Analysis and Classification of Voice Pathologies using Glottal Signal Parameters with Recurrent Neural Networks and SVM

Leonardo Forero Mendoza[1], Manoela Kohler[2], Cristian Muñoz[2], Evelyn Conceição Santos Batista[2] and Marco Aurélio Pacheco[2]

[1]*Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil*

[2]*Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Rio de Janeiro, Brazil*

Keywords: Classification of Vocal Folds Pathologies, Glottal Signal Parameters, Neural Network, Deep Learning.

Abstract: The classification of voice diseases has many applications in health, in diseases treatment, and in the design of new medical equipment for helping doctors in diagnosing pathologies related to the voice. This work uses the parameters of the glottal signal to help the identification of two types of voice disorders related to the pathologies of the vocal folds: nodule and unilateral paralysis. The parameters of the glottal signal are obtained through a known inverse filtering method and they are used as inputs to an Artificial Neural Network, RNN, LSTM, a Support Vector Machine and also to a Hidden Markov Model, to obtain the classification, and to compare the results, of the voice signals into three different groups: speakers with nodule in the vocal folds; speakers with unilateral paralysis of the vocal folds; and speakers with normal voices, that is, without nodule or unilateral paralysis present in the vocal folds. The database is composed of 248 voice recordings (signals of vowels production) containing samples corresponding to the three groups mentioned. In this study a larger database was used for the classification when compared with similar studies, and its classification rate is superior to other studies, reaching 99.2%.

1 INTRODUCTION

The diagnosis of voice pathologies currently requires invasive endoscopy procedures, such as laryngostroboscopy or surgical microlaryngoscopy. However, one wants to aid the pre-diagnosis of the vocal folds pathologies with computer-based, decision support diagnostic tools using voice signals. Two pathologies related to the vocal folds will be considered here: nodules and unilateral paralysis (Roy N et al., 2017) (Francis D. O. At al., 2014). Vocal cord nodules are growth on both vocal folds caused by their repeated and incorrect usage, which permits the developing of swollen spots on them. The nodules will become larger and stiffer the longer the vocal incorrect usage continues. Singers, teachers and announcers are examples most probably to have this kind of pathology in the vocal folds (Francis D. O. At al., 2014). Unilateral vocal fold paralysis (UVFP) occurs from a dysfunction of the recurrent laryngeal or vagus nerve innervating the larynx. It causes a characteristic breathy voice often accompanied by swallowing disability, a weak cough, and the sensation of shortness of breath. This is a common cause of neurogenic hoarseness. When this paralysis is properly evaluated and treated, normal speaking voice is typically restored (Steffen N Pedrosa V. V. and Kazuo R., Pontes P, 2009) (Behlau M, Pontes PP, 1995).

The aim here is to evaluate the use of glottal signals (signal obtained just after the vocal folds and before the vocal tract) for providing better classification models of the pathologies discussed above. The most common method for extracting voice features is directly from the voice signal (Roy N et al., 2017).

However, many researchers have looked for some characteristics extracted from the glottal signal, not only for identifying pathologies related to the vocal folds, but also to other applications, as to synthesize voice (Henrich, N., 2001)(Henrich N, d'Alessandro C., 2014) or identifying vocal aging (Mendonza L., Vellasco M., Cataldo E., 2014).

The process of obtaining the glottal signal, from the voice signal, has been facilitated due to the development of algorithms which can perform an

Mendoza, L., Kohler, M., Muñoz, C., Batista, E. and Pacheco, M.
Analysis and Classification of Voice Pathologies using Glottal Signal Parameters with Recurrent Neural Networks and SVM.
DOI: 10.5220/0007250700190028
In *Proceedings of the 11th International Conference on Agents and Artificial Intelligence (ICAART 2019)*, pages 19-28
ISBN: 978-989-758-350-6

inverse filtering from the voice signal, eliminating the influence of the vocal tract (Software Aparat).

Different methods have been used to classify diseases related to the voice, such as Hidden Markov Models (HMM) (Francis D. O. At al., 2014), Gaussian Mixture Models (GMM) and Artificial Neural Networks (Steffen N Pedrosa V. V. and Kazuo R., Pontes P, 2009), all of them using as inputs Mel-Frequency Cepstral Coefficients (MFCC) and parameters such as jitter and shimmer. However, as most voice disorders are due to some disorder on the vocal folds dynamics, it is best to work with parameters extracted from the glottal signal, since the signal is produced by the vocal folds.

