Acoustic Studies of Tremor in Pathological Voices

Eduardo Castillo-Guerra[1], Mohsen Amiri Farahani[1], Carlos A. Ferrer[2]

[1]Department of Electrical and Computer Engineering, University of New Brunswick
[2]Faculty of Electrical Engineering, Central University of Las Villas

Abstract

This work is focused on modeling the perception of tremor found in pathological voices. The main research objective is to automatically separate the different sources of tremor to estimate the magnitude of tremor perturbations using signal processing techniques. A new assessment algorithm is derived from speech recordings which combines non-linear filtering, amplitude demodulation and spectral estimation techniques. The algorithm is evaluated against the perceptual judgments provided by speech pathologists and other reported indexes. The results show that the algorithm is effective differentiating normal from pathological tremor and it is a reliable measurement of tremor perturbations with high correlation with perceptual judgments.
Index Terms: Tremor, pathological speech, dysarthria.

1. Introduction

Vocal tremor can be a symptom of neurological disorders. It has been studied with respect to various neurological conditions such as organic voice tremor (OVT) [14], spasmodic dysphonia [18], Parkinson disease (HP) [2], amyotrophic lateral sclerosis (ALS) [19] and cerebellar ataxia [7]. Darley, Aronson and Brown [7] also reported a study analyzing the relevance of tremor for eight types of dysarthria including flaccid (FD), spastic (SD), ataxic (AD), Parkinson disease (HP), OVT, chorea (HC), dystonia (HD) and ALS. They found tremor as a relevant feature for OVT and other dysarthria types such as HP, ALS and ataxic dysarthria (AD).
Voice tremor is defined as a low-frequency fluctuation in amplitude or frequency (or both) [20] caused by oscillating movements of components of the speech mechanism. Its origin is generally neurological and its frequency ranges between 1-18 Hz. The frequency and intensity of the tremor varies with the type of disorder, providing valuable information for the assessment of this type of acoustic perturbation.

There are various algorithms reported to instrumentally assess tremor perturbations. Some of the most reliable measures rely on amplitude and frequency modulation indexes that are measured in sustained vowel phonations (SVP) [14][18][2][19][7]. For instance, the frequency of the oscillations in OVT has been reported in the range of 4-8 Hz, with amplitude modulation indexes above 40% [1][13].

A spectral analysis of the demodulated speech signal was investigated in [11], where it was found that the frequency of the most prominent spectral peaks of the radiated speech signal failed to differentiate pathological tremor from normal oscillation encountered in control subjects. Only the magnitude of the 6 most prominent peaks correlated well with more severe cases of tremor (90% correlation) but did not correlate so well for control utterances (40%). This work observed that the frequency of the oscillations of normal and pathologic tremor overlaps, with mainly the magnitude of the tremor (and not its frequency) influencing in its perception.

Buder and Strand [4] proposed a combined visual analysis (called "modulogram") of the amplitude and frequency modulations (AM and FM) present in sustained vowels. They divided the frequency ranges for both, frequency and amplitude modulations, in three bands: namely "wow" (0-2Hz) "tremor" (2-10 Hz) and "flutter" (10-20 Hz). It is interesting to note that frequency spectrums of AM and FM shown as examples portray different distributions, indicating that there might be different sources of oscillations in the resulting radiated waveform. Vocal folds and vocal tract tremors can have different intrinsic oscillation patterns, and can be non-linearly combined in the radiated [16]. The simultaneous existence of different sources of tremor has been reported in [21] and the potential benefits of separated analysis.

Differences between the AM and FM demodulated signals are also shown in [8], where variable phase differences for low, normal and high pitch are reported. In the same study, it is suggested that valuable insights about the underlying mechanisms of vocal tremor can be gained by distinguishing the AM at the sound source from the AM caused by the oscillation in the resonators (vocal tract).

In this paper, a separation of both sources of oscillations is performed by means of inverse filtering techniques, under the assumption that a separate analysis can yield a better correspondence with perceptual ratings. The research is focused on modeling the perception of tremor perturbation, searching for an index that is both highly correlated with human perception and objective. An automatic algorithm to assess tremor perturbations is proposed to aid in performing acoustic evaluations of pathological speech. The performance is assessed with respect to the perception of three speech and language pathologists with experience assessing voice disorders and it is compared to other reported algorithms.

