Vocal Fold Pathology Assessment using AM Autocorrelation Analysis of the Teager Energy Operator *

Liliana Gavidia-Ceballos[‡], *John H. L. Hansen*[†], *and James F. Kaiser*[†]

[†]Dept. of Electrical Engineering, Duke University, Box 90291, Durham, North Carolina 27708-0291

[‡]Grupo de Bioingeniería y Biofísica Aplicada, Universidad Simón Bolívar, Apdo. 89000, Caracas 1081-A, Venezuela

ABSTRACT

Traditional speech processing methods for laryngeal pathology assessment assume linear speech production, with measures derived from an estimated glottal flow waveform. They normally require the speaker to achieve complete glottal closure, which for many vocal fold pathologies cannot be accomplished. To address this, a nonlinear signal processing approach is proposed which employs a differential Teager energy operator and the energy separation algorithm to obtain formant AM and FM modulations from bandpass filtered speech recordings. A new speech measure is proposed based on parameterization of the autocorrelation envelop of the AM response. Using a cubic model of the autocorrelation envelop, a three dimensional space is formed to assess changes in speech quality. This approach is shown to achieve exemplary detection performance for a set of muscular tension dysphonias. Unlike flow characterization using numerical solutions of Navier Stokes equations, this method is extremely computationally attractive, requiring only $N\log N + 8N$ multiplications and N square roots for N samples, and is therefore suitable for real time applications due to its computational simplicity. The new non-invasive method shows conclusively that a fast, effective digital speech processing technique can be developed for vocal fold pathology assessment, without the need for (i) direct glottal flow estimation or (ii) complete glottal closure by the speaker. The proposed method also confirms that alternative nonlinear methods can begin to address the limitations of previous linear approaches for speech pathology assessment.

INTRODUCTION

It is well known that vocal fold pathology (VFP) alters the mechanisms of speech production, and such disturbance is reflected in voice quality deterioration. The physics of human speech production under both healthy and especially vocal fold pathology conditions suggests that alternate production models other than traditional may be more appropriate. Although modeling actual speech production using fluid flow characteristics would be useful, the ability to characterize fluid flow properties in actual vocal fold pathology patients and maintaining a non-invasive procedure would be difficult. Instead, it is suggested that features derived using a nonlinear speech framework could reveal the potential of alternative speech models, to the traditional linear approach. In the present study, a nonlinear processing method is developed, based on the Teager energy operator, that can extract further information from the speech signal, including nonlinear excitation sources, which are hypothesized to play a crucial role in both healthy and nonhealthy speech production. The study described here represents a formulation which ultimately will address VFP assessment.

*This work was supported by grants from The Whitaker Foundation and the Venezuelan FUNDAYACUCHO-LASPAU Program.

TEAGER ENERGY OPERATOR PRINCIPLES

The ideas of vortex shedding as additional sources of acoustic energy that could influence significantly the mechanisms of sound production, were first introduced by Teager and Teager[7], through a series of experiments using hot wire anemometry. The results of their experiments strongly suggest nonlinear processes to be the primary sound producing mechanisms in the vocal tract during phonation and that separated flow and the generated flow vortexes within the confined geometry of the vocal tract are responsible for this phenomena. The inclusion of additional sources of acoustic energy may help improve intelligibility and voice quality, while at the same time may provide key elements in voice quality assessment for clinical applications.