In (Rosa I. S., 2005) (Londoño J., Llorente J., 2010), (Wang X, Zhang J, Yan Y, 2009) the Mel-frequency cepstral coefficients (MFCC) were used as input parameters to classify pathologies. A database composed of 12 recordings for men and women, resulting in a maximum performance of 80% accuracy (Londoño J., Llorente J., 2010). MFCC have also been proved to be effective in speaker recognition problems. However, their performance is not as effective in the classification of voice pathologies. In (Rosa I. S., 2005), several models for the classification of voice pathologies are compared. The best performance has been provided by a neural network based model, differing from speaker recognition applications where best results are usually obtained with GMM and HMM. This is probably because classification of voice pathologies does not fully depend on temporal features of the voice, and the pathology causes change in the voice signal (Hariharan, M., 2009).

Therefore, the main objective of this work is to evaluate the performance of voice pathologies classification models based on parameters extracted from the glottal signal. Additionally, a new database was created, with a larger number of voice recordings, which allows a better evaluation of the influence of each parameter in the classification performance.

This paper is organized as follows. Methods section explains how the glottal signal is obtained and how the features, extracted from the glottal signal, are used. Proposed Methodology section presents the three classifiers evaluated in this paper, so their performance can be evaluated in voice pathologies classification: Neural Network, Support Vector Machine and Hidden Markov Model. Results section presents the database details, results obtained and their analysis. Lastly, Conclusions are outlined in the final section.

2 METHODS

2.1 The Glottal Signal

The voice signal production, particularly the one related to voiced sounds, e.g. vowels, starts with the contraction-expansion of the lungs, generating a pressure difference between the air in the lungs and the air near the mouth. The airflow created passes through the vocal folds, which oscillate in a frequency called the fundamental frequency of the voice. This oscillation modifies the airflow coming from the lungs, changing it into air pulses. The pressure signal formed by the air pulses is quasi-periodic and it is called the glottal signal (M. D. O. Rosa., 2000).

2.2 Features Extracted from the Glottal Signal

The glottal signal is obtained performing an inverse filtering on the voice signal, which consists on eliminating the influence of the vocal tract and the voice radiation caused by the mouth, preserving the glottal signal characteristics (Pulakka H., 2005). The inverse filtering algorithm used here is the so-called PSIAIF (Pitch Synchronous Iterative Adaptive Inverse Filtering) (Mendonza L., Vellasco M., Cataldo E., 2014) (Pulakka H., 2005). It was chosen due to its high performance and ease development. There is a toolbox implementation in Matlab®, called Aparat (Software Aparat), which was constructed especially based on the PSIAIF method to obtain the glottal signal and to extract its main features or parameters. The parameters that will be used can be divided into three groups: time domain, frequency domain, and the ones that represent the variations of the fundamental frequency. More details about these parameters can be found in (Pulakka H., 2005).

2.2.1 Time-domain Parameters of the Glottal Signal

The time domain parameters which can be extracted from the glottal signal are described below (Wang X, Zhang J, Yan Y, 2009) (Pulakka H., 2005).

- Closing phase (Ko): describes the interval between the instant of the maximum opening of the vocal folds and the instant where they close (M. D. O. Rosa., 2000);
- Opening phase (Ka): describes the interval between the instant where the vocal folds start the oscillation up to their maximum opening (M. D. O. Rosa., 2000);

- Open quotient (OQ): The ratio between the total time of the vocal folds opening and the total time of a cycle (or period) of the glottal signal (T). It is inversely proportional to the intensity of the voice. The smaller it is, the higher the voice intensity (Wang X, Zhang J, Yan Y, 2009) (Pulakka H., 2005);
- Close quotient (CIQ): The ratio between the closing phase parameter (Ko) and the total length of a glottal pulse (T) (Pulakka H., 2005). It is inversely proportional to the voice intensity. The smaller it is, the higher the voice intensity;
- Amplitude quotient (AQ): The ratio between the glottal signal amplitude (Av) and the minimum value of the glottal signal derivative [16]. It is re-lated to the speaker phonation (Pulakka H., 2005);
- Normalized amplitude quotient (NAQ): It is calculated by the ratio between the amplitude quotient (AQ) and the total time length of the glottal pulse (T) (Pulakka H., 2005);
- Open quotient defined by the Liljencrants-Fant model (OQa): This is another opening quotient but calculated by the Liljencrants-Fant model for inverse filtering. Details about this model can be found in (Wang X, Zhang J, Yan Y, 2009);
- Quasi open quotient (QoQ): It is the relationship between the glottal signal opening at the exact instant of the oscillation and the closing time. It has been used in some works to classify emotions (Wang X, Zhang J, Yan Y, 2009);
- Speed quotient (SQ): defined as the ratio of the opening phase length to the closing phase length (Pulakka H., 2005);

