The Annotated Emotional Speech Database for the Romanian Spoken Language

Monica Feraru*

Horia-Nicolai Teodorescu*

*"Gheorghe Asachi" Technical University of Iasi

Overview

- Introduction
- An analysis of the current emotional speech databases (ESDs)
- The emotional speech database – SRoL
- Results
- Discussion and conclusions

Introduction

- The emotional speech database is a corpus of human speech pronounced under different emotional states.
- The speech emotion recognition and analysis is useful for:

- learning to improve communication
- human-computer speech interaction
- security
- medical applications
- video-games and interactive TV
- teachers, in the study of the Romanian language, etc.

An analysis of the current emotional speech databases (ESDs)

- A speech database is a collection of files with sounds, differently structured depending on its purposes.
- The most common emotions analyzed in existing emotional speech databases (ESDs) are fury, sadness, happiness, fear, disgust, joy, surprise, boredom, and stress.
- The English and German are the dominant languages used in the emotional speech databases.

Emotions in different languages (Abelin, Allwood, Goteborg University)

Emotion	Swedish	English	Finnish	Spanish	Mean%
Happy	92	66	65	79	76
Surprised	83	43	83	78	72
Sad	69	100	70	91	83
Angry	83	75	87	96	85
Afraid	66	42	70	78	64
Shy	45	16	39	47	37
Dominat	81	67	78	79	76
Disgusted	12	0	35	39	22

BELFAST-ESD- for English Language

- emotions: anger, fear, happiness, neutrality, and sadness.
- recordings made by 40 volunteers with age between 19 and 69 years, in a studio.
- they read 5 passages of 7-8 sentences.
- they used the FEELTRACE software developed by Cowie et al., at Queens University Belfast.
- they have two types of recordings: one is natural and one is semi-natural.
- the database is used for emotion speech synthesis [18], [19].

UG-G-ESD – for German Language

- emotions: anger, humor, indifference, stress, sadness.
- recordings made by 109 passengers in an airport.
- the goal is to determine differences in emotional speech perception between people from different countries.
- this ESD have also video capture; the database is used for emotion perception by humans [20].

Slovenian-ESD- for the Slovenian Language

- emotions: disgust, surprise, joy, fear, anger, sadness,.
- the speakers are (male and female) actors.
- recordings in a quiet environment, high quality microphone.
- the recordings are in English (186 utterances per emotion) and Slovenian Language (190 utterances per emotion).
- they recorded isolated words, affirmative and interrogative sentences (short, medium, and long) and a text passage at 48 kHz/16bit. The duration of a session is about four hours [21].
- they used a laryngograph and video analysis.
- Goal: analysis and synthesis of emotional speech.

ATR-J-ESD – for the Japanese Language

- they studied 8 emotions.
- the recordings are made by 100 native speakers (50 female, 50 male) and one professional radio speaker;
- the professional speaker read 100 neutral words.
- the ordinary people were asked to mimic the manner of the professional speaker.
- the database is used for automatic emotion recognition with ASR applications [9].

DES-ESD – for Danish Language

- the emotions are: angry, happy, sad, surprise, and neutral.
- the speakers are four actors.
- the sounds were recorded in an acoustically damped sound studio at Aarhus theatre.
- the validation commission has twenty persons (native speakers with the ages between 18 and 58 years, didn't offer more information about speakers) who identified 67% emotions (surprise and happiness states were often confused as well as neutral and sadness state).

DES-ESD – for Danish Language

- the speaker profile : name, age, sex, for how many years have worked as an actor, height, weight, smoker.
- the questionnaire of the listener: the listener name, gender, age, did you find the task of identifying the emotion, which factors made you belief that the speaker was neutral, surprised, happy, sad, angry; additional remarks to the listening test.
- the database is used for emotions speech synthesis [16].

RUSSLANA-ESD – for Russian Language

- the number of speakers is 61 (12 are native male from Russia) with ages between 16 and 28 years.
- the emotions are: surprise, happiness, anger, sadness, and fear.
- they recorded 3660 sentences.
- the studied parameters are: energy, pitch, and formants curves.
- the database is used for automatic emotion recognition with ASR applications [21].

S-ESD – for Sweden Language

- the listeners are 35 native Swedish speakers and 78 Swedish immigrants.
- they didn't offer information about what emotions they are studied, about the protocol of recordings, about the speakers [20].
- the goal of the emotional database is emotion perception by human.