2. Materials and methods

2.1. Database

The data used in this investigation contained SVP of the vowel /a/ from 127 subjects from two databases [5][12] with similar disorder distributions to the one reported by Darley, Aronson and Brown [7]. Recordings from a total of 19 healthy subjects and 108 patients were studied[1]. Perceptual judgments were obtained from three judges as described in [6] using a lineal scale between 0 and 6 with an anchored judgment system.
The frequency band of interest to this work was from 2-10Hz to include the overall reported tremor range [4] reported for OVT, HP, ALS and AD (starting at 3 Hz, ending at 8 Hz [2][19][1]).

[1] The onset and offset of all recordings were automatically removed to analyzed the steady segments where amplitude and pitch are expected to be constant or monotonically decreasing with a very smooth slope

Figure 1 shows a plot of intensity and pitch contour of a patient with OVT, selected from the database. This plot shows that tremor is present in both amplitude and pitch signals, suggesting multiple sources.

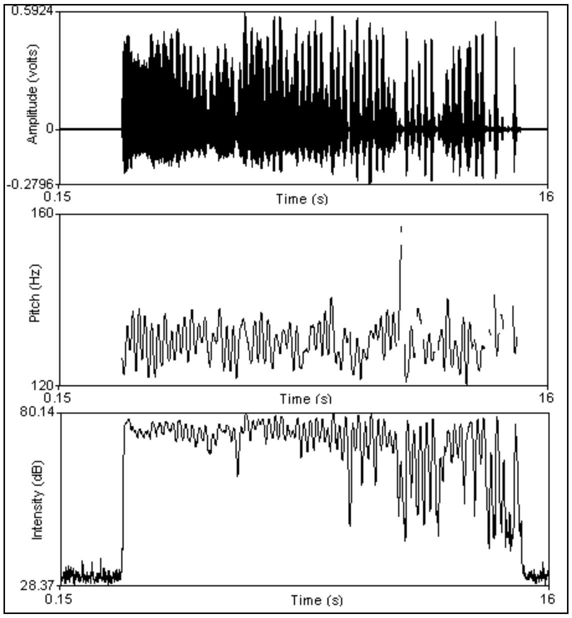

Figure. 1 Top: Sample waveform of a severe OVT subject (HO66F059A). Middle: Pitch contour. Bottom: Intensity envelope.

2.2. Tremor index

The radiated speech is, according to the source filter theory of speech production [9], the convolution of the signal generated in the glottis and the transfer function of the vocal tract. Thus, tremor in both the glottal area and the vocal tract contributes to the perceived tremor in the acoustic signal. The algorithm proposed in this work separates these different sources of tremor and combines them to determine a global tremor index based on AM parameters (Tr_{AM}). A list of the symbols and notations used in this paper to describe the calculation of the index in the following sections is summarized in Table 1.

Table 1. *Summary of notations used in the paper.*

Notation	Description
D	framed acoustic signal data matrix
k	number of frames
n	number of samples per frame
A	vocal tract filter matrix
H_z	impulse response of vocal tract filters
R	residual signal (glottal source estimation)
K	kurtosis
$\downarrow x$	downsample operation at rate x
LPF_x	low-pass filter at frequency x
PSD_x	power spectral density of signal x
F_0	pitch frequency
S_{AMD}	amplitude demodulation of speech signal
S_{AMA}	amplitude demodulation of filter response
S_{AMR}	amplitude demodulation of residual signal

2.2.1. *Signal decomposition by inverse filtering*

The isolation of tremor sources can be implemented using inverse filtering [22] as shown in Figure 2. Let the recorded signal be $D = d_{rc} \in \Re^{k_x n}$ where each column index corresponds to a sample and each row corresponds to a particular frame. Let H_z be a matrix representing the impulse response of a set of autoregressive filters with coefficients $a_{ri} \in \Re^{k_x p}$ (p: filter order) adapted based on least square of the error term. The residue $R \in \Re^{k_x n}$ signal can be expressed as:

$$R = D - \hat{D} = d_{rj} - \sum_{i=1}^{p} a_{ri} d_{r(j-i)}, \quad 1 \le j \le n, \quad 1 \le r \le k \qquad (1)$$

where R is the estimation of the glottal signal and A contains the information on the vocal tract configuration. Both terms are separate sources that contribute to the tremor observed in the recorded speech manifested in a form of amplitude/frequency modulation. Recursive Least Square algorithm provided faster convergence to optimal filter coefficient [22]. The algorithm was implemented with frames of 1s and overlapping increments of 50 ms to capture low frequency oscillations embedded in the modulated signals.

