

Correlations of the Vowel Formants in Romanian and Emotional Effects on Formants

Ștefan – Andrei Ghelțu

‘Gheorghe Asachi’ Technical University of Iași

ETTI, Iași, România

stefan.gheltu@gmail.com

Abstract – We have investigated the influence of emotions on the level of correlation of the vowel formants in the Romanian language during dynamic pronunciation. We have used in the analysis the sentence */vine mama/* (mother is coming), which is a commonly used phrase in the daily speech, and we have followed the variations of the vowel correlation, depending on the expressed emotion (joy, sadness, anger and emotional neutrality). The main conclusion that we have drawn is that, regarding emotions with negative valence (sadness and anger), the higher order formants (F3 and F4) show a significantly higher correlation level compared to lower order formants (F1 and F2). Moreover, we have found that the variation of the correlation level of the pairs of formants is more homogeneous when emotional neutrality is expressed.

Keywords - formants; vowels; emotions; fluent speech; Romanian language.

I. INTRODUCTION

Following the expansion of the performance and applicability of automatic speech recognition systems, the interest of the scientific community has focused on identifying ways to recognize individuals' emotions by analyzing the voice signal. One of the main approaches in this research direction is the correlative analysis of vowel formants from several speech types that are representative of certain emotional states of the speakers. For example, in [1] the authors performed an analysis of the changes in voice signal characteristics during psychological stress conditions, while [2] presents the voice signal fluctuations associated with depression symptoms. Moreover, [3] summarizes the theoretical contributions regarding the most common normalization procedures used for vowel formants. In the case of the Romanian language, the first studies of this nature have been carried out recently, having all been published in the last decade [4], [5], [6]. To date, some of the results [7] have showed that there are no significant differences between the correlation level of the formants for the vowels of the Romanian language in the case of female participants compared to male participants. In other words, the gender variable does not influence the correlation level of the formants. At the same time, in the case of the emphasized pronunciation, the formants of the vowels */a/*, */e/*, */o/*, */u/*, */ă/* and */â/* exhibit a significantly higher correlation level than that of the vowel */i/*. On the other hand, as it is shown in [8], the results have highlighted the fact that the correlation level of the formants is strongly influenced by the context in which the vowel is uttered. For example, it is specified that in the

structure */vine mama/* (mother is coming), the correlation level of the formants in the stressed vowel */a1/* is significantly lower, compared to the non-stressed vowel */a2/*.

A similar study [6] analyzed the extent to which changes in the variations of formants occur when the speaker's voice is associated to one of the following three fundamental emotions: joy, sadness and anger compared to emotionally neutral speech. The results obtained by applying the CORR function and calculating the linear regression coefficient [6] have led us to the following conclusions: 1) there were no significant changes to the F4 formant depending on the emotions expressed by the speakers; 2) F0 (pitch) had significantly higher values for joy, regarding all vowels; 3) in what sadness is concerned, we observed a slight decrease in fundamental frequency values. The results of a more recent study [9] have shown that, regarding the emotionally neutral speech, the third formant (F3) is largely determined by the correlation of inferior formants (F1 and F2). It has also been discovered that F1 and F2 have a lower correlation level compared to higher order formants (F3 and F4).

In this study, we investigate the extent to which the emotions of the speakers, namely joy, sadness and anger, influence the correlation of the formants for the vowels of the Romanian language within a structure frequently used in the common speech. We are interested in the influence attributed to the stressing of vowels and their position in the structure of the word, regarding the correlation degree of the formants. The article is comprised of four parts: the first one aims at summarizing the contributions of several recently published studies, the second part presents the applied method, the third section exhibits the results obtained by the research, and in the fourth section we have discussed the conclusions we have reached, certain limitations we have experienced during this study and future research directions.

II. METHOD

In the present research we used records from the SRoL [10] and [11] database, this being the practical foundation of several similar thematic studies [6], [7] and [8] that have been previously published. The files had a sampling frequency of 24 kHz, which allowed for greater accuracy in extracting the relevant parameters for subsequent statistical processing. In this study, we used 25 files (18 with female voice and 7 with male voice) which consisted of the audio recordings of the phraseology structure */vine mama/*,

in accordance with the following basic emotions: joy, sadness, anger and the state of emotional neutrality, which served as a control condition. Nine speakers aged 25-35 years, coming from the north-eastern part of Romania participated to this study, they all had a higher education degree and did not show evidence of speech pathology.