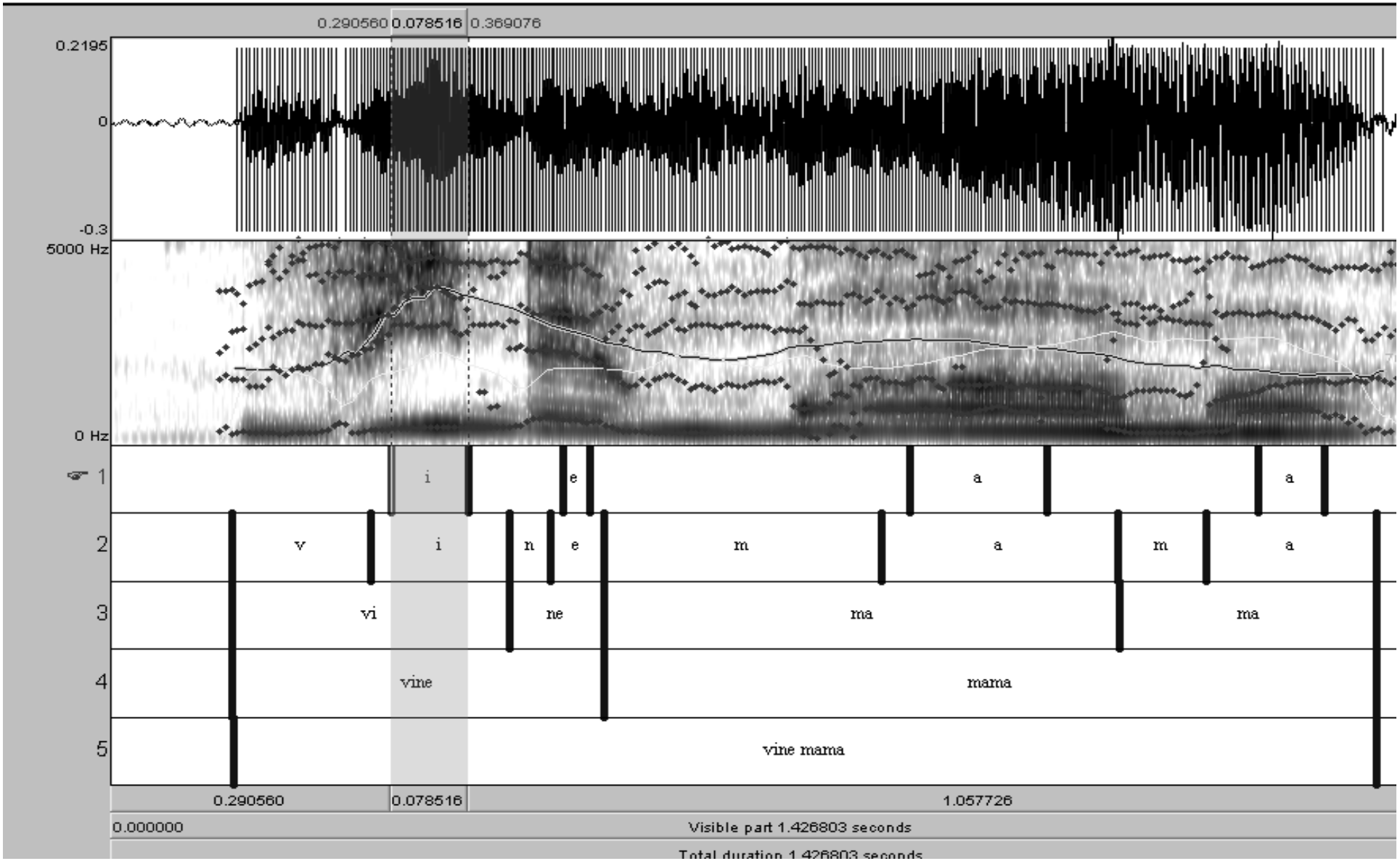
Method of Building and Content of the Emotional Speech Section

- The SRoL contains files of sounds, syllable and words, files with vowels, consonants, diphthongs, hiatuses, single and double-subject sentence, and an emotional speech database.
- the emotions are: happiness, sadness, anger, and neutral tone.
- the recordings were made by persons with ages between 25-35 years.
- the recordings are accompanied by the speaker profile and by the questionnaire concerning vocal pathology and objective factors for every speaker.

