Hz. Every speaker pronounced the sentence for three times, following the recording protocol. The sound was saved in .wav, .ogg, .txt format on 16 and 24 bits. The persons were previously informed about the objective of the project. The speaker 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 for Research Involving Human Subjects.

The database contains two types of protocols, namely the recording technical protocol and the recording documentation protocol. The recording protocol contains information about the noise, the microphone used, the soundboard, and the corresponded drivers. The documentation protocol contains information on the speaker's profile – linguistic, ethnic, medical, educational, professional information about the speaker – and a questionnaire regarding the health state of the speaker.

4.2. The emotional speech database and processing tools used in the analysis

This Section is a copy of the corresponding section in a submitted paper[1]. We recorded a set of sentences with emotional states. The sentences are: 1. *Vine mama*. (Mother is coming) 2. *Cine a facut asta*. (Who did that?) 3. *Ai venit iar la mine*. (You came back to me) 4. *Aseara*. (Yesterday evening). Only two emotions are discussed here, namely happiness and sadness. The consistency of the emotional content in the speech recordings has been verified by several listeners; the emotion confusion matrix has proved that all emotions are correctly identified, with a rate of more than 80% by the listeners.

Today, there is no standard model for the emotional annotation process [10]. The sentences have been annotated using the Praat software [16] at several levels: phoneme, syllable, word and sentence. In this paper, the analysis was made only for the sentence "Vine mama" pronounced by five persons, three times each. Our goal has been to discriminate between happiness and sadness emotional states. For that purpose, we computed the values for the formants and the duration of the vowels using several tools: Praat [17], Klatt analyzer [18], GoldWave [19], and Wasp [20]. We obtained general and particular rules which are discussed in the section on results. We have been confronted with several problems in the determination of the formants, namely with large disagreements between values provided by different applications. For example, there were cases where according to Praat™, on some segments of speech the fundamental frequency is not defined (see figure 1), while Wasp™ or Klatt analyzer™ identifies a pitch on those segments. According to Klatt Analyzer™, the F1 formant for vowel i is "missing", as exemplified in figure 2.

Notice that, in figures 3 and 4, it is difficult to visually determine the formants in these spectrograms using the GoldWave™ and Wasp™ applications. The difficulties are largely due to the imprecision of the definitions of the pitch and of the formants, especially for non-stationary signals. The nonlinear behavior of the phonatory organ, which is well documented in the medical literature as well as in the recent info-linguistic literature, [14], [15], determines a lack of significance of the parameters defined in the frame of the linear theory of speech analysis.

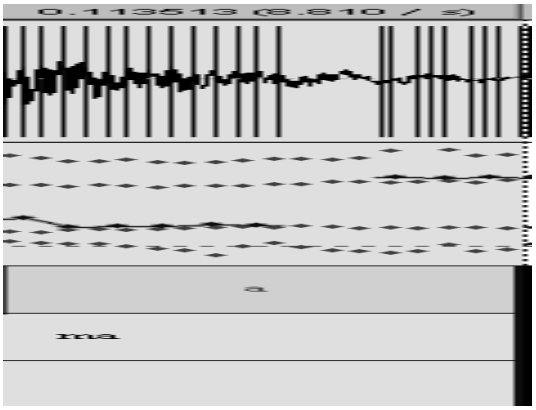


Figure 1. Determination of the F0 with Praat™ application.

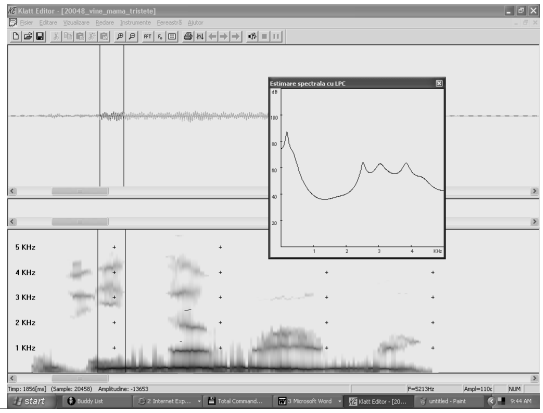


Figure 2. Determination of the formants with Klatt analyzer™.

[1] This Section is almost a copy of the corresponding section in the submitted paper TEODORESCU H.N., FERARU M., Analyzing Emotions in Spoken Romanian Language, Proceedings of the Romanian Academy, Series A, Volume 8, 2007 (to appear).