The tremor source is then analyzed in the lower-spectral band of the frequency response of the filter when it is excited with white noise.

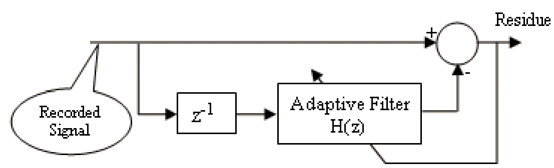

Figure 2. Block diagram for the inverse filtering process. The recorded signal is decomposed into the residue (glottal excitation) and the adaptive filter (vocal tract transfer function).

The independent AM tremor components were estimated by means of envelope detection. Let us assume that u_m is a narrow-band message signal that corresponds with the tremor observed in the recorded speech (BW_{um}<12 Hz) and that u_c is a carrier signal that corresponds with the fundamental frequency (F_0). If recorded speech is a Sustained vowel phonation (SVP) utterance in which pitch and intensity is expected to be constant, the recorded speech can be described by:

$$d(n) = C_0[1 + u_m(n)]\cos(2\pi F_0 n) \qquad (2)$$

where C_0 is a constant given by the amplitude of the carrier.

If F_0 is detected reliably using a pitch waveform matching algorithm [15] and $F_0 >> BW_{um}$ (which still holds for the case of speech with lower pitch corresponding to male speakers with average F_0 = 100-120 Hz); then, u_m can be extracted using a simple envelope (low pass) detection operation [10]:

$$u_m(n) \propto LPF_{12Hz \le f \le F_0}\{|d(n)|\} \qquad (3)$$

In this case, u_m is the narrowband modulation present in the radiated signal AM_D (the "D" subscript comes from the matrix D representing the radiated signal). This derivation can also be performed in the residual (glotal source) signal (matrix R), to obtain AM_R, the modulation in the source.

The PSD of the resulting demodulated signals (denoted S_{AMD} and S_{AMR}) were then estimated using Welch modified periodogram [17] with minimum frequency resolution of

0.0610 Hz, 1 second length Blackmann window, 25% overlap and 2^{15} points FFTs.

2.2.2. *Normal / pathological separation.*

Results in [11] showed different behaviors for the amplitude peaks of the AM spectra for normal and pathological samples, with a higher predictive value in the latter (90% vs. 40%). This fact suggested the implementation of an initial extra step in the design of the automatic measurement of tremor. This is, pathological samples should be separated from normal samples before quantifying the magnitude of the tremor perturbation. This avoids sensitivity of the tremor index to low-magnitude peaks of the spectrum of control subjects.
The SVP of the subjects from the databases that were scored by the judges with high index of tremor were analyzed together with a similar number of control subjects (19 of each type).

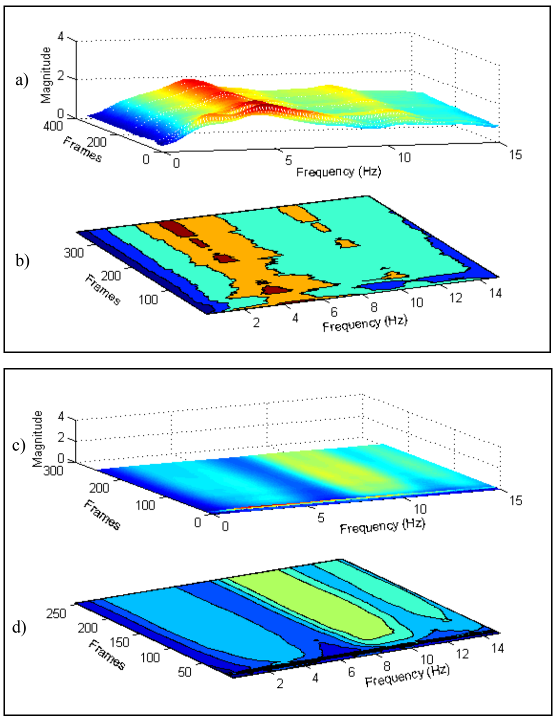

Figure 3: *Frequency response (a,c) and contour (b,d) of the amplitude-demodulated signal of the vocal tract model for: subject HO66F059A with severe OVT (top) and healthy subject PS37M086A, with no tremor.*