2.2.2 Frequency Domain Parameters

- Difference between harmonics (DH12): Also known as H1-H2, it is the difference between the values of the first and second harmonics of the glottal signal (Wang X, Zhang J, Yan Y, 2009) (Pulakka H., 2005). This parameter has been used to measure vocal quality;
- Harmonics richness factor (HRF): relates the first harmonic (H1) with the sum of the energy of the other harmonics (Hk) (Pulakka H., 2005). It has also been used to measure vocal quality;

2.2.3 Parameters that Represent Variations and Perturbations in the Fundamental Frequency

- Jitter: variations in fundamental frequency between successive vibratory cycles (Wang X, Zhang J, Yan Y, 2009) (Pulakka H., 2005). Changes in jitter may be indicative of neurological or psychological difficulties (Roy N et al., 2017);
- Shimmer: variations in amplitude of the glottal flow between successive vibratory cycles (Wang X, Zhang J, Yan Y, 2009) (Pulakka H., 2005). Changing the shimmer is found mainly in the presence of mass lesions in the vocal folds, such as polyps, edema, or carcinomas (Roy N et al., 2017);

3 PROPOSED METHODOLOGY FOR VOICE PATHOLOGIES CLASSIFICATION

The proposed model used has two stages: the first stage is the features extraction, where all the above mentioned parameters from the glottal signal are obtained; the second stage is the classification module, where four algorithms have been selected to classify different pathologies of the voice - a multilayer perceptron (MLP) neural network, a support vec-tor machine, Long short-term memory (LSTM) and a Hidden Markov Model (HMM), for comparison reasons. The proposed methodology is illustrated in Figure 1 and each model is described in the following sub-sections. A similar methodology has been already applied for classifying voice aging (Mendonza L., Vellasco M., Cataldo E., 2014), with very good results.

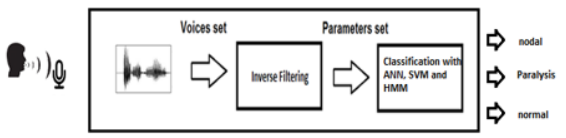

Figure 1: Methodology used for the classification of voice pathologies.

3.1 Inverse Filtering

For each vocal utterance a corresponding glottal signal is obtained by inverse filtering (PSIAIF meth-od) and the parameters are extracted using the Aparat (Software Aparat) and Praat (Software Praat) software. The following parameters are obtained:

fundamental frequency (fo), jitter, shimmer, Ko, Ka, NAQ, AQ, CIQ, OQ1, OQ2, Oqa, Qoq, SQ1, SQ2, DH12, and HRF. The parameters are separated according to the groups to which they belong. In particular, OQ was divided into OQ1 and OQ2, the open quotients calculated from the so-called primary and secondary openings of the glottal flow. The difference between OQ1 and OQ2 is that OQ1 is calculated from the closure of the glottal flow until the closure of the next glottal flow, and OQ2 is calculated from de opening until the closure of the glottal flow; SQ, as well, was divided into speed quotients calculated from the primary and secondary openings of glottal signal. It is important to mention that some parameters provide similar information, but, in this phase, all of them will be considered.

3.2 Classification Module

For the classification of voice pathologies four classifiers have been used: Artificial Neural Networks (ANN), Support Vector Machine (SVM), LSTM and Hidden Markov Models (HMM).

For the ANN classifier, a multi-layer perceptron (MLP) structure, trained with the back-propagation algorithm, was chosen, since it is a universal approximator. Different topologies were examined with different numbers of neurons in the hidden layer to seek the best generalization performance. For the SVM classifier, different kernels (polynomial, radial basis function (RBF), and sigmoid) and different values for the normalization coefficient (C) were evaluated to determine the optimal settings. Final-ly, an Estimate-Maximize (Baum Welch) approach was used to train three HMM models (one HMM for each output class), each one to maximize the likelihood of the training data with respect to the unknown parameters. To classify a sequence into one of the three classes, the log-likelihood given by each model is computed, and the most likely model defines the class that the test sequence belongs to. Left-to-right HMM models with five states and three Gaussian mixtures were trained in order to obtain an optimal classification rate.

4 RESULTS

4.1 Database

Most of the works on vocal folds diseases classification just classify speakers into two groups: speakers with disease (all kinds of disease) and speakers with normal voices (Rosa I. S., 2005), (Londoño J., Llorente J., 2010). In this work the type of disease is also identified, helping in the indication if the patient has nodule or paralysis on the vocal folds, or neither one.

The developed database is composed of 248 records consisting of voices of both genders, women and men, with different ages, and it is divided into three groups: 12 speakers with nodule on the vocal folds; 8 speakers with vocal folds paralysis; and 11 speakers with normal voices. Eight voice records were taken from each speaker. This database was obtained from a speech therapist in Rio de Janeiro among people in treatment.

