Voice Stress Detection: A Method for Stress Analysis Detecting Fluctuations on Lippold Microtremor Spectrum using FFT

Roberto Cabrera Cósetl and J. M. David Báez López
Departamento de Computación, Electrónica y Mecatrónica
Universidad de las Américas-Puebla
Cholula, Puebla, 72820
MEXICO
roberto.cabreracl@udlap.mx, david.baez@udlap.mx

Abstract

Stress detection in voice gives a great alternative for obtaining a noninvasive way to extract information about a possible deception from a person declaration. This article contains information and results of a primary work done to show how changes in Lippold microtremor can be detected through FFT signal processing when a person is under psychological pressure. The principal purpose is to obtain a tool that could help innocent people to prove their guiltlessness of having committed an offense or a crime.

1. Introduction

Stress detection in voice has become an important tool in different areas like psychology in order to detect some emotion like anger or happiness, as well, in affective computing area where it is used to achieve robust systems in voice recognition and in applications where control by voice is applied. As example, in the military field, speech technologies require integral use of speech systems for communication, command, control, and intelligence tasks. But they have problems when stress conditions are present for speech recognition, speaker verification, and synthesis and coding. These are reasons serious studies are conducted by multi-national military and non military laboratories in order to study the impact of factors such high workload, sleep deprivation, fear and emotion, confusion, psychological tension, pain, etc. on speech technology [1]

Stress detection is also present in informal applications like that in UK were people use voice stress analysis to record opinions from the customer about commercial products like clothing. They get feedback guessing a real perspective of the product through the voice analysis. Another interesting example is when police and telecommunication companies use this technology to detect frauds on emergency calls. [2, 3]

Also, an important application exists in legal areas, where Voice Stress Detection is used to validate and support true declarations from people who are innocent for having committed some illegal or criminal action. We are talking about software technology based in voice stress detection named VSA (Voice Stress Analyzers), which are based in the analysis of Lippold Microtremor theory, and offers advantages over the polygraph because of their noninvasive characteristic to detect stress. [7, 8, 9, 10]

Nevertheless, polygraph is the most popular and accepted technology that is widely and traditionally used in courtrooms in the United States of America. This technology is used to test people who are suspected to have participated in a crime. [2, 3] It consists of an integration of many medical human body lecturing devices which measure a person heart rate, blood pressure, respiratory rate and electro-dermal activity changes. The objective of the polygraph is to show fluctuations that may indicate that a person is being deceptive. These fluctuations are shown on graphics like those in figure 1. There are other technologies that aim to the study of physiological changes like Neuroscience. Nevertheless, in this paper we will concentrate on VSA and characteristics inherited from the polygraphic techniques. [6]

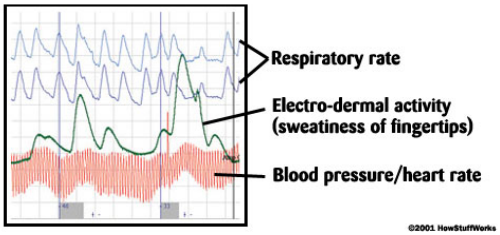


Fig. 1. Graphics from a traditional polygraph.[3]

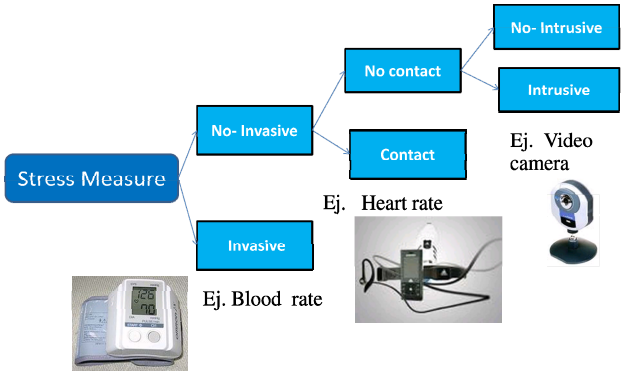


Fig. 2. VSA is a noninvasive technology.

The principal difference between the above technologies is that the polygraph measures some human responses like respiratory rate, electro-dermal activity, heart rate and other through direct contact, while VSA's just analyze voice signal with no contact measurement, see figure 2. Nevertheless, they are different technologies, both use similar protocols with particular modifications to obtain lectures like: Relevant-Irrelevant Test, Comparison Question (Control Question) Test, Zone Comparison Test, and other. This means that each test session must follow a protocol in order to perform a realistic study of stress signals and develop them under a controlled environment. If no guideline is followed both analysis are exposed to errors caused by other stressors like fear to fail the test giving incorrect results [5].