Figure 3 shows 3-dimensional plots of the estimated frequency response of the amplitude-demodulated signal derived for the vocal tract filter model of a subject with OVT (a) and normal (c). The spectrum was obtained as indicates in previous section. Plot a) exhibits high-amplitude oscillations in the frequency range reported in [1] for OVT, with the higher intensity between 4 and 5 Hz. In contrast, the plot b) shows small-magnitude oscillation with peaks in the 8-11Hz band (range reported for tremor in normal subjects [23]). All signals were normalized to account for differences in energy, down-sampled and time-limited to an average duration of 6s. The contour plots (b, d) reflect a more accurate representation of the bands with higher energy with respect to the frequency axis.

Two indexes were evaluated to separate healthy from pathological tremor recordings. One was the ratio of the energy in the 3-8 Hz band to the energy of the 0-25Hz band:

$$B_{rX} = \frac{\sum_{3-8Hz} S_{AMX}}{\sum_{D-25Hz} S_{AMX}} \tag{4}$$

where subscript "X" can be any of the three power spectral densities obtained (D-radiated, R-residual, A-filter response). B_{rX} index should be higher for utterances with tremor due to higher relative energy in the 3-8 Hz band.

The other index is the kurtosis (K) of the SAMA values in the 0-25 Hz range. The kurtosis is defined as:

$$K_X = \frac{\sum_{i=1}^{n} (S_{AMX\ i} - \overline{S_{AMX}})^4}{(n-1)\sigma^2} - 3 \tag{5}$$

where $\overline{\bullet}$ is the mean operation, σ is the standard deviation and n is the number of samples of the estimated spectrum per frame. The rationale for the use of K_X is that the peak excursions in S_{AMX} should increase the value of K_X, while spectra of healthy subjects should have a closer-to-normal distribution.

Threshold values to separate normal from pathological samples were established in terms of equal error rates (EER) [24] assuming normal distributions for the values of B_{rX} and K_X in the two groups.

2.2.3. *Quantification of tremor perturbation magnitude*

After an abnormal tremor is detected, the quantification of the magnitude of the perturbation is required. A peak ratio algorithm was used to assess the magnitude of the tremor perturbation observed in the studied disorders. A peak peaking algorithm was implemented to find the most relevant peaks in the frequency band of interest for the amplitude-demodulated power spectrum of the original radiated speech (S_{AMD}), the amplitude demodulated power spectrum of the residual signal (S_{AMR}) and the amplitude demodulated power spectrum of the vocal tract filters (S_{AMA}).

A peak matching algorithm was later performed between the S_{AMD} and the other two spectra. The peaks that matched within a range of ±0.25Hz were considered to derive an energy ratio index that quantified the magnitude of the tremor perturbations. The index was derived as the ratio between the energy of the matched peaks within the 3-8Hz frequency band in S_{AMD} and the total energy of the S_{AMD} spectrum below 25 Hz. A weigh factor of 1.3 times was used emphasize peaks that are perfectly aligned in the three spectra.

Figure 4 shows the signal required to derive the tremor index for a subject with OVT. It is observed that a good separation of the two sources of tremor is achieved since the product of S_{AMR} and S_{AMA} (plot e) closely resembles the original spectrum of the amplitude demodulated spectrum. However, both spectra provide different information of the manifested tremor. On one side S_{AMA} matches more closely the amplitude modulation information due to the resemblance with the amplitude demodulation spectrum S_{AMD} (highest peak located at 4.7Hz). On the other side S_{AMR} matches more closely the frequency demodulated spectrum (S_{FM}) shown for comparison in plot b (highest peak located at 4.1Hz). S_{FM} was estimated with the average pitch frequency derived with a waveform matching algorithm [15] for each frame of D. These observations were consistently observed in subjects with perturbation judgments above 3 in the perceptual scale.

