International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 23 (2017) pp. 13582-13588

Recognition of Organic, Neurologic and Functional Diseases Using Cepstral Features Extracted From The Speech Multiscale Product

Chekili Salma[1], Belhaj Asma[2] and Bouzid Aicha[3]

[1,2,3] Tunis el Manar University, National Engineering School of Tunisia, Laboratory of Signal, Images and Information Technology, Tunisia.

[1]Orcid: 0000-0001-5712-992X

Abstract

In recent years, the objective assessment of voice pathologies has become popular in many research studies. Automatic voice-pathology detection and classification systems effectively contribute to the evaluation of voice pathologies, to help clinicians for the detection of the existence of any voice diseases and the type of pathology that patients suffer from it in the early stage. The primary aim of this paper is to investigate new parameters for the classification of various types of pathologies which are organic, neurologic and functional. It's about calculating Mel Frequency Cepstral Coefficients MFCC not among the speech signal directly but among the speech multiscale product (MP), the MP provides a derived speech signal which is more straightforward to be analyzed. In fact, it has a periodic structure for voiced sounds with high peaks at the glottal closure instants and null elsewhere. In this study, we adopt the Support Vector Machine classifier SVM and we use the Massachusetts Eye and Ear Infirmary MEEI database. The experimental results show that the classification rates obtained using MFCC features extracted from the multiscale product MP give better results than those derived from the speech signal.

Keywords: MFCC; SVM; Multiscale product; MEEI, Organic, Functional, Neurologic

INTRODUCTION

In the recent history of medical researchers, a significant interest in voice analysis was shown; it's about a medical study of the patient's voices with problems in the vocal cords. To analyze the patient's voice, a specialist needs extensive experience and training. But in some situations, it's not possible, because generally, voice problems derive from Laryngeal muscle or the vocal folds, which controls the voice production.

The vocal fold analysis is physically challenging and not efficient, thus for the diagnosis of voice pathology, there is a need for a device or equipment. This requirement produces consciousness among the researcher to provide some technology to help therapists in detecting the voice pathology. Such technological support can be made base on the complete comprehension or study of the most common vocal disorders, their causes, symptoms and the side effects of diseases.

To solve these problems, researchers decided to create a detection system of voice pathology which will be a consistent automatic system, modest and non- invasive used for the vocal cords pathologies' detection. The acoustic parameters have a principal effect of the discrimination between the normal voices and the pathological ones. Based on the parameters obtained from the acoustic features, the classification system analyzes and decides if the voice signal produced is affected by one or more diseases or not.

This paper explores an efficient voice pathology detection system and we propose new parameters for the classification of various types of pathologies which are: organic, functional and neurologic. The symptoms in the voice of patients with laryngeal pathologies, i.e. organic pathologies are mostly associated with breathy voice, hoarseness, and abnormal vibration of the vocal folds due to the presence of polyps and/or nodules [1].

Besides, the hypernasality is one of the voice disorders with functional origins, for example, in the voice of sufferers with cleft lip and palate (CLP), the hypernasality is the principal feature. This disease causes the production of the voice with too much nasalization, which results from inappropriate control of the velum, generating abnormal resonances in the vocal and nasal cavities [2-3]. Regarding the neurological disorders, they are caused by abnormalities of the brain or unusual control of the voice box muscles, palate, tongue, throat, jaw or lips resulting in a variety of speech and/or voice problems such as: paralysis, stroke, Parkinson's disease etc. [4-5].

The main idea is to calculate the MFCC not among the speech signal but among the multiscale product (MP) and to study their effects on the classification. Then, we try to analyze and to discriminate pathological voice and normal voice using classification method, which is a famous classification model, i.e., Support Vector Machine Algorithm. This paper indicates the literature review in section (II), a detailed description of the

proposed research work, methodology and techniques was discussed in section (III). Experiments and results were discussed in section (IV). Finally, the conclusion is noted in the section (V) and plan for future work is discussed in section (VI).

LITERATURE REVIEW

Several research works in the identification of pathological voice, discriminate between normal and pathological voices. The commonly used classifiers in the speech recognition are considered in the pathological voices classification as: the hidden Markov model (HMM), neural networks, the Gaussian mixture model (GMM), and the support vector machines (SVM). A great number of studies has proposed a binary classification: normal/pathological voices [6-8].

