Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 138 (2018) 280–285

Procedia Computer Science

www.elsevier.com/locate/procedia

CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies, CENTERIS/ProjMAN/HCist 2018

Harmonic to Noise Ratio Measurement - Selection of Window and Length

Joana Fernandes[a], Felipe Teixeira[a], Vitor Guedes[a,b], Arnaldo Junior[b], João Paulo Teixeira[a,c,*]

[a]Instituto Politécnico de Bragança, Bragança 5300, Portugal
[b]Universidade Tecnológica Federal do Paraná, Câmpus Medianeira, Brasil
[c]Research Centre in Digitalization and Intelligent Robotics (CEDRI), Applied Management Research Unit (UNIAG), - (IPB), Bragança 5300, Portugal

Abstract

Harmonic to Noise Ratio (HNR) measures the ratio between periodic and non-periodic components of a speech sound. It has become more and more important in the vocal acoustic analysis to diagnose pathologic voices. The measure of this parameter can be done with Praat software that is commonly accept by the scientific community has an accurate measure. Anyhow, this measure is dependent with the type of window used and its length. In this paper an analysis of the influence of the window and its length was made. The Hanning, Hamming and Blackman windows and the lengths between 6 and 24 glottal periods were experimented. Speech files of control subjects and pathologic subjects were used. The results showed that the Hanning window with the length of 12 glottal periods gives measures of HNR more close to the Praat measures.

Selection and peer-review under responsibility of the scientific committee of the CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies.

Keywords: HNR measure; window length; Hanning window; Hamming, Blackman.

* Corresponding author. Tel.: +351 273 30 3129; fax: +351 273 30 3051.
E-mail address: joaopt@ipb.pt

1877-0509 © 2018 The Authors. Published by Elsevier Ltd.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/)
Selection and peer-review under responsibility of the scientific committee of the CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies.
10.1016/j.procs.2018.10.040

1. Introduction

Voice pathologies or laryngeal pathologies are those related to some larynx disorder, and can be caused by minimal, structural and / or functional lesions of the larynx [1].

There are several examinations like laryngoscopy or stroboscopic exams that can be done to detect voice-related pathologies [2]. However, they are invasive procedures that causes discomfort to the patient.

Through vocal acoustic analysis it is possible to determine the vocal quality of the individual in a non-invasive way through the acoustic parameters of the signal - periodicity, amplitude and spectral composition [3] [4].

The HNR is a parameter where the ratio of harmonic and noise components provides an indication of the general frequency of the speech signal by quantifying the relationship between the periodic component (harmonic part) and the aperiodic component (noise), expressed in dB [1] [5] [6]. The general value of a signal varies because different vocal tract configurations imply different amplitudes for harmonics. This is what happens in the vowels / a /, / i / and / u /, where the vowel / u / is different from the vowels / a / and / i / [7]. An example of this is a healthy speaker which produces a sustained / a / or / i / sustained with a hamonicity of about 20 dB and an / u / of about 40 dB, the difference being from high frequencies in / a / and / i /, the low frequencies in / u /, resulting in a higher sensitivity of HNR [8].

In this work we intend to make a comparative study between the HNR values obtained in the algorithm developed by Gonçalves and Teixeira [9] [10] and the values obtained in Praat software [8], with the aim of obtaining values close to Praat and thus improve the algorithm with the objective of developing an automatic system for the diagnosis of laryngeal pathologies in the long term. This algorithm is intended to be integrated into a tool for the automatic diagnosis of pathologies. It is used as a comparison of values, Praat software, which is commonly accepted by the scientific community as a precise measure and can be obtained free of charge. It is a software used for speech analysis that was developed by Paul Boersma and David Weenink [8], of the Institute of Phonetic Sciences, University of Amsterdam. This algorithm allows the extraction of 9 parameters - absolute jitter; relative jitter, jitter rap, jitter ppq5, absolute shimmer, relative shimmer, shimmer apq3, shimmer apq5 and HNR, however, only HNR will be used in this study, since other studies have already been done [9].

This article is organized as follows: Section 2 describes the calculation of HNR, the pathologies used, the database used and the number of subjects that were used for the study; section 3 describes the studies done and the results obtained and, finally, in section 4 the conclusions obtained.