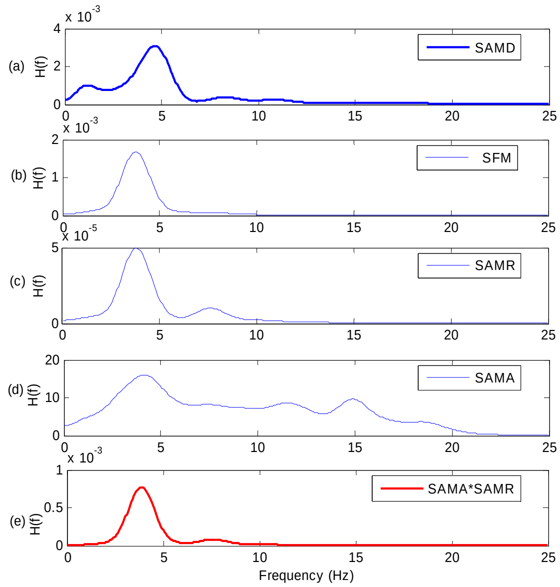

Figure 4: *PSD of the amplitude-demodulated (a,c,d,e) and frequency (b) signals used quantify the tremor perturbations (subject HO66F059A). (a,b) S_{AMD} and S_{FM} derived from the radiated signal (c) S_{AMR} derived from the residual signal (d) S_{AMA} derived from the vocal tract model. (e) estimated S_{AMD} signal from S_{AMR} and S_{AMA}.*

3. Results and Discussion

3.1. Normal/Pathological separation:

Results showed that only the kurtosis operator of S_{AMA} (from $0 \leq f \leq 25$Hz) (K_A) showed different means for both groups at the 95% significance level.

The average normal subjects exhibited less variability having higher K_A values (over 9.2). Subjects with tremor exhibited kurtosis values closer to 0. A detection threshold of 8.2 was found as optimal according to the EER criterion.

The fact that only K_A significantly separated normal from pathological samples imply that the main differences between normal and pathological tremor should be given by the tremor perturbations made by the vocal tract. The tremor observed in normal subjects would be mainly due to oscillation in the glottal area, with smaller contribution of the vocal tract. The solution to the normal/pathological separation requires the separation of tremor sources and the assessment of the information contained in the vocal tract model (matrix A).

3.2. Tremor magnitude estimation

The new algorithm was compared to commonly-used tremor indexes previously reported such as: the frequency demodulation index (Fo_{TRI}) from Kay Elemetrics [12], the tremor intensity index (A_{TRI}) [12] and the AM-modulation and frequency modulation indexes reported in [23]. The results of analyzing the data with all tremor indexes was compared to the perceptual judgments (PJ) provided by the clinicians (intra-rater correlation 0.9283). The new measure, Tr_{AM}, showed the highest correlation with respect to the PJ exhibiting a coefficient of 0.8723 ($p<0.001$), followed by A_{TRI} with 0.6531 ($p<0.001$) and next by Fo_{TRI} with 0.4968 ($p<0.01$). The other two measurements yielded smaller correlation coefficients with p values below the 95% confidence level. The differences in performance between the Tr_{AM} and A_{TRI} algorithms were attributed to the independent estimation of the different sources of tremor which enabled the peak alignment and peak weighting algorithms.

It was noticed that the frequency demodulation parameter Fo_{TRI} did not correlate significantly with perceptual judgments, denoting a higher influence of the amplitude modulation component over the perceptual judgments. Figure 6 shows the correspondence of the PJ and the objective measures obtained with Tr_{AM} and A_{TRI}. It is observed a higher correlation between both types of judgments with PJ in the extremes of the scale while in the correlation in the middle is weaker. The best fit of a second order polynomial is shown to highlight a nonlinear trend between both types of judgments.

The sequence of operations for computing Tr_{AM} is given in Table 2 and graphically represented in Figure 5.

