ORIGINAL PAPER

DESCRIPTIVE STATISTICS AND CROSS CORRELATION OF SOME VOCAL AND ACOUSTIC PARAMETERS INVOLVED IN LIVE BROADCASTIG

VALENTIN GHISA[1], SORIN MORARU[2]

Manuscript received: 02.04.2015; Accepted paper: 27.04.2015;
Published online: 30.06.2015.

Abstract. *For the analyze of communication we need to study the main parameters that describe the vocal sounds from the point of view of information content transfer efficiency. In this paper we analyze the physical quality of the "on-air" information transfer, according to the audio streaming parameters and from the particular phonetical nature of the human factor. Applying this statistical analyze we aim to identify and recording the correlation level of the acoustical parameters with the vocal ones and the impact which the presence of this cross-correlation can have on the communication structures improvement.*

Keywords: *correlation analyzes, vocal parameters, SPSS, live broadcasting.*

1. INTRODUCTION

Speech analyze are becoming more and more useful because of the results obtained in the last period of time, especially thanks to some important developed applications. In the 90^s a lot of phonetical theories were in the attention of the researches having the implications in the improvement of prosodiacal aspect of vocal structures. Those created the premises of entering in a new level of synthesis sistems of vocal recognition, by the vocal signal description from prosodiacal point of view and emotional states [1]. In the mean time for a theoretical survey of the acoustical signals we often need the math modellation of the main parameters that determine the evolution of this phenomenon. In this respect, an essential part is the one which analyse the vocal signal processing, a component that concentrates on the development and implementation of the methods and algoritms for pattern extraction, the interpretation transformation and encoding. The first level in this research refers to the vocal recognition of the words which knows two different ways of realisation, regarding the dependence or independence by the speaker. The second method of vocal recognition contains the selection and identification of a certain set of spectral parameters which are specific to the speech content transmitted by communication. In the respect of sistemic theory of dynamic transfer process, from the technical point of view, sound signals can be definited as representing the physical support of information transmition between internal parts of the system. For being processed in a proper way the acoustical signal from the vocal source are recorded and coverted in electric signals with the help of some specific electronic devices: microphone, amplifiers, filters, digital-analogical convertors etc. The majority of vocal signals which come from the communicational enviroment have a continous variation in time and in

[1] Transilvania University of Brasov, Department of Electrical Engineering and Computer Science, 500036 Brasov, Romania. E-mail: valentin.ghisa@unitbv.ro.
[2] Transilvania University of Brasov, Department of Electrical Engineering and Computer Science, 500036 Brasov, Romania. E-mail: smoraru@unitbv.ro.

this respect there are used analogical systems for their processing [2]. As an acoustic phenomenon the language is presented in the shape of some continous phonetical sequences which are separated by intervals. In this paper we analyze the physical quality of the information transfer „*on-air*" according to different enviromental factors, the audio streaming parameters and by the particular phonetical caracteristics of the human factor. The impact from the physical point of view depends by the different acoustical sizes of the vocal signal: efficient acoustical pressure, sound energy, audio intensity and reverberation time. From the phonetical point of view the study is focused on the some phonic elements through which the communication is materialized: *speech clarity, speech intelligibility* and *speed of speech* etc [3]. Regarding the speaking, no matter it is in the shape of a discourse or a conversation, all these communication deeds are effectively defined by some coherent elements, composed by lexical units, which generates a semantic content. This can be amplified or, on the contrary, attenuated by the effects of an association between some vocal acoustic parameters and by the factors that are tied with the accuracy and the technique of speaking [4]. Through the application of some statistic methods, we intend to identify and measure the level of the correlation of the vocal and acoustic parameters and the impact that can be obtained from this cross-correlation in order to improve the communication structures.

