Jitter, Shimmer and HNR Classification within Gender, Tones and Vowels in Healthy Voices

Article *in* Procedia Technology · December 2014

DOI: 10.1016/j.protcy.2014.10.138

2 authors:

João Paulo Teixeira
Polytechnic Institute of Bragança

Paula Odete Fernandes
Polytechnic Institute of Bragança

Available online at www.sciencedirect.com

ScienceDirect

Procedia Technology 16 (2014) 1228 – 1237

Procedia Technology

CENTERIS 2014 - Conference on ENTERprise Information Systems / ProjMAN 2014 - International Conference on Project MANagement / HCIST 2014 - International Conference on Health and Social Care Information Systems and Technologies

Jitter, Shimmer and HNR classification within gender, tones and vowels in healthy voices

João Paulo Teixeira[a,b]*, Paula Odete Fernandes[a,b,c]

[a]Polytechnic Institute of Bragança, Campus de Sta Apolónia, 5301-857 Bragança, Portugal
[b]UNIAG, Portugal
[c]NECE (UBI), Portugal

Abstract

A statistical analysis of the Jitter, Shimmer and Harmonic to Noise Ratio parameters was applied to classify and compare genders, vowels and tones of healthy voices. Different type of speech records has used for the comparison, namely records with sustained vowels /a/, /i/ and /u/ at High, Low and Neutral tones. A gender comparison has made denoting differences only in Jitter parameter. The parameters determined in recorded vowels /a/, /i/ and /u/ has also compared and the Kruskal Wallis statistical test showed differences for parameters rap, Shim, ShdB, apq3, apq5 and HNR. High, Low and Neutral tones has compared using the same statistical test denoting statistical differences for all Jitter, Shimmer and HNR parameters. A statistical classification of the mean and standard values for these parameters on healthy voices is also presented.

Keywords: Jitter, Shimmer, Harmonic-to-Noise Ratio (HNR), Vocal Acoustic Analysis.

1. Introduction

The speech analysis of patient's voice is nowadays very valuable technique for speech pathology detection [1, 2, 3, 4], because voice disorder's can be noticeable by the analysis of several acoustic signal parameters.

Various techniques have been used to assess the patient's voice quality. One of them consists in the auditory perceptual analysis; however these may lead to different results depending on the experience of the practitioner

* Corresponding author. Tel.: +351 273 303129; fax: +351 273 313051.
E-mail address: joaopt@ipb.pt

Peer-review under responsibility of the Organizing Committee of CENTERIS 2014.
doi:10.1016/j.protcy.2014.10.138

involved. In the medical field this is a subjective assessment technique which leads to the lack of consensus among professionals. Therefore it became necessary to search for an objective assessment, in which the voices were analyzed by devices which are capable of measuring several acoustic parameters, as stated by Almeida [4]. Using speech signal processing it is possible to extract a set of parameters of the voice that may allow detecting pathologies of the vocal cords in individuals by comparing the data of patients with certain pathology with the data of persons considered with healthy voice.

The parameters obtained by the acoustic analysis have the advantage of describing the voice objectively. With the existence of normative databases characterizing voice quality or using intelligent tools combining the various parameters, it is possible to distinguish between normal and pathological voice or even identify or suggest the pathology. These tools allow the monitoring of clinical standpoint and reduce the degree of subjectivity of perceptual analysis, as Teixeira, et al. [5].

Currently, acoustic parameters commonly used in applications of acoustic analysis as well as the most referenced in the literature, are the fundamental frequency (F_0), jitter, shimmer, HNR and frequency formants.

The measure of these parameters is performed in a recorded speech signal with the patient/control producing a long steady state vowel.

Measurements of F0 disturbance jitter and shimmer, has proven to be useful in describing the vocal characteristics. Jitter is defined as the parameter of frequency variation from cycle to cycle, and shimmer relates to the amplitude variation of the sound wave, as Zwetsch et al. [2] and [5, 6, 7]. In Fig. 1 the jitter and shimmer are represented.

