Cured Database of Sustained Speech Parameters for Chronic Laryngitis Pathology

Joana Fernandes, Instituto Politécnico de Bragança (IPB), Bragança, Portugal, a29423@alunos.ipb.pt

Felipe Teixeira, Instituto Politécnico de Bragança (IPB), Bragança, Portugal, a26171@alunos.ipb.pt

Paula Odete Fernandes, Applied Management Research Unit (UNIAG), Instituto Politécnico de Bragança (IPB), Bragança, Portugal, pof@ipb.pt

João Paulo Teixeira, Research Centre in Digitalization and Intelligent Robotics (CEDRI), Instituto Politécnico de Bragança, Applied Management Research Unit (UNIAG), Bragança, Portugal, joaopt@ipb.pt

Abstract

This paper reports the construction and organization of a database of speech parameters extracted from a speech sound database. The database is freely available on internet and the paper intends also theirs advertise for the research community. The database includes the parameters extracted from the sound of sustained vowels produced by a group of Chronic Laryngitis patients and a group of control subjects with similar characteristics concerning gender and age. The set of parameters of this database consists in the Jitter, Shimmer, Harmonic to Noise Ratio (HNR), Noise to Harmonic Ratio (NHR) and Autocorrelation extracted from the sound of sustained vowels /a/, /i/ and /u/ at low, neutral and high tones.

Keywords: Cured database, Chronic laryngitis, Speech parameters.

Introduction

Vocal acoustic analysis is receiving more and more attention from scientific researchers as an advantageous alternative to diagnose pathologies. Vocal pathologies were the first receiving contributions of the acoustic analysis for diagnose pathologies, since vocal apparatus is used to the production of speech and damages can interfere in the speech quality.

Vocal acoustic analysis techniques are often used for voice disorders assessment (Brockmann-Bauser, 2011; Bielamowicz et al., 1996; Salhi et al., 2010), but also other types of diseases related with the central nervous system or neurodegenerative diseases such Alzheimer (Meilán et al., 2014) and Parkinson (Chaitanya and Shruti, 2017) diseases are receiving attention from research community as alternative or complementary information to make early diagnose.

This technique rely on the non-invasive character of diagnose when compared with, for example, laryngoscopy exams.

An additional advantage of using the speech for early diagnose is the possibility of have the diagnose without subjects look for it. Although the ethical questions that should be considered this can also be carefully taken as an opportunity for non-suspicious cases.

This paper describes the set of parameters extracted from the speech of patients with chronic laryngitis and a control group. It also describes the organization of the processed database of speech features. It is also the purpose of this publication to disseminate the availability of the database for the scientific community.

The database starts with the features for the chronic laryngitis. It is intended to extend the database with other features such as MFCC and formant frequencies (Cordeiro, 2016) and for other pathologies.

Laryngitis disease occurs when the vocal fold gets irritated or inflamed. This is a very common condition that often causes hoarseness and/or loss of voice. The laryngitis can be acute or chronic. Both cases have similar symptoms. The acute laryngitis comes suddenly and disappears in days or one week. The chronic laryngitis takes longer periods of time like several weeks or months. The chronic laryngitis can be caused by smoke for long periods of time, gastroesophageal reflux, infections (bacterial, viral, fungal), bronchitis, autoimmune, irritative, traumatic or allergic factors, pneumonia, excessive exposure to toxic chemicals, complications of influenza or chronic cold (Tusaliu et al, 2016).

Sustained Speech Parameters

The parameters on the database were extracted from sustained speech sounds of the vowels /a/, /i/ and /u/ at low, normal and high tones of the the Saarbrücken Voice Database (SVD) (Barry and Pützer). Each segment of speech consists in a steady state sustainable pronunciation of the respective vowel and tone.

The speech signal of the SVD database was sampled at 50 kHz ratio and 16 bit of resolution.

Parameters related with the variations of the periodicity such as jitter and shimmer were registered, namely the absolute and relative measures of jitter and also the absolute and relative measures of shimmer. The parameters that measures the harmonic and unharmonic components were also extracted and registered, namely the harmonic to noise ratio, noise to harmonic ratio and autocorrelation.