Table 2. *Computation of Tr_{AM}.*

	Algorithm description
1	compute ↓2kHz and $LPF_{1.2*F0}$
2	construct matrix *D*
3	derive matrices *R* and *A* using eq. (1)
4	derive S_{AMA} with (3) as in section 2.1.1 ($0 \leq f \leq 25$Hz)
5	derive *K* using (4) → normal/abnormal decision
6	obtain S_{AMD} with (3) as in section 2.1.1 ($0 \leq f \leq 25$Hz)
7	derive S_{AMR} with (3) as in section 2.1.1 ($0 \leq f \leq 25$Hz)
8	perform peak peaking on S_{AMA}, S_{AMD} and S_{AMR}
9	perform peak matching and weight function algorithms
10	obtain energy ratio from prominent peaks

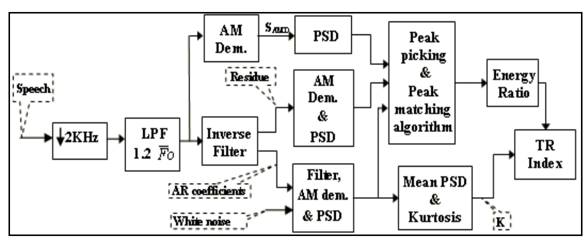

Figure 5: *Block diagram to obtain the tremor index.*

Figure 7 shows the magnitude of TR_{AM} for all the disorders in the databases. The algorithm is able to differentiate OVT from the rest of the groups (3 subjects did not present large amounts of tremor). It is also noticeable that other groups such as HP, HD and ALS show small non-zero averages of the index.

4. Conclusions

This paper reported an alternative automatic measurement to assess tremor perturbations in SVP utterances with high correlation with perceptual judgments. It was shown that estimating independent sources of tremor was beneficial to differentiate normal from pathological tremor and contributed positively to achieve objective judgments of tremor perturbations.

The results obtained in this work indicate that the reported measurement performed better than the other tremor indexes studied by at least 26% when measurements are correlated with the perception of tremor perturbations. The performance of the measurement was enhanced by the independent estimation of tremor sources that enabled a two-tier analysis.

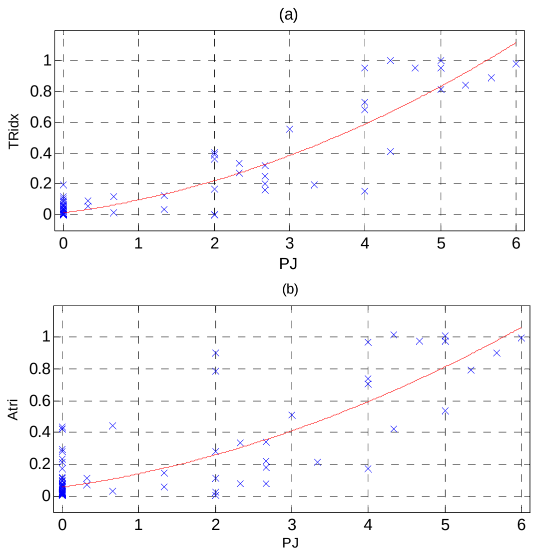

Figure 6: *Plot of TR index vs. PJ.(a)TR_{AM} index. (b) A_{TRI} index. Best fit of a 2^{nd} order polynomial is shown.*

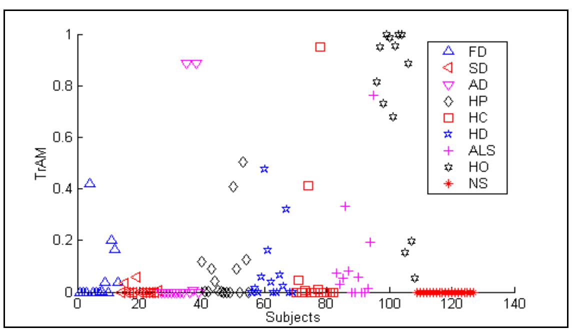

Figure 7: *Results of TR_{AM} index for all subjects.*

5. Acknowledgements

The authors acknowledge N. Mendez R. Gutierrez, and R. Espinosa (Department of Neurology of AMC Hospital, Santa Clara) for their contribution to this study. They also acknowledge the support obtained from the National Science and Engineering Council of Canada.

