

# THE EMOTIONAL SPEECH DATABASE CREATION AND PROCESSING TOOLS

Horia-Nicolai TEODORESCU, Monica FERARU

Technical University of Iasi

We recorded a set of sentences with two emotional states: happiness and sadness. 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) [8].

By now, there is no standard model for the emotional annotation process [1]. The sentences have been annotated using the Praat software [2] at several levels: phoneme, syllable, word and sentence. In this paper, the analysis was made only for the sentence “Vine mama” pronounced by three persons, three times each. Our goal has been to make the difference 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 [2], Klatt analyzer [3], Goldvawe [4], and Wasp [5]. The purpose is to discriminate based on these values, the emotional states of happiness and sadness. We obtained general and particular rules which are discussed in the section on results. We have been confronted with several problems in determination the formants, namely with large disagreements between values provided by different applications

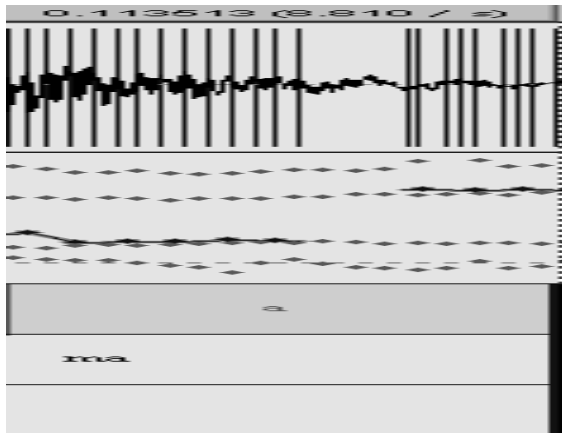


Figure 1. Determination of the F0 with Praat™ application analyzer™

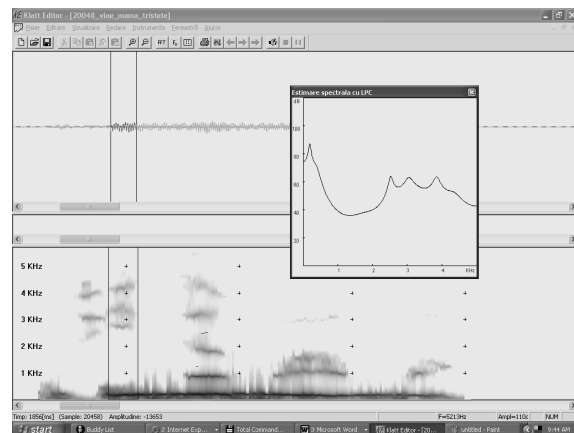


Figure 2. Determination of the formants with Klatt

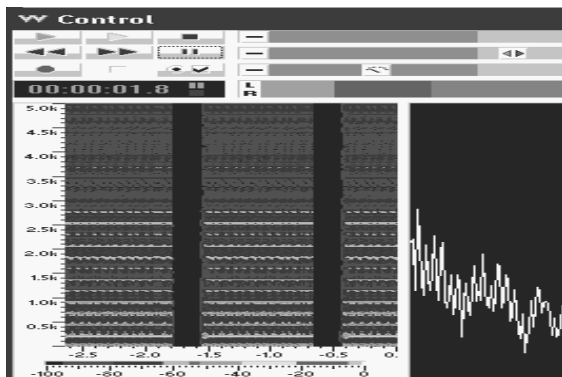


Figure 3. Determination of the F0 according to GoldWave™ Wasp™

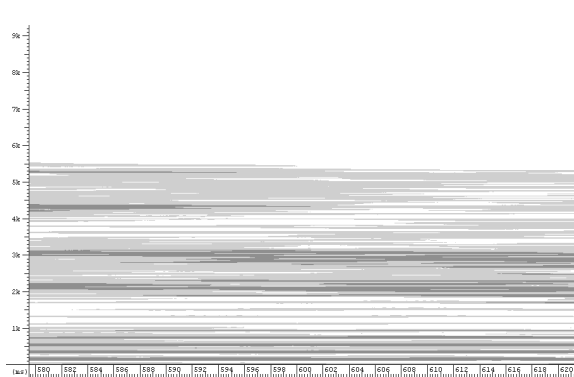


Figure 4. Determination of the formants according to

For example, there were cases where according to Praat™, on some segments there the fundamental frequent is not defined (see figure 1) while Wasp™ or Klatt analyzer™ identifies a pitch on those segments. According to Klatt analyzer™, we couldn't see the F1 formant, as presented in figure 2.

In the 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 are well documented in the medical literature as well as in the recent info-linguistic literature, [6] [7], determines a lack of significance of the parameters defined in the frame of the linear theory of speech analysis. The tools used reflect these limits. This is one reason why the results we report should be considered preliminary.

### Emotion analysis

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 five tables and our notations used are: - means decrease the obtained values in sadness compared with happiness, + means increase, ± means constant.

- 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.

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

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 and didn't help to 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 using Wasp™ application for sentence “Vine mama”, 20048f

Parameters	Happiness			
	i	e	a1	a2
F0	100-200	400-450	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 using Klatt analyzer™ application for sentence “Vine mama”, 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 55555f person

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 01312f person

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™ application.
- 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”.

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

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

- For the person 123456f (table 6), the values of F1 formant for vowel “e” and the two vowels “a” decrease according to all applications.
- 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 77777m person

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

## REFERENCES (selected)

- [1] ODERLMAN, R., POEL, M., HEYLEN, D., Emotion annotation in the AMI project, February, 2005  
 [2] <http://www.praat.org>  
 [3] <http://www.speech.cs.cmu.edu/comp.speech/Section5/Synth/klatt.kpe80.html>  
 [4] <http://www.goldwave.com>  
 [5] <http://www.wasp.dk>  
 [6] 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.iaa.upf.es/mtg/publications/DAFX04-aloscos.pdf>, accesat 12 nov. 2006)  
 [7] SUN, X., Pitch Determination and Voice Quality Analysis Using Subharmonic-to-Harmonic Ratio. <http://www.ling.northwestern.edu/~jbp/sun/sun02pitch.pdf>. (accesat 12 nov. 2006)  
 [8] [http://www.etc.tuiasi.ro/sibm/romanian\\_spoken\\_language/index.htm](http://www.etc.tuiasi.ro/sibm/romanian_spoken_language/index.htm)

*This material is a section of the paper “STUDIES ON THE PROSODY OF THE ROMANIAN LANGUAGE: THE EMOTIONAL PROSODY AND THE PROSODY OF DOUBLE SUBJECT SENTENCES”, authors: H.N. Teodorescu, M. Feraru, D. Trandabat, to be presented within the Conference Speech Technology and Human – Computer Dialogue, SpeD 10-11 May 2007.*

### Copyright

The results presented here were obtained, in cooperation with Horia-Nicolai Teodorescu, by Monica Feraru, on the problem of emotional prosody, as proposed by the first author, who also proposed the methodology. The results can be used only with the writing agreement of the authors. You can contact us on the addresses [hnteo2004@yahoo.com](mailto:hnteo2004@yahoo.com) and [mferaru@etc.tuiasi.ro](mailto:mferaru@etc.tuiasi.ro) in order to demand an agreement. Without the writing agreement, we reserve the right to apply the law