A VSA can be defined as a tool that shows measurements, without body contact, of the involuntary psychology answer from a person voice that is under stress. The measurement can be either in a graphical or non graphical form. The theory that supports VSA is the one whose subject matter is the assumption that it had been discovered that human body has involuntary responses that are associated with deceptive answers and/or stressful situations. Some of these responses are the lack of Lippold microtremor or infrasonic frequency modulation in voice. [5]

An analyzer of stress measures the inexistence of micro oscillations that modulate human voice when the person under test is under stress. The phenomenon can be explained with the following: the muscles in the throat, which mainly modulate voice, get rigid suppressing the frequency modulation on the voice, especially infrasonic modulation, which means that fundamental frequency of the voice does not show infrasonic frequency modulation.

Two types of voice change are directly consequential of stress. The first of these is referred to as the gross change which usually occurs only as a result of substantially stressful situation. This change manifest itself in audio perceptible changes in speaking rate, volume, voice tremor, change in spacing between syllables, and changes in fundamental pitch or frequency of voice. The second type of voice change is that of voice quality and is not discernible to the human ear, but is an unconscious manifestation of the slight tensing of vocal cords under even minor stress, resulting in dampening of selected frequency variations. It is also known that a third signal category exists in the human voice and that this third signal category is related to the second type of voice change. This is an infrasonic, or subsonic, frequency modulation, which is present, in some degree, in both the vocal cord sounds and in the formant sound. This infrasonic signal is one of the more significant voice indicators of psychological stress. It has been determined that during a relatively relaxed state a natural muscular undulation occurs typically at the 8-12 Hertz range. This undulation causes a slight variation in the tension of the vocal cords and causes shifts in the basic pitch frequency of the voice. These shifts are about a central frequency and constitute frequency modulation of the central carrier frequency. In order to observe this frequency modulation any one of several existing techniques for the demodulation of frequency modulation can be employed, bearing in mind, of course, that the modulation frequency is the nominal 8-12 Hertz and the carrier is one of the bands within the voice spectrum[7][8].

From the previous paragraph extracted from [7][8], we can realize that there exists a manner to detect stress using FFT. Of course, a demodulation process has to be performed before we obtain the psychological tremor. Now, to demodulate the frequency carrier one can take fundamental frequency f_0 or pitch as the carrier frequency, expecting a signal with frequency components between 8-12 Hz. But according to the results obtained in the voice analysis, we find that the main voice frequency components suffer the infrasonic modulation.

2. Method to obtain Corpus

Participants and Design:
16 women and 10 men from the Universidad de las Américas, Puebla were interviewed with questions from three dynamics, where each interview was recorded. The *primary dynamic* consisted in writing down five different foods that the participant likes the most and another five that dislikes, also five characteristics the participant considers valuable from other people and five characteristics that they detest from people, and five activities they never would like to perform (i.e. prostitute, drinking, etc). Then, the interviewer asked participant for answering with lies. The second dynamic, consists of a certain number of quick questions that become more difficult and stressful to answer as time passes due to the personal nature of questions. Finally, the third dynamic the participants are asked for remembering a strong situation where they had to lie and they feel ashamed of having done so. Then, they are asked to lie while the interviewer asks them for details of the situation they related. This way the participants are induced to remember the situation and forced to lie again feeling ashamed for lying.
Another study was developed over four real cases where people suspected to have committed a specific crime or abuse were interviewed and audio-recorded by authorities in the United States of America. Two recordings per case were available to analyze, where each couple of recording belongs to

a single suspicious person. People are native English speakers. Finally, results obtained through our analysis are compared with those registered by a Diogenes VSA apparatus.

Materials

The material used to perform studies over the voice is: a Sony IC Recorder (ICD-MS515) with a microphone ECM-DM5P, with a frequency response 100Hz to 15 kHz in parallel with a PC 140 Sennheiser microphone with a frequency response 80Hz to 15 kHz, to record all the interviews. In addition, in case of records made with the IC Recorder, we used the Digital Voice Editor V.3 software to convert from MSV (Memory Stick Voice File) to WAV(Waveform Audio Formant) file 16 bits, 44.1 kHz, stereo. A Gold Wave Software was used to convert and extract answers from the interviews. The resultant files are WAV's 16bits, 32 kHz, Mono. Finally, we used Matlab in order to program GUI´s (Graphical User Interfaces) that help us to extract and process answers.