The record processing procedure consisted of the extraction of vowels /i/, /e/ and /a/ from the sentence /vine mama/. Their segmentation was done manually by using the Praat software package [12]. The database was first organized by separating the recordings according to the four emotions, then the recordings of the vowels /i/, /e/, /a1/ and /a2/ were extracted separately for each of the four emotions. Subsequently, the formants F1, F2, F3 and F4 were extracted from each vowel. The size of the sliding windows was 6ms (0.006s), the number of samples oscillated depending on the vowel position in the word structure and depending on the speaker.

The Pearson correlation coefficient for a pair of formants is calculated by applying the formula (1)

$$C_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})^2}} \quad (1)$$

where: x represents the value of the first formant in the pair, y represents the value of the second formant and n represents the total number of values.

Once concatenation was done, a time series file was obtained for each vowel and each corresponding emotion. All speakers (participants) were included in the file in the order of the numbering in which their records were processed. Thus, 64 time series were obtained as a result of the association of the formants (F1-F4) of the vowels /i/, /e/, /a1/ and /a2/ from the structure /vine mama/, one for each of the four emotions used in the analysis. This result was obtained by applying the formula (2)

$$F_n(i(E = N)) \quad (2)$$

where: F_n – n order formant, i – vowel, E – emotion.

We calculated the level of correlation for all couples of vowel formants. The correlation coefficients resulted from applying the formula below to all time series couples (two by two), keeping the order of concatenation of all vowels corresponding to the emotions analyzed (3).

$$\text{Corr}(F_1(i(E = N))F_2(i(E = N))) \quad (3)$$

The values obtained for calculating the correlation coefficients of formant pairs of vowels /i/, /e/, /a1/ and /a2/ are reported in the following sections.

III. RESULTS

Subsequent to the statistical processing of the time series, the value of the correlation coefficient of the pairs of formants (F1-F4) was calculated according to the specific autocorrelation function of the Excel tool.

Table I shows the values of correlation coefficients for the formant pairs of the vowel /i/ depending on the three emotions analyzed and the state of emotional

neutrality. We found that for the vowel /i/, the correlation level of F1-F2, F1-F3, F1-F4 formant pairs remains low, regardless of the type of emotion. At the same time, the resulting correlation degree is similar to that reported in the study we have done previously [7]. Therefore, regarding the vowel /i/, the type of emotion does not influence the correlation level of the pairs of formants.

TABLE I. THE RELATIVE VALUES OF THE FORMANT PAIRS FOR THE VOWEL /i/

Vowel /i/				
	Joy	Anger	Sadness	Neutral
Correl(F1,F2)	0.163	0.250	0.073	-0.168
Correl(F1,F3)	0.224	0.200	-0.088	0.298
Correl(F1,F4)	0.073	0.186	-0.013	0.245
Correl(F2,F3)	0.804	0.596	0.788	0.627
Correl(F2,F4)	0.639	0.348	0.330	0.476
Correl(F3,F4)	0.675	0.675	0.332	0.593

As shown in Table II, in the case of the vowel /e/, the pairs of formants have much lower correlation values for joy, anger and sadness than for the state of emotional neutrality. The degree of correlation of the F1-F2, F1-F3 and F1-F4 formant pairs is significantly lower than that reported for F2-F3, F2-F4 and F3-F4 pairs.

TABLE II. THE RELATIVE VALUES OF THE FORMANT PAIRS FOR THE VOWEL /e/

Vowel /e/				
	Joy	Anger	Sadness	Neutral
Correl(F1,F2)	0.240	-0.183	0.158	0.568
Correl(F1,F3)	0.341	0.128	0.000	0.748
Correl(F1,F4)	0.323	0.220	0.315	0.771
Correl(F2,F3)	0.729	0.395	0.541	0.706
Correl(F2,F4)	0.601	0.261	0.701	0.733
Correl(F3,F4)	0.763	0.709	0.498	0.908

The results of a study conducted on the same database and with a similar research theme [8] showed that when the speaker's voice is neutral, the stressed vowel in the word /mama/, namely /a1/, has a lower correlation level of the formants, in comparison with the non-stressed vowel /a2/. Tables III and IV summarize the results obtained following the correlational analysis of the formants, an analysis which was developed in the research of this study. We can see that we have obtained similar results, in the sense that the non-stressed vowel shows a higher degree of correlation of the formants, compared to the stressed vowel, but only in the case of sadness and emotional neutrality.