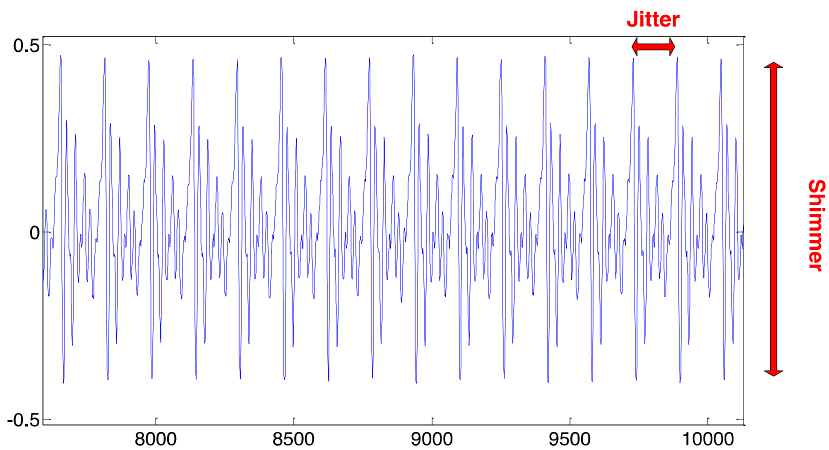

Fig.1. Jitter and Shimmer perturbation measures in speech signal [6].

The jitter is affected mainly by the lack of control of vibration of the vocal cords; the voices of patients with pathologies often have a higher percentage of jitter.

The shimmer changes with the reduction of glottal resistance and mass lesions on the vocal cords and is correlated with the presence of noise emission and breathiness.

Diseases that affect larynx cause changes in the patient's vocal quality. Early signs of deterioration of the voice due to vocal malfunctioning are normally associated with breathiness and hoarseness of the produced voice. The most common signs that may indicate changes in the larynx relate hoarseness, breathiness and roughness. The transient hoarseness may result from abuse of the voice or the casual flu. But when the hoarseness persists and becomes a characteristic voice, is indicative of pathology of the larynx. Hoarseness can also be an early symptom of cancer of the larynx, Teixeira, et al. [5]. The most common pathologies affecting voice are vocal nodules, the laryngitis, the paralysis, polyps, cysts and Reinke's Edema. Other pathologies of the larynx that may lead to dysphonic speech are ulcers of contact, as stated by Lopes [8].

2. Methods and Methodology

2.1. The Saarbrücken Voice Database

The Saarbrücken Voice Database (SDB) [9] was used in this study. The part of the SDB for healthy voices consisting in 34 female and 7 male has used. Only 7 male control subject were available in the data base. These healthy voices will be used in further studies as a control voices. For each voice one segment of speech record was used for sustained vowels /a/, /i/ and /u/ for High, Low and Mid/Neutral tones in a total of 9 speech segments. The average and standard deviation ages for female controls is 23.8±7.4, and for male controls 31.3±14.3.

Each segment of speech consists in a steady state sustainable pronunciation of the respective vowel.

For each speech segment a set of jitter, shimmer and HNR parameters, detailed below, was determined using the Praat software [10].

2.2. Data Analysis

In this paper a statistical analyses of healthy voices is presented. A statistical mean and standard deviation values for male and female voices for each parameters is presented. A comparison between male and female voices for each parameter is performed. Concerning the tones, three different tones (High, Low and Neutral) are analyzed with the healthy voices. Namely, a presentation of mean values for each parameter and a comparison of the tones for each parameter is presented. For the vowels a similar study was conducted for vowels /a/, /i/ and /u/.

In order to search for differences between genders, tones and vowels for each parameter inferential statistical analyses was carried out.

It must be mentioned that when it was not possible to use the parametric tests because one of the assumptions were violated (i- sample size is 30 or greater or the variables follows a normal distribution, ii- the variance in the independents samples should be approximately equal) the non-parametric tests were used. In all analysis it was used a significance level of 0.05.

3. The Parameters

The set of jitter and shimmer parameters and the Harmonic-to-Noise Ratio (HNR) is presented below according to several authors [5, 6, 7, 11, 12, 13].

3.1. Jitter

The values for Jitter can be measured in different parameters, such as absolute, relative, relative average perturbation (rap) and the period perturbation quotient (ppq5).