2. Methodology

2.1. Harmonic to Noise Ratio

By definition the HNR is a parameter in which the relationship between harmonic and noise components provides an indication of the voiced components of the speech signal by quantifying the relationship between the periodic and the aperiodic component, expressed in dB [1] [5] [6]. The overall HNR value of a signal varies because different vocal tract configurations imply different amplitudes for harmonics. This unit of measurement relates the energy conveyed by the voiced signal through the glottal impulses, and the energy of the glottic noise fraction after being filtered through the vocal tract. This noise arises with the turbulence generated when the airflow passes through the glottis during phonation, occurring when, for example, the vocal cords close inappropriately [11] [12].

There are different approaches to automatically determine the HNR, for instance [13] [14] used the Cepstrum to measure the harmonic and noise componentes, meanwhile [6] and [11] used the autoccorrelation. In mathematical terms a voiced signal with harmonic structure in the frequency domain can be express by equation 1.

$$X(w) = H(w) + N(w) \tag{1}$$

Where X(w) corresponds to the speech signal in the frequency domain, H(w) to the harmonic component and N(w) to the noise component.

The HNR is a logarithmic measure of the energy ratio that is associated with the harmonic and noise component. Through equation 2 it is possible to integrate the spectral power over the audible range of frequencies [11].

$$HNR = 10 \times \log_{10} \frac{\int_w |H(w)|^2}{\int_w |N(w)|^2} \quad (2)$$

In this algorithm the HNR was implemented taking into account the studies published by Boersma [6]. In this study Boersma uses a procedure based on the properties of the autocorrelation function to obtain the separation of components described previously. Autocorrelation consists of the correlation of a signal with itself. If we consider a voice signal x(t), the autocorrelation function $r_x(\tau)$ is represented in equation 3.

$$r_x(\tau) \equiv \int x(t)x(t+\tau)dt\,. \quad (3)$$

In this function there is a global maximum for τ=0. If there are peak values outside 0, the signal is periodic and there is a phase shift T_0, called period, so that all these maxima are placed in the offset nT_0, for each integer n, with $r_x(nT0)=r_x(0)$. The fundamental frequency F0 of this periodic signal is defined by F0=1/T0. If there are no global maxima other than 0, there may be maximum peaks. If the largest of them is in the displacement τmax, and if its height $r_x(\tau_{max})$ is sufficient, the signal is designated as having a periodic part, and its harmonic force R0 is a number between 0 and 1, equal to the local maximum $r'_x(\tau_{max})$ of normalized autocorrelation (equation 4).

$$r'_x(\tau) \equiv \frac{r_x(\tau)}{r_x(0)} \quad (4)$$

The total autocorrelation of the signal is the sum of the autocorrelation of its harmonic and noise components as can be seen in equation 5.

$$r_x(0) = r_H(0) + r_N(0) \quad (5)$$

If the noise is white (no correlation possible), the local maximum is τ_{max}=T0 with height $r_x(\tau_{max})=r_H(T0)=r_H(0)$ [6].

With this, the autocorrelation function of a sustained speech signal displays local maxima for multiple values of τ, all multiples of the fundamental period. Thus, to determine the HNR, it is only necessary to calculate the autocorrelation function of the speech signal and to identify the first local maximum that will correspond to the harmonic component. The value of the noise component is determined by equation 6 and the HNR by equation 6 [11].

$$HNR(dB) = 10 \times \log_{10} \frac{r'_x(\tau_{\max})}{1 - r'_x(\tau_{\max})} \quad (6)$$

2.2. Pathologies of the Larynx

Lesions in the vocal chords alter the process of phonation, since the patterns of vibration during the opening and closing phases of the vocal chords are irregular [15]. For this study, 3 laryngeal diseases were used, chronic laryngitis, dysphonia and vocal cord paralysis.

Chronic Laryngitis - corresponds to a persistent inflammation of the laryngeal mucosa, sometimes with many years of evolution, usually provoked by repeated acute infections [15].

Dysphonia - is a communication disorder and makes vocal production difficult, with an impediment to voice production [16].

Vocal Cord Paralysis - occurs when the laryngeal muscles cannot perform their function, paralysis of one of the vocal cords or of the two [15];

2.3. Database

The German Saarbrucken Voice Database (SVD) was used. This database is available online by the Institute of Phonetics at the University of Saarland [17].

The database consists of voice signals of more than 2000 subjects with disease and controls. Each person has the recording of phonemes /a/, /i/ and /u/ in the low, normal, high tones, varying between tones, and the German phrase "Guten Morgen, wie geht es Ihnen?" ("Good morning, how are you?"). The size of the sound files is between 1 and 3 seconds and have a sampling frequency of 50 kHz.

