Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 64 (2015) 466 – 473

Procedia Computer Science

Conference on ENTERprise Information Systems / International Conference on Project MANagement / Conference on Health and Social Care Information Systems and Technologies, CENTERIS / ProjMAN / HCist 2015 October 7-9, 2015

Acoustic Analysis of Vocal Dysphonia

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

[a]*Polytechnic Institute of Braganaça – Campus Sta. Apolónia 5301 857 Bragança - Portugal*
[a]*UNIAG – Portugal,* [b] *NECE (UBI) - Portugal*

Abstract

Voice acoustic analysis is becoming more and more useful in diagnosis of voice disorders or laryngological pathologies. The facility to record a voice signal is an advantage over other invasive techniques. This paper presents the statistical analyzes of a set of voice parameters like jitter, shimmer and HNR over a 4 groups of subjects with dysphonia, functional dysphonia, hyperfunctional dysphonia, and psychogenic dysphonia and a control group. No statistical significance differences over pathologic groups were found but clear tendencies can be seen between pathologic and control group. The tendencies indicates this parameters as a good features to be used in an intelligent diagnosis system, moreover the jitter and shimmer parameters measured over different tones and vowels.

Peer-review under responsibility of SciKA - Association for Promotion and Dissemination of Scientific Knowledge

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

1. Introduction

The voice analysis as becoming nowadays a very valuable technique for voice disorders diagnose [1-4], because voice disorder's can give relatively different values in some parameters than in healthy voices.

Patient's voice quality can be diagnosed over an auditory perceptual analysis. However these analyzes may lead to different results depending on expertize of the practitioner involved. The subjective assessment technique leads to

* Corresponding author. Tel.: +351 273303129; fax: +351 273313050.
E-mail address: joaopt@ipb.pt

Peer-review under responsibility of SciKA - Association for Promotion and Dissemination of Scientific Knowledge
doi:10.1016/j.procs.2015.08.544

lack of consensus among professionals. Therefore it became very important 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]. Speech signal processing allows the extraction of a set of voice parameters that may be used to diagnose several pathologies of the vocal cords in individuals by comparison with healthy voice. But the differences between healthy and pathological voice parameters must be known. This study is a contribution to enrich that knowledge.

The parameters obtained by the acoustic analysis have the advantage of describing the voice objectively rather than subjective perceptual analysis. With the existence of normative databases characterizing voice quality or using intelligent tools combining several parameters, it is intended 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 stated by 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 F_0 disturbance jitter and shimmer, has proven to be useful in describing the vocal characteristics.

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 [6]. This study will focus on a 4 groups of dysphonic pathologies.

2. Methods and Methodology

2.1. The Saarbrücken Voice Database

The Saarbrücken Voice Database (SDB) [7] was used in this study. 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. 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 [8].

One part of the SDB for healthy voices consisting in 34 female and 7 male has used as the control group. This control group were already analyzed by the authors [9].

Voices of subjects grouped by dysphonia pathology were also used in this analyzes. Namely for functional dysphonia consisting in 22 female and 7 male subjects; hyperfunctional dysphonia with 23 female and 6 male subjects; psychogenic dysphonia with 21 female and 8 male subjects; dysphonia with 15 female and 11 male subjects.

2.2. Dysphonia

Dysphonia is a descriptive medical term meaning disorder (dys-) of voice (-phonia). There are many causes of dysphonia [2]. Basically, dysphonia is a communication disorder characterized by difficulty in vocal production, presenting an impediment in the natural voice production. Can be caused by a dysfunction, vocal abuse or misuse of the voice, is more common in individuals that use their voice abundantly all the days at an incorrect manner [1].

People with dysphonia may present with hoarseness and a sore or dry throat. A singer may notice that he or she is no longer able to sing in the upper range. There may be other associated symptoms such as a continuous drip at the back of the throat (nasal catarrh) and heartburn.

There are a relation between vocal health, voice disorders (dysphonia) and working conditions. Dysphonia can manifest itself through a series of changes: difficulty maintaining voice; vocal fatigue; variations in the usual frequency; hoarseness; lack of volume and projection; loss of vocal efficiency and low resistance when speaking.