For the recordings is used a computer, the Doctor speech software and an omnidirectional microphone. The voices were recorded in a doctor's office.

The speakers belonging to the pathology groups (nodule and paralysis) have different categories of the disease in each group, as described in Tables 1 and 2. The following tables describe the speakers in more details.

Table 1: Speakers with Nodule on the Vocal Folds (F – Female, M – Male).

Speaker	Gender	Age	Description of the disease
Speaker 1	F	42	Bilateral nod-ules causing a small irregular vocal cord chink
Speaker 2	F	38	Bilateral nod-ules with mid-posterior chink
Speaker 3	F	24	Vocal nodules with moderate and severe anterior and posterior irregular chinks
Speaker 4	F	53	Vocal nodules with an irregular vocal cord chink
Speaker 5	F	53	Vocal nodules with an irregular vocal cord chink

Table 1: Speakers with Nodule on the Vocal Folds (F – Female, M – Male) (Cont.).

Speaker 6	F	38	Vocal nodules with mid-posterior chink
Speaker 7	F	34	Vocal nodules with mid-posterior chink
Speaker 8	F	32	Fibrous nodules - mid-posterior chink - great vocal effort
Speaker 9	F	29	Vocal nodules with mid-posterior chink
Speaker 10	F	33	Vocal nodules with an irregular vocal cord chink
Speaker 11	F	28	Vocal nodules with a slight irregular vocal cord chink
Speaker 12	F	28	Vocal nodules with mid-posterior chink

Table 2: Speakers with Vocal Folds Paralysis (F – Female, M – Male).

Speaker	Gender	Age	Description of the disease
Speaker 13	M	50	Right vocal fold paralysis with scar re-traction in the middle 1/3 - anterior spin-dle chink (lar-yngeal trauma sequel)
Speaker 14	M	50	Right hemilarynx idiopathic paralysiswith slight vocal cord bowing
Speaker 15	M	24	Right vocal cord paralysis with spindle chink
Speaker 16	F	69	Right vocal cord paralysis in paramedian posi-tion with a slight bowing and a slight spindle chink - paralytic falsetto
Speaker 17	F	45	Left vocal cord paralysis in the left median and paramedian positions – no chinks
Speaker 18	F	43	Right hemilarynx idio-pathic paralysis - para-median position
Speaker 19	M	66	Left vocal cord paralysis with a slight bow-ing (intubation trauma)
Speaker 20	M	53	Right vocal cord paralysis in paramedian position - left vocal fold stiffness

Table 3: Speakers with No Disease (F - Female, M – Male).

Speaker	Gender	Age
Speaker 21	F	56
Speaker 22	M	30
Speaker 23	F	41
Speaker 24	M	46
Speaker 25	F	61
Speaker 26	M	35
Speaker 27	M	63
Speaker 28	M	48
Speaker 29	M	26
Speaker 30	F	56
Speaker 31	F	56

4.2 Analysis of the Parameters for Classification

In order to evaluate the influence of each input parameter in the classification of voice diseases, the box-plot [24] function was used. The boxplot was constructed for each of the parameters extracted from the glottal signal, in order to see their influence in each type of pathology or normal voices and to compare their behavior. The three boxplots for each group are related to Nodule, Paralysis and Normal Voices, respectively. To facilitate the analysis and better understand the parameters variation, their boxplots were grouped and analyzed by each type of parameter: Time-domain parameters, Frequency-domain parameters and Parameters that Represent Variations and Perturbations in the Fundamental Frequency, as described in the following sub-sections and in Figures 2 to 15.

4.2.1 Time-domain Parameters of the Glottal Signal

The following figures shows the corresponding boxplots for the so-called time-domain parameters extracted from the glottal signal, where some interesting observations can be extract-ed. The parameter Ko, which shows the closing phase of the vocal folds, is higher in normal voices than in voices with the pathologies considered (Figure 2). OQ1, OQ2, CIQ, AQ and NAQ parameters (Figure 4 to 8) imply that normal voices have more intensity and better voice quality when compared with pathologies. The values of the parameters SQ1 and SQ2 are lower in normal voices, which indicate a shortening in the structure of the vocal folds when one has these diseases, especially paralysis (Figures 11 and 12).

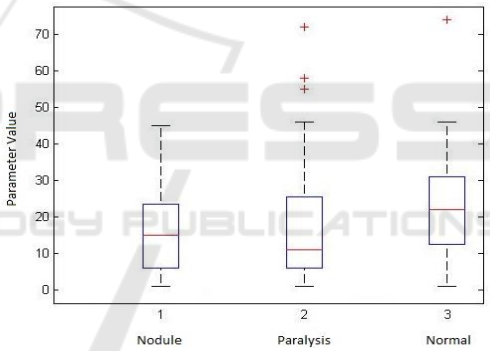

Figure 2: Closing phase(Ko).