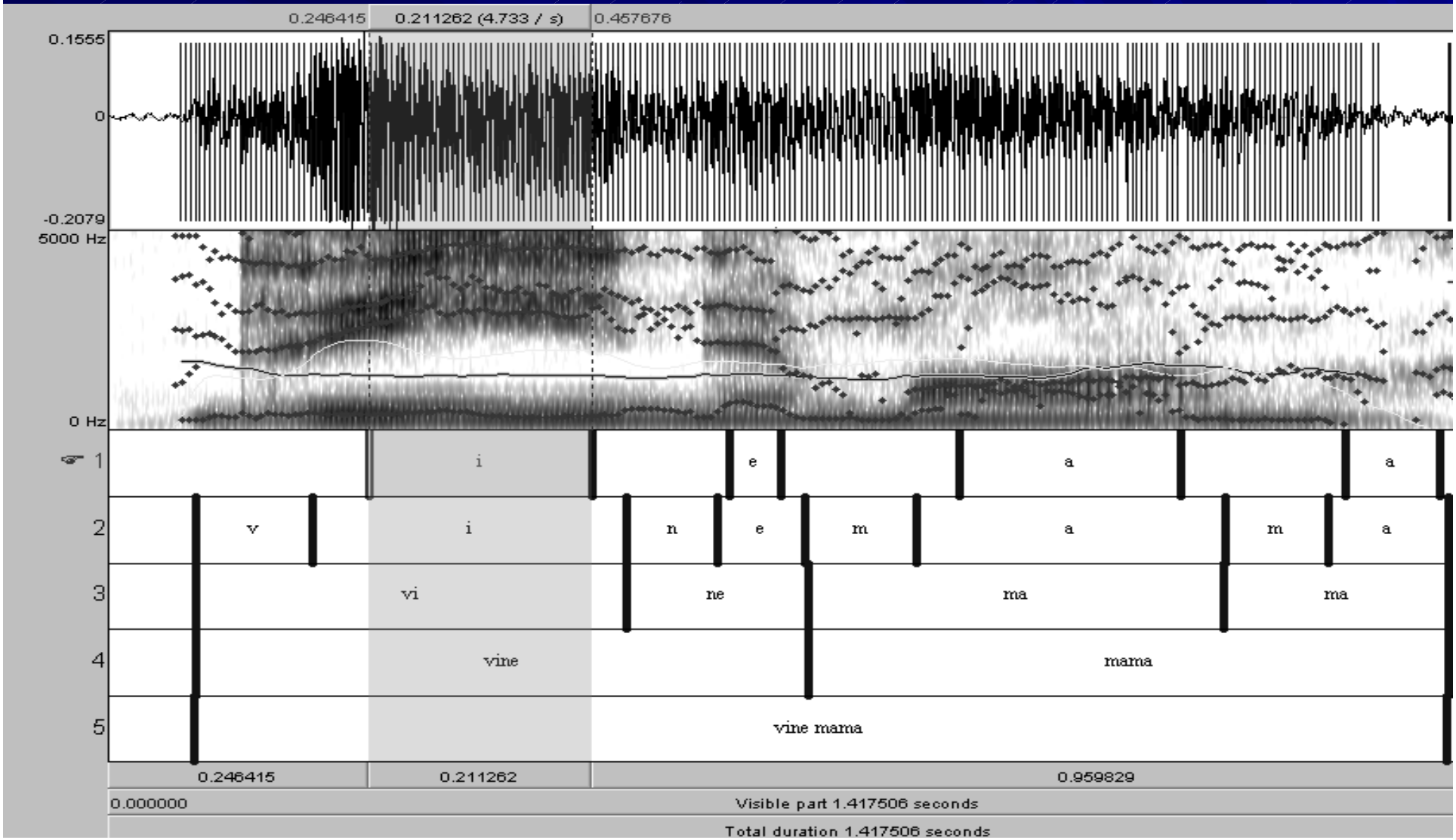
Emotional Speech Section in SRoL

- standard prosody which include the following intonations: interrogative, exclamatory and neutral tone;
- three basic emotional states (joy, sadness and fury) and neutral tone, and from this page we can go to other page with more emotions (hate, optimism, pessimism, sorrowed, etc);
- double subject (DS), simple subject (SS) and appositions (APP);
- documentations.

Example of the annotation – the happiness state



Example of the annotation – the sadness state



Results for happiness compared with sadness states

Subject	Fo			F1			F2		
	a	ea	ă	a	ea	ă	a	ea	ă
20048f	-	-	-	-	-	-	-	+	-
01312f	-	-	-	+	-	-	-	-	-
55555f	-	-	-	±	+	-	+	-	-
123456f	-	-	-	-	-	-	-	-	±
77777m	-	-	-	-	-	-	-	-	-
263315m	-	-	-	-	-	-	-	-	-
14411f	-	-	-	-	-	+	-	±	-
26653	-	-	-	-	-	-	-	-	±

Comments on the Method and Results (I)

- didn’t have a stable validation commission;
- didn’t use any video capture;
- didn’t make more analyses like EEG, EMG, electroglottogram, etc;
- we choose sentences that can express various emotional states.

Comments on the Method and Results (II)

- The states of happiness and sadness, on one side, and fury from sadness, on the other side can be easily distinguished in all cases.
- There is confusion between exclamation and happiness state, between pessimism and sadness state, and between the happiness and the joy state. It is more difficult to distinguish between happiness and fury.

Comments on the Method and Results (III)

- In the sadness state, the majority are speaking slowly, and in the fury states the persons are speaking in hurry.
- In the sadness and in the happiness states, a few speakers make short pauses between words, or say the word slowly.
- In fury state, there is no pause between words, and the talking is quick and strong, with enforced, accentuated vowels, highly energetic.
- The durations of the emotional states are in the following order: the fury state, the happiness state, the neutral tone and the sadness state – speaking in the last state being the longer, for the same word

Discussion and conclusions

- The site is useful for doctors, linguists, behaviorists, for the psychologists, persons who don't know the Romanian language, etc.
- The database may be helpful in improving voice recognition systems based on acoustical features.
- The content of the site can be freely 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.

Conclusions and further work

- As a general conclusion, we think that there is similarity of emotion representation in European languages, disregarding their particular roots.
- Further work is needed to add new recordings to the database, to refine the analysis and to document the emotion evaluation panel.
- The statistical information is growing day by day and will be adding on the website.

Thank you.