2. MATERIALS AND METHODS

The measurements of vocal and acoustic parameters were realized in a radio transmission studio, in the recording room, having the dimension: 8×6×3 [m], so a volume of V=144 [m³]. The reverberation time was determined RT_{60} = 0.376 [sec]. The value of the sound speed in this room at a temperature of 20^0C it was 343 [m/s]. The distortion coefficient of the sound δ = 0.18 %, being proper for a normal transmission of the signal. The error in the signal detection in a level of reliability of 95%, was ±0.94 dB in presence of a certain filter and ±0.96 dB without this, both for voices and noise. The background sound intensity of the recording room was situated at the value of: $N_s \approx 20$ dB. The former experiments underlined the reality that for obtaining a high precision in the vocal recognition of the lexical units in Romanian language, a constant and intensified speaking is needed. In order to solve this aspect, we use a voice processor Tascam TA-1VP, formed by a pre-amplifier with microphone, a compressor, waves deesser and correction pitch. This is wired to another processor of reverb and multi-effect TC Electronic M3000. The microphones AT 2020 USB+ condenser assure the conversion from analogical system in the digital one at 16 biti, with a sampling rate of 48 KHz and frequency response in the interval between 20-20000 Hz. The emission mixer was Traktor Kontrol 22, which is provided with the encoders of high quality, 3 sizes converter and a filter for each channel. The recording and signal processing system has a strong recorder Tascam TM – SSCDR 200 for wav, mp3, serial RSB2C and parallel recording. In the same time, the system has an audio converter, multichannel Xlogic Alpha-Link Audio that has 64 digital channels at 48 kHz, 24 I/O analogic and 12 I/O digital stereo AES/EBU.

The recording of some certain texts reading by 59 persons (30 men and 29 women) was made. Every person had to uttering on the microphone a number of three texts in Roumanian language, belongs in terms of semantics of the different literary species. The condition was that each subject was recorded reading one of these three texts only for one minute. The vocal spectrum was recorded on the hard memory of a PC and it was processed in audio 3D QSound Pro 9.0 SSMS product by Sony. The soft allowed the auto-play of the materials on every media player program on the PC [5].

3. WORK TECHNIQUES AND DATA

In order to underline the possible association between some vocal and acoustic parameters, and to emphasize the relevance of these connections on the informational structures, we choose to analyze the phenomenon using an IBM SPSS 20.0. All the 59 audio recording were processed in the program Q Sound and organized in Data Base that contained the stochastic variable (Table 1):

Code. Var.	Explained variable	Variable type	[unit]
SC	Speech clarity	parametric	[dB]
SI	Speech intelligibility	parametric %	[undim.]
VF	Voice frequency	parametric	[Hz]
VL	Voice loudness	parametric	[dB]
SS	Speech speed *(*word/minute*)	parametric	[wpm]
ID	Information density	parametric %	[undim.]
NSE	Number of speech errors**	parametric	[n.u.]

* it also exist the choice [sps] syllables/sec.

** contains partially superposed words, incorrect words emphasis, speech errors etc

The software IBM SPSS 20.0 under Windows is an interactive and very useful software package which is designed for data analyzes and includes multiple facilities and techniques of the statistical nature. Through these facilities we find great distributive options, automatic models, ability to work with server versions of IBM SPSS Statistics Base, a syntax editor, Microsoft Office integration etc [6]. In the present article, for the accurate representation of variable evolution we will follow also the determination of the scattering and central tendency indicators, unifactorial analysis of variance ANOVA, analysis of covariance as well as simple linear regression analysis with the SPSS program.

4. RESULTS AND DISCUSSION

4.1. DESCRIPTIVE STATISTICS (**Table 2**) *AND DISTRIBUTION VALUES* (**figs. 1-7**)

Variable	SC	SI	VF	VL	SS	ID	NSE
N	59	59	59	59	59	59	59
Mean	-1,9678	44,22	1293,58	44,92	154,83	78,24	5,58
Std. Error of Mean	,23497	1,070	48,028	1,393	2,529	1,405	,393
Median	-1,9300	44,00	1227,00	43,00	155,00	81,00	5,00
Mode	-2,88	38[a]	741[a]	32	172	85	2[a]
Std. Deviation	1,80484	8,217	368,907	10,699	19,426	10,790	3,018
Variance	3,257	67,520	136092,662	114,458	377.350	116,425	9,110
Skewness	-,030	-,001	,183	,468	-,067	-,456	,368
Kurtosis	-,630	-,557	-1,181	-,886	-,739	-,825	-,847