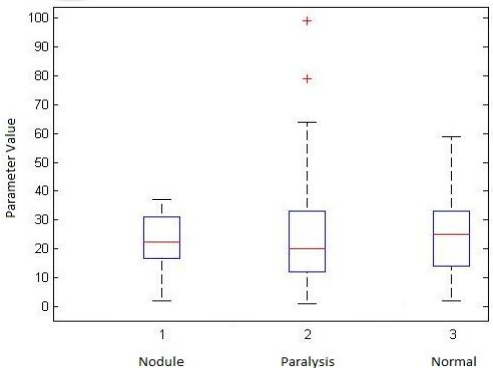

Figure 3: Opening phase(Ka).

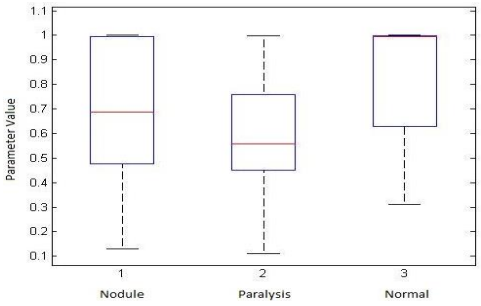

Figure 4: Open quotient(OQ1).

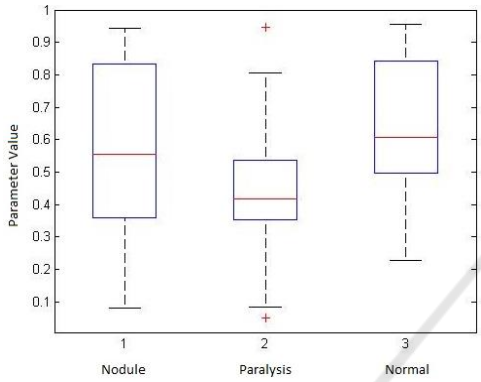

Figure 5: Open quotient (OQ2).

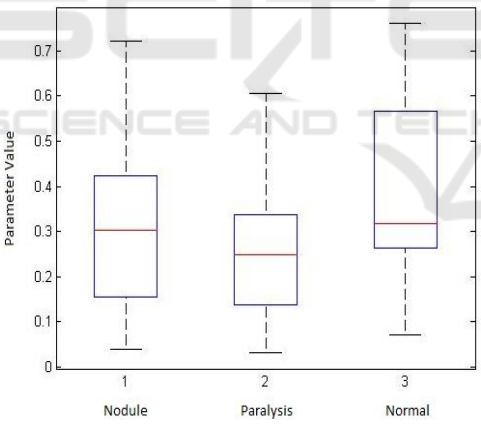

Figure 6: Close quotient(CIQ).

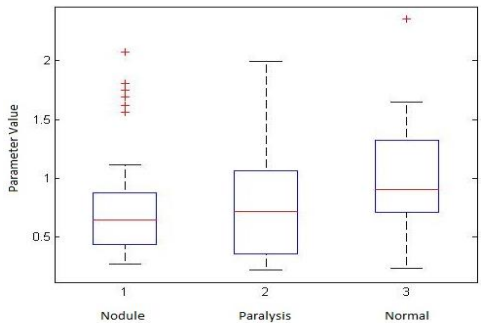

Figure 7: Amplitude quotient (AQ).

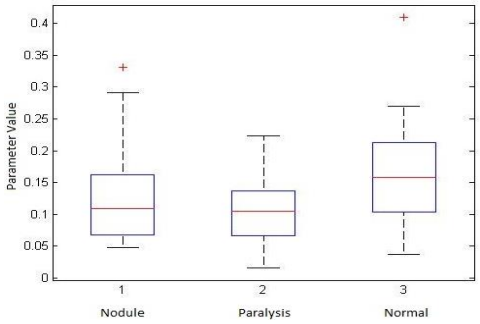

Figure 8: Normalized amplitude quotient (NAQ).

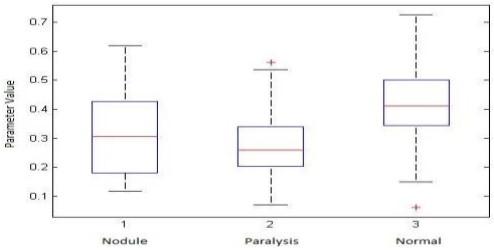

Figure 9: Open quotient defined by the Liljencrants-Fant model (OQa).