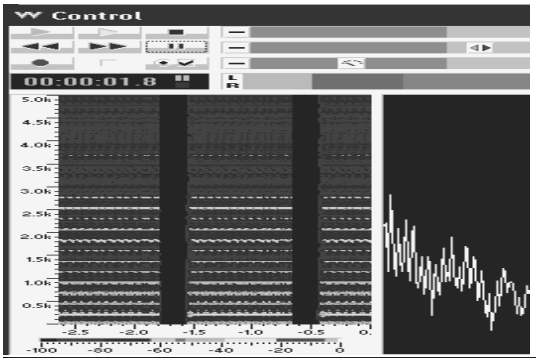


Figure 3. Determination of the F0 according to GoldWave™.

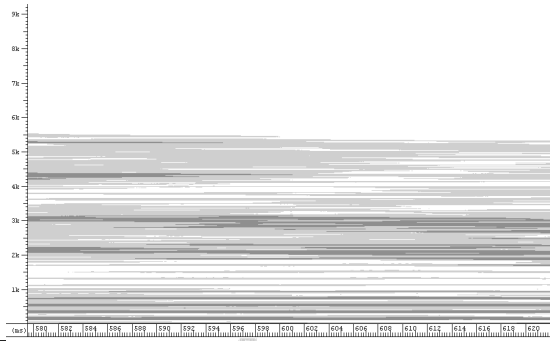


Figure 4. Determination of the formants according to Wasp™.

The differences in the results obtained with various tools reflect the theoretical limits of the formant parameters, as well as the capabilities of the various approximation methods used in the tools. These inconsistencies are one reason why the results we report should be considered preliminary, although we made every effort to obtain the results according to the best present knowledge.

4.3. Analysis and annotation of double subject sentences[2]

A hypothesis proposed by the first author is that most languages use specific syntactic structures to emphasize the emotional attitude of the speaker with respect to the listener, to the topic discussed, or to the context. For example, such a construction is the apposition. In the Romanian language, there is a special manner to carry meaning and subtle emotional attitude: the double subject construction. This type of syntactic structure does not exist in other European languages, but it is found in some Asian languages, like Japanese and Korean.

In double subject constructions, the semantic arguments of a predicate (the subject, the direct object and the indirect object) can be doubled, in the Romanian language. While the objects are commonly doubled by clitic pronouns (the doubling is sometimes mandatory, like in ***L***-am văzut ***pe Ion***), the subjects receive, occasionally, and mainly colloquially, a doubling pronoun (not only in Romanian, as [9] shows[3]). The doubling of the subject for the Romanian language is a controversial phenomenon: after having long been considered an apposition, Alexandra Cornilescu [3] has reopened the double-subject problem, Verginica Barbu [1] has modeled it using HPSG instruments, but until today, there is neither consensus on the topic, nor a detailed research, especially from a phonologic point of view. What supplementary information the pronouncing brings, from a descriptive perspective, in double subject phrases, remains an open question.

Some examples of sentences with double subject are:

(a) Vine ea mama!

(b) „A trecut el aşa un răstimp" (Sadoveanu M.)

We propose several interpretations for the double subject sentences, depending on the prosody:

- a neutral pronunciation indicates a non-determination of the time interval.
- a pronunciation accentuating the pronoun "el" indicates that the speaker has an idea about the time interval duration, and that the focus is on the passing of that time, and not on the duration.
- if the sentence is further developed, it can bring a further specification of the interval. For example, in the development „A trecut el aşa un răstimp de lung, încât...", the duration of the interval is specified in a certain way.

(c) O şti el careva cum să rezolve asta.

Different pronunciations may mark either the fact that the speaker does not know who is the person mentioned („el"), either that he knows, but has no intention of telling to the audience (when the accent is on

[2] This Section is almost a copy of the corresponding section in the submitted paper TEODORESCU H.N., FERARU M., Analyzing Emotions in Spoken Romanian Language, Proceedings of the Romanian Academy, Series A, Volume 8, 2007 (to appear)

[3] There is no definite explanation why not all languages accept the double subject structure. For these languages, in most of the cases, the doubling of the subject is realized as an apposition. Romanian language considers both double subject and apposition structures.

„careva”), or clearly specifies, by an apposition, who is envisaged, if the sentence is developed as „O ști el careva, Ionică, cum să rezolve asta”.