TABLE III. THE RELATIVE VALUES OF THE FORMANT PAIRS FOR THE VOWEL /a1/, STRESSED

Vowel /a1/				
	Joy	Anger	Sadness	Neutral
Correl(F1,F2)	0.444	0.514	0.510	0.592
Correl(F1,F3)	0.439	0.772	0.709	0.404
Correl(F1,F4)	0.307	0.522	0.696	0.479
Correl(F2,F3)	0.546	0.529	0.489	0.325
Correl(F2,F4)	0.286	0.630	0.412	0.295
Correl(F3,F4)	0.675	0.567	0.697	0.752

TABLE IV. THE RELATIVE VALUES OF THE FORMANT PAIRS FOR THE VOWEL /a2/, NON-STRESSED

Vowel /a2/				
	Joy	Anger	Sadness	Neutral
Correl(F1,F2)	0.458	0.254	0.689	0.661
Correl(F1,F3)	0.667	0.476	0.736	0.679
Correl(F1,F4)	0.397	0.286	0.421	0.657
Correl(F2,F3)	0.454	0.336	0.782	0.820
Correl(F2,F4)	0.398	0.524	0.584	0.685
Correl(F3,F4)	0.293	0.587	0.584	0.806

Regarding the emotion of joy, Table V and Figure 1 highlight that there are no statistically significant differences between the correlation degree of the formant pair /a1/ and /a2/, with the exception of the F3-F4 pair, for which the correlation coefficient is lower for the non-stressed vowel /a2/.

TABLE V. THE RELATIVE VALUES OF THE FORMANT PAIRS FOR JOY

Joy				
	i	e	a1	a2
Correl(F1,F2)	0.163	0.240	0.444	0.458
Correl(F1,F3)	0.224	0.341	0.439	0.667
Correl(F1,F4)	0.073	0.323	0.307	0.397
Correl(F2,F3)	0.804	0.729	0.546	0.454
Correl(F2,F4)	0.675	0.763	0.675	0.486
Correl(F3,F4)	0.675	0.763	0.675	0.293

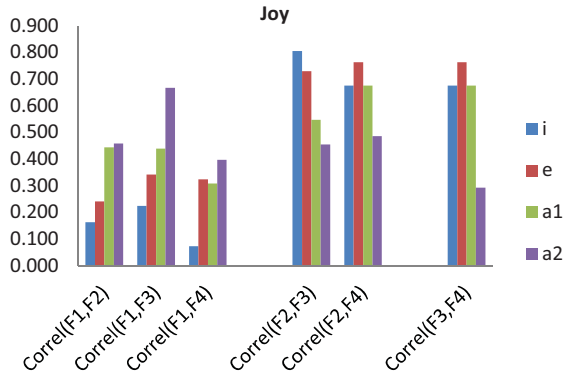


Figure 1. Graphic representation of the formants for joy

As for anger, the correlation values are illustrated in Table VI and Figure 2, the stressed vowel /a1/ of the word /mama/ has a correlation level of the pairs of formants significantly higher than that of the non-stressed vowel /a2/, which is surprising since both in the previous studies [8] and in the case of the other three emotions, the formants of the non-stressed vowels were those who correlated more strongly. Therefore, we can see that the voice signal characteristics for anger facilitated the reversal of the correlation ratio between the stressed and the non-stressed vowel.

TABLE VI. THE RELATIVE VALUES OF THE FORMANT PAIRS FOR ANGER

Anger				
	i	e	a1	a2
Correl(F1,F2)	0.250	-0.183	0.514	0.254
Correl(F1,F3)	0.200	0.128	0.772	0.476
Correl(F1,F4)	0.186	0.220	0.522	0.286
Correl(F2,F3)	0.596	0.395	0.529	0.336
Correl(F2,F4)	0.675	0.709	0.567	0.587
Correl(F3,F4)	0.675	0.709	0.567	0.587

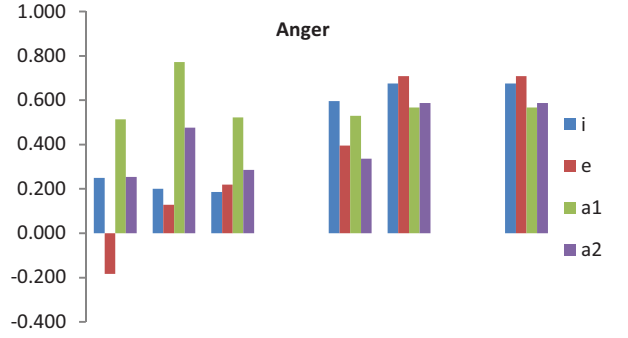


Figure 2. Graphic representation of the formants for anger

In regards to sadness, referred to in Table VII and Figure 3, it has been observed that vowel /a/, be it stressed or non-stressed, shows a significantly higher correlation level of pairs of formants, compared to the correlation level of formant pairs of the vowels /i/ or /e/. For example, for the pair (F1, F3), the calculated correlation coefficient is 0.73 for the vowel /a2/ and is obviously higher than 0.08, which is the correlation coefficient value of the formant pair for the vowel /i/.