[a] There are multiple mode

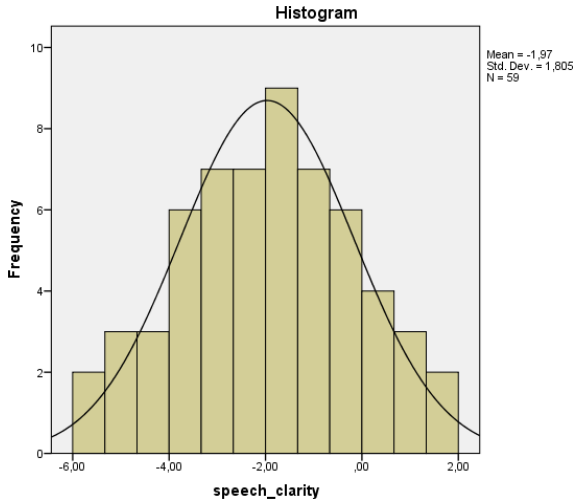

Figure 1

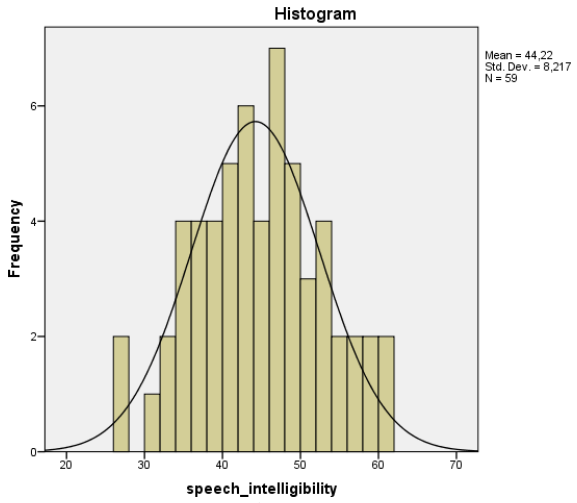

Figure 2

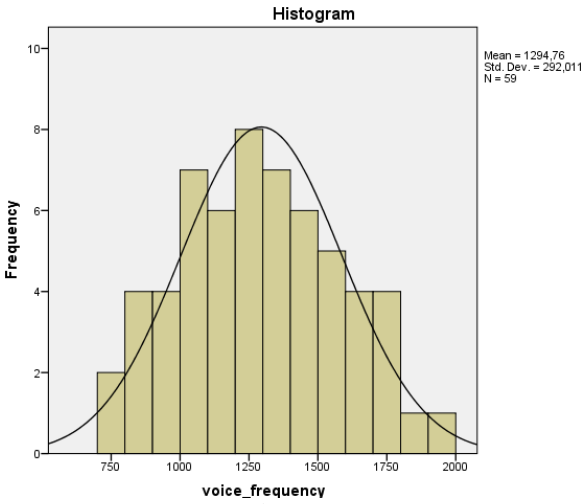

Figure 3

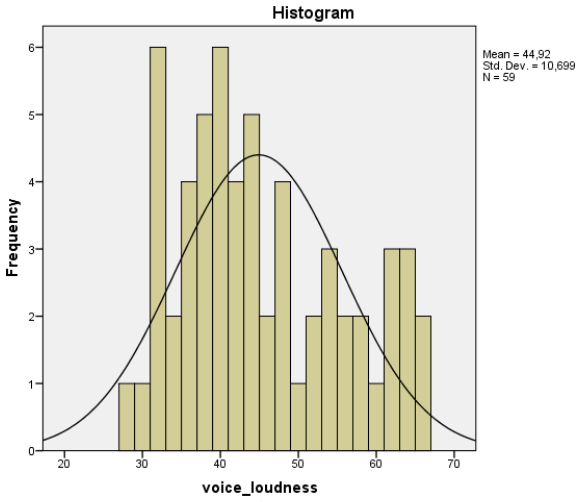

Figure 4

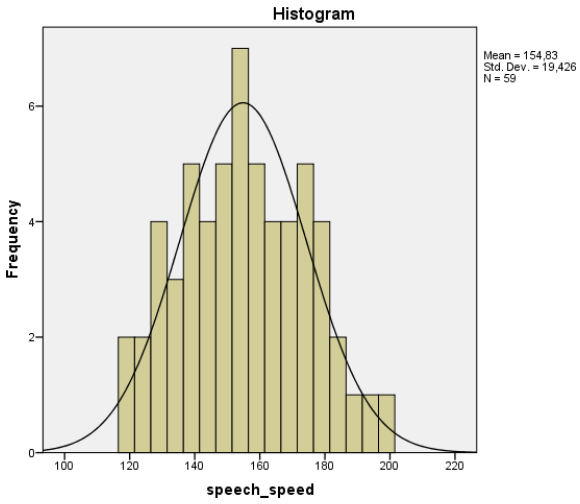

Figure 5

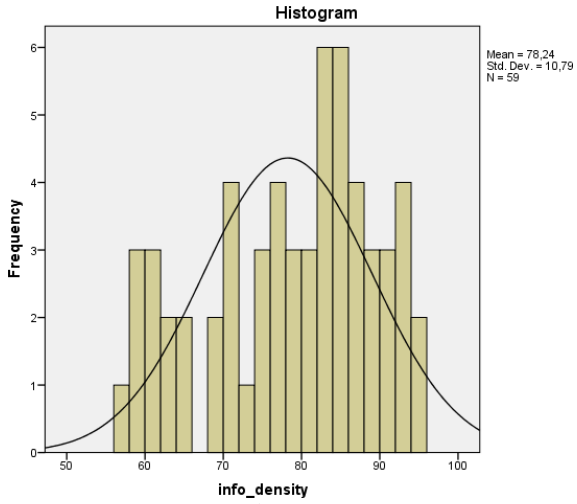

Figure 6

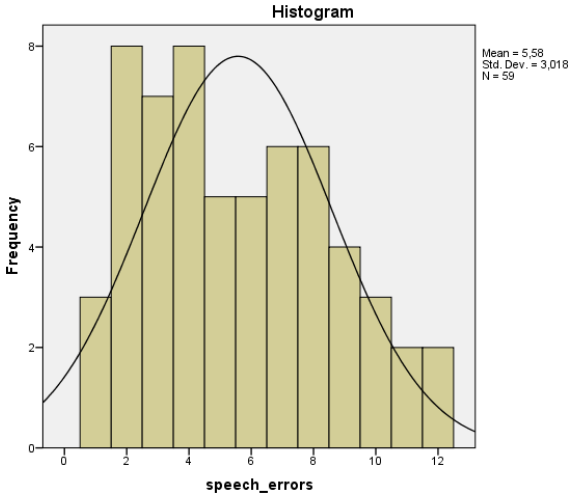

Figure 7

4.2. VARIABLE CORRELATIONS

Table 3. Correlation matrix

Correlations		SC	SI	VF	VL	SS	ID	NSE
SC speech_clarity	Pearson Correlation	**1**	**-,141**	**,118**	**-,215**	**,019**	**,217**	**,215**
	Sig. (2-tailed)	,000	**,287**	**,372**	**,101**	**,887**	**,098**	**,102**
SI speech_intelligibility	Pearson Correlation	-,141	**1**	**-,714****	**-,117**	**-,893****	**,113**	**,100**
	Sig. (2-tailed)	,287	,000	**,000**	**,378**	**,000**	**,395**	**,449**
VF voice_frequency	Pearson Correlation	,118	-,714**	**1**	**,035**	**,617****	**-,044**	**-,074**
	Sig. (2-tailed)	,372	,000	,000	**,792**	**,000**	**,739**	**,577**
VL voice_loudness	Pearson Correlation	-,215	-,117	,035	**1**	**,199**	**-,995****	**-,903****
	Sig. (2-tailed)	,101	,378	,792	,000	**,130**	**,000**	**,000**
SS speech_speed	Pearson Correlation	,019	-,893**	,617**	,199	**1**	**-,201**	**-,199**
	Sig. (2-tailed)	,887	,000	,000	,130	,000	**,127**	**,131**
ID info_density	Pearson Correlation	,217	,113	-,044	-,995**	-,201	**1**	**,911****
	Sig. (2-tailed)	,098	,395	,739	,000	,127	,000	**,000**
NSE no._speech_errors	Pearson Correlation	,215	,100	-,074	-,903**	-,199	,911**	**1**
	Sig. (2-tailed)	,102	,449	,577	,000	,131	,000	,000

**Correlation is significant at the 0.01 level (2-tailed).

In conformity with the correlation values from the table 3, we observe a multiple and important relationship between the next variables: SI (speech_intelligibility), VF (voice_frequency) and also SS (speech_speed). The correlation coefficient between SI and VF (r = -0.714) is a moderate one, very close to the superior limit (-0.75) and between the SI și SS is a strong correlation (r = -0.893). An important correlative cluster is formed by the next variables VL (voice_loudness), ID (info_density) and NSE (no_speech_errors). Here, in the sense of the determination cause-effect, the dependent variable is the informational density ID, the predictors being the parameters VL and NSE.