For the examples b) and c), the interpretation is that the information must be partially known by the auditorium (knowledge at the generic level, but not at the level of instantiation with a concrete individuality).

(d) Mama vine și ea mai târziu.

(e) Mama știe ea ce face.

Examples d) and e) are considered by some linguists [1] as constructions with doubled subject, while other authors [3] consider them particular structures of the Romanian language. We intend to compare these two sentences with equivalent sentences with single subject, in order to determine if there are differences in the prosodic realizations of the two groups. The existence of marked prosodic differences may be an argument supporting the concept of double subject. A more complete support would be offered by comparing such sentences with sentences including appositions. If, again, there would be marked differences between sentences with appositions and sentences with double subject, a “phonetic proof” of the double subject will be completed. The comparison of the prosody for double subject sentences with the prosody with corresponding simple subject sentences will be detailed in another paper.

We recorded a set of sentences bearing doubled subject for a comparative analysis of the prosody in sentences with doubled and simple subject and to observe the modifications involved by the doubling of the subject. The main objectives of our study are:

- comparing the prosody for simple subject and double subject sentences;
- clarifying the prosodic aspects and differences, if any, between the standard double subject constructions (examples (a)-(c)) and the non-standard structures (examples (d) and (e));
- study the modifications induced by the doubling of the subject in the sentence prosody;
- correlating the semantic charge with the pronunciation (different accentuation of the sentences with doubled subject);
- determine if the spoken language brings distinctions that may change the sentence behavior closer to a simple subject construction or a double subject one.

To perform a correlation between the semantic charge of a sentence and the representation of its subject, the speakers have recorded several variants of the sentences: neutral tone, accentuation of the doubling pronouns, focuses on the words next the pronouns, or the development of the sentences. The sentences have been annotated using the Praat™ software at several levels: phoneme, syllable, word, sentence, subject position, and articulation type.

In order to distinguish the differences between the spelling of the doubled subject and the spelling of the simple subject sentences, we calculated the values for the formants and the duration of the vowels using several programs: Praat™, Klatt analyzer™, GoldWave™ and Wasp™. The obtained results are discussed in the next section.

5. RESULTS[4]

5.1. Results of the analysis of speech with manifest emotions

The main general rules that we obtained based on the reduced number of cases we analyzed are listed below. The results are shown in the next tables; the notations are: “-“ means decrease the obtained values in sadness compared with happiness, “+” means increase, “±” means fluctuant, no conclusion can be derived. The symbol a1 represents the first a vowel in the word “mama”; a2 represents the second *a* in the same word.

- The obtained values for the F0 formant for all the persons decrease in sadness state compared with the happiness state (table 1). We notice that the Klatt Analyzer™ application “is not seeing” the F1 formant, we can distinguish easily with the application GoldWave™ the F0 and F1 formants, and with the application Wasp™, the F2 and F3 formants.

[4] This Section is based on the submitted papers TEODORESCU H.N., FERARU M., Analyzing Emotions In Spoken Romanian Language, Proceedings of the Romanian Academy, Series A, Volume 8, 2007 (to appear) and TEODORESCU H.N., TRANDABĂȚ D., The Prosody of the Double-Subject Sentences in Romanian, submitted, 2007, Revue Roumaine de Linguistique.

Table 1. The tendency for the F0, F1, F2 formants for the five persons (- =increase, + =decrease, ± =fluctuant)

Subject	F0			F1			F2		
	e	a1	a2	e	a1	a2	e	a1	a2
20048f	-	-	-	-	-	-	-	±	-
01312f	-	-	-	+	-	±	-	-	-
55555f	-	-	-	±	±	-	+	-	-
123456f	-	-	-	-	-	-	-	-	±
77777m	-	-	-	-	-	-	-	-	-

- The accentuated vowels (like the vowel ”i” from the word “vine” and the vowel “a”, first ”a” from the word “mama”) don’t offer important information compared with non-accentuated vowels (vowel ”e” from the word “vine’ and vowel “a”, the last from the word “mama”).
- The vowel “i” from the word “vine” has random values of the formants; therefore, it does not help in the emotion recognition.
- The obtained values of the formant F2, for the vowel “a” (the last “a” from the word “mama”) decrease in sadness compared with happiness and the values of the F1 formant have the tendency to decrease too.
- The obtained values of the formants F1 and F2, for the vowel “a” (the first “a” from the word “mama”) have the tendency to decrease in sadness compared with happiness states.
- The obtained values of the formants F2, for the vowel “e” (from the word “vine”) have the tendency to decrease in sadness compared with happiness states.