Speech formants are characterized by the poles of the vocal tract transfer function, when a linear model is used. Teager was convinced that the speech resonances can change rapidly both in frequency and amplitude even within a single pitch period, possibly due to separated airflow in the vocal tract [5]. The nonlinear differential Teager energy operator can detect formant AM-FM modulations by estimating the product of their time-varying amplitude and frequency. The Teager operator is considered a high-resolution energy estimator. The AM-FM model proposed by Maragos, Kaiser and Quatieri [5] represents a single speech resonance $R(t)$ as an AM-FM signal

$$R(t) = a(t)\cos(2\pi[f_c t + \int_0^t q(\tau)d\tau] + \theta) \qquad (1)$$

where f_c is the formant frequency value, $q(t)$ is the frequency modulating signal, and $a(t)$ is the time-varying amplitude. The instantaneous formant frequency of the signal is defined as $f_i(t) = f_c + q(t)$. To demodulate a speech resonance $R(t)$ into its varying amplitude $|a(t)|$ and instantaneous frequency $f_i(t)$, the energy separation algorithm (ESA) (developed by Maragos, Kaiser and Quatieri [5]) is applied to the signal resonance $R(t)$ obtained after filtering the speech signal around the formant under consideration. The ESA is based on the Teager-Kaiser

energy tracking operator, with discrete time domain representation of the form,

$$\Psi[x(n)] = x^2(n) - x(n+1)x(n-1) \quad (2)$$

and the ESA frequency and amplitude estimates using the DESA-2 algorithm (Maragos, Kaiser and Quatieri [5]) are

$$f(n) \approx \frac{1}{4\pi T} \arccos(1 - \frac{\Psi[x(n+1) - x(n-1)]}{2\Psi[x(n)]}) \quad (3)$$

$$|a(n)| \approx \frac{2\Psi[x(n)]}{\sqrt{\Psi[x(n+1) - x(n-1)]}} \quad (4)$$

In this study, we developed a nonlinear speech processing framework that uses the Teager energy operator and the energy separation algorithm to extract the first formant AM modulation characteristic from bandpass filtered speech recordings. Thia response is used for vocal fold pathology assessment based upon parameterization of features that we believe to be correlated to the regularity of vocal fold vibratory movement in a healthy condition and to the asymmetry and irregular structure for a vocal fold pathology condition.

ALGORITHM FORMULATION

Fig. 1 shows the flow diagram of the nonlinear procedure used. A 10th order LPC spectrum analysis was performed on the speech signal, in order to extract the first formant (F_1) location. The first formant was specifically chosen because this formant can be associated with the region just above the vocal folds. A finite impulse response (FIR) bandpass filter was applied to the original speech signal, around the first formant of the original speech, and the filter bandwidth was chosen to be 450Hz.

An estimate of the AM and FM modulation of F_1 through time was obtained from the Teager energy operator of both the filtered speech and its derivative, using the energy separation algorithm developed by Maragos, Kaiser, and Quatieri [5, 4]. A 21-point median smoothing was applied to both AM and FM components, at those locations where the value of the signal exceeded the overall median value by 40%.

Pitch information was extracted from the AM signal by finding the location of the first maximum in the AM autocorrelation function. A window of 3 pitch periods was used to calculate the mean and extract it from the AM component.

The autocorrelation function of the AM envelope was obtained by performing a circular correlation. In a circular correlation, a periodic extension of the sequence is used and only one period of the result is taken. By adequate zero padding, a circular correlation can be made identical to a linear correlation. If a sequence $s(n)$ is of length N, the resulting linear autocorrelation has length $N-1$, so $N-1$ zeros need to be appended to $s(n)$ to yield identical results from a circular autocorrelation. A circular correlation can be performed through discrete Fourier transforms. The details of this procedure can be found in Strum [6] (pp. 430-433).

The envelope of the AM component was obtained, using a peak-picking turning point algorithm. The turning point (TP) algorithm is described elsewhere (Tompkins and Webster, 1981)[8]. A simple slope calculation is performed to determine the instants where changes in slope sign occur, which identify the peaks (turning points) of the signal.