Jitter absolute is the cycle-to-cycle variation of fundamental frequency, i.e. the average absolute difference between consecutive periods, expressed as:

$$jitta = \frac{1}{N-1}\sum_{i=1}^{N-1}\left|T_i - T_{i-1}\right| \tag{1}$$

Where Ti is the extracted glottal period lengths and N is the number of extracted glottal periods.

Relative Jitter or local Jitter is the average absolute difference between consecutive periods, divided by the average period. It is expressed as a percentage:

$$jitter(relative) = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|T_i - T_{i-1}\right|}{\frac{1}{N}\sum_{i=1}^{N} T_i} \times 100 \tag{2}$$

Jitter (rap) is defined as the Relative Average Perturbation, the average absolute difference between a period and the average of it and its two neighbors, divided by the average period. It is expressed as a percentage:

$$rap = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|T_i - \left(\frac{1}{3}\sum_{n=i-1}^{i+1} T_n\right)\right|}{\frac{1}{N}\sum_{i=1}^{N} T_i} \times 100 \tag{3}$$

Jitter (ppq5) is the five-point Period Perturbation Quotient, computed as the average of it and its four closest neighbors, divided by the average period. It is also expressed as a percentage:

$$ppq5 = \frac{\frac{1}{N-1}\sum_{i=2}^{N-2}\left|T_i - \left(\frac{1}{5}\sum_{n=i-2}^{i+2} T_n\right)\right|}{\frac{1}{N}\sum_{i=1}^{N} T_i} \times 100 \tag{4}$$

3.2. Shimmer

Shimmer (dB) is expressed as the variability of the peak-to-peak amplitude in decibels, i.e. the average absolute base-10 logarithm of the difference between the amplitude of consecutive periods, multiplied by 20:

$$ShdB = \frac{1}{N-1}\sum_{i=1}^{N-1}\left|20 * \log\left(\frac{A_{i+1}}{A_i}\right)\right| \tag{5}$$

Where A_i is the extracted peak-to-peak amplitude data and N is the number of extracted fundamental frequency periods.

Shimmer relative is defined as the average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude, expressed as a percentage:

$$Shim = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|A_i - A_{i+1}\right|}{\frac{1}{N}\sum_{i=1}^{N}A_i}\times 100 \tag{6}$$

Shimmer (apq3) is the three-point Amplitude Perturbation Quotient, the average absolute difference between the amplitude of a period and the average of amplitudes of its neighbors, divided by the average amplitude. It is expressed in percentage:

$$apq3 = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|A_i - \left(\frac{1}{3}\sum_{n=i-1}^{i+1}A_n\right)\right|}{\frac{1}{N}\sum_{i=1}^{N}A_i}\times 100 \tag{7}$$

Shimmer (apq5) is defined as the five-point Amplitude Perturbation Quotient, the average absolute difference between the amplitude of a period and the average of the amplitudes of its four closest neighbors, divided by the average amplitude. It is also expressed in percentage:

$$apq5 = \frac{\frac{1}{N-1}\sum_{i=2}^{N-2}\left|A_i - \left(\frac{1}{5}\sum_{n=i-2}^{i+2}A_n\right)\right|}{\frac{1}{N}\sum_{i=1}^{N}A_i}\times 100 \tag{8}$$

3.3. Harmonic to noise ratio

The Harmonic to noise ratio provides an indication of the overall periodicity of the voice signal by quantifying the ratio between the periodic (harmonic part) and aperiodic (noise) components. This parameter is usually measured as an overall characteristic of the signal, and not as a function of frequency. The overall value of the HNR of the signal varies because different vocal tract configurations involve different amplitudes for the harmonics [14, 15, 16, 17].
HRN is given by following equation according to Boersma, P [14]:

$$HNR = 10 * \log_{10}\frac{AC_V(T)}{AC_V(0) - AC_V(T)} \tag{9}$$

Were $AC_v(0)$ is the autocorrelation coefficient at the origin consisting in the all energy of the signal. The $AC_v(T)$ is the component of the autocorrelation corresponding to the fundamental period. The difference between to all energy and the fundamental period energy is assumed to be the noise energy.

