THE IMPORTANCE OF VOCAL PARAMETERS CORRELATION FOR INFORMATION PROCESSES MODELLING

Valentin GHISA*
Nicoleta GHISA**

*Transylvania University, Department of Electrical Engineering and Computer Science, Romania
**Radio Romania Cluj, Chief Editor, Department Of News and Broadcasting Production, Romania

To analyze 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 phonetic nature of the human factor. Applying this statistical analysis we aim to identify and record the correlation level of the acoustical parameters with the vocal ones and the impact which the presence of this cross-correlation can have on communication structures' improvement.

Key words: correlation analysis, vocal parameters, SPSS

1. INTRODUCTION

The data transfer between different systems, processes, social and economic entities or subsystems creates a large field for analysis, the so called the communication problem.

The interactive information flow model needs new possibilities of improving communication capabilities and the rise of the quantitative and qualitative level of useful information. This fact drives to a better dynamic of the communication process.

The practical efficiency of a message does not consist in reaching a high information level, but in the visible effects of the message form and content through the listeners. What is very important in this respect is message semantics and the probability of understanding its content, along with the integration of these salient features with the effectiveness and form of the utterance itself.

There are three ways of operating on information content: morpho-syntactic (strongly tied with rigorous abidance by linguistic norms and rules), pragmatic (a nature of the utility of information for the receivers) and semantic (the significance which is shaped and submitted by the formal language).

As an acoustic phenomenon, language is a set of phonetic continuous sequences which are separated by intervals. Beyond the deterministic tackling of signals processing, in which the signal is studied as a local manifestation, statistical analysis considers that the voice sign is an information keeper identified and strictly classified after some parameters related to stochastic variables [1].

The deviation from the values of the parameters from a mean value are quantified for a set of information components, through simple indicators (amplitude, median, mode), but more than that, through synthetic indicators, such as dispersion (σ^2), mean square deviation (σ), the variation coefficient or the linear mean deviation ($\bar{d}$).

All these descriptive indicators show the level of the variables for each information component, and the way in which all these present deviation from the mean value. At the same time, the coefficient of variation shows the spread in a direct relationship with the average.

Its nature (which does not depend on the value of indicators) recommends such an indicator for the analysis of parametric convergence. Generally speaking, through the dependence of implied variables in a communication process, we understand just one type of dependence, total or deterministic.

On the other hand, there are some situations in which two stochastic variables x and y are connected through a probabilistic relationship, when, for example, we know that the probability of x, and y depends on the repartition law, being dependent by x.

In a strong relationship with the stochastic variables, the dependence can be determined through regression or correlation.

The correlation is a descriptive statistic method because it shows what happens in a group of results. The concept of correlation shows the concomitant variation of two variables, allowing a reciprocal prediction procedure of the evolution of those variables.

That means there may be no relationship between some stochastic variables. The coefficient of correlation (r) shows the degree/intensity of a relationship between two variables.

In the correlational highest point, all the values from the graphic of both variables are on the right part of the regression. In the case of the lowest point of correlation, the distance between these points and the right part of the regression shows the error between the associations of the variables [2].

The estimation of the cross-correlation between the different stochastic variables - the parameters of the information system, plays a very important role in the analysis of the features of the communication process in its physical form.

Through the application of some statistical analysis methods, we try to identify and quantify the level of the correlation of some acoustic and voice parameters and the impact of this cross-correlation on the optimization of the communication structures.

2. SOFT & WORK TECHNIQUES

Previous experiments in the field of voice recognition underlined that for obtaining a high precision in the vocal recognition of the lexical units in the Romanian language, constant and intensified speaking is needed.

The vocal spectrum was recorded on the hard memory of a PC and it was processed in the 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 [3].

Before that, data preprocessing was made by GoldWave v5.69 software, a highly rated, professional digital audio editor fully capable to do a wide range of operations, from the simplest recording and editing to the most sophisticated audio processing, restoration, enhancement and conversion [4].

59 persons (30 men and 29 women) were recorded reading certain texts. Every person had to read out aloud three texts of their own choice in the Romanian language on a microphone located 30 cm in front of their mouth.

The condition was for each subject of the experiment to be recorded reading each of these three texts for one minute. The parameters values targeted for analysis were averaged and stored in the final Data Base.

The IBM SPSS 20.0 application was employed for the analysis of possible associations between specific broadcasting parameters, and in order to emphasize the relevance of these correlations on the information structures.