Dysphonia is actually only one pathology in several different disorders and symptoms, sometimes manifesting as secondary symptom, either as principal.

Dysphonia can be organic or functional. Organic dysphonia is due to an anatomical change in the vocal fold, like nodules or benign tumors. Functional dysphonia is assumed when no anatomic changes are known. Between this to cases it can be considered the organic functional dysphonia that is generally initiated with a functional dysphonia not treated and progress to a secondary vocal fold lesions.

In this work 4 types of dysphonia voices were analyzed and compared with healthy voices. The organic dysphonia pathology called dysphonia from now on, the generic functional dysphonia, and two variations of the functional dysphonia: the hyperfunctional dysphonia and the psychogenic dysphonia.

The hyperfunctional dysphonia is an excessive involuntary muscle contraction, as a consequence of improper phonation. Results in a hoarse or strained voice.

The psychogenic dysphonia as the result of strong emotions influence that causes changes in voice. Strong emotions like anger, joy, anxiety or fear cause psychogenic dysphonia.

2.3. Parameters

Jitter, Shimmer and HNR parameters were extracted with Praat software [8] and were analyzed in this work.

Jitter is defined as the periodic variation from cycle to cycle, and shimmer relates to the amplitude variation of the wave sound, as Zwetsch et al. [2] and [5] and [10-11]. Fig. 1 shows a perspective of jitter and shimmer.

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.

The jitter and shimmer can be measured usually with 4 different forms as can be found in [5] and [9-14]. Jitter can be measured as absolute, relative, relative average perturbation (rap) and the period perturbation quotient (ppq5). Shimmer can be measured as absolute value in dB, as relative value, as Amplitude Perturbation Quotient in 3 cycles (APQ3) and as Amplitude Perturbation Quotient in 5 cycles (APQ5).

Jitter absolute is the cycle-to-cycle variation of fundamental frequency, i.e. the average absolute difference between consecutive periods, expressed by Eq.1.

Relative Jitter or local Jitter is the average absolute difference between consecutive periods, divided by the average period. It is expressed as a percentage (Eq. 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 (Eq. 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 (Eq. 4).

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

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

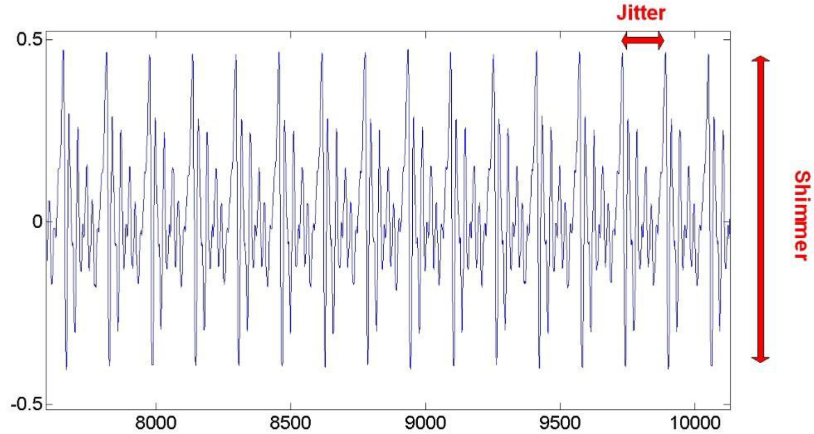

Fig.1. Jitter and Shimmer perturbation measures in speech signal.

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

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

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

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 (Eq. 5).

Shimmer relative is defined as the average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude, expressed as a percentage (Eq. 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 (Eq. 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 (Eq. 8).

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

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

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

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

Where Ai is the peak-to-peak amplitude and N the number of periods.

The **Harmonic to Noise Ratio** (HNR) 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 [15-18].

HRN is given by following equation according to Boersma, P [15]:

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

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.