In this analysis, a random sample of 10 control subjects (5 male and 5 female) and 10 subjects (5 male and 5 female), of the 3 diseases were used (4 subjects with chronic laryngitis, 2 male and 2 female; 4 with dysphonia, in which two were males and the other two females, and 2 with paralysis of the vocal cords, one of each gender).

The control subjects had between 40 and 65 years of age, with a mean of 55 years and a standard deviation of 8 years, and the pathologic subjects had between 16 and 77 years of age, with a mean of 51 years and standard deviation of 19 years.

3. Analysis of Experimental Results

In this work three vowels, /a/, /i/ and /u/, each with three different tones, high, low and normal were used to measure the HNR with three windows with different lengths. The measures were compared with the Praat measure of HNR. The windows of Hanning, Hamming and Blackman were used and for each window the length of the window was changed between 6, 11, 12, 13 and 24 glottal periods. The choice of this window length is based on the values used by Boersma [6] (12 glottal periods) and the experiments already taken by Gonçalves [11] that had experimented 5, 10, 20 and 50 glottal periods and selecting the 10 glottal periods length. However the results with 10 glottal periods were slightly different than the ones of Praat. The values of 6, 24, 12, 11 and 13 are half, double, equal, minus 1 and plus one glottal periods as the 12 used in [6].

The window overlap was also experimented, however, the values were similar to those that did not have a window overlap and therefore it was decided to use the algorithm without overlapping the windows.

3.1. Results for Real Speech

Table 1 presents the data referring to the average measures of the vowel /a/ for the high tone for the used speech files (10 controls and 10 pathologic). The results are presented and analyzed for the vowel /a/ high tone because according to [16] the vowel /a/ has generally lower HNR than the /i/ and /u/ vowels, and the high tone has higher HNR than the other tones. For control group this vowel/tone has 50% of its values between 24 and 26 dB measured by Praat. For pathologic groups the variation is higher. Anyhow, the same process was applied to the other vowels and tones and differences are presented below.

With the result of the means, a subtraction was made, in module, with the Praat value. With the subtraction values a new mean was made for each window and length, where, for example, for the hamming window of 6 glottal periods, all values with this window and this length were averaged. Table 2 shows the results of this analysis.

From the data in table 2 it can be observed that for the patients the best window is Hanning with 11 glottal periods and for the control we obtain two windows with the same result, the Hamming window with 11 glottal periods and the Hanning window with 12 glottal periods. Since it is not possible to use two different lengths and two different windows, it was concluded that it would be best to use the Hanning window with 12 glottal periods, since this value for patients, even though it is not the closest, is lower than the control and considering the difference in both groups 11 and 12 glottal periods have the lower sum of differences.

Comparing the means difference to Praat of all vowels and all tones with this window the one which differs most in the control group is the vowel /u/ low tone with 3,40 dB, and for patients group it is the vowel /a/ normal tone with 3,12 dB. The value that approximates Praat's mean value, also, for this window, in control, is the vowel /i/ high tone at 0,10 dB, and for patients, it is the vowel /i/ high tone in 0,09 dB.

If we compare this study with the study done by Gonçalves [11] it can be seen that this makes the algorithm more close to the Praat measures, since in his study the algorithm does not present a range of measures as extensive as Praat do. The difference between the studies is the windows applied, the window length and the number of subjects used, but in both cases the best result was the Hanning window.

Table 1 - Average HNR (in dB) with window variation for vowel /a/ high tone

Vowel	Window	Glottal Periods	Average HNR			
			Patients		Control	
			Algorithm	Praat	Algorithm	Praat
/a/ high	HAMMING	6 GP	15,18	20,95	15,56	25,26
		11 GP	23,47		25,74	
		12 GP	24,27		27,05	
		13 GP	25,12		28,22	
		24 GP	28,64		34,82	
	HANNING	6 GP	13,78		13,93	
		11 GP	22,47		24,39	
		12 GP	23,34		25,75	
		13 GP	24,29		26,98	
		24 GP	28,39		34,23	
	BLACKMAN	6 GP	11,20		10,99	
		11 GP	20,58		21,88	
		12 GP	21,55		23,31	
		13 GP	22,63		24,63	
		24 GP	27,80		32,78	

Table 2 – Mean HNR differences from each window and glottal periods to Praat measures