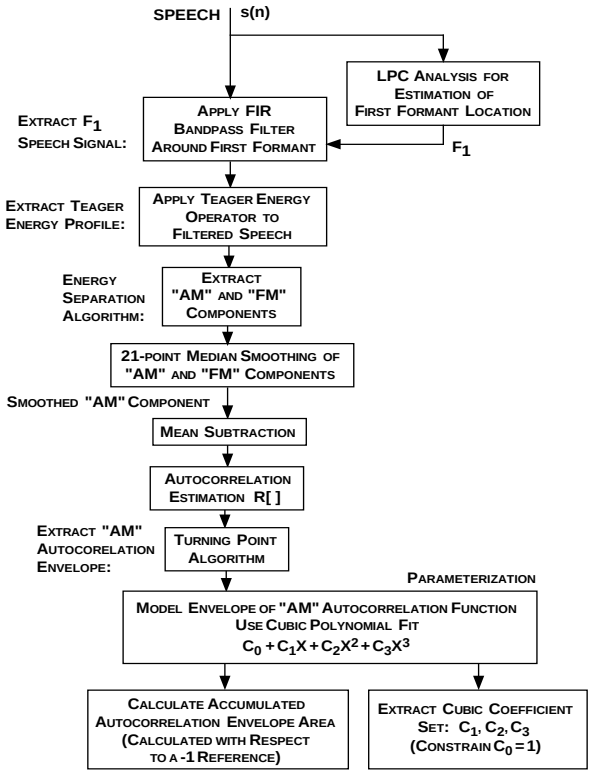

Figure 1: Flow diagram of the nonlinear speech processing algorithm.

The envelope of the AM component was modeled using a cubic polynomial fit $p(X) = C_0 + C_1X + C_2X^2 + C_3X^3$. The coefficients C_1, C_2, and C_3 were obtained, while C_0 was constrained to 1, since this is the value of the autocorrelation function at $X = 0$. The polynomial fit was used to calculate the accumulated autocorrelation envelope area. The area was calculated with respect to a -1 reference value.

To evaluate the performance of the proposed nonlinear speech processing algorithm, the procedure was tested on speech data obtained from the Scottish Rite Children's Medical Center[1]. The speech sample consisted of eleven adult speakers (ten female, one male) producing the vowel /e/ in the phrase 'he is', extracted from the 'Grandfather passage', a text fragment commonly used by speech therapists to assess voice quality on their patients. The vowel /e/ was specifically chosen because its first and second formants are widely separated (over 1000 Hz separation), and therefore the AM-FM modulations of the second formant will not influence the first formant AM-FM modulations measured with this nonlinear speech processing technique. From the eleven speakers, ten suffered from muscular tension dysphonias (MTD) and one was completely healthy. For some speakers, recurrent laryngitis was also present as a result of vocal abuse. The results of this evaluation are discussed in the following section.

EVALUATION & DISCUSSION

An evaluation of the AM and FM components obtained for the eleven patients with muscular tension dysphonias pre and post voice therapy treatment revealed that the AM component