3. Analysis of Results

Three levels of analyze for the parameters were performed. A comparison between the voices for the three tones (H-high, L-low and N-normal), than a comparison for the three vowels (/a/, /i/ and /u/) and finally a general comparison between the four groups of subjects (Functional Dysphonia - FD, Hyperfunctional Dysphonia - HD, Psychogenic Dysphonia - PD and Dysphonia (organic) - D) with control group - C.

No comparison between genders was made because the number of male and female subjects were not equilibrated. Furthermore, as state in [9], for healthy voices, only the absolute Jitter presented differences between gender, because the male glottal period are generally longer than female glottal period been natural to have also higher absolute variation. This difference vanish with the relative jitter (jitt(%)), RAP and PPQ5.

3.1. Analyzes over tones for vowel /a/

A comparison along the three tones is presented in Fig.2 for relative jitter - jitt(%), relative shimmer - shim(%) and HNR for vowel /a/. Similar results were obtained for /i/ and /u/ vowels but, as stated in [9], for healthy voices, there are no significant statistical differences between vowel /a/ and vowel /i/ needier between vowel /a/ and vowel /u/, although there are differences between vowel /i/ and /u/. The same analyzes were performed for the four jitter parameters but again no significant differences were found between jitter parameters beside the lower values for absolute jitter at L tones due, again, to higher glottal period for this tone. For shimmer parameters no significant differences were found between absolute shimmer, relative shimmer, APQ3 and APQ5.

Considering the results presented in the error bar plots of Fig. 2 for jitter parameter, there are a tendency of higher jitt for L tone followed by N tone and lower jitt for H tone for the 5 groups under consideration. Comparing the groups there is a statistically significant difference between control group and the other pathologic groups for the three tones but no significant difference between the pathologic groups. It is noticeable a higher variance of jitt in the dysphonia group and also higher values.

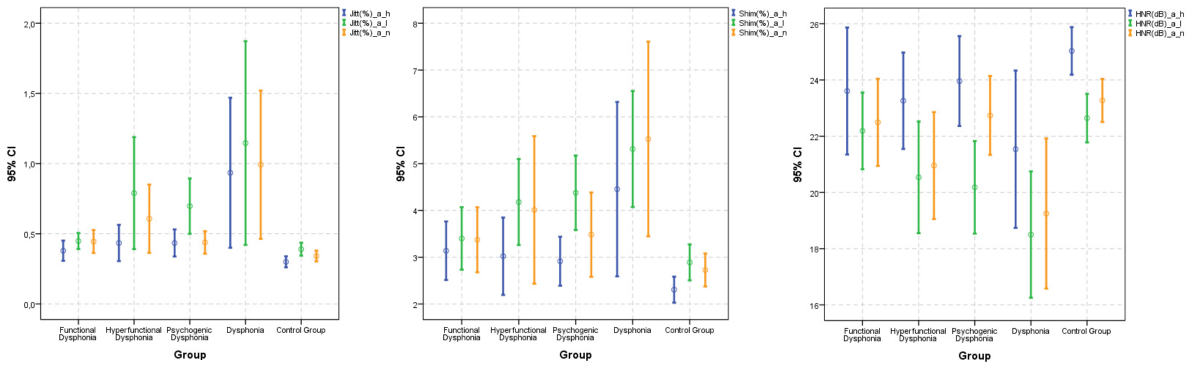

Fig. 2. Error bar for the three tones of the parameters Jitt (left), Shim (center) and HNR (right) for the 5 groups.

The analyzes of the boxplot for shim parameter of Fig. 2 shows a tendency for lower shim for H tone, less variance and lower shim at the control group, and for L tone there is a significant difference between C group and HD, PD and D groups. There is a significant difference between control group and FD group only at H tone and a difference between control and D group for N tone.

The error bar plot for HNR of Fig. 2 shows higher HNR for H tone followed by N tone and less HNR for L tone, a tendency for higher HNR and lower variance at control group. Again D groups presents higher variance. There is a significant difference between control group and PD and D groups at L tone and between control group and D group at N tone. No significant differences exist between pathologic groups.