3.PARAMETER DESCRIPTIONS

All the 59 audio recordings/ speaker were processed and organized in a Data Base that contains the following basic variables:

Table 1. Basic variables

Abbreviation	Parameters	Var. type	[unit]
TH	*tonal height*	discrete	[mel]
VL	*voice loudness*	discrete	[dB]
SS	*speech speed* (there is the option syllables/sec. [sps])	discrete	[wpm]
VT	*voice timbre*	discrete	[Hz]
NWW	*number of wrong words* (Incomplete words, incorrect words emphasis, utterance errors, language disorders, etc.)	discrete	[wwpm] (wrong word per minute (wwpm)

The table below lists the vocal derivatives parameters.

Table 2. The vocal derivatives parameters

Abbreviation	Parameters	Var.type	[unit]
SF	*speech fracture*	discrete	[wwpm/wpm]
UF	*utterance flexibility*	discrete	[mel/dB]
IE	*information energy*	discrete	[undim.]

The *tonal height* of the voice TH refers to the propriety of being more profound or sharper.

This height must be related to the age and gender of the speaker. The experience shows that this subjective acoustic parameter varies straight with the frequency of sound oscillations. The height of the voice unit of measurement is the *mel*.

The *mel* scale is a feeling scale for TH tonal height measurement, obtained by experimental drawing of the curve fitting which illustrates TH variations by its frequency.

The *voice loudness* VL must be related with the intention of the communication, with the time of communication or the receivers of it, and with the place. This means that the VL varies with these pragmatic parameters [6].

The *utterance flexibility* UF refers to the relationship between the tonal height TH and voice loudness VL, in order to underline the emphasis, semantic differences, emotional feature of the speaking act;

$$UF = \frac{TH}{VL} \text{ [mel/dB]} \qquad (1)$$

The *speech fracture* SF is a derived measurement, being considered as a rapport between NWW and SS. It is the opposite to the fluency of speaking. We will consider it as follows:

$$SF = \frac{NWW}{SS} \text{ [\%]} \qquad (2)$$

For the stochastic discrete variable X whose values represent the states of a certain information system, and with the distribution: $p_k \geq 0$, we refer to *information energy* as delineated by Onicescu, *IE* and, in accordance with the stochastic discrete distribution of the variable X, the following formula is used [7]:

$$E^2(X) = \sum_{k=1}^{n} p_k^2 \qquad (3)$$

Information energy $E^2(X)$ appears to be an overall value attached to a stochastic distribution and of the same nature as the Shannon information entropy.

We call *information correlation* of two random variables X_1 and X_2 or a distribution $p_k \geq 0$ and $q_k \geq 0$

$$\sum_k p_k = \sum_k q_k = 1$$

in the Onicescu case the expression:

$$O^2[X_1, X_2] = \sum_{k=1}^{n} p_k q_k$$

and the *information correlation coefficient* of their the rapport:

$$R^2[X_1, X_2] = \frac{O^2[X_1, X_2]}{\sqrt{E^2(X_1)E^2(X_2)}} \qquad (4)$$

The random variables X_1 and X_2 can be two arbitrary random variables or can represent the same random feature but for two different statistical populations. If the discrete IE - i*nformation energy*, $E^2(X)=1$, we have an absolute information concentration namely $p_k = 1$, $p_j = 0$, $j \neq k$, $1 \leq j \leq n$ and the information energy falls when the indeterminate …..or uniformity increases.

4. PARAMETRIZED CORRELATION METHOD

The purpose of a Parametrized Correlation Method is to determine a relationship between two sets of stochastic discrete variables. In statistics, the "correlation" concept has a very specific meaning.

Statistical correlation means that, given two discrete variables X and Y measured for each case in a sample, variation in X corresponds (or on the contrary, does not) to variation in Y.

As a rule, when we make an analysis of linear correlation, we want to know how powerful or intensive the relationship is between these two variables.

The extreme values of X are related with extreme values of Y, and less extreme X values with analogues Y values. The correlation coefficient (Pearson **r** for the parametric variables) measures the degree of this connection.

If one variable causally influences a second variable, then we would expect an intensive correlation between them. However, an intensive correlation may mean, for example, that they are both causally influenced by a third variable.