The best classifier used in problems is the (SVM) classifier[9]. For example, [10] suggests using a set of features containing 11 MFCC coefficients, GNE(Glottal to Noise Excitation), HNR (Harmonic to Noise Ratio), NNE (Normalized Noise Energy), Energy, and their first derivatives.

In [11], the classification rates obtained by the SVM classifier, using features extracted from wavelet transform of speech samples to discriminate between normal and pathological voices, are 97.5% for normal voices and 100% for pathological ones.

In [12], to discriminate between normal and pathological voices, the SVM classifier allows good rates which are respectively 98.26% for normal voices and 99.65% for pathologic ones. These results are obtained with feature vectors containing the MFCC and their variations only or combined with the energy and the open quotient.

So the SVM classifier using specific parameters, have achieved the best performance. However, the classification between the pathologies is operated in few works [13] and the results are not sufficiently efficient.

In [14], SVM classifier and spectral modulation features are used to recognize polyp pathology among normal voices and three different pathologies which are vocal nodules, keratosis leukoplakia and adductor spasmodic dysphonia from the MEEI corpus [15] using the vowel \a\. The system achieves 90% as an average recognition rate.

In [16], it's a comparison between the results obtained using MFCC parameters and those obtained in previous work [14]. The results achieved are almost 25% lower than the results produced by spectral modulation.

In [17], the SVM classifier using the classic feature as the cepstral features, the fundamental frequency and a new parameter called the open quotient is considered. They noticed that MFCC coefficients are the most significant. However, the primary frequency allows distinguishing between the diseases and the open quotient allows differentiating pathological from normal voices.

PROPOSED RESEARCH METHODOLOGY

The main idea in this work is to calculate the MFCC parameters not from the speech signal but from the speech multiscale product.

The summary of the stated research work to detect and classify the voice pathology is illustrated in figure 1.

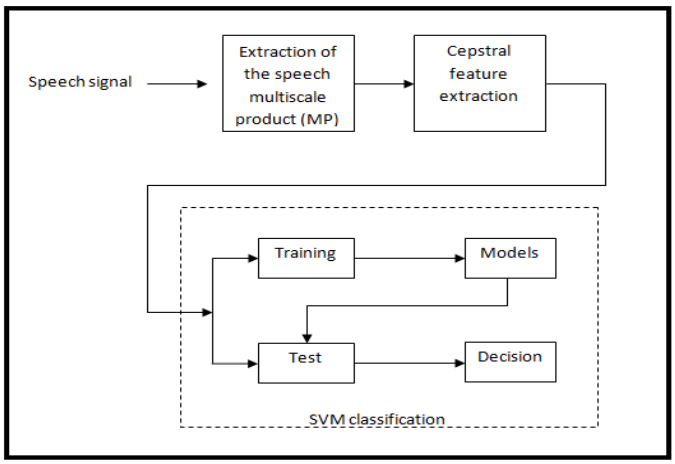

Figure 1: Research plan for proposed work.

A. Extraction of MFCC from the speech multiscale product:

The Multi-scale product was introduced for signal edge detection by Sadler et al. [18-19]. It has shown that the scale multiplication achieves better results than any scale especially on the localization performance [20-21].

In this work, we are interested in the use of the MP because it presents a derived speech signal with a simple structure; it is quasiperiodic for voiced sounds and almost zero for unvoiced sounds having a shape very close to the electroglottograph signal: a signal describing the vocal source activity [12].

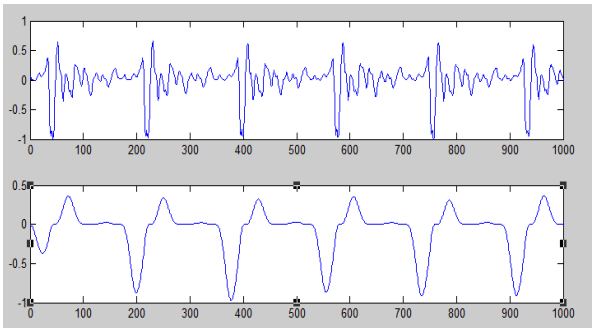

Figure 2: Speech pathological voice corresponding to a sustained vowel /a/ extracted from Polyp CXR13AN pronounced by a female speaker and its MP.