TABLE VII. THE RELATIVE VALUES OF THE FORMANT PAIRS FOR SADNESS

Sadness				
	i	e	a1	a2
Correl(F1,F2)	0.073	0.158	0.510	0.689
Correl(F1,F3)	-0.088	0.000	0.709	0.736
Correl(F1,F4)	-0.013	0.315	0.696	0.421
Correl(F2,F3)	0.788	0.541	0.489	0.782
Correl(F2,F4)	0.332	0.498	0.697	0.584
Correl(F3,F4)	0.332	0.498	0.697	0.584

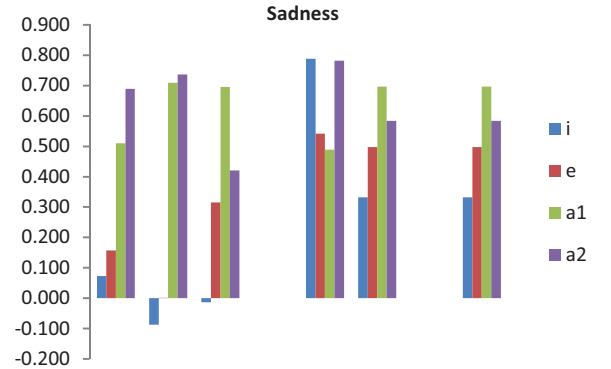


Figure 3. Graphic representation of the formants for sadness

The results we have obtained and which are synthesized in Table VIII and illustrated in Figure 4 show that, regarding the state of emotional neutrality, the correlation level of the pairs of formants corresponding to the four vowels is high (on average there are values of the correlation coefficient higher than 0.5). We have also discovered that, unlike the correlation level of the formants reported for vowels associated with emotions, in the case of neutral tone, the correlation of the formants is more homogeneous, with no extreme oscillations from one vowel to the other.

TABLE VIII. THE RELATIVE VALUES OF THE FORMANT PAIRS FOR NEUTRALITY

Emotion neutrality				
	<i>i</i>	<i>e</i>	<i>a1</i>	<i>a2</i>
Correl(F1,F2)	-0.168	0.568	0.592	0.661
Correl(F1,F3)	0.298	0.748	0.404	0.679
Correl(F1,F4)	0.245	0.771	0.479	0.657
Correl(F2,F3)	0.627	0.706	0.325	0.820
Correl(F2,F4)	0.593	0.908	0.752	0.806
Correl(F3,F4)	0.593	0.908	0.752	0.806

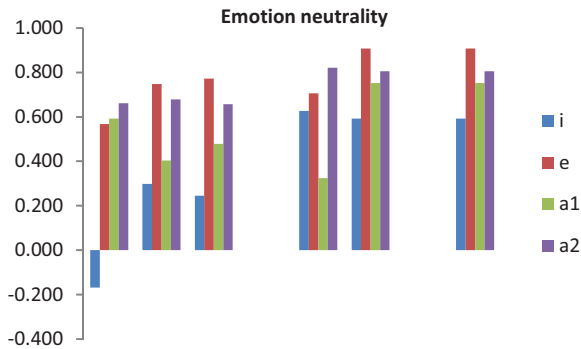


Figure 4. Graphic representation of the formants for neutrality

IV. DISCUSSION AND CONCLUSIONS

The three main conclusions are: (1) emotions are a variable with a strong influence on the level of correlation of vowel formants in the common speech of individuals; (2) in the case of emotions with negative valence (sadness and anger), the higher order formants (F3 and F4) have a significantly higher correlation level compared to the lower order formants (F1 and F2) and (3) regarding the state of emotional neutrality, the variation of the correlation level is more homogeneous, in the sense that fewer oscillations have been observed from extremely low values for a particular vowel to extremely high values.

