Low-frequency components analysis in running speech for the automatic detection of Parkinson's disease

T.Villa-Cañas[1], *J.D. Arias-Londoño*[1], J.R. Orozco-Arroyave[1,2], *J.F. Vargas-Bonilla*[1], *Elmar Nöth*[2]

[1]Faculty of engineering, Universidad de Antioquia UdeA, Calle 70 No. 52-21, Medellín, Colombia.
[2]Friedrich Alexander Universität, Erlangen-Nürnberg, Germany

tatiana.villa@udea.edu.co

Abstract

This paper explores the analysis of low-frequency components of continuous speech signals from people with Parkinson's disease, in order to detect changes in the spectrum that could be associated to the presence of tremor in the speech. Different time-frequency (TF) techniques are used for the characterization of the low frequency content of the speech signals, by paying special attention on the ability to work in non-stationary frameworks, due to the need for the analysis of long enough time segments, where the assumptions of stationary can not be met. The set of variables extracted from the TF representations includes centroids and the energy content of different frequency bands, along with entropy measures and nonlinear energy operators, which are used as features for the automatic detection of people with Parkinson's disease vs healthy controls. The discrimination capability of the estimated features is evaluated using three different classification strategies: GMM, GMM-UBM, and SVM. Furthermore, the information provided by different TF techniques is combined using a second classification stage. The results show that the changes in the low frequency components are able to discriminate between people with Parkinson's and healthy speakers with an accuracy of 77%, using one single sentence.

Index Terms: Parkinson's disease, time-frequency analysis, continuous speech characterization, pathological voice detection.

1. Introduction

Parkinson's disease (PD) is a neurological disorder characterized by the progressive loss of dopaminergic neurons in the substantia nigra of the midbrain [1], and affects different functions performed in the basal ganglia, including the control of voluntary movements, procedural learning, and movements of the jaw and eyes [2]. As consequence, the voice of patients with PD is affected in many ways: reduced or monoloudness, hoarse or strangled phonation, monopitch intonation and variable articulation rate. In this sense, recent studies [3], [4], [5] suggest that PD can affect different subsystems of speech production including respiration, phonation, articulation, and prosody. Moreover, due to the deficits in motor speech, one of the commons effects for all of the subsystems of speech production, is the introduction of low frequency components known as tremor [6, 7], which, if detected in an early stage of the disease, which could turn into a relevant bio-marker for diagnosis and treatment assessment.

Several works in the state of art have been focused on the automatic detection of voice pathologies based on acoustic measures extracted from pronunciation of sustained vowels [8], [9]. However, if the speech signal loses part of its quasi-periodic behaviour due to the presence of pathology, and such affectation of the voice involves not only the phonatory processes, but also the dimensions of articulation and prosody (as in the case of PD), the analysis of speech signals must include additional exercises as word pronunciation, phrases and/or monologues. This is because sustained vowels alone are not suitable for assessing voice quality and communication skills [10], and also because they do not incorporate dynamic aspects of continuous speech (e.g. co-articulations, onset and offset effects etc.) [11].

In general, techniques based on the analysis of continuous speech samples rely on some segmentation procedure to identify the voiced, unvoiced, and silence periods. This is due to the fact that measures that quantify periodicity and regularity (e.g. Harmonic-to-Noise ratio, cepstral peak prominence, and pitch amplitude) are valid only in the voiced regions [12]. Additionally several acoustic measures are based on methods that make assumptions of signal stationarity, which are not fully satisfied in continuous speech signals due to, for instances, the variations of pitch periods, rhythm, intonation and other suprasegmental features [13]. Moreover, the mixture of voiced, unvoiced, and silent periods itself leads to non-stationary conditions [12].

Several works in the literature have addressed the problem of detection of PD using different kind of features and dimensions (phonation, articulation and prosody) [4], [5], [14], reporting successfully results; however, most of the works that analyzed continuous speech used voiced/unvoiced segmentation, which for the fact of resolution in time, prevent the capture of low frequency changes in the speech signals that for people with PD, could be associate to tremor [15].