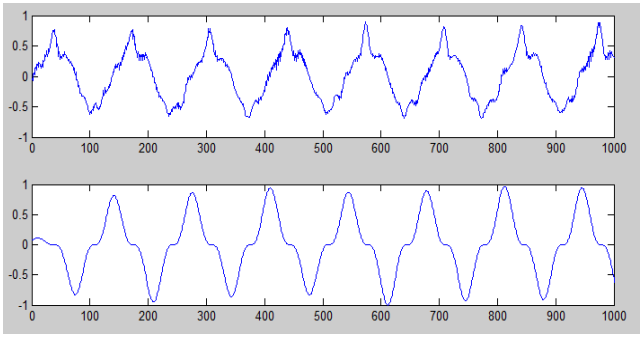

Figure 3: Speech pathological voice corresponding to a sustained vowel /a/ extracted from Edema HLM24AN uttered by a female speaker and its MP.

In this work, we used the wavelet called the quadratic spline function. Interested by the efficiency of the MP for edge detection improvement, the method is applied to pathological signal [20]. The product of three levels of wavelet decomposition is optimal and allows detection of small peaks.

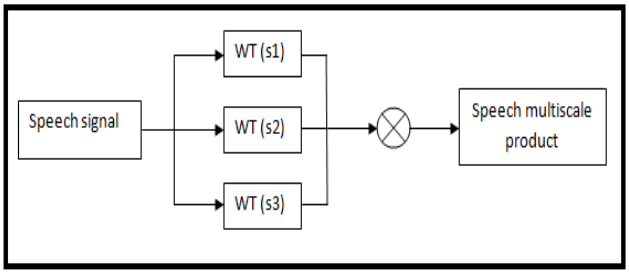

Figure 4: Specch multiscale product extraction

B. Cepstral feature extraction: MFCC calculating

One of the most widely used cepstral representations for speech is obtained by computing the Mel- Frequency Cepstral Coefficients (MFCCs). They are used for instance in speech recognition [22] or synthesis [23]. The MFCCs provides several advantages: human perception is taken into account by considering a perceptive scale of frequencies; the MFCCs are not correlated thanks to the DCT operation. The spectral envelope of the speech frame is summarized into a limited number of coefficients which are computed for each speech frame by weighting the magnitude spectrum by a mel-filterbank, computing the log of each filter output and finally computing the Discrete Cosinus Transform (DCT) of the log-mel-spectrum. The MFCCs are the resulting coefficients of this DCT operation.

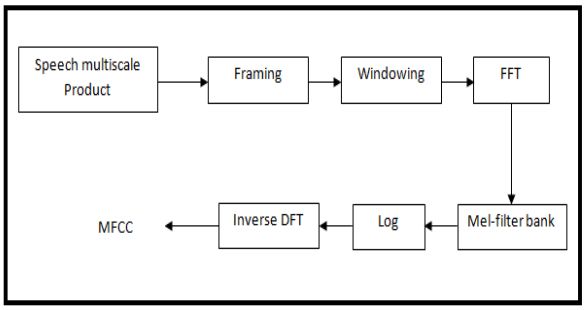

Figure 5: MFCC calculation from the speech multiscale product

Figures 6 and 7 depict respectively the MFCC calculate from the speech signal and its MP.

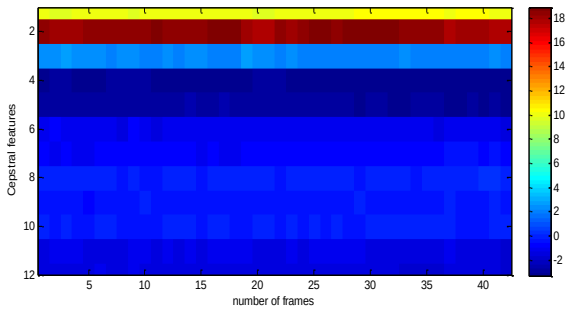

Figure 6: Cepstral features extracted from the speech signal.

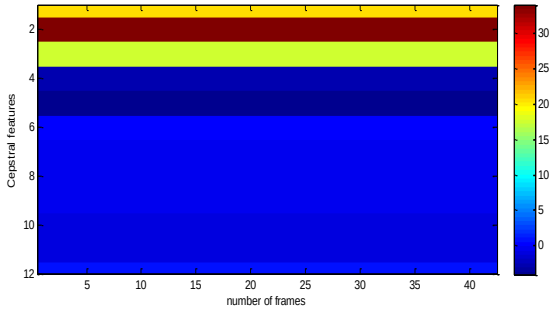

Figure 7: Cepstral features extracted from the speech multiscale product

We can see clearly the difference between the two signals, we notice that the cepstral parameters obtained from the multiscale product are more impressive than those derived from the speech signal and especially the first four parameters.