All answers were not only Spanish "Sí" or "No" but also English "Yes" and "No". Although not all answers are true, we part from the base that we know real answers so we are able confirm the information delivered from the computer processing.

3. GUI and Program Processing

In order to process the signal, a program has been created over the Matlab platform. The next figure describes the manner the GUI operates, and the path that follows the data to be processed.

The corpus obtained in the interviews described previously was recorded and saved as a new answer. The new record can be opened and analyzed in the same manner that corpus. As well, there is a tool to perform an interview with N questions, this way one can store the total interview in a wav file, and similarly, each answer in its own wav file. Also, at the end of interview one can decide whether to atomically analyze all answers and keep results in an Excel file. If not recorded, then the files are stored and we can analyze any answers manually any other time. All files are saved to a predefined folder (Refer to figure 4). Figure 5 shows the general process the signal follows to be analyzed. Pitch is the frequency of human voice produced by the vibration of the vocal cords, which in turn, is a product of partially closing the glottis and forcing air through the glottis by contraction of the lung cavity and the lungs. The frequencies of these vibrations can vary generally between 100 and 300 Hz, depending upon the sex and age of the speaker and upon intonations the speaker applies. [8]

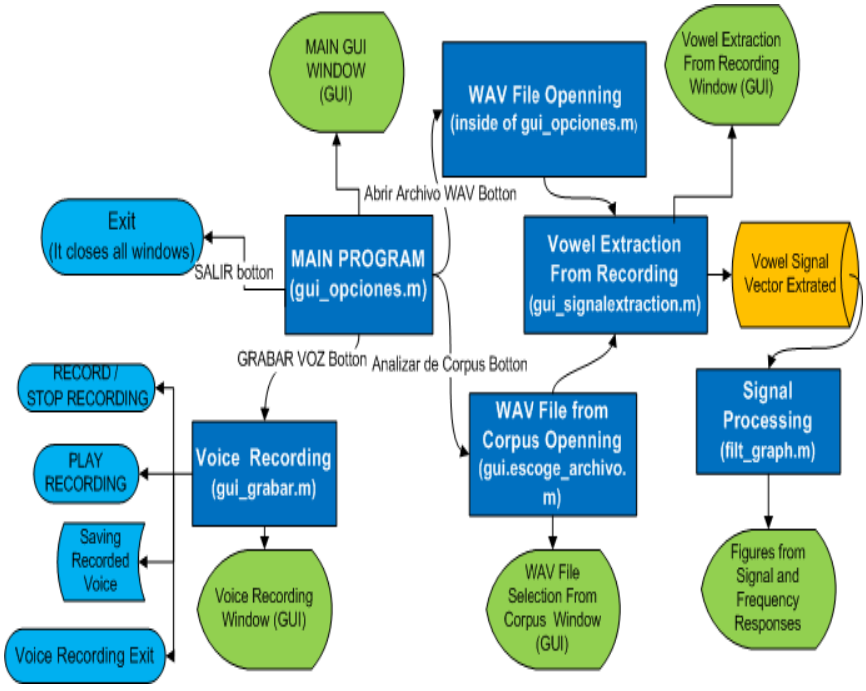


Fig. 3. Diagram shows the GUI's (Graphical Interface User) communication and processing mode

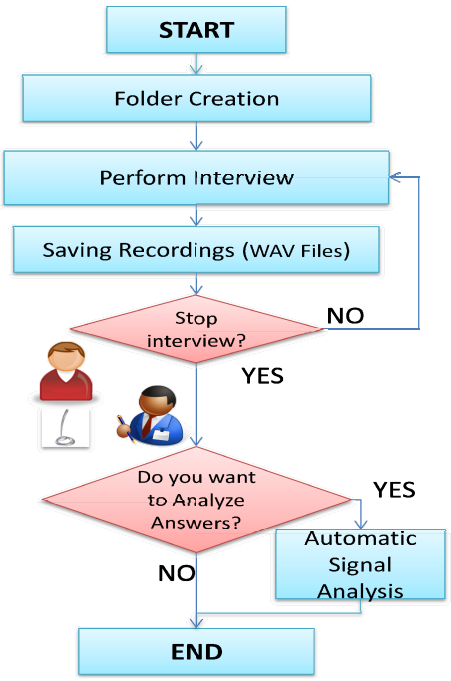


Fig. 4. Questioning Signal Process.