A different way to extract features from continuous speech signals is to use a non-stationary technique directly, thereby avoiding the need for any segmentation procedures and allowing the analysis of longer frames making more feasible to detect low-frequency changes. In this sense, in recent years, several time-frequency (TF) techniques of analysis have been applied successfully in the context of speech. Multi-resolution analysis based on wavelet theory is one of the most widespread methods. Wavelet transform (WT) is particularly well suited for TF analysis and for the characterization of singularities in non-stationary signals, and has been effectively used in developing measures for the screening of pathologies in speech [16]. In [17] it was proposed a voice disorders identification algorithm by using the energy of coefficients obtained from a discrete wavelet transform (DWT) to feed a Neural Network classifier. The authors reported classification rates about 90%. A more recent work presented in [18], reports good results (around 100%) in the context of pathological voice quality assessment, using a characterization based on WT and a Support Vector Machine (SVM) as pattern recognizer. Another kind of TF techniques used in the context of speech include modulation and Wigner-Ville-based spectra. In [19], a recent approach used the spectral modulation to detect changes in the acoustic spectrum of pathological voices yielding to detection rates around 94.08%. On the other hand, Wigner-Ville distribution (WVD) is a valuable and effective method for the analysis of non-stationary signals. It allows a better simplicity and characterization of the time-dependent spectra in comparison to STFT, and it has also been successfully used in the context of speech [20].

One of the drawbacks of the TF analyses is that the spectra generated by the different techniques contain a large amount of data, that prevent their direct usage in pattern recognition systems, and therefore in most of the cases, a dimensionality reduction or sub-characterization stages is required. In this sense, this work analyzes the low-frequency components of continuous speech signals uttered by PD patients and healthy controls (HC), using three different TF techniques, in order to determine whether characteristics associated with low-frequency components of the speech signals can be used to detect the presence of PD. The spectra are in turn characterized based on subband energy analysis and spectral subband centroids [21], [22] and the automatic detection is performed using Gaussian Mixture Models (GMM), GMM-Universal Background Model (GMM-UBM) and SVM based detectors. Furthermore, the information provided by the different TF techniques is combined using a second classification stage as proposed in [23].

The paper is organized as follows: section 2 presents the proposed methodology and a brief description of the methods applied in this work; section 3 provides details about the experimental framework and section 4 shows the results obtained. Finally, in section 5 some conclusions arisen from the results are presented.

1.1. Methodology

Figure 1 shows a schematic of the methodology used for the classification system. Three different TF techniques of analysis are used: Wavelet Packet Transform (WPT), modulation spectra (MS) and WVD. The speech signals were computed on a

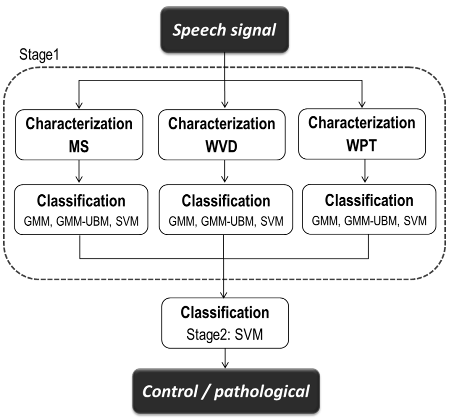

Figure 1: General diagram of the proposed system.

frame-by-frame basis using $262\ ms$ windows shifted by $64ms$ as proposed in [19]. The spectra obtained by MS and WVD were characterized with centroids and the energy content of different frequency bands. In the case of WPT, the Teager-Kaiser energy operator (TKEO) and entropy measures were computed in decomposition bands [24]. Besides, due to the computational cost of some the TF techniques, the speech signals were downsampled to $11025Hz$. The performance of the proposed system was evaluated in two stages: first, the features extracted from TF techniques were used separately to feed pattern classifiers based on GMM, GMM-UBM and SVM. After the first stage, the outputs of the best classifiers were merged together using a SVM classifier to take the final decision. In the following, every stage of the process will be briefly described.