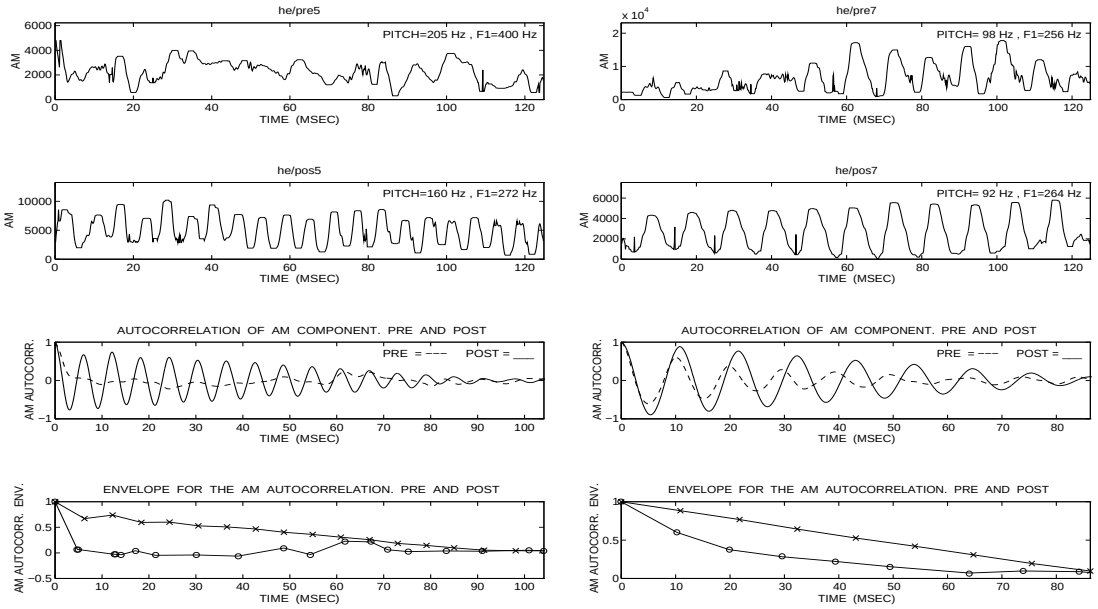

Figure 2: First formant AM modulation pre and post voice therapy treatment for both a female and a male speakers with muscular tension dysphonia. Also shown are the pre and post AM autocorrelation functions and AM autocorrelation envelopes. Pre and post AM autocorrelation peaks are marked with 'o' and 'x', respectively.

was better able to convey consistent information, believed to be associated with the quality of speech under a muscular tension dysphonia A complete set of the signals obtained in this evaluation can be found in Gavidia-Ceballos[2]. Note in the two examples given in Fig. 2, the marked periodicity and regularity of the AM component after successful voice therapy treatment, as compared to that obtained pre treatment.

In order to quantify this distinctive feature, an autocorrelation of the AM component was performed. The decay in the successive peak values of the AM autocorrelation function represents a measure of the regularity and periodicity of the signal. There is an inherent decay that arises from the fact that the autocorrelation estimate is the biased estimate, so the total number of points used to calculate the autocorrelation decreases for each successive lag. However, this fact does not affect the performance of the processing technique, since its effect on both pre and post signals is the same. If an unbiased estimate is used instead, the envelope of the AM autocorrelation is flatter, but the separation between both pre and post voice therapy treatment groups is still present, regardless of the autocorrelation technique used. The biased autocorrelation estimate has the added advantage that it can be obtained using the fast Fourier transform, which saves computational time.

Fig. 2 shows two examples of AM components obtained pre and post voice therapy treatment, for both a female and a male speaker, respectively. Also shown are the AM autocorrelation, and the AM autocorrelation envelope, obtained after extracting the local maximas of the autocorrelation function, and connecting them using a piece-wise linear approximation. Note that in general, the AM autocorrelation envelope for all the post cases can be approximated by a straight line, and its amplitude is noticeably greater than those corresponding to pathology (pre-treatment). Also note the irregularity in the shape of the autocorrelation envelope for pathology.

The envelope of the AM component was modeled using a cubic polynomial fit $p(X) = C_0 + C_1 X + C_2 X^2 + C_3 X^3$. The second and third order coefficients C_2 and C_3 are expected to be very small for the post therapy cases, since the shape follows primarily a linear trend. The coefficient C_0 was set to 1, since the value of the autocorrelation for lag zero is always one.

Table 1 shows the set of coefficients C_1, C_2, and C_3 obtained for the cubic fit of the autocorrelation envelope for the eleven patients considered in this study. Note the smaller order of magnitude of the coefficients for the post (healthy) cases. A linear approximation discarding C_3 and C_2 in the post cases is still adequate to represent the envelope. This is not the case for the pre (pathology) cases.