The particular rules obtained are:

- For the person 20048f (table 2 and table 3), the values for the F1 formant have the tendency to decrease in the sadness state compared with the happiness state, for all the considered vowels.

Table 2. The values of formants, in Hz, using Wasp™ application for sentence “Vine mama”, person #20048f

Parameters	Happiness			
	i	e	a1	a2
F0	100-200	300-400	200-300	200-400
F1	400-600	600-800	600-700	800-1000
F2	2400-2600	900-1000	900-1000	1400-1600
F3	3300-3500	1800-2200	2500-2600	3000-3300

Parameters	Sadness			
	i	e	a1	a2
F0	100-200	200-300	150-250	200-300
F1	300-400	600-700	600-700	800-900
F2	2200-2400	1800-2000	1200-1400	1300-1500
F3	3000-3200	3000-3200	2900-3200	3200-3400

- The values of F2 formant for the vowel “e” (from the word “vine”) and for the vowel “a” (the last “a” from the word “mama”) have the tendency to decrease in the sadness state compared with the happiness state.

Table 3. The values of formants, in Hz, using Klatt™ application for sentence “Vine mama”, person #20048f

Parameters	Happiness			
	i	e	a1	a2
F0	264	352	264	176
F1	???	951	951	1021
F2	2659	2290	1444	1462
F3	3293	3311	3311	3223

Parameters	Sadness			
	i	e	a1	a2
F0	176	229	176	141
F1	???	845	933	1021
F2	2747	1955	1374	1444
F3	3399	3082	1744	3276

- For the "a" vowel (the first "a" from the word "mama") the values of F2 formant decrease according to GoldWave™, increase according to Praat™, and are constant with Wasp™ and Klatt analyzer™.
- For the person 55555f (table 4), the values of the F1 formant for the vowel "a" (the last "a" from the word "mama") decrease in sadness compared with happiness, except for the values obtained with Praat™.
- For vowel "e" and the first "a" (from the word "vine") the values of F1 seem constant according to GoldWave™ and Klatt analyzer™, decrease according to Wasp™ and increase in Praat™.
- For the last "a", the values of F1 formant decrease according to all applications except for Praat™, where the values for F1 increase.

Table 4. The tendency for the F0, F1, F2 formants for the person #55555f

55555f	F0			F1			F2		
	e	a1	a2	e	a1	a2	e	A1	a2
GoldWave™	-	-	-	±	±	-	+	±	-
Wasp™	-	-	-	-	-	-	+	-	-
Klatt analyzer™	-	-	-	±	±	-	±	-	-
Praat™	-	-	-	+	+	+	-	-	+

- The values of F2 formant for the vowels "a" (the first and the last "a" from the word "mama") decrease in the sadness state compared with happiness state according to Wasp™ and Klatt analyzer™ and increase in Praat™ for the last "a".
- For vowel "e", the values of F2 formant increase according to GoldWave™ and Wasp™ application and decrease according to Praat™ application.
- For the person 01312f (table 5), for the vowel "e", the values of F1 formant increase according to GoldWave™, Praat™ and Klatt Analyzer™ application and decrease according to Wasp™ application.

Table 5. The tendency for the F0, F1, F2 formants for the person #01312f

01312f	F0			F1			F2		
	e	a1	a2	e	a1	a2	e	a1	a2
GoldWave™	-	-	-	+	-	±	-	-	-
Wasp™	-	-	-	-	±	±	+	-	-
Klatt analyzer™	-	-	-	+	-	-	-	+	-
Praat™	-	-	-	+	-	-	-	-	-

- The values of F2 formant for vowel "e" have the tendency to increase according to Wasp™ application and to decrease according to GoldWave™, Praat™, and Klatt Analyzer™ applications.
- For the vowel "a" (the first and the last "a" from the word "mama"), the values of F2 formant have the tendency to decrease in the sadness state compared with happiness state, except for the values obtained with Klatt Analyzer™ for the first "a".
- For the person 123456f (table 6), the values of F1 formant for vowel "e" and the two vowels "a" decrease according to all applications.