1.2. Wavelet packet transform (WPT)

The WPT is a generalization of DWT that provides accurately detailed information across time and frequency domain of a speech signal. WPT structure gives ability to advance on the wavelet packet tree nodes where each subband is divided into two smaller subbands with equal frequency range [18]. WT-based methods have advantages over traditional Fourier transforms for representing functions that have discontinuities and sharp peaks, and for accurately deconstructing and reconstructing finite, non-periodic or non-stationary signals [18]; hence WT has the advantage of using a variable length window for different frequency components. In this work, the Daubechies10 wavelet was used, and all the coefficients from second to fifth level of decomposition were considered for the characterization. The set of features includes the following measures [24]:

- *The normalized log energy:* this measure is calculated according to (1):

$$E(k) = log \left[\frac{\sum_{n=1}^{N_k} [d(k,n)]^2}{N_k} \right] \quad (1)$$

where $d(k, n)$ is the coefficient of the k-th level of decomposition in the n-th sample, and N_k is the number of coefficients [24].

- *The normalized log energy of TEO:* It is a nonlinear operator used for taking advantage of the existence of multiple components on a signal $\boldsymbol{x} = \{x(1), x(2), ...\}$, which in this case corresponds to the sequence of coefficients obtained from each of the k decomposition levels [24]. TEO is used here to model the changes in the speech signal due to non-linear effects in vocal tract and it is defined as:

$$TEO(\boldsymbol{x}, n) = x(n)x^*(n) - x(n+1)x^*(n-1) \quad (2)$$

where $(*)$ corresponds to complex conjugate. For a whole signal segment, the logarithm of average TEO is used to obtain a better resolution of the characteristic; it is computed as:

$$E_{TEO}(k) = log \left[\frac{\sum_{n=1}^{N_k} |TEO(\boldsymbol{x}, n)|}{N_k} \right] \quad (3)$$

- *Shannon and log energy entropy:* Entropy is a measure of the unpredictability of a random variable. In this work, two entropy measures were computed on each of the decomposition levels: the non-normalized Shannon entropy and the "log energy" entropy described by:

$$SE(k) = -\sum_{j=1}^{b} [|p(j)|^2 \cdot log \left|p(j)^2\right|] \quad (4)$$

$$LEE(k) = -\sum_{j=1}^{b} log\left|p(j)^2\right| \quad (5)$$

where b is the number of bins used for the definition of the probability mass function, and $p(j)$ is the probability that one of the coefficients falls into j-th bin.

1.3. Modulation Spectra (MS)

Modulation spectra may be seen as a non-parametric way to represent the modulation present in the speech which is introduced by the presence of pathologies [19]. The most common modulation frequency analysis for a discrete signal, initially employs a short-time Fourier transform, and subsequently, the magnitude of the spectrum is used for a second frequency analysis with another Fourier transform [19]. Figure 2 shows a representation of MS obtained from a normal and a pathological speech signal (uttered by a patient with PD). It is possible to observe that for

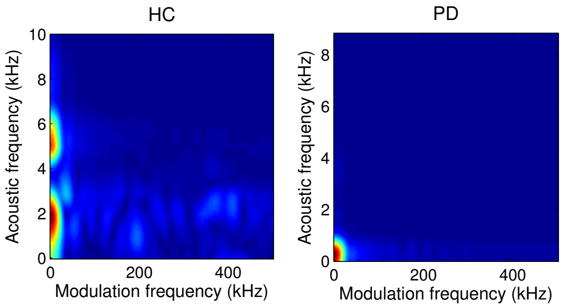

Figure 2: Spectrogram of MS for normal and pathological voice.

pathological speech signals, most of the energy at the modulations corresponding to fundamental frequency and its harmonics, is localized in the lower acoustic frequencies.

1.4. Wigner-Ville distribution (WVD)

The Wigner Distribution is a popular method for TF analysis of mono-component signals. Several derived applications of this method have been proposed in different domains [25]. WVD provides a high resolution in both TF planes, but only for mono-component signals. In multi-component cases the technique is not suitable due to the cross-term artifacts of the aliasing. To overcome this problem, a widely accepted window function such as the Hamming, can be applied to the WVD to smooth the cross-terms (smooth windowed WVD or smoothed pseudo WVD (SPWVD)). For the sake of comparison, figure 3 shows a representation of WVD obtained from a normal and a pathological speech signal (PD).