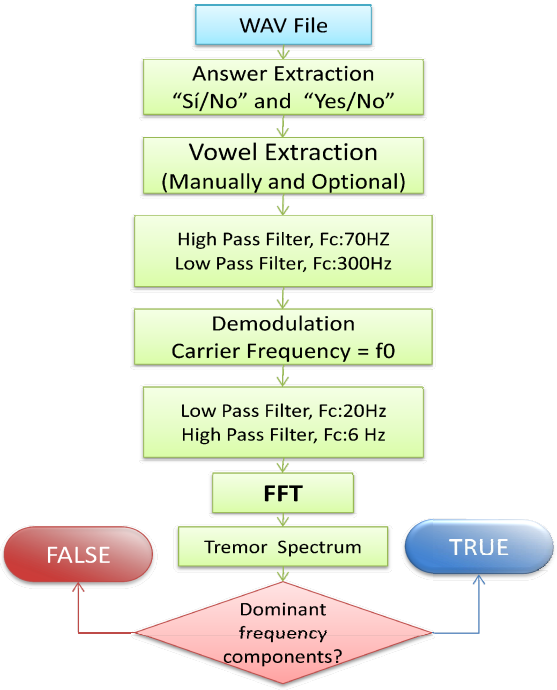
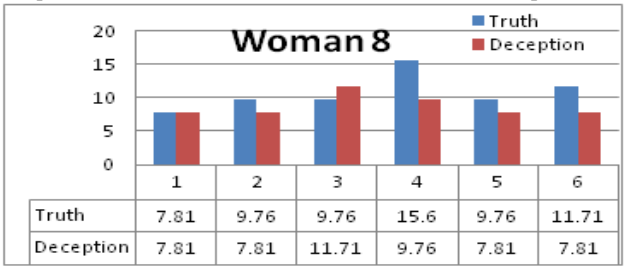


Fig. 5. Voice Signal processing

4. Results

Through the FFT application over the demodulated voice signal, we were able to observe that the frequency components between 8 to 12 Hz present a magnitude diminution when a person is under stress. To detect these changes visually we compare deceptive answers' tremor spectrum against non deceptive answers' tremor spectrum. Here, we can affirm that differences were observed from these two results types. Nevertheless, differences do not always appear where variations are expected to be. Additionally, another way to detect stress was reached. We were able to detect stress automatically with the program which detects dominant frequency components between 6-20Hz, and delivers whether an answer is false or true.

Table1. Highest frequency (Hertz) component from 10 different questions, half of them are true and the rest are deceptive.



Truth	7.81	9.76	9.76	15.6	9.76	11.71
Deception	7.81	7.81	11.71	9.76	7.81	7.81

In table 1, there are the dominant frequency components of five deceptive answers (red and left bar) and five true answers (blue). One can observe that true answers keep a bigger magnitude in frequencies major to 7.81Hz. In the analysis it is considered a true or false stressful answer when dominant frequency components are present between 8-12 Hz. Results are more remarkable on hard questions when people are evidently stressed and do not really want to answer trustily.

From the interviews performed of native English speakers mentioned before, results shows signals of stress over some answers. These stress detections coincides with some outcomes of Diogenes apparatus. In table 2 and 3 there are results from the analysis applied over the wav archives recorded from Angela, woman suspected to have stolen a digital camera. In the table, in the first column are the results the software produced, in second column are the questions of the interviewer and the answers of Angela, third column corresponds to the number of samples of the answer, fourth column are some attributes extracted from the analysis of the archive wav of the answer, and the fifth column are the values that corresponds to the attributes of fourth column. Differences can be observed when one compares the table 2 and 3.

We can observe that some simple answers of an irrelevant question like the number four, in the second interview, shows less stress and the software produces a “true” as a result. Also, question 10 is “false” in both interviews according to the analysis. Nevertheless, answer from the question 5 is “false” only in the second interview. It is important to know that both interviews were made applying MZOC-GC of Diogenes Company.

This method uses first-question stress (FQS), irrelevant, relevan,t and control questions in order to determine when an answer can be taken as no deceptive using a specific order in the interview. For more analysis, it is possible to record an interview and get an automatic result following some of a formal protocol like MZOC-GC of Diogenes Company to develop an interview for stress detectors [11].