The way to determine linear correlation between the two variables is called the Pearson correlation coefficient (**r**). The formula for computing the Pearson *product-moment correlation* **r** is as follows:

$$r = \frac{1}{n-1} \sum_i \frac{(x_i - \overline{X})(y_i - \overline{Y})}{s_x s_y} \quad (5)$$

The value of r ranges within [-1, 1] interval. We can say that if:

- $r > 0$ indicates a positive connection of X and Y: as one gets larger, the other gets larger.
- $r < 0$ indicates a negative connection: as one gets larger, the other gets smaller.
- $r = 0$ indicates no connection

We can also calculate the correlation between more than two variables. Given variables x, y and z, we define the *multiple correlation coefficient*:

$$R_{z,xy} = \sqrt{\frac{r_{x/z}^2 + r_{y/z}^2 - 2 r_{x/z} r_{y/z} r_{x/y}}{1 - r_{x/y}^2}} \quad (6)$$

where, for example, $r_{x/y}$ is defined as:

$$r_{y/x} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{\left[n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right] \cdot \left[n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right]}} \quad (7)$$

and the relative correlation rapport:

$$R_{y/x} = \sqrt{1 - \frac{\sum_i \left(y_i - Y_{x_i}\right)^2}{\sum_i \left(y_i - \bar{y}\right)^2}} \quad (8)$$

If $(r_{y/x}) = (R_{y/x})$ then we can see a powerful, direct and intensive connection between the involved variables.

The IBM SPSS 20.0 software is an interactive and very useful software package which is designed for data analyses and includes multiple statistical facilities and techniques. 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. [5]. For the accurate representation of variable evolution we follow the determination of the scattering and central tendency indicators as well as simple correlation analysis with the SPSS program.

5. RESULTS AND DISCUSSION

The correlation matrix shows that TH (tonal_height) parameter presents a positive and strong relationship with VT parameter (vocal_timbre), (r=0.995). The same parameter reveals a strong and positive relationship with UF (utterance_flexibility), (r=0.815) but a positive and more than mean relationship with SS (speech_speed), (r=0.617), all those having a very good significance threshold, under p = 0.01.

In the case of VL (voice_ loudness) variable, we have strong negative correlations with SF (speech_fracture) parameter and NWW (nr._wrong_words), (r=-0.853 and r=-0.903) and a relatively medium negative association with UF (utterance_flexibility), (r=-0.508).

At the same time, the derivative parametric variable UF presents low to medium positive correlation with NWW and SS variables (r=0.464 and r=0.422) and a weak connection with the other variable SF (speech_ fracture), (r≈0.300) under a significant threshold of p=0.05. The same parameter has a strong positive relationship with VT parameter (r=0.816).

The derivative parameter SF is observed to have a very strong relationship with NWW (r=0.963) and the one of low to mean level and negative (r=0.432) with SS (speech_speed), both under a 0.01 significance threshold. Another significant correlation issues between VT (voice_timbre) and VR (speech_ speed), (r=0.604).We further note two other multiple and significant cross-correlation clusters about the derivative variables.

The first of them is where UF is the dependent variable and the variable pairs cluster (TH-VT), (VL-NWW) and also SS are exogenous/generator variable. The second group has as a main point the dependent variable SF (speech fracture), (VL-NWW) pair and also the independent parameter SS.

Based on the obtained values, we can study the models of multiple regression for UF and SF derivative parameters. We also determined that the information energy IE correlates intensely and directly with NWW (r=0.91) and also with VL parameter which is linked pretty consistently (r=0.697). Those two endogenous variables, independent, VL and NWW determine about 85% the evolution of information energy IE.

Table 3. Descriptive statistics

VARIABLE	TH	VL	SS	VT	NWW	SF	UF
Mean	1093.29	44.92	154.83	2658.27	5.58	3.7105	25.6614
Std. Error of Mean	48.031	1.393	2.529	47.747	.393	.27882	1.44945
Median	1012.00	43.00	155.00	2581.00	5.00	3.7500	22.5900
Mode	684	32	172	2104[a]	2[a]	2.32	10.20[a]
Std. Deviation	368.935	10.699	19.426	366.753	3.018	2.14167	11.1334
Variance	136113.17	114.458	377.350	134507.89	9.110	4.587	123.954
Skewness	.186	.468	-.067	.180	.368	.452	1.073
Kurtosis	-1.181	-.886	-.739	-1.174	-.847	-.713	.833