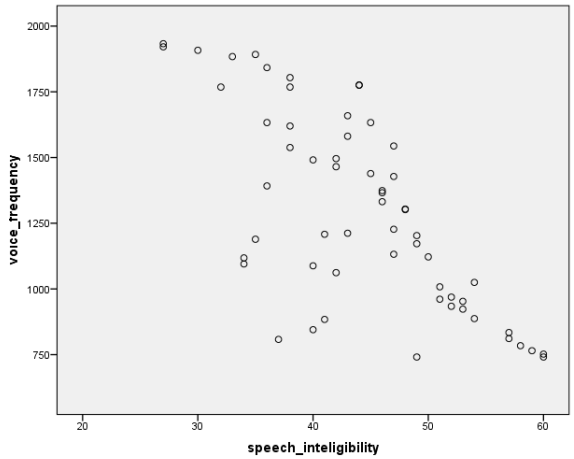

Figure 8

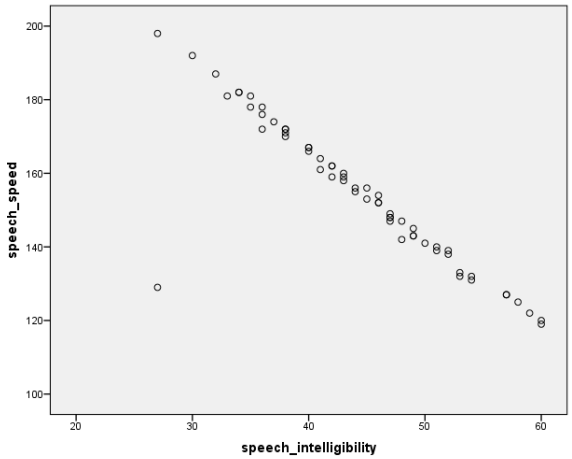

Figure 9

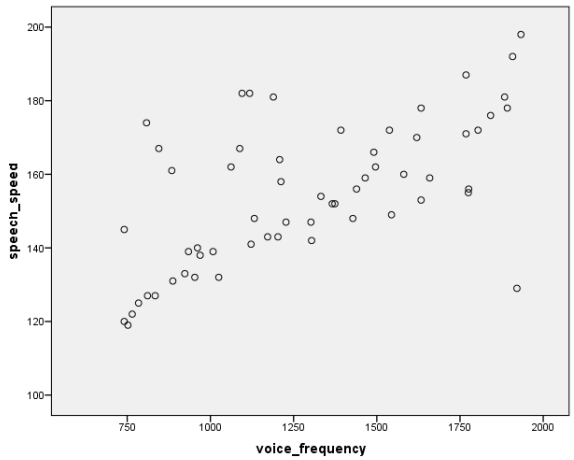

Figure 10

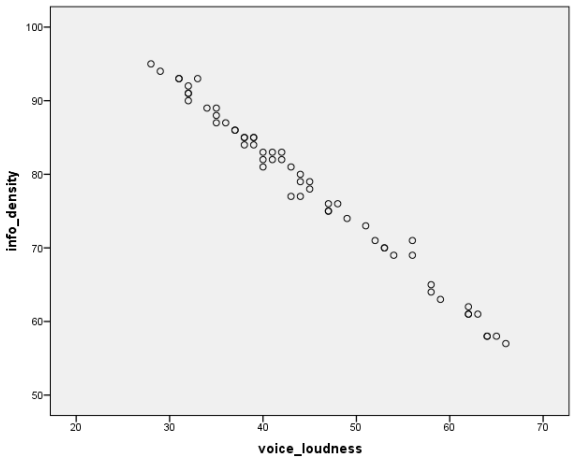

Figure 11

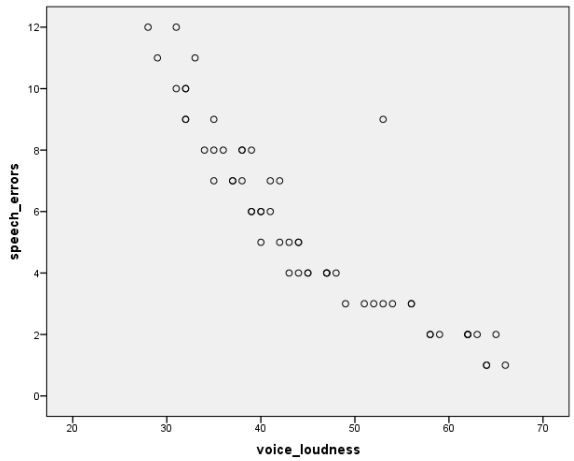

Figure 12

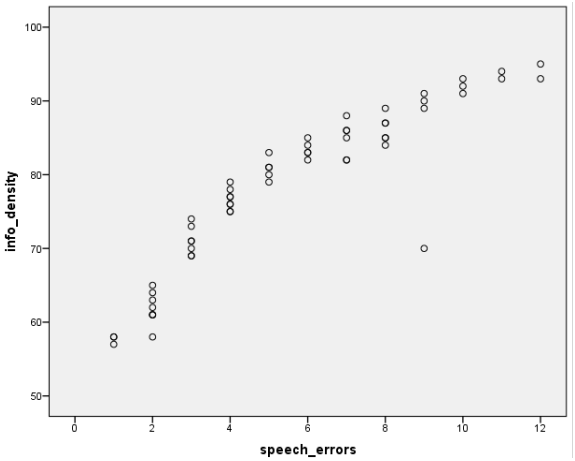

Figure 13

The previous charts (figs. 8-13) show the most important calculated correlations between vocal and acoustic pairs of measured variables.

5. CONCLUSIONS

From the correlations matrix, we observe the SC variable (*speech clarity*) presents relatively low associations with the other variables, between 0.019 and 0.217. In this situation we can`t discuss about a total independence of SC variable, with the exception of SS (*speech speed*). In the respect with the correlation values from the table 3, we observe a multiple consistent connection between next variables: SI (*speech intelligibility*), VF (*voice frequency*) and also SS (*speech speed*). The correlation coefficient between the SI and VF (r = -0.714) is a moderate one, close to the superior limit (-0.75) and between SI and SS is a strong connection (r = -0.893). In this condition, we can study the possibility of a certain multiple linear regression in the respect of an informational meaning the dependent variable is the SI while the VF and SS are considered independent ones. Regarding the VF (*voice frequency*), we distinguish a positive correlation coefficient (r = 0.617) with the variable SS (speech speed). This correlation coefficient reveals a middle relationship under a very good level of significance p = 0.01. If we consider that between VF and SS, the voice frequency is a dependent one, than it depends 38% by SS variable.

Further, the variable VL (*voice loudness*) reveals two strong association with the variable ID (*informational density*), r = 0.995 and NSE (*number of speech errors*), r = 0.903. This values of the correlation coefficients, in the significance limits under p = 0.01, show very intense connections between these two parameters, appearance which leads to try an analyze of the possibility to exists a linear regressional model composed of those three