1.5. Characterization of dynamic features

Unlike WPT, MS and WVD spectra were characterized with spectral centroids and the energy content at different frequency bands. These features are considered effective methods of combining the frequency and magnitude information from the power spectrum [22]. The centroid of the power spectrum $X(f)$ with frequency f can be estimated by:

$$C_v = \frac{\int_0^{F_s/2} f H_v(f) X^{\gamma}(f) df}{\int_0^{F_s/2} H_v(f) X^{\gamma}(f) df} \quad (6)$$

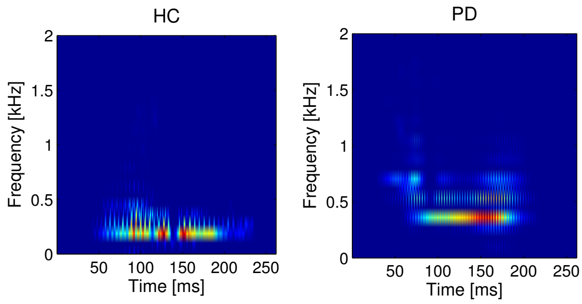

Figure 3: Spectrogram of WVD for normal and pathological voice.

where F_s is signal sampling frequency, $H_v(f)$ is the filter centered at the subband v, and γ is a parameter that decides the dynamic range of the spectrum used in the computation of the centroid. The expression (6) can be implement in the discrete time n by

$$C_v\left[n\right] = \frac{\sum_{w=1}^{W} w H_v[w] \left|X^{\gamma}\left[n,w\right]\right|^2}{\sum_{w=1}^{W} H_v[w] |X^{\gamma}\left[n,w\right]|^2} \quad (7)$$

where $H_v[w]$ is the discrete version of the filter bank $H_v(f)$. The energy around each centroid with fixed bandwidth Δw is computed by means of:

$$\hat{E}_v[n] = \sum_{W=C_v[n]-\Delta w}^{C_v[n]+\Delta w} \left|X\left[n,w\right]\right|^2 \quad (8)$$

1.6. Classification

The classification was performed individually per each TF technique using three different standard classifiers in the context of speech: GMM along with the GMM-UBM strategy [26], and SVM. The information provided by the different TF techniques was combined using a second classification stage where the outputs of the best classifiers in the first stage were used as inputs for the second stage [23]. In all the cases, the decision was taken by averaging the outputs of the classifiers per each frame of a speech recording, in order to have a global decision per speaker.

2. Experimental framework

2.1. Corpus of Speakers

The database used in this work is fully described in [27]. It includes several speech exercises such as sustained vowels, isolated words, several sentences and monologues. The voice samples were uttered by 50 PD patients and their age and gender-matched healthy controls (HC). The recordings were captured in noise controlled conditions, in a sound proof booth that was built at the Cínica Noel, in Medellín, Colombia; the registers were sampled at 44.1 KHz with a resolution of 16 bits. The participants are Colombian Spanish native speakers and all of the patients were diagnosed by neurologist experts and none of people in the HC group has history of symptoms related to PD or any other kind of movement disorder syndrome. The mean values of their neurological evaluation according to the UPDRS-III and Hoehn & Yahr scales [28] are 36.7 ± 18.7 and 2.29 ± 0.8, respectively. For this work, the Spanish sentence: "Los libros nuevos no caben en la mesa de la oficina" was used for the experiments.

2.2. Experimental setup

The validation was carried out using a cross-validation strategy with 10 folds. Training and testing samples were chosen randomly but maintaining the balance in age and gender. As pointed out before, three methods of classification were used during the experiments. For the GMM and GMM-UBM strategies, there were evaluated different numbers of Gaussian components (G) (from 2 to 6), with diagonal covariance matrix. For the SVM, the regularization parameter C and RBF kernel parameter γ were adjusted. The grid values for C included [0.1 1 10 100 1000] and for γ included [0.0001 0.001 0.01 0.1 1 10 100]. The performance of the system was measured in terms of accuracy (Acc), sensitivity (Sens) and specificity (Spec). Also the ROC curves and Area under the curve (AUC) are provided.