5. Conclusions

Although no psychological micro tremors were clearly observed after the voice demodulation, changes in its frequency composition occurs and we are able to detect it by using the FFT together with an algorithm to detect dominant frequency components.

FFT is a very used method for signal processing, and it was applied to show that in effect there occur changes in the frequency components of a demodulated voice signal in a rank of 8 to 12 Hz.

The majority of people, who were interviewed for this paper, gave their answers knowing that no critical consequences could derive from their answers.

Table 2. Results from the test applied over the first suspicious person Angela, stolen digital camera. First interview.

RESULT	Nombre de Archivo	# Samples	Atributes	F0
TRUE	1 FQS_Is your first name Angela_YES.wav	2658	Frequency	175.931233
			Magnitud	4.656851629
			Energy	3.39252E+17
			Tremor Hz	8.074951
			Tremor Mag.	421938800
TRUE	2 IR - C_Are you over 23 years of age_NO.wav	3023	Frequency	209.0493048
			Magnitud	156.1293355
			Energy	3.4087E+17
			Tremor Hz	8.074951
			Tremor Mag.	355499400
TRUE	3 Rel (Y)_Your story regarding Larry is true_YES.wav	3503	Frequency	160.2174422
			Magnitud	1.392417808
			Energy	5.80863E+17
			Tremor Hz	8.074951
			Tremor Mag.	696183400
FALSE	4 IR -C_Do you live in the United States_YES.wav	1887	Frequency	197.550359
			Magnitud	8.166705369
			Energy	4.08207E+16
			Tremor Hz	5.383301
			Tremor Mag.	179221600
TRUE	5 Rel (Y)_Did you take the digital camera_No.wav	2827	Frequency	206.650599
			Magnitud	142.1371314
			Energy	8.74053E+16
			Tremor Hz	8.074951
			Tremor Mag.	195355100
TRUE	6 C_Did you ever commited a serious crime but were never caugth_NO.wav	4325	Frequency	201.39301
			Magnitud	243.6162276
			Energy	8.56584E+17
			Tremor Hz	12.11243
			Tremor Mag.	551882900
TRUE	7 IR-C_Did you ever go to school_YES.wav	3578	Frequency	127.7889384
			Magnitud	8.857150502
			Energy	3.81924E+17
			Tremor Hz	8.074951
			Tremor Mag.	422938900
FALSE	8 Rel (K)_Do you know who took the digital camera_NO.wav	6670	Frequency	163.3170636
			Magnitud	1.08722313
			Energy	6.19804E+17
			Tremor Hz	6.729126
			Tremor Mag.	377332700
TRUE	9 Rel (S)_Do you suspect anyone taking the digital camera_NO.wav	4255	Frequency	191.6161868
			Magnitud	4.750750947
			Energy	7.21188E+17
			Tremor Hz	9.420776
			Tremor Mag.	488687600
FALSE	10 C_Do you now remember ever committing a crime for what you were never caugth_YES.wav	3284	Frequency	186.9379281
			Magnitud	4.359971883
			Energy	2.94803E+17
			Tremor Hz	16.1499
			Tremor Mag.	294311000
TRUE	11 Rel (K)_Do you know where the digital camera is_NO.wav	5934	Frequency	169.5268764
			Magnitud	9.857965923
			Energy	1.14999E+18
			Tremor Hz	8.074951
			Tremor Mag.	540931900

Table 3. Results from the test applied over the second interview of the first suspicious person Angela.