C. SVM classification

The support vector machine (SVM) approach was introduced by Vapnik in 1995 [24] as an alternative for the classification. Initially planned to solve classification problems with two

classes, today there are generalizations about multi-classes [25].

In order to classify normal and pathological voices, a classification algorithm is applied; in this case, SVM is used to organize the voice signals. SVM is used for both classification and regression; it's a supervised Machine Learning algorithm.

In supervised algorithms, there's use of the training sequences and then the real-time test data set will be processed. Here the final voice signal is maintained into the system as a test data, based on the training set data model; it divides the voice signal as normal voice or pathological one.

EXPERIMENTS AND RESULTS

The results of our work will be presented for MEEI databases. Besides, we operate three classifications: a binary classification to provide a preliminary diagnosis concerning paralysis or hyperfunction, a classification into three classes polyp/edema/nodule and a third classification combining these vocal pathologies: paralysis/ hyperfunction/ polyp/ edema/ nodule. For every classification, we compare the results obtained using features extracted from the MP and those derived from the speech signal directly.

A. MEEI database

The MEEI database [26] contains 53 healthy subjects and 724 subjects with voice disorders. For these two set of subjects, the sustained vowel /a/ and a continuous speech excerpt, "rainbow passage" are available.

Of all subjects with pathologies, only 477 have the information about the pathology. The signals were recorded with a sampling rate of 25 kHz or 50 kHz.

In this work, three sets of pathologies with one disease were defined. A first set was containing subjects diagnosed with neurological diseases: paralysis, a second set comprising subjects diagnosed with functional diseases: hyperfunction and a third set were created with subjects diagnosed with organic diseases: polyp, edema and nodule.

B. Experiment parameters

In this paper, we calculate the cepstral features from the speech multiscale product which is obtained from the product of the speech wavelet transforms at three different scales; the scales (s1 = 3, s2=4, s3= 5) for men and a second (s1'= 2, s2' = 5/2, s3 '= 3) for women.

The cepstral features are computed using the melcepst function provided by the voicebox toolbox [27].

The speech signal is divided into frames of 46.44 ms and with a half recovery, if the sampling frequency is 25 kHz; the window contains 1161 samples with an overlap of 581 samples. Else if it is 50 kHz, the window includes 2322 samples with an overlap of 1161 samples.

C. Classification of neurological and functional diseases

The table 1 and 2 present the classification of neurological (paralysis) and functional (Hyperfunction) diseases, using features extracted from the speech multiscale product and features derived from the speech signal respectively.

TABLE 1: MFCC-MP

Parameters	Paralysis	hyperfunction
MFCC	100	100
MFCC+ Δ	100	100
MFCC+ ΔΔ	100	100
MFCC+ Δ + ΔΔ	100	100
MFCC+*E*+ Δ + ΔΔ	100	100

TABLE 2: MFCC-SPEECH

Parameters	Paralysis	hyperfunction
MFCC	53.57	66.28
MFCC+ Δ	52.38	66.28
MFCC+ ΔΔ	53.57	66.28
MFCC+ Δ + ΔΔ	52.38	66.28
MFCC+*E*+ Δ + ΔΔ	52.38	48.84

The Paralysis/ hyperfunction classification is performed with features extracted from PM using various combinations of parameters.

In table 1 where the features are extracted from the multiscale product, it's concluded that the MFCC coefficients used alone or with other parameters give the best results so that the recognition of paralysis and hyperfunction increases to 100%.

However using features extracted from the speech signal we note that the best results are obtained using the MFCC parameters used alone or with other parameters, expressed by 53.57% for paralysis and 66.28% for hyperfunction.

D. Classification of Organic diseases

Tables 3 and 4 consider the classification of three organic diseases: Polyp, Edema and Nodule, using features extracted

from the speech multiscale product and features derived from the speech signal respectively.

TABLE 3: MFCC-MP

Parameters	Polyp	Edema	Nodule
MFCC	100	68.66	100
MFCC+ Δ	100	68.66	100
MFCC+ ΔΔ	100	68.66	100
MFCC+ Δ + ΔΔ	100	68.66	100
MFCC+E+ Δ + ΔΔ	100	68.66	100

TABLE 4: MFCC-SPEECH

Parameters	Polyp	Edema	Nodule
MFCC	70.24	67.91	98.86
MFCC+ Δ	70.24	67.91	98.86
MFCC+ ΔΔ	70.24	67.91	98.86
MFCC+ Δ + ΔΔ	70.24	67.91	98.86
MFCC+E+ Δ + ΔΔ	85.83	67.16	98.86

The best recognition rates are obtained from the MFCC-MP showing the effecting of the MP on the classification between pathologies.