3. Results and discussions

Table 1 shows the results of the first stage of classification obtained with the three different TF techniques, where each feature set is considered separately. The accuracy obtained from the three methods is very similar, being WPT the best technique yielding to 73% of recognition rate. It is worth to note that while WPT provides the best accuracy and specificity, MS and WVD achieved better sensitivity than WPT, which is very valuable in the medical context. This result suggests that the information provided by the three TF techniques could be complementary for the evaluation of PD patients, therefore in the following the systems with the best performance per each of the TF techniques, were merged together into one system. Table 2

Table 1: Results obtained for the TF techniques in stage 1

Features	classifier	% Acc	% Sens	% Spec	Parameters
	GMM	64.0±16.3	58.0±22.0	60.1±24.4	G=2
MS	GMM-UBM	60.0±10.9	64.0±18.4	63.5±18.4	G=2
	SVM	**68.0±10.7**	86.0±16.4	50.0±17.0	C=10, γ=0.1
	GMM	70.0±16.1	62.0±28.9	71.3±20.2	G=4
WVD	GMM-UBM	71.0±16.6	62.0±31.9	72.1±18.5	G=6
	SVM	**71.0±10.4**	82.0±14.7	60.0±21.1	C=0.1, γ=0.1
	GMM	63.0±14.2	70.0±19.4	60.5±25.8	G=3
WPT	GMM-UBM	63.0±14.1	68.0±21.5	64.0±16.5	G=3
	SVM	**73.0±9.0**	72.0±23.5	74.0±21.2	C=100, γ=0.001

shows the best results of the first stage of classification along with the combination stage. In this case, the overall performance of system was better than any of the former individual evaluation, achieving a classification rate of 77% and reducing the confidence interval. The specificity of the whole system was also better in comparison with those reported in Table 1. Moreover, the sensitivity and specificity of the whole system show quite balance. On the other hand, and with the aim of present the results in a more compact way, the ROC curves for each of the results in Table 2 with their corresponding AUC are shown in figure 4. It is worth to note that, according to the ROC curves, WPT does not present better accuracies than MS and WVD. In this case, WVD is the technique with better expected performance confirming the need for the fusion of the information provided by the different TF representation. As expected, the best AUC was obtained with the fusion of MS, WVD and WPT.

Table 2: Results obtained with the best classifier in stage 1 and the fusion using a SVM classifier

Feature	% Acc	% Sens	% Spec	Parameters
MS	68.0±10.7	86.0±16.4	50.0±17.0	C=10. γ=0.1
WVD	71.0±10.4	82.0±14.7	60.0±21.1	C=0.1.γ=0.1
WPT	73.0±9.0	72.0±23.5	74.0±21.2	C=100. γ=0.001
Fusion	**77.0±6.4**	78.0±14.8	76.0± 22.7	C=1000. γ=0.0001

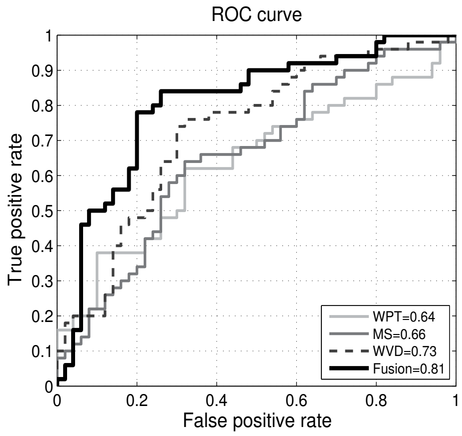

Figure 4: ROC curves for the different methods of TF analysis with their corresponding AUC.

4. Conclusions

This paper undertook different approaches to the analysis of low-frequency components of continuous speech signals by using three different TF techniques. The results shown that, for the database used, the changes in the energy components can differentiate between normal and Parkinsonian speech signals, with an accuracy around to 77%, using a single sentence. The MS analysis shown that, unlike normal voices, the energy components of speech from people with PD are mainly located at very low acoustic and modulation frequencies.