RESULT	Nombre de Archivo	# Samples	Atributes	F0
TRUE	1 FQS_Is your first name Angela_YES.wav	3466	Frequenc	184.024692
			Magnitud	7.726457229
			Energy	6.71798E+17
			Tremor Hz	8.074951
			Tremor Mag.	636041300
TRUE	2 IR - C_Are you over 23 years of age_NO.wav	3477	Frequency	212.3904169
			Magnitud	146.454903
			Energy	5.13954E+17
			Tremor Hz	10.7666
			Tremor Mag.	486563500
TRUE	3 Rel (Y)_Your story regarding Larry is true_YES.wav	5382	Frequency	149.2964579
			Magnitud	4.331902492
			Energy	6.41427E+17
			Tremor Hz	14.80408
			Tremor Mag.	359467000
TRUE	4 IR -C_Do you live in the United States_YES.wav	2517	Frequency	201.2355854
			Magnitud	23.10054684
			Energy	2.11156E+17
			Tremor Hz	8.074951
			Tremor Mag.	384802700
FALSE	5 Rel (Y)_Did you take the digital camera_No.wav	3520	Frequency	211.2809467
			Magnitud	154.0133557
			Energy	2.40935E+17
			Tremor Hz	13.45825
			Tremor Mag.	309588400
TRUE	6 C_Did you ever commited a serious crime but were never caugth_NO.wav	3287	Frequency	210.7099283
			Magnitud	182.2197063
			Energy	3.96476E+17
			Tremor Hz	10.7666
			Tremor Mag.	509781600
TRUE	7 IR-C_Did you ever go to school_YES.wav	4046	Frequency	141.009271
			Magnitud	8.35118064
			Energy	2.51631E+17
			Tremor Hz	8.074951
			Tremor Mag.	414173200
TRUE	8 Rel (K)_Do you know who took the digital camera_NO.wav	3282	Frequency	206.0360655
			Magnitud	140.3763268
			Energy	4.74489E+17
			Tremor Hz	10.7666
			Tremor Mag.	497086700
TRUE	9 Rel (S)_Do you suspect anyone taking the digital camera_NO.wav	3036	Frequency	214.571562
			Magnitud	44.27827715
			Energy	5.05458E+17
			Tremor Hz	10.7666
			Tremor Mag.	559774900
FALSE	10 C_Do you now remember ever committing a crime for what you were never caugth_YES.wav	2748	Frequency	190.9640172
			Magnitud	5.546043948
			Energy	3.04191E+17
			Tremor Hz	13.45825
			Tremor Mag.	353637800
TRUE	11 Rel (K)_Do you know where the digital camera is_NO.wav	3335	Frequency	192.2886598
			Magnitud	21.82009462
			Energy	4.77629E+17
			Tremor Hz	10.7666
			Tremor Mag.	515072600

This explains why, in some cases, no stress is detected. In order to obtain more clearly results, it is proposed to perform recordings to interviews sessions over people that are in jail. These people naturally will be under real pressure and then when the answers are analyzed we would obtain better results of stress detection.

Voice stress analysis is not just for the detection of deception, but also for the detection of anomalies in people, who is under aggressive work environments. There also exists work that take studies of voice stress to the speech recognition.

Even though work has been done to detect deception there is not a sophisticated procedure for detecting deception can warrant a 100% of accuracy because of the presence of some erroneous signs of deception when a true answer is analyzed. That is why some literatures recommend using VSA as an auxiliary tool to detect some signals of stress from the interviewed person.

Finally, [8] discusses the investigation done by the Air Force Research Laboratory (AFRL) which has been tasked by the National Institute of Justice to investigate voice stress analysis (VSA) technology and evaluate its effectiveness for both military and law enforcement applications. This study concludes that VSA technology can identify stress better than polygraph systems, but that experience and training improves the accuracy of results.

6. References

[1]. C. Vloeberghs et. al. "The Impact of Speech Under Stress on Military Speech Technology". Research and Technology Organization/North Atlantic Treaty Organization 2000.

[2]. Kevin Boson. "How Lie Detectors Work" http://people.howstuffworks.com/lie-detector.htm

[3]. BBC News (2003)." *Lie detectors cut car claim'"* http://news.bbc.co.uk/1/hi/uk/3227849.stm

[4]. L F Lowenstein PhD. "Possible Signs of Deception and How to Detect Them,_Part II". Southern England Psychological Services.

[5]. The Polygraph and Lie Detection (2003) Board on Behavioral, Cognitive, and Sensory Sciences and Education (BCSSE) Committee on National Statistics (CNSTAT)

[6]. Kenneth R. Foster. "Building Better Lie Detectors With Neuroscience?" Spectral Lines.

[7]. Lippold, Olof "Oscillations In The Stretch Reflex Arc And The Origin Of The Rhythmical 8-12 C/S Component Of The Physiological Tremor." The Journal Of Physiology, February 1970.

[8]. CliffordS Hopkins et. al "Evaluation of Voice Stress Analysis Technology", 38ª Conferencia Internacional de Ciencias de Sistemas Hawai. 2005

[9]. Bell Jr. et al. "Physiological Response Analysis Method and Apparatus." 1979. United States Patent.

[10]. Nunally Patrick O´Neal, "Audio Psychological Stress Indicator Alteration Method and Apparatus." 2003

[11]. Rules in reading the 14 question MZOC-GC test. Diogenes Company Rights Reserved 1997.