3.2. Analyzes over vowels for Normal tone

A comparison of the same three parameters along the vowels /a/, /i/ and /u/ for normal tone is presented in Fig. 3. Again, similar analyzes were done for the other tones with similar results confirming the presented results of [9] for healthy voices. The same study [9], showed no significant differences between normal tone and low or high tones for healthy voices.

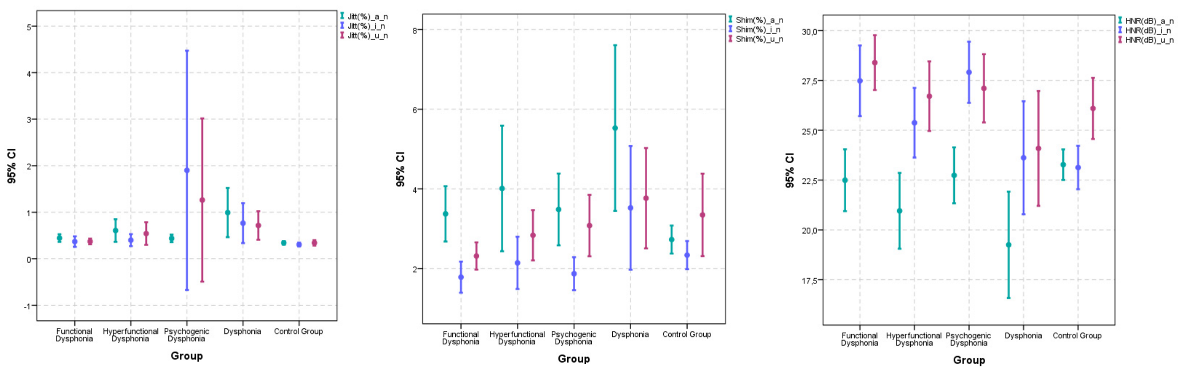

Fig. 3. Error bar for the three vowels of the parameters Jitt (left), Shim (center) and HNR (right) for the 5 groups.

Considering the error bar plot for jitt parameter in Fig. 3: there is no significant difference between vowels; the control group has lower variance and tendency to lower values of jitt. The difference between control group and the pathologic groups are confirmed for the 3 vowels, except for PD group with /i/ and /u/ vowels.

For shim parameter in Fig 3, shows a tendency of lower shim for vowel /i/ and higher for vowel /a/, except in control group. Only the vowel /a/ has significant difference between control group and D group.

For HNR shows a tendency for lower HNR for the /a/ vowel and higher for /u/ vowel.

3.3. General analyzes of pathologic and control groups

Table 1 present the mean and standard deviation values for the jitt, shim and HNR parameters for vowels /a/, /i/ and /u/ considering the three tones in the upper half part, and the means and standard deviation for the H, L and N tones along the three vowels, at the lower part for the 5 groups.

The control group has a general lower jitt and shim for the three vowels and higher HNR only for vowel /a/.

Considering the average for each tone, again the control group present a general lower jitter and shimmer. HNR has no clear tendency between control and pathologic groups.

Table 1: Summary of descriptive statistics for each parameter, by vowel and tone.