4. Results and Discussion

4.1. Gender Comparison

A general comparison of the 9 above mentioned parameters (jitta, jitter, rap, ppq5, ShdB, Shim, apq3, apq5 and HNR) is made in a first instance between female and male genders using the records for the three vowels and the three tones in a total of 9 different records for each voice, in total 369 observations under study (306 for male and 63 for male).

An analysis of mean and standard deviation for male and female is presented in Table 1 for each parameter. It is possible to observe the existence of differences in the values of mean and standard deviation only for the **jitta** parameter. The standard deviation for this parameter is relatively high that means a high dispersion of the values for this parameter.

Right part of Table 1 presents the results of the application of t-student analysis (parametric test). This test shows that only for the jitta parameter is statistically evident the differences between male and female values (p-value <0.001 for a significance level of 0.05). This result confirms the differences in the mean values for male and female healthy voices.

Table 1: Resume of the gender descriptive statistics for each parameter and t- Student test.

Parameters	**Gender**	**Descriptive Statistics**			**t- Student test**		
		n	**Mean**	**Std. Deviation**	**test value**	**df**	**p-value**
Jitta (µs)	Female	306	14,10	8,674	-7,747	68,019	< 0,001
	Male	63	32,11	18,028			
Jitt (%)	Female	306	,3324	,17168	-1,238	367	,216
	Male	63	,3619	,17637			
rap (%)	Female	306	,2014	,20217	,962	367	,337
	Male	63	,1762	,10883			
ppq5 (%)	Female	306	,1965	,12613	-,858	367	,391
	Male	63	,2111	,10490			
Shim (%)	Female	306	2,7458	2,31167	1,527	367	,128
	Male	63	2,2873	1,26242			
ShdB (dB)	Female	306	,2389	,19675	1,393	367	,165
	Male	63	,2032	,11355			
apq3 (%)	Female	306	1,3516	1,13736	1,403	367	,161
	Male	63	1,1429	,69601			
apq5 (%)	Female	306	1,6614	1,51868	1,184	367	,237
	Male	63	1,4286	,78708			
HNR (dB)	Female	306	24,9474	4,48382	1,518	367	,130
	Male	63	24,0095	4,36920			

This result can be explained because male voices have lower fundamental frequency (F_0), and therefore higher glottal periods than female voices. Thus in a longer glottal period it is expected longer perturbation in the glottal periods.

For the all remaining parameters there are no statistical evidences of differences for male and female voices.

Fig. 2 presents the error bars for the parameter of jitter (jitta, jitt, rap and ppq5), shimmer (Shim, ShdB, apq3 and apq5) and HNR for the gender comparison. The parameters are grouped in three graphs because of the scale of each parameter.

Error bars are a graphical representation of the variability of data and are used on graphs to indicate the error, or uncertainty in a reported measurement. They give a general idea of how accurate a measurement is, or conversely, how far from the reported value the true (error free) value might be.

The analyses of the error bar for the jitta parameter shows again that this parameter has distinctive results for each gender because the vertical lines for each gender never have the same value (in vertical scale). On the other way the vertical lines for each of the remaining parameters always cross in vertical scale meaning no statistical significance between male and female.