For all the TF techniques used, the best performance was obtained when the classification was carried out using SVM. Furthermore, by taking into account the second classification stage, an absolute improvement of 4% in the accuracy rate, along with a lower variance was reached. This fact suggest that the information provided by the different techniques is complementary and also that more reliable predictions can be made when the fusion of classifiers is considered.

The use of TF techniques allows to find low-frequency components of continuous speech that could be associated to an indicator of the presence of PD. Further additional analyses should include other sentences with different phonetic and prosodic content, in order to get a better characterization of the low-frequency components under different efforts. Also, the combination of the features derived from TF analysis must be combined with other kind of characterizations, in order to establish the real contribution of these kind of techniques as part of a multi-task system for the detection and assessment of PD patients.

5. Acknowledgements

Juan Rafael Orozco Arroyave is under grants of Convocatoria 528 para estudios de doctorado en Colombia, generación del bicentenario, 2011 financed by COLCIENCIAS. This work was granted by COLCIENCIAS, project N111556933858.

6. References

[1] O. Hornykiewicz, "Biochemical aspects of parkinsons disease," *Neurology*, vol. 51, pp. S2–S9, 1998.

[2] A. Alm, "Stuttering and the basal ganglia circuits: a critical review of possible relations," *Journal of Communication Disorders*, vol. 37, no. 4, pp. 325–369, 2004.

[3] J. Rusz, R. Cmejla, H. Ruzickova, J. Klempir, V. Majerova, J. Picmausova, J. Roth, and E. Ruzicka, "Acoustic analysis of voice and speech characteristics in early untreated parkinson's disease." in *MAVEBA*, 2011, pp. 181–184.

[4] T. Khan, J. Westin, and M. Dougherty, "Classification of speech intelligibility in parkinson's disease," *Biocybernetics and Biomedical Engineering*, vol. 34, no. 1, pp. 35–45, 2014.

[5] J. Rusz, R. Cmejla, H. Ruzickova, and E. Ruzicka, "Objectification of dysarthria in parkinsons disease using bayes theorem," *Age (year)*, vol. 61, no. 12.60, pp. 58–08, 2011.

[6] A. L. Merati, Y. D. Heman-Ackah, M. Abaza, K. W. Altman, L. Sulica, and S. Belamowicz, "Common movement disorders affecting the larynx: a report from the neurolaryngology committee of the aao-hns," *Otolaryngology-Head and Neck Surgery*, vol. 133, no. 5, pp. 654–665, 2005.

[7] K. Perez, L. Ramig, M. Smith, and C. Dromey, "The parkinson larynx: tremor and videostroboscopic findings," *Journal of Voice*, vol. 10, no. 4, pp. 354–361, 1996.

[8] J. B. Alonso, F. Díaz-de María, C. M. Travieso, and M. A. Ferrer, "Using nonlinear features for voice disorder detection," in *ISCA Tutorial and Research Workshop (ITRW) on Non-Linear Speech Processing*, 2005.

[9] N. Sáenz-Lechón, J. Godino-Llorente, V. Osma-Ruiz, and P. Gómez-Vilda, "Methodological issues in the development of automatic systems for voice pathology detection," *Biomedical Signal Processing and Control*, vol. 1, pp. 120–128, 2006.

[10] J. Godino-Llorente, R. Fraile, N. Sáenz-Lechón, V. Osma-Ruiz, and P. Gómez-Vilda, "Automatic detection of voice impairments from text-dependent running speech," *Biomedical Signal Processing and Control*, vol. 4, no. 3, pp. 176–182, 2009.

[11] B. Hammarberg, B. Fritzell, J. Gaufin, J. Sundberg, and L. Wedin, "Perceptual and acoustic correlates of abnormal voice qualities," *Acta oto-laryngologica*, vol. 90, no. 1-6, pp. 441–451, 1980.