6. References

[1]. Ackermann H. and Ziegler W., "Cerebellar voice tremor: An acoustic analysis", J. of Neurol., Neurosur. and Psych, 41:74-76, 1991.

[2]. Antal A., Dibó G., et al., "P300 components of visual event-related potentials distinguishes patients with idiopatic Parkinson disease and essential tremor", J. Neural Trans., 107:787-798, 2001.

[3]. Aronson A.E., "Dysarthrias: differential diagnosis", Rochester, MN: Mentor Seminars (audio tape), 1993.

[4]. Buder, E.H., Strand, E.A. "Quantitative and graphic acoustic analysis of phonatory modulations: the modulogram" J. Speech, Lang. and Hearing Research. 46. 475-490. 2003.

[5]. Castillo-Guerra E., "A modern approach to dysarthria classify-cation", Ph.D. Thesis at University of New Brunswick, 2002.

[6]. Castillo-Guerra, E., Ruiz, A., "Automatic modeling of acoustic perception of breathiness in pathological voices," IEEE Trans. on Biom. Eng., 56(4), 2009.

[7]. Darley F.L, Aronson A.E. and Brown J.R., "Differential Diagnosis of Dysarthria", JSHR, 12:246-269, 1969a.

[8]. Dromey, C., Warrick, P., Irish, J. "The influences of pitch and loudness changes on the acoustics of vocal tremor" J. Speech, Lang. and Hearing Research. 45. 879-890. 2002.

[9]. Fant, G. "Acoustic Theory of Speech Production" Mouton. The Hague. 1960.

[10]. Haykin S., "Communication systems", John Willey & Sons Inc., 1994.

[11]. Jiang J., Lin E. and Hanson D.G., "Acoustic and airflow spectral analysis of voice tremor", JSLHR, 43:191-204, 2000.

[12]. Kay Elemetrics, "Multi-dimensional voice program" model 4305, Lincoln park, NJ: Kay Elemetric Corporation, 1993.

[13]. Kent R.D., and Ball M.J., "The Handbook of Voice Quality Measurement", Singular Publishing Group Inc., Feb. 1999.

[14]. Ludlow C., Bassich C., Connor N. and Coulter D., "Phonatory characteristics of vocal fold tremor", J. of Phonetics, 14:509-515, 1986.

[15]. Medan Y., Yair E. & Dan C., "Super resolution pitch determination of speech signals", IEEE Trans. Sig. Proc., 39:40-48, 1991.

[16]. Porter, R.J. and Hogue, D.M. "Nonlinear dynamical systems in speech perception and production" Nonlinear Dynamics, Psychology, and Life Sciences. 2. 95-129. 1998.

[17]. Proakis, J.G., and Manolakis, D.G., "Digital Signal Processing", Upper Saddle River, NJ: Prentice-Hall, 1996.

[18]. Raming L.A. and Shipp T., "Comparative measures of vocal tremor and vibrato", J. of Voice, 1:162-167, 1987.

[19]. Raming L.A., Scherer R.C., Klasner E.R., Titze I.R. and Ringel S.P., "Acoustic analysis of voice in amyotrophic lateral sclerosis; a longitudinal case study", JSLHR, 55:2-14, 1990.

[20]. Titze, I. R. Summary Statement. *Workshop-on acoustic voice analysis*. National Center of Voice and Speech. Iowa. 1995

[21]. Warrick, P., Dromey, C., Irish, J., Durkin, L., Pakiam, A. and Liang, A. Botulinum toxin for essential tremor of the voice with multiple anatomical sites of tremor: A crossover design study of unilateral versus bilateral injection. Laryngoscope. 111. 1366-1374. 2000.

[22]. Widrow B. and Stearns S.D., "Adaptive signal processing", Prentice-Hall, Englewood Cliffs, N.J., 1985.

[23]. Winholtz W. S. and Ramig L.O., "Vocal tremor analysis with the Vocal Demodulator", JSLHR, 35:562-573, 1992.

[24]. Kar-Ann T., Jaihie Kand Sangyoun L.," Biometric scores fusion based on total error rate minimization," Pattern Recognition, Vol. 41(3):1066-1082, 2008.