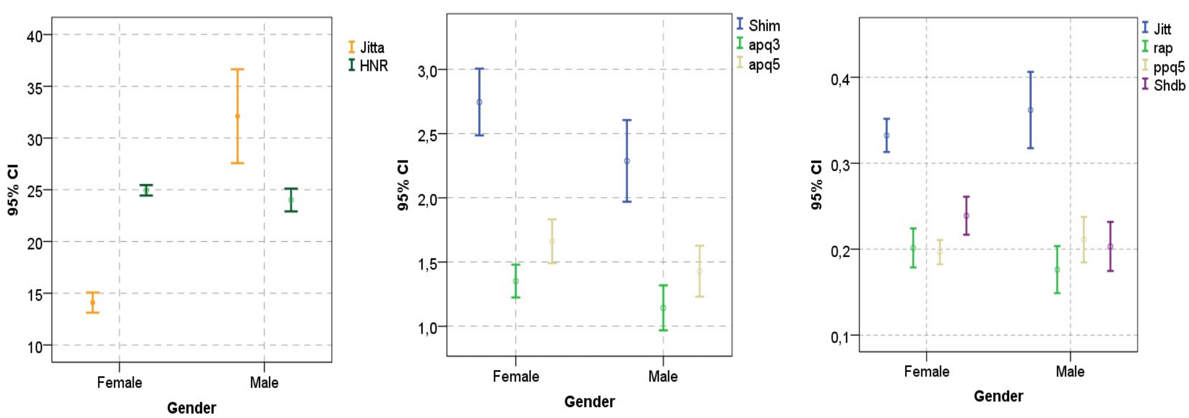

Fig. 2- Error bars for gender comparison.

4.2. Tone and vowel comparison

Grouping the parameters correspondent to each of the three tones (High, Low and Neutral) and then grouping the parameters by vowels (/a/, /i/ and /u/) the mean values of each parameter is presented in Table 2.

Table 2: Mean values of the parameters for H, L and N tones and for vowels /a/, /i/ and /u/.

Tone/Vowel	jitta (μs)	jitt (%)	rap (%)	ppq5 (%)	ShdB (dB)	Shim (%)	apq3 (%)	apq5 (%)	HNR (dB)
High	10.0	0.3	0.2	0.2	0.2	2.3	1.1	1.3	26.7
Low	17.9	0.4	0.2	0.2	0.3	3.1	1.5	1.9	24.0
Normal	14.3	0.3	0.2	0.2	0.3	2.9	1.5	1.8	24.2
/a/	14.3	0.3	0.2	0.2	0.2	2.6	1.4	1.6	23.9
/i/	13.0	0.3	0.2	0.2	0.2	2.2	0.9	1.2	24.2
/u/	15.0	0.4	0.2	0.3	0,3	3.5	1.7	2.2	26.7

For the statistical analyses by tones and by vowels, once the normality assumption was violated, the non-parametric test of Kruskal-Wallis was realized as alternative to the ANOVA test. This test intends to verify the existence of differences in the distributions by tones first and then by vowels, considering the 9 parameters under study. The results of the test are presented in table 3.

Table 3: Kruskal Wallis test for comparison of tones and vowels.

		Jitta	Jitt	rap	ppq5	Shim	Shdb	apq3	apq5	HNR
Tone	Chi-Square	54,184	15,259	7,852	15,465	27,313	25,465	19,493	31,515	26,899
	df	2	2	2	2	2	2	2	2	2
	p-value	< 0,001	< 0,001	,020	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
Vowels	Chi-Square	3,289	3,695	7,635	3,431	18,151	8,005	40,724	18,215	49,157
	df	2	2	2	2	2	2	2	2	2
	p-value	,193	,158	,022	,180	< 0,001	,018	< 0,001	< 0,001	< 0,001

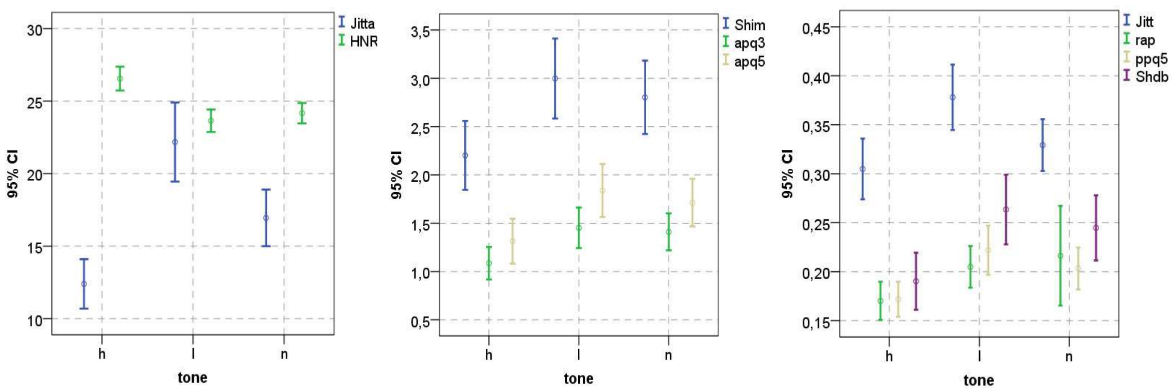

Fig. 3- Error bars for tone gender comparison.