The parameters were extracted using the Praat software (Boersma P, Weenink D).

Jitter and Shimmer Parameters

Jitter is the measure of the variation of the glottal period between successive cycles of vocal fold vibration. Subjects who cannot control vocal chords vibration tend to have higher jitter values. The jitter can be measured in four different ways (Teixeira and Gonçalves, 2014). However, only two of these forms were registered, relative jitter and absolute jitter. The other two measures are relative average perturbation (rap) and the period perturbation quotient (ppq5) that measures the same variability within a window of 3 and 5 glottal periods. Previous statistical analysis (Teixeira and Fernandes, 2015) showed that relative jitter has similar results as rap and ppq5, therefore only the jitta of absolute jitter, and the jitter of relative jitter were registered.

Absolute jitter (jitta) is the glottal period variation between cycles, that is, the mean absolute difference between consecutive periods, expressed by Eq. 1. The values in the database are in micro seconds (µs). It should be noted that the sample period of the speech signal is 20 µs, therefore a jitter lower than 20 µs is less than the sample period.

The relative jitter (jitter) is the average absolute difference between consecutive glottal periods relative to the average period. It is expressed as a percentage (Eq. 2).

Shimmer parameter is related to the magnitude variation along the glottal periods. A reduction in glottal resistance and lesions may cause variations in glottal magnitude period. This phenomenon is correlated with breathiness and noise emission, giving rise to higher shimmer values. The shimmer can be measured in four different ways (Teixeira and Gonçalves, 2014), however, only two of them are registered in the database, relative shimmer (Shim) and absolute shimmer (ShdB). The other two measures are Amplitude Perturbation Quotient along 3 cycles (APQ3) and Amplitude Perturbation Quotient in 5 cycles (APQ5) that measures the same variability within a window of 3 and 5 glottal periods, respectively. Previous statistical analysis (Teixeira and Fernandes, 2015) showed that relative shimmer has similar results as APQ3 and APQ5.

The absolute shimmer (ShdB) is expressed as the peak-to-peak magnitude variation in decibel, that is, the base 10 logarithm of the absolute mean of the magnitude ratio between consecutive periods multiplied by 20. It is expressed in decibel (Eq. 3).

The relative shimmer (Shim) is defined as the mean absolute difference between magnitudes of consecutive periods, divided by the mean magnitude, expressed as a percentage (Eq. 4).

In equations 1-4 Ti is the time of the glottal period i. Ai is the magnitude of the glottal period i. N is the total number of glottal periods.

$$jitta = \frac{1}{N-1}\sum_{i=1}^{N-1} |T_i - T_{i-1}| \qquad (1)$$

$$Jitter = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1} |T_i - T_{i-1}|}{\frac{1}{N}\sum_{i=1}^{N} T_i} \times 100 \qquad (2)$$

$$ShdB = \frac{1}{N-1}\sum_{i=1}^{N-1} \left|20 * log\left(\frac{A_{i+1}}{A_i}\right)\right| \qquad (3)$$

$$Shim = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1} |A_{i+1} - A_i|}{\frac{1}{N}\sum_{i-1}^{N} A_i} \times 100 \qquad (4)$$

Harmonic Parameters

The harmonic characteristics of the voice can be measured into three parameters, HNR (Harmonic to Noise Ratio), NHR (Noise to Harmonic Ratio) and Autocorrelation. The HNR measures the relationship between harmonic and noise components. It provides an indication of overall periodicity of the speech signal by quantifying the relation between the periodic component (harmonic part) and aperiodic components (noise). The overall HNR value of a signal varies because different vocal tract configurations imply different amplitudes for harmonics. Therefore different vowel can have different HNR. The u vowel has higher frequency components, therefore it is expected to have higher HNR. Since the HNR is sensitive to the vowel, HNR for different vowels shouldn't be compared.