Table 6. The tendency for the F0, F1, F2 formants for the person #123456f

123456f	F0			F1			F2		
	e	a1	a2	e	a1	a2	e	a1	a2
GoldWave™	-	-	-	-	-	-	+	-	-
Wasp™	-	-	-	-	-	-	-	±	-
Klatt analyzer™	-	-	-	-	-	±	-	-	+
Praat™	-	-	-	-	-	-	-	-	+

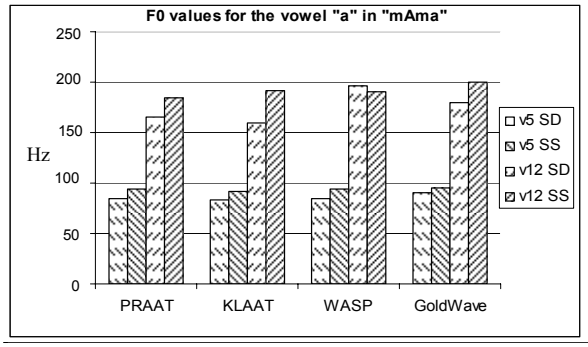
- For the first “a” and the vowel “e”, the values of F2 formant have the tendency to decrease in the sadness state compared with happiness state, except for the values obtained with GoldWave™ for the vowel “e” and the values obtained with Wasp™ for “a”.
- The values for the last “a” decrease with GoldWave™ and Wasp™ and increase with Praat™ and Klatt Analyzer™.
- For the person 77777m (table 7), the values of F1 formant for all the vowels decrease with all four applications except for the values obtained with Praat™ for the second “a”.
- The values of F2 formant for all the vowels decrease with all four applications except for the values obtained with Praat™ for the second “a”.
- The values of the F0 formant for the second “a” is undefined according with Praat™.

Table 7. The tendency for the F0, F1, F2 formants for the person #77777m

77777m	F0			F1			F2		
	e	a1	a2	e	a1	a2	e	a1	a2
GoldWave™	-	-	-	-	-	-	-	-	-
Wasp™	-	-	-	-	-	-	-	-	-
Klatt analyzer™	-	-	-	-	-	-	-	-	-
Praat™	-	-	undefined	-	-	+	-	-	+

5.2. Phonetic analysis of sentences with double subject

As we already said, the hypothesis that motivated this analysis is that the double-subject constructions are related in a specific way to the emotion and inter-relationship representation. To test this hypothesis, we have analyzed therefore the values of the formants and duration of the vowels for five subjects (three female and two male) from our database[5] for the sentence “`Vine mama`” (simple subject) vs. “`Vine ea mama`” (doubled subject). The recorded subjects belong to the same age bin (25-35 years), have higher education, and came from the same geographic area, as already mentioned in section 4.1.



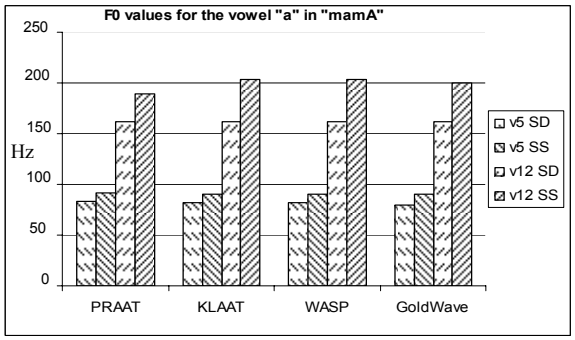
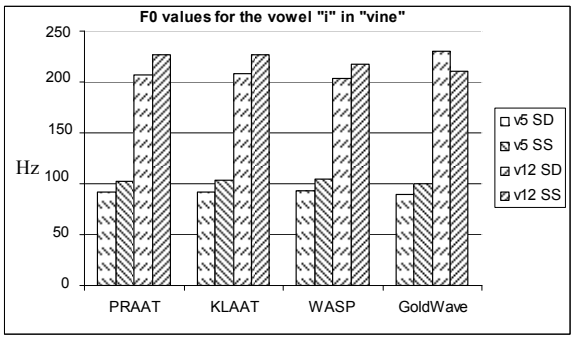


Figure 5a. The F0 evolution for the vowel “i” in “Vine mama”. **Figure 5b.** The F0 evolution for the vowel “e” in “Vine mama”. (SD = double subject; SS = single subject)

Figure 5 presents the analysis of the F0 for two subjects (subject 5 – male and subject 12 – female). The values have been computed using four programs already mentioned (Praat™, Klatt Analyzer™, GoldWave™ and Wasp™). In the graphics in figure 5, the first two bars for each program represent the values for the male subject (double subject sentence vs. simple subject sentence), while the last two are the F0 values for the female subject. When looking at the F0 values for the vowels of the analyzed sentence (namely the vowel *i* in Figure 5a, *e* in Figure 5b, the first *a* in mama in Figure 5c and the second *a* in Figure 5d), we noticed that all the four programs show an increasing of the F0 values, for all the vowels in the sentences with simple subject.