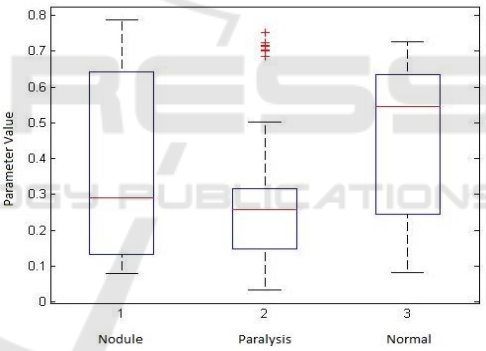

Figure 10: Quasi opening quotient (QoQ).

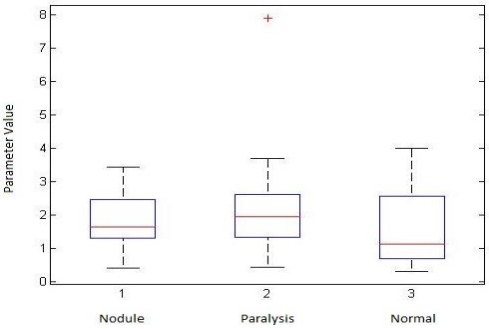

Figure 11: Speed quotient (SQ1).

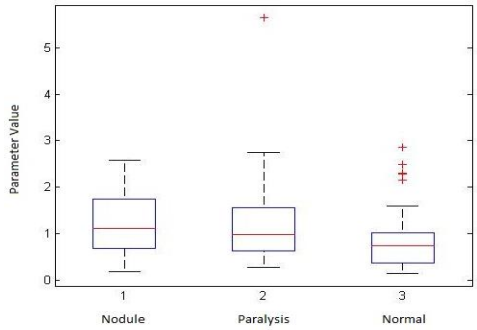

Figure 12: Speed quotient (SQ).

4.2.2 Frequency Domain Parameters

Figures 13 to 15 show the corresponding boxplots related to the frequency domain parameters.

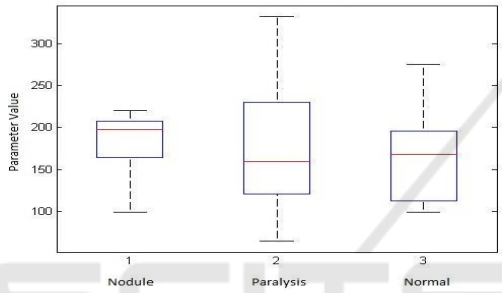

Figure 13: Fundamental Frequency (F0).

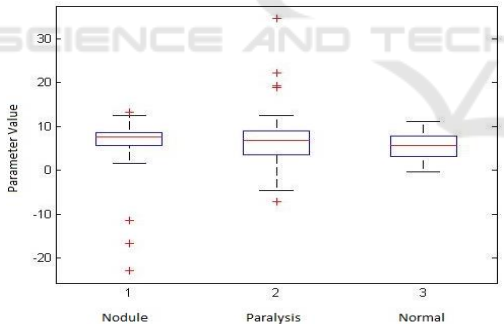

Figure 14: Difference between harmonics (DH12).

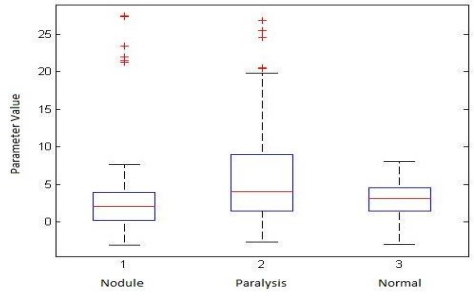

Figure 15: Harmonics richness factor (HRF).

The fundamental frequency (Fig. 13) has a wide variation for voices with unilateral paralysis, showing a greater disturbance in the vocal folds. Voices with nodules have less variation of fundamental frequency when compared with normal voices. Harmonics richness factor (Fig. 15) changes a lot for unilateral paralysis.

4.2.3 Parameters That Represent Variations and Perturbations in the Fundamental Frequency

Figures 16 and 17 present the boxplots of the parameters directed related to the variations and perturbations of the fundamental frequency.

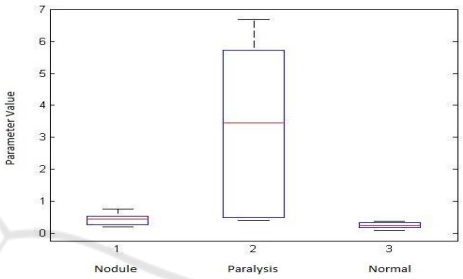

Figure 16: Jitter.