variables. It also noted that the VL parameter, on both relations association, describes the same sense trends (see the charts). A last outstanding correlation is observed between ID (*info density*) and NSE (*number of speech errors*) parameters. In this case, the value of correlation coefficient is r = 0.911, on the < p = 0,01 significance threshold. Taking in consideration the causality relationship between these two parameters, the ID variable as a dependent one, we can say the NSE parameters explain/determine about 83% from the content of the parameter ID. We must specify the fact that through some *speaking errors*, we understand the wrong pronounced words or accents, incomplete words, some improper breaks and lack of coherence in the semantic of the message transmission.

Acknowledgement*: This paper is supported by the Sectoral Operational Programme Human Resources Development (SOP HRD), ID134378 financed from the European Social Fund and by the Romanian Government.*

REFERENCES

[1] Furui S., *50 years of progress in speech recognition technology -- Where we are, and where we should go*, Tokyo Institute of Technology, International Conference on Acoustics, Speech, and Signal Processing (ICASSP), which will be held at the Hawai‘i Convention Center in Honolulu, April 15 - 20, 2007

[2] Teodorescu H.N., *Aproposed theory in prosody generation and perception: the multi-dimensional contextual integration principle of prosody*, SpeD 2005 - 3th Conference on Speech Technology and Human Dialogue, Eds. C. Burileanu, Trend in Speech Technology, Ed. Academiei Romane, Buc., 2005

[3] Kochanski G., Grabe E., Coleman J., Rosner B., *Loudness Predicts Prominence; Fundamental Frequency Lends Little,* J. Acoustical Society of America 11(2), 2005
[4] Apopei V., Jitca D., Grigoras F. (2003a), *Studiul trasaturilor acustice necesare pentru evidentierea evenimentelor fonetice in vederea segmentarii semnalului vocal,* Simpozionul Sisteme de Inteligenta Artificiala SIA 2003, Iasi, septembrie 2003
[5] http://www.qsound.com/technology/overview.htm
[6] Field A., *Discovering Statistics Using SPSS,* SAGE Publications Ltd; Third Edition edition, London, 2009