Analyzing table 3 for tones (three independent samples are presented: High tone, Low tone and Neutral tone) there are statistically relevant differences on the values of the independent samples for all the 9 parameter. This means that at least one independent sample is different than the others. This can be seen because for each parameter the p-value is less than the significance level of 0.05.

This result can be complemented with the error bars displayed in Fig. 3 for the tones. For jitta parameter none tone bar share the same vertical space, meaning that each tone has statistically significant differences from the other. This can be explained because lower tones have longer glottal periods been comprehensive to have longer period variations. A similar analysis can be made for the other parameters. Therefore, for jitt, rap, ppq5, Shdb, Shim, apq3 and apq5 parameters, the high tone is different from low tone. For the HNR the high tone is different from low and normal tones.

Considering the results presented in table 3 for vowels (three independent samples are presented: /a/, /i/ and /u/) there are statistically relevant differences on the values of the independent samples for the parameters rap, Shim, ShdB, apq3, apq5 and HNR. This means that at least one independent sample is different than the others for these parameters. For the other parameters the values are independent of the vowel. This can be seen in the table because the p-value of these parameters is less than the significance level of 0.05.

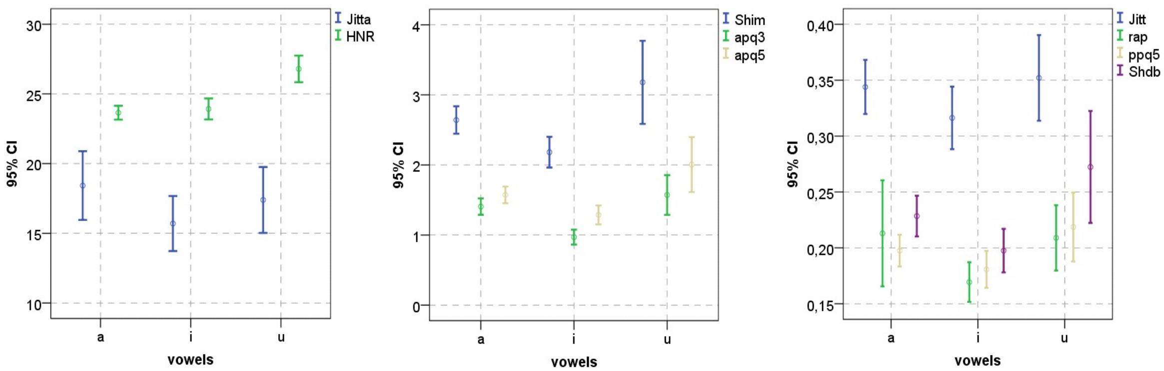

Fig. 4- Error bars for vowel comparison.

The error bars presented in Fig. 4 complement this result for the vowels. For jitta, jitt and ppq5 there is a vertical space shared by the three vertical bars. For the rap the vowel /i/ is different from vowel /u/; for Shim, ShdB and apq3 the vowel /i/ is different from vowels /a/ and /u/; for the parameter apq5 the three vowels are different; and for HNR the vowel /u/ is different from vowels /a/ and /i/.

5. Conclusion

In this paper the Saarbrücken Voice Database [9] was used to compare gender, tones and vowels for Jitter, Shimmer and HNR parameters of healthy voices.

The mean and standard deviation values for jitta, jitt, rap, ppq5, ShdB, Shim, apq3, apq5 and HNR for male and female voices were documented. Also the mean values for the same set of parameters for High, Low and Neutral tones and for vowels /a/, /i/ and /u/ were presented.

Considering the results of the gender comparison, only the jitta parameters registered statistically significant differences between male and female voices, being higher for male voices, naturally.

Considering the tone comparison, the jitta parameter again is different for each tone, being higher for low tones and lower for high tones. The explanation is the same as for male and female voice differences. For HNR parameter the high tone is different from low and normal tones. For the remaining parameters the values for high tone are different from low tone.