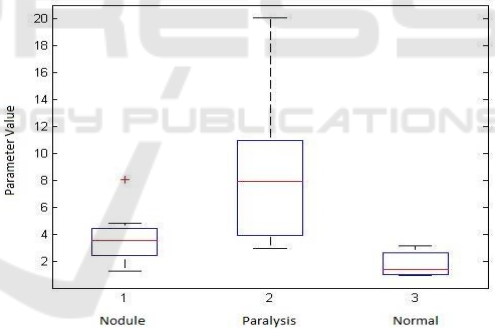

Figure 17: Shimmer.

As can be seen from these boxplots, in the pathologies cases, the function of the vocal folds is greatly compromised, which is indicated by jitter and shimmer parameters, as shown in Figures 16 and 17. Jitter and shimmer parameters vary the most in the voice when paralysis occurs. Jitter and Shimmer are very high in voices with paralysis, proving to have affected the most the vocal folds.

4.2.4 Analysis of the Classification Results

Classification of the pathologies was performed using four different classifiers: ANN, SVM, LSTM and HMM. For each classifier, three cases were considered for the input parameters: (i) only the

parameters extracted from the glottal signal, (ii) only the MFCCs, and (iii) a combination of (i) and (ii). The results for each input configuration are presented in the following sub-sections.

Classification Results with the Parameters of the Glottal Signal. In this case, the inputs of the classifiers are 16 parameters of the glottal signal. The original database was divided into training, validation and test sets, where 70% of the database was used for training, 20% for validation, and 10% for testing. For ANN (after performing lots of tests varying the number of the neurons in the hidden layer) the best result was obtained with 8 processors in the hidden layer.

Considering the SVM as the classifier, the best result was achieved when a RBF kernel was used with a regularization constant of C=1 and a Gaussian standard deviation of σ=1.

Our model is a deep recurrent neural network with two layers of 100 LSTM cells each. The bottommost layer is the input layer where we inject each time frame of an individual example at each time step. The layer contains 13 units that would contain the coefficients of the time frames. The next layer is two layers are LSTM recurrent layers.

An Estimate-Maximize (Baum Welch) approach was used to train three HMM models (one HMM for each class), each one to maximize the likelihood of the training data with respect to the unknown parameters. To classify a sequence into one of the three classes, the log-likelihood given by each model is computed, and the most likely model defines the class that the test sequence belongs to. Left-to-right HMM models with five states and three Gaussian mixtures were trained in order to obtain an optimal classification rate.

The classifiers has three outputs: speakers with nodule on the vocal folds, containing 93 voice records, speaker with vocal folds paralysis, containing 67 records, and speaker with normal voice, containing 88 records.

4.2.5 Classification Results with Mel-Frequency Cepstral Coefficients (MFCCS)

Mel-Frequency Cepstral Coefficients (MFCCs) are coefficients that collectively make up an MFC and are derived from a type of cepstral representation of the audio clip (a nonlinear "spectrum-of-a-spectrum"). MFCCs are common in speaker recognition, which is the task of recognizing people from their voices. 12 MFC coefficients were used, the number most often used in the literature (Rosa I. S., 2005) (Londoño J., Llorente J., 2010).

The inputs of the classifiers are, therefore, 12 MFC coefficients in this case. The original database was divided into training, validation and test sets, where 70% of the database was used for training, 20% for validation, and 10% for testing, as in the previous case. After lots of tests varying the number of the neurons, the best result was achieved with 6 processors in the hidden layer. Considering the SVM as the classifier, the best result was achieved when a RBF kernel was used with a regularization constant of C= 0,8 and a Gaussian standard deviation of σ=1. HMM configuration is the same as above.

LSTM had the best performance with 100 cells.

4.2.6 Classification Results with Combining MFCCs and Glottal Signal Parameters

In this third configuration, the input vector of the classifiers is composed of 12 MFC coefficients and 16 parameters of the glottal signal. The original database was also divided into training, validation and test sets, where 70% of the database was used for training, 20% for validation, and 10% for testing. After lots of tests, the best ANN configuration was obtained with 9 processors in the hidden layer. Considering the SVM as the classifier, the best result was achieved when a RBF kernel was used with a regularization constant of C= 2 and a Gaussian standard deviation of σ=1. HMM and LSTM configuration is the same as above.

4.2.7 Discussion

Table 4 presents a summary of the results obtained with all three classifiers and all four configurations of input signals. As can be seen from the results in Table 4, the classification was successful with the glottal signal parameters, despite having an imbalanced database (fewer samples for voices with paralysis) and factors such as gender and age difference between speakers, reaching the conclusion that these parameters are good discriminators for classifying voice disorders.