[5] We realize that an analysis over five subjects can have no claims on generality, but it represents a good start for the pioneering Romanian double subject phonetic analysis.



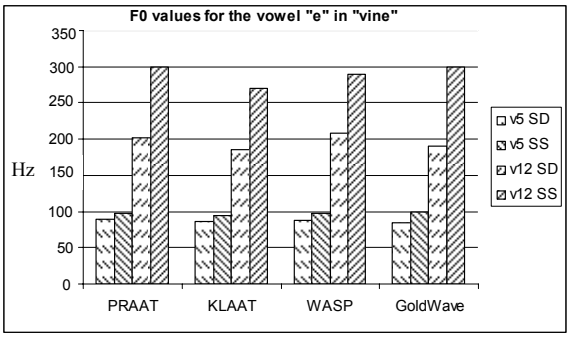


Figure 5c. The F0 evolution for the vowel "a1" in "Vine mama". **Figure 5d.** The F0 evolution for the vowel "a2" in "Vine mama".

We have computed the mean of the values presented in figure 5 and the standard deviation, in order to see how significant the increasing of the pitch is. The obtained values are presented in table 8. We observed that the major differences in the mean pitch values were for the unaccented vowels. Thus, the vowel *e* in *vine* decreases in the simple subject structures by 93.5 Hz for subject 12 and by 10.3 Hz for subject 5, while the last *a* in *mama* decreases by 36.8 Hz for the female subject and by 8.8 Hz for the male speaker. In the accented vowel case, the decreasing is lower (8.3 Hz for the *i* in *vine* for subject 12 and 10.1 Hz for subject 5, and respectively 16.6 Hz for the first a in mama for subject 12 and 8.3 Hz for subject 5). A possible cause that deserves a further detailed analysis may be the location of the unaccented vowels at the end of the words. The growing tendency of the F0 values is evident also for the other subjects. For the same sentences, the mean values obtained for the pitch for the vowel *a* are shown in table 9.

Table 8. The mean for the F0, in Hz, and the standard deviation

	Vine ea mama				Vine mama			
	I	e	mAma	mamA	i	e	mAma	mamA
	F0	F0	F0	F0	F0	F0	F0	F0
subiect 12 ST DEV	12	10	17	0	8	14	6	7
subject 12 MEAN	212	196	175	161	220	290	192	199
subiect 5 ST DEV	1	2	3	1	2	2	1	0
subject 5 MEAN	92	87	85	82	102	98	94	91

Table 9. Values for the last two vowels in the subject of "`Vine mama`" vs. `Vine ea mama`"; frequency in Hz and time in s

Subject	**Vine ea mama**				**Vine mama**			
	a1 in mAma		**a2 in mamA**		**A1 in mAma**		**a2 in mamA**	
	F0	Duration	F0	Duration	F0	Duration	F0	Duration
subject 1	200	0.086	215	0.082	211	0.103	223	0.098
subject 2	189	0.101	179	0.137	215	0.067	206	0.098
subject 12	162	0.099	162	0.135	188	0.127	196	0.136
subject 5	84	0.094	83	0.084	93	0.122	91	0.138
subject 7	76	0.080	71	0.079	77	0.089	82	0.070

The data recordings we have annotated are not sufficient to draw statistically pertinent conclusions for the vowels duration changes. For now, we can only say that the tendency to increase or to decrease the duration of the vowels seems to be preserved similar in both construction types. Thus, if, for example, in a double subject construction subject 12 has increased the duration of the last vowel *a*, this increasing is to be found also in the simple subject structure. This goes similar for subject 1 with the decreasing tendency. However, subject 5 disobeys this rule, and the values obtained for subject 7 are too close to be relevant.