Window	Glottal Periods	Average HNR difference	
		Patients	Control
Hamming	6	7,77	10,45
	11	1,61	**1,70**
	12	2,35	1,99
	13	3,39	2,26
	24	8,86	8,36
Hanning	6	9,29	12,07
	11	**1,38**	1,90
	12	1,58	**1,70**
	13	2,41	2,01
	24	8,39	7,66
Blackman	6	12,02	14,96
	11	2,11	4,22
	12	1,51	2,85
	13	1,46	1,79
	24	7,24	6,36

4. Conclusion

A comparative analysis of the window and its length was performed between the HNR values measures by the algorithm and the Praat software. The Hanning, Hamming and Blackman windows and length of 6, 11, 12, 13 and 24 glottal periods were experimented to see which window was best and how long it was. Signals from control and pathologic subjects of the SVD were used. Since the objective of the algorithm is to be used in the vocal pathologies diagnose, male and female subjects with dysphonia, chronic laryngitis and vocal cords paralysis were used.

It was concluded that the best window was the Hanning window with a length of 12 glottal periods, but 11 glottal period length has very similar results.

Acknowledgements

This work is supported by the Fundação para a Ciência e Tecnologia (FCT) under the project number UID/GES/4752/2016.

References

[1] H. Cordeiro, (2016). Reconhecimento de Patologias da Voz usando Técnicas de Processamento da Fala, Universidade Nova de Lisboa: PhD Thesis.

[2] M. Brockmann-Bauser, (2011). Improving jitter and shimmer measurements in normal voices, PhD Thesis. Newcastle: Institute of Cellular Medicine, Medical School, Newcastle University.

[3] J. P. Teixeira, F. Teixeira, J. Fernandes e P. O. Fernandes, (2018) "Acoustic Analysis of Chronic Laryngitis - satistical Analysis of Sustained Speech Parameters,".

[4] A. M. G. A. M. &. A. Z. Al-nasheri, (2016) "Investigation of Voice Pathology Detection and Classification on Different Frequency Regions Using Correlation Functions," *Journal of Voice,* vol. 31(1), pp. 3-15.

[5] N. Alves, (2016). Diagnóstico Inteligente de Patologias da Laringe, Instituto Politécnico de Bragança: Master Thesis.

[6] P. Boersma, (1993) "Accurate short-term analysis of the fundamental frequency and the harmonic-to-noise ratio of a sample sound," *IFA Proceeding,* vol. 17, pp. 97-110.

[7] J. P. Teixeira and P. O. Fernandes, (2014) "Jitter, Shimmer and HNR classification within gender, tones and," in *Procedia Technology, 16, 1228-1237*.

[8] P. Boersma e D. Weenink, "Praat: doing phonetics by computer," Phonetic Sciences, University of Amsterdam, [Online]. Available: http://www.fon.hum.uva.nl/praat/. [Accessed at 24 04 2018].

[9] J. P. Teixeira e A. Gonçalves, (2016) "Algorithm for Jitter and Shimmer Measurement in Pathologic Voices," *Procedia Computer Science, 100, 271-279*.

[10] J. P. Teixeira e A. Gonçalves, (2014) "Accuracy os Jitter and Shimmer Measurements," *Procedia Technology, 16, 1190-1199*.

[11] A. A. Gonçalves (2015), Patologias da Laringe com Análise Acústica Vocal, Instituto Politécnico de Bragança: Master Thesis.

[12] M. Hirano, (1974) "Morphological structure of the vocal cord as a vibrator and its variations," *Folia Phoniatrica,* vol. 26(2), pp. 89-94.

[13] P. Murphy, (2000) "A CEPSTRUM-BASED HARMONICS-TO-NOISE RATIO," in *6th International Conference on Spoken and Language Processing*, Beijing - China.

[14] P. Murphy and O. Akande, (2005) "Cepstrum-Based Estimation of the Harmonics-to-Noise Ratio for Synthesized and Human Voice Signals," em *Nonlinear Analyses and Algorithms for Speech Processing*, LNAI 3817, Springer.

[15] J. P. Teixeira, D. B. Ferreira e S. M. Carneiro, (2011) "Análise Acústica Vocal - Determinação do Jitter e Shimmer para Diagnóstico de Patologias da Fala," em *VI Congresso Luso Moçambicano de Engenharia*.

[16] J. P. Teixeira e P. Fernandes, (2015) "Acoustic Analysis of Vocal Dysphonia," *Procedia Computer Science, 64, ,* pp. 466-473.

[17] W. Barry e M. Pützer, "Saarbruecken Voice Database," Institute of Phonetics at the University of Saarland, 2007. [Online]. Available: http://www.stimmdatenbank.coli.uni-saarland.de. [Accessed at 24 April 2018].