In fact, we obtain the following rates: 100% for a polyp, 68.66% for edema and 100% for nodule with the MFCC-MP coefficients versus 85.83% for a polyp, 67.16% for edema and 98.86% for nodule using the MFCC-speech coefficients.

E. Classification of organic, neurological and Functional diseases

Tables 5 and 6 show the classification of organic, neurological and functional diseases, using features extracted from the speech multiscale product and features derived from the speech signal respectively.

TABLE 5: MFCC-MP

Parameters	Polyp	Edema	Nodule	Paralysis	Hyper
MFCC	66.67	68.66	100	66.67	100
MFCC+ Δ	66.67	68.66	100	66.67	100
MFCC+ ΔΔ	66.67	68.66	100	66.67	100
MFCC+ Δ + ΔΔ	66.67	68.66	100	66.67	100
MFCC+E+ Δ + ΔΔ	66.67	68.66	100	66.67	100

TABLE 6: MFCC-SPEECH

Parameters	Polyp	Edema	Nodule	Paralysis	Hyper
MFCC	38.89	37.31	74.61	32.54	33.06
MFCC+ Δ	38.89	37.31	75.38	32.54	33.06
MFCC+ ΔΔ	38.89	31.34	82.31	32.54	33.06
MFCC+ Δ + ΔΔ	38.89	37.31	75.38	32.54	33.06
MFCC+E+ Δ + ΔΔ	40.48	35.82	73.08	32.54	33.06

As the number of the classes increases, the accuracy of the classification decreases.

However, the MFCC-MP gives always the best rates comparing with MFCC-speech and remains performant for nodule and hyperfunction diseases.

CONCLUSION

In this work, we present new parameters for the classification of several types of pathologies: organic, neurologic and functional.

The idea consists of calculating the MFCC not from the speech signal but from the multiscale speech product. This approach is sustained by the fact that the MP generates a signal having a shape close to that of the derivative source signal.

The multiscale speech product is the product of the speech wavelet transforms at three different scales.

Three classifications are considered in this work: Classification between paralysis and hyperfunction pathologies, classification between polyp, edema and nodule pathologies and a classification between polyp, edema, nodule, paralysis and hyperfunction pathologies.

The classification is performed by an SVM multiclass system according to one against all approach using the Gaussian kernel and the MEEI database of pathological voices.

For each classification, we compute and compare the rates obtained using features extracted from the speech signal and those derived from the speech MP.

We conclude that the classification rates using MFCC extracted from the speech-MP give the best results.

FUTURE WORKS

Our future work will be about the hierarchical classification of neurologic and organic diseases in the MEEI database and extend this work to other databases.

REFERENCES

[1] J. Hillenbrand and R. A. Houde, "Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech," J. Speech Hearing Res., vol. 39, no. 2, pp. 311–321, 1996.

[2] G. Henningsson, D. P. Kuehn, D. Sell, T. Sweeney, J. E. Trost-Cardamone, and T. L.Whitehill, "Universal parameters for reporting speech outcomes in individuals with cleft palate," Cleft Palate-Craniofacial J., vol. 45, no. 1, pp. 1–17, 2008.

[3] J. Orozco, E. Belalcazar, D. Arias-London, F. Vargas-Bonilla, S. Skodda, J. Rusz, K. Daqrouq, F. Honig, and E. Noth, "Characterization Methods for the Detection of Multiple voice disorders: Neurological, Functional, and Laryngeal diseases", IEEE journal of biomedical and health informatics, vol. 00, n°. 0, 2015.

[4] J. Logemann, H. Fisher, B. Boshes, and E. Blonsky, "Frequency and cooccurrence of vocal tract dysfunctions in the speech of a large sample of parkinson patients," J. Speech Hearing Disorders, vol. 43, pp. 47–57, 1978.

[5] A. K. Ho, R. Jansek, C.Marigliani, J. L. Bradshaw, and S. Gates, "Speech impairment in a large sample of people with Parkinson's disease," Behavioral Neurol., vol. 11, pp. 131–137, 1998.