The values of the first formant are presented in figure 6. Notice that, for the first *a* in the sentence (figure 6a), three subjects have increased their F1 values, while two have decreased them. For the second *a* (figure 6b), there is an inverse tendency: three values decrease, while two increase. We can therefore determine any tendency, until more subjects are analyzed. However, we notice that the increasing or the decreasing intra-speaker tendency is preserved by the speakers for the two vowels, with the exception of the last two speakers (male). We need to determine through for further analysis if this exception – that looks unnatural – regarding male speakers is valid.

The values of the other formants show no regularities. We envisage therefore two directions for further analysis of the upper formants: one involves collecting more data and continuing the presented approach, the other intends to use a decision instrument to observe if there are regularities that haven't been noticed by human annotators.

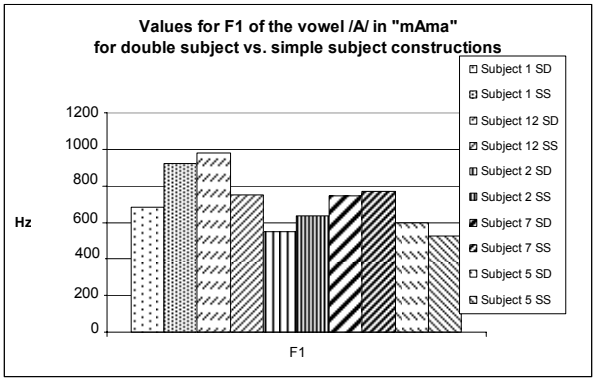


Figure 6a. The F1 evolution for the vowel "a1" in "Vine mama".

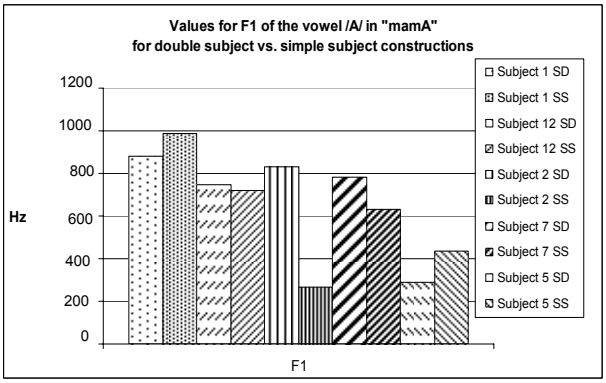


Figure 6b. The F1 evolution for the vowel "a2" in "Vine mama".

After analyzing several double and simple subject constructions, we believe that the hypothesis we have started with is at least partly proved. The inter-relationship between the speaker and the auditorium becomes visible by the remark that the speaker has already a pattern (referring to the pith contour) when beginning to pronounce a structure (higher pitch for simple subject structures, lower pitch for double subject).

5. CONCLUSIONS

The reported research had the general but somewhat diffuse aim of determining whether there are prosodic features that support the interrelationship theory of language. The choice of the linguistic and paralinguistic features selected for the analysis has been motivated by the analysis of manifest, intentional emotions and of the indirectly expressed emotional charge in specific constructions, namely in sentences with double subject.

For sentences uttered with manifest emotional charge in the Romanian language, we found that most informative regarding the emotions is the change of the pitch. This conclusion is compatible with some findings reported for other languages. In contrast, we found that the accented vowels do not carry significantly more emotional information than the non-accented vowels; rather, the opposite is true. This conclusion is a departure from findings by other authors, for different languages. We need to further analyze this issue to determine its validity for a larger number of sentences and subjects. We also found that some higher formants, F1 and F2, in both accented and non-accented vowels, are also essential in conveying emotional information, at least from the perspective of the voice personality of some speakers.

To test the hypothesis of emotional attitude information conveyed by specific syntactic structures, we have analyzed the influence of the double subject construction on the prosody in the Romanian language. The analysis involved short sentences which are parallel in the sense that they are identical up to the use of single or double subject constructions. The main conclusion which can be derived from this preliminary research is that the pitch is changed in a consistent way in the two types of sentences. Namely, the pitch is lower in double subject constructions than in single subject sentences. Also, the frequency of the first formant is different in the two constructions, but the way of changing and the change amplitude depend significantly on the speaker. The vowel duration also may change, but there is no a single type of change; yet, the intra-subject change tends to be consistent. These findings show a good degree of support to the idea that the constructions some researchers argue to be double subject sentences are semantically different compared to single subject constructions. Future analysis will be devoted to contrast the prosody of parallel sentences with double subject constructions, appositions, and simple (i.e., no apposition, no double subject) constructions in the Romanian language.