[a]There are multiple mode

Table 4. Correlation matrix

Correlations		TH	VL	UF	SF	NWW	VT	SS
TH	Pearson	**1**	**.035**	**.815****	**-.225**	**-.074**	**.995****	**.617****
	Sig. (2-tailed)	.000	**.790**	**.000**	**.087**	**.577**	**.000**	**.000**
VL	Pearson	.035	**1**	**-.508****	**-.853****	**-.903****	**.024**	**.199**
	Sig. (2-tailed)	.790	.000	**.000**	**.000**	**.000**	**.858**	**.130**
UF	Pearson	.815**	-.508**	**1**	**.299***	**.464****	**.816****	**.422****
	Sig. (2-tailed)	.000	.000	.000	**.021**	**.000**	**.000**	**.001**
SF	Pearson	-.225	-.853**	.299*	**1**	**.963****	**-.220**	**-.432****
	Sig. (2-tailed)	.087	.000	.021	.000	**.000**	**.094**	**.001**
NWW	Pearson	-.074	-.903**	.464**	.963**	**1**	**-.070**	**-.199**
	Sig. (2-tailed)	.577	.000	.000	.000	.000	**.597**	**.131**
VT	Pearson	.995**	.024	-.816**	-.220	-.070	**1**	**.604****
	Sig. (2-tailed)	.000	.858	.000	.094	.597	.000	**.000**
SS	Pearson	.617**	.199	.422**	-.432**	-.199	.604**	**1**
	Sig. (2-tailed)	.000	.130	.001	.001	.131	.000	.000

**Correlation is significant at the 0.01 level (2-tailed)
*Correlation is significant at the 0.05 level (2-tailed).

The next charts (figs.1-8) show the most important cross-correlations between vocal parameters:

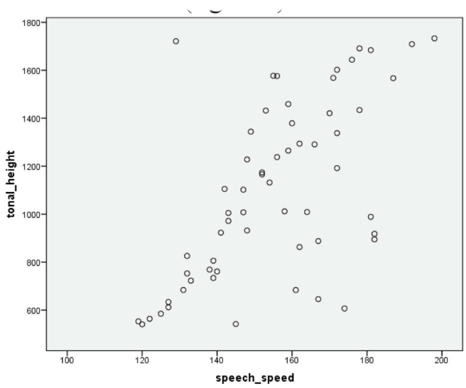

Figure no. 1

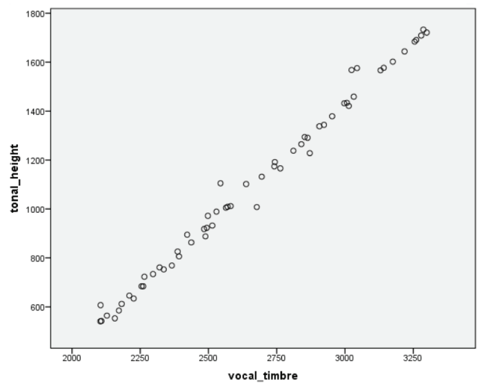

Figure no. 2

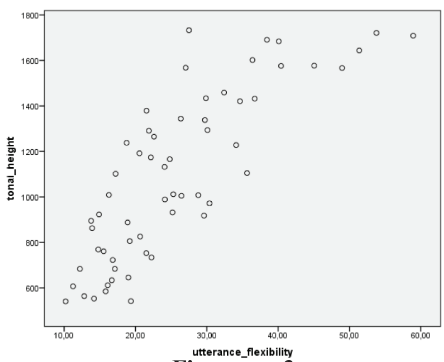

Figure no. 3

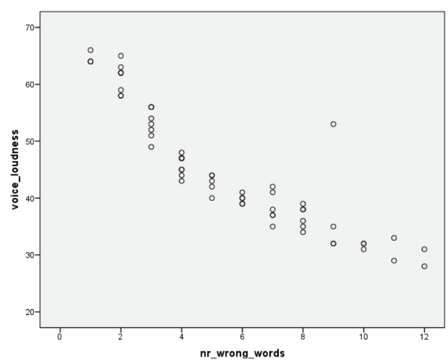

Figure no. 4

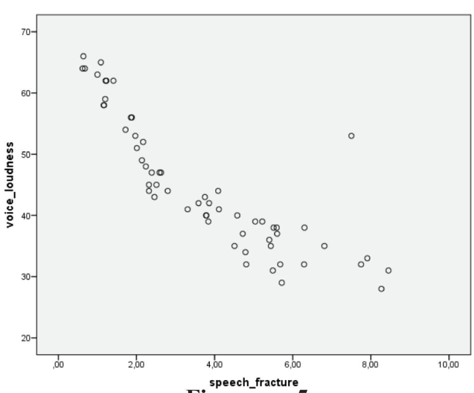

Figure no. 5

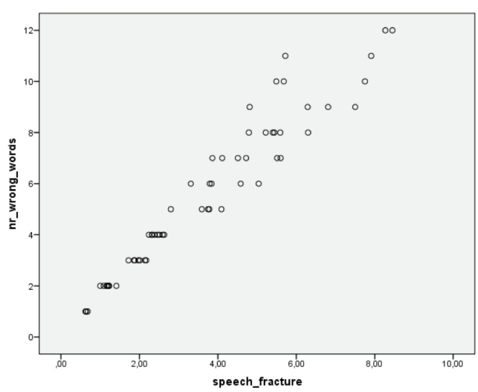

Figure no. 6

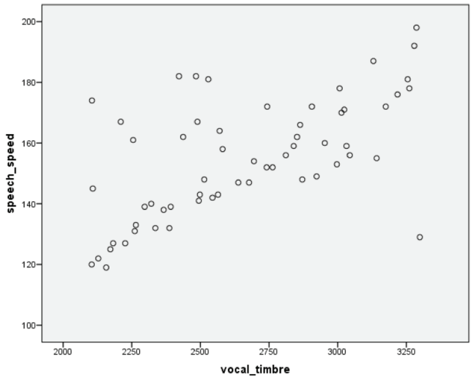

Figure no. 7

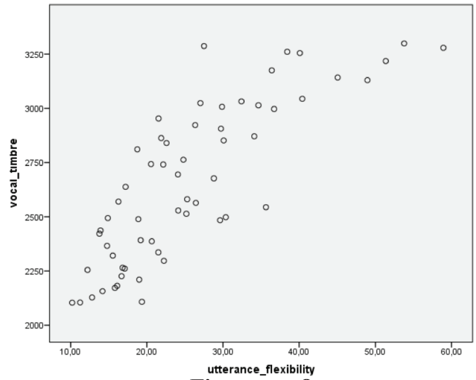

Figure no. 8

6. CONCLUSIONS

The positive strong correlation between TH (tonal_height) and VT (vocal_timbre) is normal because of the height of the voice which is often confused with the frequency of the voice tone itself, and it depends at the same time on the stylistics of every person, as a result of a rigorous control of intonation and articulation, the status of the body and particularly, the vocal tract system.

Another strong relationship is developed between VL (voice_loudness) and NWW (nr._wrong_words). This correlation is a negative one and shows us that, the more the number of speaking/uttering errors, the weaker the voice intensity of the speakers. Further, the correlational pairs (VT-SS) and (SS-TH) underline direct but significant relationships of medium level, between 60÷62%.

Concerning the first pair we can say that about 60% from the maintenance of a constant level of vocal timbre is due to an adequate text utterance. All too naturally, because of the direct and very strong relationship between TH and VT it is natural that the intensity of it to be partly transferred to SS-TH pair.

So, concerning the pair SS-VT a certain speed of text utterance leads to a particular level of voice sound intensity appearance. The faster is the speed of speech and the less the metrics of uttering, the bigger are the values of VT-TH.

About the derivative parameter UF (utterance_flexibility) we observe a massive concentration of a functional dependence of this parameter by the group TH-VL-NWW-VT-SS. In this parametric cluster we can distinguish those two pairs of correlations which were analyzed earlier: TH-VT and VL-NWW.

If we consider the relationship as a definition (1), we observe from the point of view of coefficients' sign that UF (utterance_flexibility) rises at the same time with the development of TH parameter (r=0.815) and decrease with the increasing of VL (r=-0.508).

So, the subjective and emotional component of the discourse is enhanced by an increase of the vocal sound height and the vocal timbre too, at the same time with the decreasing of the voice intensity.

On the other hand, increasing the speed of speech on texts, SS is explained about 42÷44 % percentage variable growth UF, respectively decreasing the speech fracture SF.

The negativity of the correlation coefficient between SS and SF is in accordance with the formula of the SF parameter shown in (2). Obviously, at the same time with the increasing number of speaking errors, the speech fracture also significantly increases.

REFERENCES

[1] Kochanski G., Grabe E., Coleman J., Rosner B. (2005) Loudness Predicts Prominence; Fundamental Frequency Lends Little, J. Acoustical Society of America 11(2).

[2] Opariuc D. (2009) Statistica Aplicata In Stiintele Socio-Umane (Romanian for Applied Statistics in Socio Humanities), Asociatia de Stiinte Cognitive din Romania Publishing House.

[3] http://www.qsound.com/technology/overview.htm.

[4] http://www.goldwave.com.

[5] Field A. (2009) Discovering Statistics Using SPSS, SAGE Publications Ltd; Third Edition Edition.

[6] A.E. Aronson, D.Bless (2009) Clinical Voice Disorders, Thieme, NY, 4th Edition.

[7] I. Purcaru (1988) Informație şi corelație (Romanian for Information and Correlation), Scientific and Encyclopedic Publishing House.