Some authors propose their own way to measure the HNR (Boersma, 1993; Shama et al, 2007). The measure presented by Boersma, 1993 is the one used by the Praat software that was used to extract the values of the parameters. This method consists in measure the energy of the first peak of the normalized autocorrelation and consider that this is the energy of the harmonic component of the signal, and consider the remaining energy as the noise energy given by the difference between 1 and the harmonic energy, as represented in equation 5. In this equation H is the harmonic component given by the energy of the first peak of the normalized autocorrelation of the signal. The final value of HNR is the average HNR along the successive segments of 2 glottal periods of the entire signal.

The NHR tends to be the inverse of the HNR, anyhow once the measure is made at the logarithmic domain (dB), their values tend to move in opposite directions but the values are not exactly the inverse nor the symmetric, because the final measure is an average along the signal.

The Autocorrelation function gives a measure of the similar parts of speech repeated along the signal. As higher the autocorrelation value higher is the repetitions of similar events along the signal. Autocorrelation is given by equation 6.

$$HNR = 10 \times log_{10} \frac{H}{1-H} \qquad (5)$$

$$AutoCorr(n) = \sum_{m=-N}^{N} x(n+m)x(m) \qquad (6)$$

Organization of Cured Database

The total number of the pathologic group is 70 subjects divided by 40 male subjects and 30 female subjects. The control group was selected to have a similar distribution between male and female and a similar average age. Table 1 displays the details of the sample concerning control/pathologic, male/female and their ages.

Table 1 : Characterization of Sample

	Male		**Female**	
	Control	Pathol	Control	Pathol.
Length of sample	33	40	59	30
Average age	50,2	52,5	47,9	49,2
Standard deviation of ages	14,9	12,6	14,3	13,4

Since the number of subjects in the SVD with chronic laryngitis pathology alone is not so extensive, it was also included subjects with chronic laryngitis and other pathologies. This allowed to increase the length of the pathologic group with chronic laryngitis, from 25 to 40 for male, and from 16 to 30 female subjects. Table 2 presents the characterization of the pathologic group by gender and the number of subjects with other pathologies beside chronic laryngitis. In this table laryngitis means chronic laryngitis.

Table 2 : Characterization of Pathologic group

Pathology	**Male**	**Female**
Laryngitis	25	16
Laryngitis + Dysphonia	2	1
Laryngitis + Reinke's Edema	-	4
Laryngitis + Leukoplakia	8	4
Laryngitis + hyper functional dysphonia	1	1

Laryngitis + Polyp	1	1
Pachydermia laryngis	1	-
Laryngitis + Carcinoma in the epiglottis	-	1
Laryngitis + recurrent laryngeal nerve palsy	-	1
Laryngitis + case study	1	-
Laryngitis + Carcinoma	1	-
Laryngitis + hyper functional dysphonia l + leucoplakia	-	1

The total number of samples in the cured database is 70 pathologic subjects by 9 samples given a total of 630 samples for each parameter in the pathologic group. Similarly, the total number of 828 samples for each parameter in the control group is available in the cured database.

The database is organized in 3 parts. Each part displays the values for the all set of parameters: jitter, jitta, Shim, ShdB, Autocorrelation, NHR and HNR. The first part presents the data for pathologic and control group sub-sets by subject. The second part presents a statistical analysis of the parameters, and the third part presents the values for each parameter aggregated by parameter and pathologic or control group. This organization allows the user to access the data easier as they need the data organized by subjects or parameter by pathologic/control group.

Pathologic/Control Group Data by Subject

The first part of the database contains the set of parameters for the group of pathologic voices at left and for the group of control subjects at right. In this part each subject is identified with his ID number of the SVD, and their information about gender and age. For each subject the files corresponding to the vowels /a/, /i/ and /u/ are identified as well as the three tones high, neutral and low, in a total of 9 lines of parameters for each subject, as can be seen in Figure 1.

Statistical Analysis

The second part of the database shows a simple statistical analysis made over the parameters of each pathologic/control group. The statistical analysis present the mean, minimum, maximum, mode, median, 1º quartile, 3º quartile, lower limit of diagram and upper limit of diagram, as presented in Figure 2. These statistical parameters can be used to produce whiskers or boxplot graphic.