The results for the vowel comparison showed no differences between vowels for jitta, jitt and ppq5 and differences for rap, Shim, ShdB, apq3, apq5 and HNR. For rap there are differences between /i/ and /u/, for Shim, ShdB and apq3 vowel /i/ is different from /a/ and /u/, for apq5 the three vowels are different, finally for HNR vowel /u/ is different from vowels /a/ and /i/. The difference of HNR for the vowels can be explained because each vowel has its own harmonic components with different energy.

References

[1] Bielamowicz, S.; Kreiman, J.; Gerratt, B.; Dauer, M.; Berke, G. Comparison of Voice Analysis Systems for Perturbation Measurement. Journal of Speech and Hearing Research, 1996; 39, 126-134.

[2] Zwetsch, I., Fagundes, R., Russomano, T., Scolari, D.. Digital signal processing in the differential diagnosis of beningn larynx diseases, Porto Alegre, 2006.

[3] Wertzner, H.; Schreiber, S.; Amaro, L. Analysis of fundamental frequency, jitter, shimmer and vocal intensity in children with phonological disorders. Rev Bras Otorrinolaringologia 2005; 71, 5, 582-88.

[4] Almeida, N. Sistema Inteligente para Diagnostico da Patologias na Laringe Utilizando Maquinas de Vetor de Suporte. Msc., Universidade Federal Rio Grande do Norte – Natal – Brasil, 2010.

[5] Teixeira, J. P.; Ferreira, D.; Carneiro, S.. Análise acústica vocal - determinação do Jitter e Shimmer para diagnóstico de patalogias da fala. In 6º Congresso Luso-Moçambicano de Engenharia. Maputo, Moçambique, 2011.

[6] Teixeira, J. P; Oliveira, C. and Lopes, C,. Vocal Acoustic Analysis – Jitter, Shimmer and HNR Parameters. Procedia Technology. Elsevier, Vol. 9, 2013; 1112-1122

[7] Meike Brockmann, Michael J. Drinnan, Claudio Storck, and Paul N. Carding. Reliable Jitter and Shimmer Measurements in Voice Clinics: The Relevance of Vowel, Gender, Vocal Intensity, and Fundamental Frequency Effects in a Typical Clinical Task. Journal of Voice, Vol. 25, No. 1, pp. 44-53. 2011.

[8] Lopes, J.. Ambiente da análise robusta dos principais parametros da voz. Msc. University of Porto, 2008.

[9] Barry, W.J., Pützer, M.: Saarbrücken Voice Database, Institute of Phonetics, Univ. of Saarland, http://www.stimmdatenbank.coli.uni-saarland.de/

[10] Boersma, Paul and Weenink, David. Praat: doing phonetics by computer. Phonetic Sciences, University of Amsterdam. http://www.fon.hum.uva.nl/praat/

[11] Brockmann-Bauser, M. Improving jitter and shimmer measurements in normal voices. Phd Thesis of Newcastle University 2011.

[12] M. Farrús, J. Hernando e P. Ejarque, “Jitter and Shimmer Measurements for Speaker Recognition,” Barcelona.

[13] Vasilakis M.; Stylianou, Y. Spectral jitter modeling and estimation. Biomedical Signal Processing and Control 2009; 129

[14] Boersma, P. Accurate short-term analysis of the fundamental frequency and the harmonic-to-noise ratio of a sample sound. IFA Proceedings 1993; 17, 97-110.

[15] Murphy, P. and Akande, O. Cepstrum-Based Estimation of the Harmonics-tonoise Ratio for Synthesized and Human Voice Signals. In Nonlinear Analyses and Algorithms for Speech Processing. Barcelona, LNAI 3817, Springer, 2005.

[16] Oller, L. L. “Analysis of voice signals for the Harmonics-to-noise Crossover Frequency,” KTH - School of Computer Science and Communication. Department of Speech, Music and Hearing. Barcelona, june 2008.

[17] K. Shama, A. Krishna e N. U. Cholayya, “Study of Harmonics-to-Noise Ratio and Critical-Band Energy Spectrum of Speech as Acoustic Indicators of Laryngeal and Voice Pathology,” Manipal, 13 January 2006.