		Functional Dysphonia (n=29) [Female=22; Male=7]		Hyperfunctional Dysphonia (n=29) [Female=23; Male=6]		Psychogenic Dysphonia (n=29) [Female=21; Male=8]		Dysphonia (n=26) [Female=15; Male=11]		Control Group (n=41) [Female=34; Male=7]	
		Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Jitt (%)	/a/	.42	.141	.61	.640	.52	.224	1.02	1.249	.34	.106
	/i/	.36	.195	.46	.336	.88	2.260	.75	.781	.32	.124
	/u/	.38	.201	.50	.340	.71	1.539	.82	.802	.35	.167
Shim (%) s	/a/	3.30	1.295	3.74	2.557	3.59	1.405	5.10	3.720	2.64	.829
	/i/	1.97	1.568	2.20	1.647	1.79	.817	3.30	2.837	2.18	.926
	/u/	2.33	1.077	2.91	1.628	3.21	1.841	3.69	2.553	3.18	3.052
HNR (dB)	/a/	22.76	3.767	21.59	4.329	22.29	2.985	19.76	5.705	23.65	2.152
	/i/	27.77	3.667	25.38	4.252	27.99	3.193	24.17	6.881	23.92	3.440
	/u/	29.02	3.005	27.13	4.519	27.04	4.326	24.95	6.313	26.79	4.471
Jitt (%)	Tone H	.32	.171	.40	.242	.36	.185	.85	1.290	.30	.129
	Tone L	.44	.222	.65	.688	.56	.295	.92	1.072	.38	.125
	Tone N	.40	.170	.52	.377	1.20	2.652	.82	.852	.33	.107
Shim (%)	Tone H	2.31	1.265	2.51	1.885	2.26	1.048	3.48	3.878	2.20	1.318
	Tone L	2.80	1.446	3.33	1.959	3.53	1.760	4.33	2.271	3.00	1.437
	Tone N	2.49	1.072	3.00	2.008	2.81	1.385	4.27	3.427	2.80	1.427
HNR (dB)	Tone H	27.85	3.389	26.38	4.457	27.59	3.862	25.27	7.825	26.55	3.391
	Tone L	25.58	2.900	23.37	4.560	23.81	3.954	21.29	5.513	23.65	2.979
	Tone N	26.12	3.274	24.35	3.912	25.92	3.139	22.32	5.717	24.17	2.565

4. Conclusion

The paper presents a statistical analyzes of the jitter, shimmer and HNR parameters measured for three vowels and three tones over four groups of dysphonia pathologic voices and a control group.

Only the results of the relative jitter and shimmer parameters were presented although the conclusions are the same for the remaining jitter and shimmer parameters.

Considering the comparison between tones there are tendency of high jitt for L tones and lower jitt for H tones. For the three tones the jitt is significantly lower in control group than for the pathologic groups. The H tone are tendency for lower shim. Control group also present lower variance and lower values for shim than pathologic groups. For L tone there is significant different shim between control group and HD, PD and D groups. The difference of shim between control group and FD group only exist for the H tone. The HNR are higher for H tone, followed by N tone and lower for L tone. The control group has tendency to have higher HNR. There is significant difference between control group and PD and D groups at L tone and between control group and D group at N tone.

The comparison between vowels confirmed lower values and variance of jitt in control than in pathologic groups for the three vowels. The vowel /i/ presented a tendency to lower values of shim than vowel /a/. Finally, vowel /a/ has tendency to lower HNR and vowel /u/ to higher HNR.

A more general and less detailed conclusion resulted from analyzes of Tab 1, showing lower jitt and shim for all vowels and tones for the control group and higher HNR only within vowel /a/.

As a final conclusion, HNR parameter didn't show strong discriminant capacity, meanwhile jitter and shimmer seems to be relevant parameters to be used in an intelligent diagnosis system of dysphonia pathologies. Moreover if this parameters can be measured first at three tones and second within the three vowels.

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

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

[8] Boersma, P., Weenink, D.. Praat: doing phonetics by computer. Phonetic Sciences, University of Amsterdam. http://www.fon.hum.uva.nl/praat/

[9] Teixeira, J., Fernandes, P. Jitter, Shimmer and HNR classification within gender, tones and vowels in healthy voices. *Procedia Technology*, 2014, 16:1228-1237.

[10] Teixeira, J., Oliveira, C., Lopes, C,. Vocal Acoustic Analysis – Jitter, Shimmer and HNR Parameters. *Procedia Technology*., 2013, 9:1112-1122

[11] Brockmann, M., Drinnan, M., Storck, C., Carding, P.. 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*, 2011, 25(1):44-53.

[12] Farrús, M., Hernando, J., Ejarque, P. 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] Brockmann-Bauser, M. Improving jitter and shimmer measurements in normal voices. Phd Thesis of Newcastle University 2011.

[15] 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.

[16] Oller, 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] Shama, K., Krishna, A., Cholayya, N. 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.

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