Regarding the available speech analysis tools, we conclude that no tool provides irrefutable results. While we used four tools and compared the results, no one is significantly better than the others. We have indicated a methodology to choose a stable section of the vowels for the analysis, to improve consistency in measurements, but even using this methodology, the lack of good formant extractors restricts today possibilities of obtaining high confidence in the results.

Overall, the individual findings and conclusions reached in this research support the theory of interrelationship expression through prosodic and paralinguistic information.

Acknowledgments. We acknowledge the partial support of the Romanian Academy "priority research" grant "Sisteme cognitive" and of the CEEX grant "Sistem automat de diagnostic paraclinic în sindromul disfuncțional al sistemului stomatognat".

We thank the referees for the very helpful corrections.

REFERENCES

1. BARBU V., Construcții cu subiect dublu în limba română actuală. O perspectivă HPSG, în volumul Aspecte ale dinamicii limbii române actuale, **Vol**. II, Ed. Universității Bucureşti, pp. 73-79, 2003.
2. BULUTI M., NARAYANAN S.S., SYRDAL A.K., Expressive speech synthesis using a concatenative synthesizer, Proc. ICSLP, 2002.
3. CORNILESCU A., The double subject construction in Romanian. Notes on the syntax of the subject, Revue Roumaine de Linguistique, no. 3-4, 1997.
4. DOUGLAS-COWIE E. , CAMPBELL N., COWIE R. and ROACH P., Towards a new generation of databases, Speech Communication, **Vol**. 40, pp. 33-60, 2003.
5. FAKOTAKIS N., Corpus design, recording and phonetic analysis of Greek emotional database, Proc. The Language Resources and Evaluation Conference, LREC'04, 2004.
6. IRIONDRO I., GAUS R., RODRIGUES A., *et al*., Validation of an acoustical modeling of emotion expression in Spanish using speech synthesis techniques (http://serpens.salleurl.edu/intranet/pdf/239.pdf).
7. http://kom.aau.dk/~tb/speech/Emotions/des.pdf.
8. http://pascal.kgw.tu-berlin.de/emodb/.
9. MASAHIRO O., Analyzing Japanese double-subject construction having an adjective predicate, Proceedings of the 16th Conference on Computational linguistics - Volume 2, Copenhagen, Denmark, pp. 865–870, 1996.
10. ODERLMAN R., POEL M., HEYLEN D., Emotion annotation in the AMI project, February, 2005.
11. TEODORESCU H.N. (coordinator), TRANDABĂȚ D., FERARU M., GANEA R., PISTOL L., ZBANCIOC M., VERBUȚĂ A., HNATIUC M., The research project "Sunetele Limbii Române", http://www.etc.tuiasi.ro/sibm/romanian_spoken_language/index.htm.
12. TEODORESCU H.N., A proposed theory in prosody generation and perception: the multi-dimensional contextual integration principle of prosody", Proc. The 3rd Conference on Speech Technology and Human-Computer Dialogue "SpeD 2005", Cluj-Napoca, Romania, May 13-14, 2005, Burileanu, C. (Coordinator), Trends in Speech Technology, Romanian Academy Publishing House, Bucharest, Romania, pp. 109-118, 2005.
13. TEODORESCU H.N., FERARU M., TRANDABAT D., Nonlinear assessment of the professional voice "pleasantness", Biosignal, 28-30 June 2006, Brno, Czech Republic, pp. 63-66, 2006.
14. LOSCOS A., BONADA J., Emulating rough and growl voice in spectral domain, Proc. of the 7th Int. Conference on Digital Audio Effects (DAFX-04), Naples, Italy, October 5-8, 2004. (http://www.iua.upf.es/mtg/publications/DAFX04-aloscos.pdf, accessed 12 Nov. 2006).
15. SUN X., Pitch determination and voice quality analysis using subharmonic-to-harmonic ratio. http://www.ling.northwestern.edu/~jbp/sun/sun02pitch.pdf. (accessed 12 Nov., 2006).
16. www.praat.org.
17. www.speech.cs.cmu.edu/comp.speech/Section5/Synth/ klatt.kpe80.html.
18. www.goldwave.com.
19. www.wasp.dk.