[6] J.I. Godino-Llorente, P. Gomez Vilda, N. Saenz-Lechon1, M. Blanco-Velasco, F. Cruz-Roldan, and M. Angel Ferrer-Ballester, "Support Vector Machines Applied to the Detection of Voice Disorders", Springer-Verlag, Berlin Heidelberg, 2005. pp.219-230.

[7] J. Iagnacio Godino-Llorente , Member, IEEE, P. Gomez Vilda , Member, IEEE, M. Blanco-Velasco, Member, IEEE, "Dimensionality Reduction of a Pathological Voice Quality Assesment System Based on Gaussian Mixture Models and Short-Term Cepstral Parameters", In: IEEE 2006 Transactions on Biomedical Engineering, October 2006; 5.

[8] A.A. Dibazar, T.W. Berger, and S.S. Narayanan, "Pathological Voice Assesment", In: IEEE 2006 EMBS 2006; New York.

[9] C. Bishop, "Pattern Recognition and Machine Learning", Springer, New York, NY, USA, 2006.

[10] J.I. Godino-Llorente, P. Gomez-Vilda, N. Saenz-Lechon, M. Blanco-Velasco, F. Cruz Roldan, and M.A. Ferrer, "Discriminative methods for the detection of voice disorders", In: NOLISP 2005 International Conference on Non-Linear Speech Processing, April 2005; Barcelona, Spain.

[11] P. Kukharchik, I. Kheidorov, E. Bovbel, and D. Ladeev, Image and signal processing, In: Speech Signal Processing Based on Wavelets and SVM for Vocal Tract Pathology Detection, Lecture Notes in Computer Science Springer; Berlin, Germany, pp. 192–199, 2008.

[12] A. Belhaj, A. Bouzid, N. Ellouze, "Normal and Organic Pathology Classification of Female Voices Using SVM Classifier", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 15 (2016) pp 8702-8707.

[13] J. Iagnacio Godino-Llorente, Member, IEEE, P. Gomez Vilda, Member, IEEE, M. Blanco-Velasco, Member, IEEE, "Dimensionality Reduction of a Pathological Voice Quality Assesment System Based on Gaussian Mixture Models and Short-Term Cepstral Parameters", IEEE 2006 Transactions on Biomedical Engineering, October 2006; 5.

[14] M. Marakaki, Y. Stylianou, " Using modulation spectra for voice pathology detection and classification", In: Proc. IEEE EMBC 2009, Minneapolis, pp. 2514-2517, 2009.

[15] Key Elemetrics, Elemetrics Disordered Voice Database, 1994.

[16] M. Marakaki, Y. Stylianou,"Voice Pathology Detection and Discrimination Based on Modulation Spectral Features IEEE Transactions on Audio, Speech, and Language Processing 19, 1938–1948, 2011.

[17] S. Chekili, A. Belhaj, A. Bouzid, N.Ellouze," Recognition of pathological voices", Multi-Conference on Systems, Signals & Devices (SSD), 2014, 11th International, Barcelona.

[18] B.M. Sadler, T. Pham, L.C. Sadler," Optimal and wavelet-based shock wave detection and estimation", J. Acoust. Soc. Am., 2, pp. 955-963, 1998.

[19] B.M. Sadler, A. Swami, "Analysis of multiscale products for step detection and estimation", IEEE Trans. Inform. Theory, 45, 3, pp. 1043-1051, 1999.

[20] L. Zhang, P. Bao, "Edge detection by scale multiplication in wavelet domain", Pattern Recognition Letters, 27,9, pp. 1485-1490, 2005.

[21] Y. Xu, J.B. Weaver, D.M. Hea, J. Lu, "Wavelet transform domain filters: a spatially selective noise filtration technique", IEEE Trans. Image processing, 3,6, pp. 747-758, 1994.

[22] R. W. Schafer. Springer handbook of speech processing, chapter Homomorphic Systems and Cepstrum Analysis of Speech, pages 161–180. Springer, 2008. 103.

[23] K. Tokuda, H. Zen and A. W. Black. An HMM-Based Speech Synthesis System Applied To English. In Proceedings of 2002 IEEE Workshop on Speech Synthesis, pages 227–230,2002. 103.

[24] Key Elemetrics, Elemetrics Disordered Voice Database, 1994.

[25] V. VAPNIK,"The nature of statistical Learning Theory

springer, 1995.

[26] A. Bouzid, N. Ellouze, "Voice Source Parameter Measurement Based on Multi-Scale Analysis of Electroglottographic Signal, Speech Communication, vol. 51, pages 782-792, 2009.

[27] http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html