When using only MFCC parameters, the best result is obtained with the LSTM classifier, since its stochastic behaviour can better handle temporal samples.

The combination of MFCCs and glottal signal parameters provided the best classification results, with an increase of 1% in the average performance when compared with the results with only glottal signal parameters. The best classification performance was obtained with the LSTM classifier,

with over 98,3% accuracy. The results were obtained by Intel® Optimization for TensorFlow. The LSTM network was run on a gold Xeon processor showing faster speed than on a 1080 nvidea graphics board in 30% under the same conditions.

Table 4: Classification of Voice Pathologies.

Parameters	ANN	HMM	LSTM	SVM
Parameters of the glottal signal	95.8%	82%	97%	96.2%
MFCCs	75.2%	87%	94%	80%
Glottal signal parameters and MFCCs	96.6%	92%	98.3%	97.2%

LSTM	**Xeon**	**1080 Nvidia**
Training time	**527 sec**	742 sec

5 CONCLUSIONS

The aim of this work was the classification of two voice diseases: nodule and unilateral paralysis and the evaluation of the impact of parameters from the glottal signal on this identification. Three different classifiers have been used, to compare their performance: an Artificial Neural Network, a Support Vector Machine, LSTM and a Hidden Markov Model.

From the results obtained, it can be verified that glottal signal parameters are more relevant to discriminate pathologies of the vocal folds than MFCC's, when they are evaluated individually. This is the case even when the database is composed of individuals with different genders and ages, providing an average accuracy over 99%.

ACKNOWLEDGMENTS

This work was supported by Intel Corporation.

REFERENCES

Roy, N., Holt, K. I., Redmond, S., Muntz, H., 2007, *Behavioral characteristics of children with vocal fold nodules*. J Voice. 21(2):157-68.

Francis, D. O.; McKiever, M. E.; Garrett, G., Jacobson, B.; Penson, D. F., 2014, *Assessment of Patient Experience with Unilateral Vocal Fold Immobility: A Preliminary Study*, Journal of voice 28 (5), 636-643.

Steffen, N., Pedrosa, V. V., Kazuo, R., Pontes, P., 2009, *Modifications of Vestibular Fold Shape from Respiration to Phonation in Unilateral Vocal FoldParalysis*, Journal of Voice, Vol. 25, No. 1, pp. 111-113.

Behlau, M., Pontes, P. P., 1995, *Avaliação e tratamento das disfonias. São Paulo: Lovise, in unilateral vocal fold paralysis*, Journal of Voice 13(1):36-42.

Henrich, N., 2001, *Étude de la source glottique en voix parlée et chantée modelisation et estimation, mesures acoustiques et electroglottographiques, perception*, Thèse de doctorat de l'Université Paris 6 (PhD Thesis).

Henrich, N., d'Alessandro, C., Doval, B., Castellengo, M. 2005, *Glottal Open quotient in singing: Measurements and correlation with laryngeal mechanisms, vocal intensity, and fundamental frequency*, Journal of the Acoustical Society of America 117(3), pp 1417-1430.

Mendonza, L., Vellasco, M., Cataldo, E., 2014, *Classification of Vocal Aging Using Parameters Extracted From the Glottal Signal J Voice.* 21(2):157-68.

Software Aparat, http://aparat.sourceforge.net/index.php/Main_Page, Helsinki University of Technology Laboratory of Acoustics and Audio Signal Processing.).

Rosa, I. S., 2005, *Análise acústica da voz de indivíduos na terceira idade*, Tese de mestrado Universidade de São Paulo, São Carlos (in portuguese).

Londoño, J., Llorente, J., 2010, *An improved method for voice pathology detection by means of a HMM-based feature space transformation Pattern Recognition*, Volume 43, Issue 9, September 2010.

Wang, X., Zhang, J., Yan, Y., 2009, *Glottal Source biometrical signature for voice pathology detection, Speech Communication*, 51 759-781.

Hariharan, M., Paulraj, M. P., Yaacob, S., 2009, *Identification of vocal fold pathology based on Mel Frequency Band Energy Coefficients and singular value decomposition Signal and Image Processing Applications* (ICSIPA), volume 514 – 517.

Rosa, M. D. O., Pereira, J. C., Grellet, M., 2000, *Adaptive Estimation of Residue Signal for Voice Pathology Diagnosis*, IEEE Trans. Biomedical Eng., Vol. 47, No. 1, Jan. 2000.

Pulakka H., 2005, *Analysis of Human Voice Production Using Inverse Filtering, High-Speed Imaging, and Electroglottograph,*. University of Technology Helsinki.

Software Praat, http://www.fon.hum.uva.nl/praat/, University of Amsterdam.