DATA BASE

Pathology Associated to this Data Base: Chronic Laryngitis

* This parameters were extracted from speech sound of the database "Saarbruecken Voice Database", and can be accessed by the link: http://www.stimmdatenbank.coli.uni-sa

Praat Free Software Programm was used to extract the parameters

Pathologic Group

Patient			Speech sound		JITTER		SHIMMER		Harmonic Parameters		
ID number	gender	age	Vowel	Tone	jitter %	jitta (µs)	Shim	ShdB	Autocorr.	NHR	HNR
498	F	26	A	Low	1,081	70,624	5,159	0,446	0,963	0,0392	14,81
				Normal	0,323	15,945	2,711	0,235	0,993	0,0075	21,44
				Hight	0,218	7,821	2,494	0,217	0,995	0,0046	23,51
			I	Low	0,342	18,419	3,019	0,264	0,994	0,0057	22,61
				Normal	0,588	26,668	4,351	0,373	0,990	0,0100	20,07
				Hight	0,277	9,554	1,570	0,370	0,997	0,0026	26,03
			U	Low	0,542	34,840	5,820	0,515	0,992	0,0079	21,89
				Normal	0,308	15,194	3,908	0,342	0,995	0,0048	23,54
				Hight	0,333	11,277	3,760	0,328	0,998	0,0022	27,14

Control Group

Subject			Speech sound		JITTER		SHIMMER		Harmonic Parameters		
ID number	gender	age	Vowel	Tone	jitter %	jitta (µs)	Shim	ShdB	Autocor.	NHR	HNR
23	F	26	A	Low	0,323	16,806	1,764	0,156	0,996	0,0043	24,49
				Normal	0,270	10,947	2,383	0,216	0,997	0,0035	25,02
				Hight	0,161	5,404	1,694	0,149	0,999	0,0011	30,47
			I	Low	0,180	9,273	3,201	0,281	0,994	0,0057	22,75
				Normal	0,210	8,369	2,374	0,211	0,996	0,0045	23,87
				Hight	0,165	5,502	1,852	0,163	0,998	0,0021	27,31
			U	Low	0,286	14,501	2,889	0,249	0,997	0,0029	25,90
				Normal	0,270	10,813	1,977	0,170	0,999	0,0012	29,80
				Hight	0,140	4,662	0,981	0,085	1,000	0,0003	35,66

Figure 1 : Header of the first part of the database

Parameters of Pathologic/Control Group Data

This third part of the database displays the values of each parameter without patient or subject identification neither vowel nor tone. And for each parameter the values of the pathologic and control groups are side by side. This organization is more useful to automatically extract the set of parameter values for each of the pathologic/control group.

Statistical values for the parameters of each phatologic/control group															
	JITTER				SHIMMER				Harmonic Parameters						
	jitter %		jitta (µs)		Shim		SHdB		Autocorr.		NHR		HNR	
	phath	*control*	*phath*	*control*	*phath*	*control*	*phath*	*control*	*phath*	*control*	*phath*	*control*	*phath*	*control*
mean	0,903	0,381	65,035	23,465	4,555	2,447	0,406	0,217	0,969	0,993	0,0432	0,0083	21,68	26,32
minimum	0,112	0,043	1,095	0,164	0,251	0,022	0,043	0,037	0,564	0,592	0,0004	0,0001	1,12	1,77
maximum	11,544	4,776	838,174	395,262	42,121	17,901	3,327	1,528	1,000	1,000	0,7886	0,7641	35,84	42,34
mode	0,323	0,279	63,170	13,470	4,949	2,152	0,235	0,157	0,997	0,998	0,0026	0,0016	26,65	24,49
median	0,493	0,305	30,942	16,510	3,030	2,009	0,270	0,176	0,992	0,997	0,0082	0,0031	22,28	26,28
1ºquartile	0,320	0,216	18,521	9,997	1,962	1,451	0,175	0,127	0,977	0,993	0,0034	0,0015	18,14	23,01
3ºquartile	0,822	0,429	57,350	27,388	5,469	2,911	0,494	0,258	0,997	0,998	0,0247	0,0067	26,20	29,75
lower limit of diagram	-0,432	-0,104	-39,724	-16,090	-3,299	-0,740	-0,303	-0,070	0,947	0,986	-0,0286	-0,0063	6,05	12,90
upper limit of diagram	1,574	0,749	115,594	53,474	10,730	5,101	0,973	0,455	1,027	1,006	0,0567	0,0145	38,28	39,86