The first conclusion is in line with the results of the previous studies and tends to become a certainty rather than a hypothesis requiring empirical testing. The research was directed towards creating and testing models for the recognition and interpretation of emotions by analyzing the statistical parameters of the voice signal. The applicability of these systems used for analyzing and recognizing emotions in the voice signal is relevant for speech analysis and synthesis, for optimization of call center applications and pathological speech detection [13], by providing the right context for raising quality standards regarding human-computer interaction.

The main limitation of this study is represented by the low number of speakers who were included in the analysis and by the fact that the gender equality ratio was not respected. Another limitation is that the approach of the study borders the area of descriptive statistics (data presentation and comprehension) rather than the area of inferential statistics which has a much greater explanation and prediction power over the investigated aspects.

As future research directions, we intend to increase the number of participants and to include

representative groups of speakers in the study, so that the results obtained can be generalized to a much larger number of individuals in the target population. On the other hand, we suggest a comparative study on variations in the correlation level of the vowels of the Romanian language to be carried out, taking into account the comparison between normal speech and pathological speech.

ACKNOWLEDGMENT

I acknowledge that Horia-Nicolai Teodorescu suggested the topic, the research plan and the method, helped deriving the analysis results and corrected several versions of the article.

REFERENCES

- [1] P. Boersma and D. Weenink, Praat for Windows 64-bit edition: praat6046_win64.zip, from www.fon.hum.uva.nl/praat/download_win.html, (Univ. of Amsterdam)
 - [2] M. Stanek, M. Sigmund, "Finding the most uniform changes in vowel polygon caused by psychological stress." *Radioengineering*, Vol. 24, No. 2, pp. 604-609, Jun 2015.
 - [3] N. Flynn, "Comparing vowel formant normalisation procedures." *York Papers in Linguistics*, Series 2, no. 11, pp.1-28, March 2011.
 - [4] H.N. Teodorescu, and M. Feraru, "A study on Speech with Manifest Emotions", 10th International Conference on Text, Speech and Dialogue", TSD 2007, Pilsen, Czech Republic, September 3-7, 2007, Lecture Notes in Computer Science, Springer Verlag, vol. 4629/2007, ISBN 978-3-540-74627-0, pp. 254-262.
 - [5] H.N. Teodorescu, M. Feraru, and D. Trandabăț, "Studies on the prosody of the Romanian language: the emotional prosody and the prosody of double-subject sentences", C., Teodorescu, H.N. (Eds.) *Advances in Spoken Language Technology*, The Publishing House of the Romanian Academy, Bucharest, România, 2007, ISBN 978-973-27-1516-1, pp. 171-182.
 - [6] M.D. Zbancioc, H.N. Teodorescu, S.M. Feraru, "Statistical Characteristics of the Formants of the Romanian Vowels in Emotional States", 6th Conference on Speech Technology and Human-Computer Dialogue (SpeD), IEEE, Brasov, Romania, May 18-21 (2011).
 - [7] Ș.A. Ghelțu, "Correlation of the Formants of the Romanian Vowels", IEEE International Conference – 10th Edition (ECAI), Jun 28-30, 2018.
 - [8] S.A. Ghelțu, H.N. Teodorescu, "Correlations and dependencies of the formants during speech in Romanian." *Proc. IEEE ECAI 2018 Int. Conference – 10th Edition*, 28-30 June, 2018, Iași, România.
 - [9] H.N. Teodorescu, "How much are formants correlated and what that means?" "Proc. 14-th International Symposium on Signals, Circuits and Systems, IEEE ISCS, 2019.
 - [10] H.N. Teodorescu, M. Feraru, M. Zbancioc, D. Trandabăț, et al., "Romanian Speech Database - SRoL", 2014, http://www.etc.tuiasi.ro/sibm/romanian_spoken_language, SRoL, http://iit.academiaromana-is.ro/srol/ro/arhiva_sunete.htm.
 - [11] S.M. Feraru, H.N. Teodorescu, M.D. Zbancioc, "SRoL - Web-based Resources for Languages and Language Technology e-Learning". *International Journal of Computers Communications & Control*, Vol. 5, no. 3, pp. 301-313, Sep 2010.
 - [12] J.R. Williamson et al., "Detecting depression using vocal, facial and semantic communication cues." *Proc. AVEC'16*, pp.11-18, Oct 2016, Amsterdam, Netherlands, ACM. ISBN 978-1-4503-4516-3/16/10, DOI: [http://dx.doi.org/10.1145/2988257.2988263](https://doi.org/10.1145/2988257.2988263).
- P.H. Dejonckere et al., "A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques." *Eur Arch Otorhinolaryngol* 258, pp. 77-82, 2001.