[12] K. Umapathy, S. Krishnan, V. Parsa, and D. G. Jamieson, "Discrimination of pathological voices using a time-frequency approach," *Biomedical Engineering, IEEE Transactions on*, vol. 52, no. 3, pp. 421–430, 2005.

[13] A. Fourcin and E. Abberton, "Hearing and phonetic criteria in voice measurement: Clinical applications," *Logopedics Phonatrics Vocology*, vol. 33, no. 1, pp. 35–48, 2008.

[14] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig, "Accurate telemonitoring of parkinson's disease progression by noninvasive speech tests," *Biomedical Engineering, IEEE Transactions on*, vol. 57, no. 4, pp. 884–893, 2010.

[15] G. Deuschl, J. Raethjen, R. Baron, M. Lindemann, H. Wilms, and P. Krack, "The pathophysiology of parkinsonian tremor: a review," *Journal of neurology*, vol. 247, no. 5, pp. V33–V48, 2000.

[16] M. A. Little, P. E. McSharry, S. J. Roberts, D. A. Costello, I. M. Moroz *et al.*, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," *BioMedical Engineering OnLine*, vol. 6, no. 1, p. 23, 2007.

[17] J. Nayak and P. Bhat, "Identification of voice disorders using speech samples," in *TENCON 2003. Conference on Convergent Technologies for the Asia-Pacific Region*, vol. 3, Oct 2003, pp. 951–953 Vol.3.

[18] M. K. Arjmandi and M. Pooyan, "An optimum algorithm in pathological voice quality assessment using wavelet-packet-based features, linear discriminant analysis and support vector machine," *Biomedical Signal Processing and Control*, vol. 7, no. 1, pp. 3–19, 2012.

[19] M. Markaki and Y. Stylianou, "Using modulation spectra for voice pathology detection and classification," in *Proceedings of IEEE EMBC09,Minneapolis, MN*, 2009, p. 25142517.

[20] B. Boashash and P. Black, "An efficient real-time implementation of the wigner-ville distribution," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 35, no. 11, pp. 1611–1618, 1987.

[21] B. Gajic and K. Paliwal, "Robust feature extraction using subband spectral centroid histograms," in *In Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'01)*, 2001, pp. 85–88.

[22] K. Paliwal, "Spectral subband centroids as features for speech recognition," in *In Proceedings IEEE Workshop on Automatic Speech Recognition and Understanding*, 1997, pp. 124–131.

[23] J. Arias-Londoño, J. Godino-Llorente, M. Markaki, and Y. Stylianou, "On combining information from modulation spectra and mel-frequency cepstral coefficients for automatic detection of pathological voices," *Logopedics Phoniatrics Vocology*, vol. 36, no. 2, pp. 60–69, 2011.

[24] A. B. Kandali, A. Routray, and T. K. Basu, "Vocal emotion recognition in five native languages of assam using new wavelet features," *International Journal of Speech Technology*, vol. 12, no. 1, pp. 1–13, 2009.

[25] G. Matz and F. Hlawatsch, "Wigner distributions (nearly) everywhere: time–frequency analysis of signals, systems, random processes, signal spaces, and frames," *Signal Processing*, vol. 83, no. 7, pp. 1355–1378, 2003.

[26] D. Reynolds, T. Quatieri, and R. Dunn, "Speaker verification using adapted Gaussian Mixture Models," *Digital Signal Processing*, vol. 10, no. 1-3, pp. 19–41, 2000.

[27] J. Orozco-Arroyave, J. Arias-Londoño, J. Vargas-Bonilla, M. González-Rátiva, and E. Nöth, "New spanish speech corpus database for the analysis of people suffering from parkinsons disease," in *Proceedings of the 9th Language Resources and Evaluation Conference (LREC)*, 2014.

[28] C. G. Goetz, W. Poewe, O. Rascol, C. Sampaio, G. T. Stebbins, C. Counsell, N. Giladi, R. G. Holloway, C. G. Moore, G. K. Wenning *et al.*, "Movement disorder society task force report on the hoehn and yahr staging scale: status and recommendations the movement disorder society task force on rating scales for parkinson's disease," *Movement disorders*, vol. 19, no. 9, pp. 1020–1028, 2004.