Figure 2 : Second part of the database

Values organized by parameters and group (pathologig or control), for all vowels and tones													
JITTER				SHIMMER				Harmonic Parameters					
jitter %		jitta (µs)		Shim		SHdB		Autocorr.		NHR		HNR	
phath	*control*	*phath*	*control*	*phath*	*control*	*phath*	*control*	*phath*	*control*	*phath*	*control*	*phath*	*control*
1,081	0,323	70,624	16,806	5,159	1,764	0,446	0,156	0,963	0,996	0,0392	0,0043	14,81	24,49

Figure 3 : Header of the third part of the database

Conclusions

The paper presents the database of speech parameters extracted with PRAAT software from the Saarbrücken Voice Database of speech sounds corresponding to the steady pronunciation of the vowels /a/, /i/ and /u/ at high, mid (neutral) and low tones by subjects with the chronic laryngitis pathology and a control group with similar age.

This database consists of the speech parameters of jitta, jitter, Shim, ShdB, Autocorrelation, HNR and NHR extracted with Praat from the above mentioned speech sounds. They are organized into 3 part as described in section 3. This organization allows easy extraction for further research.

The database is made available for the scientific community research.

In the future it is intended to extend the database increasing the number of parameters, namely some parameters from continuous speech such as MFCC, LSF, MLSF since they are been used also by other researchers (Cordeiro, 2016). The inclusion of other patients from different databases is also under perspective. The inclusion of the parameters for other pathologies is already under way for dysphonia, and will be included soon.

Acknowledgment

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

References

Barry, W.J., Pützer, M. Saarbrücken Voice Database, Institute of Phonetics, Univ. of Saarland, http://www.stimmdatenbank.coli.unisaarland.de/

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

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

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

Brockmann-Bauser, M., 2011. Improving jitter and shimmer measurements in normal voices. Institute of Cellular Medicine, Medical School, Newcastle University.

Chaitanya Gupte and Shruti Gadewar, 2017. Diagnosis of Parkinson's Disease using Acoustic Analysis of Voice. Int. J. Sc. Res. in Network Security and Communication. Volume-5, Issue-3, 14-18.

Cordeiro, Hugo T, 2016. Reconhecimento de Patologias da Voz usando Técnicas de Processamento da Fala. PhD thesis at Universidade Nova de Lisboa.

Meilán JJ, Martínez-Sánchez F, Carro J, López DE, Millian-Morell L, Arana JM., 2014. Speech in Alzheimer's disease: can temporal and acoustic parameters discriminate dementia? Dement Geriatr. Cogn. Disord; 37(5-6): 327-34.

Salhi, L., Mourad, T., Cherif, A., 2010. Voice Disorders Identification Using Multilayer Neural Network. The International Arab Journal of Information Technology, Vol. 7, No. 2, 177-185.

Shama, K., Krishna, A. and Cholayya, N., 2007. Study of Harmonics-to-Noise Ratio and Critical-Band Energy Spectrum of Speech as Acoustic Indicators of Laryngeal and Voice Pathology. EURASIP Journal on Advances in Signal Processing, Vol. 2007.

Teixeira, J. P., Fernandes, P. O., 2015. Acoustic Analysis of Vocal Dysphonia. Procedia Computer Science. Elsevier 64 466 – 473.

Teixeira, J. P., Gonçalves, A., 2014. Accuracy of Jitter and Shimmer Measurements. Procedia Technology. Elsevier, Volume 16, 1190-1199.

Tusaliu, M., Dragu, A., Goanta, M., Mihalcea, G., Ionita, C., Luca, S., Ghiuzan, L. & Budu, A., 2016. Chronic laryngitis in adults. Archives of the Balkan Medical Union, V. 51, Issue 1, pp 34-36.