Fig. 3 shows a spatial distribution on a 3-D plane of the pre and post cases. The coefficients C_1, C_2, and C_3 are represented as the projection on the x, y and z axes, respectively. All healthy cases tend to group in a cluster, whereas the pathology cases are scattered throughout the parameter space. This illustrates how healthy and non-healthy speech cases are distributed in the C_1, C_2, C_3 space. We point out however that without independent subjective assessment methods, it is not possible at this time to quantify how overall speech quality is improved as the response moves from pathology locations (i.e., marked with 'x'), towards healthy (marked with 'o').

Fig. 4 shows the autocorrelation envelopes of all pre and post cases, using a piece-wise linear approximation and a cubic fit polynomial, respectively. From the figures, it is evident that to be able to model the entire 25 msec segment using one single polynomial fit in each case, an order higher than linear is required to model pathology, whereas a linear fit is adequate to model the healthy cases. Again, note the clear separation between healthy and pathology groups. Further results using other parameters extracted from the AM autocorrelation envelope can be found in Gavidia-Ceballos[2], and Hansen, Gavidia-Ceballos, and Kaiser[3].

C_1		C_2		C_3	
PRE (e-02)	POST (e-03)	PRE (e-04)	POST (e-05)	PRE (e-07)	POST (e-08)
-1.7398	-2.5816	0.9250	1.2603	-1.6502	-5.3053
-2.0952	-4.9701	1.3713	2.7528	-3.2492	-7.5690
-2.9916	-6.5875	3.0268	5.0556	-8.9327	-16.995
-1.3582	-4.8052	0.9225	2.3760	-2.0465	-6.0506
-3.7094	-8.1328	3.8540	6.6220	-11.973	-17.987
-3.5276	-1.7012	3.6167	-0.5126	-10.765	2.4104
-0.6088	-1.4213	0.1469	0.1090	-751.69	-0.0363
-3.2641	-2.3366	3.6416	0.5349	-11.592	-1.3707
-1.6690	-1.0490	1.6775	-0.4798	-4.9659	1.0523
-1.6472	-2.3363	0.9068	0.8410	-1.6023	-3.0186
-3.6123	-2.0386	4.1223	-1.2266	-13.059	4.3244
COEFFICIENT MEANS					
-2.3839	-3.4509	2.2010	1.5757	-6.3556	-4.6248

Table 1: Coefficients C_1, C_2, and C_3 of the cubic fit polynomial to the first formant AM modulation of eleven subjects pre and post voice therapy treatment. Also shown are the means for each coefficient.

CONCLUSIONS

In this study, we have considered a nonlinear speech processing technique which was proposed to extract features that correlate well with the presence of vocal fold pathology. In particular, evaluation of this algorithm in the analysis of speech with muscular tension dysphonias for both pre and post voice therapy treatment showed that the analysis of the first formant AM modulation characteristic allows the extraction of features that we believe to be correlated to the regularity of vocal fold vibratory movement in a healthy condition and to the asymmetry and irregular structure for a vocal fold pathology condition. The total number of multiplications required with this procedure is $N\log N + 8N$, and N square roots for N samples. Therefore, this procedure could be suitable for real time applications due to its computational simplicity. Further research is suggested on a larger speech database, to extract the adequate parameters based on age, gender, and phoneme. In addition, further studies to assess the degree of speech quality improvement using the parameters obtained with this procedure are suggested. An example would be an in-depth study on how the response moves from healthy in the C_1, C_2, C_3 space in association with speech quality subjectively assessed. This measure could also potentially provide a better way of measuring speech quality for patients under speech therapy.

[1] We would like to thank Dr. J. Riski of Scottish Rite Children's Medical Center, Atlanta, GA for his help in providing speech data used in this study.

References

[1] Aronson A.E., "Motor Speech Signs of Neurologic Disease", in *Speech Evaluation in Medicine,*

[2] Gavidia-Ceballos L., "Analysis and Modeling of Speech for Laryngeal Pathology Assessment", Ph.D. Thesis, Dept. Biomedical Engineering, Robust Speech Processing Lab. Dept. Electrical Engineering, Duke Univ., Aug. 1995.

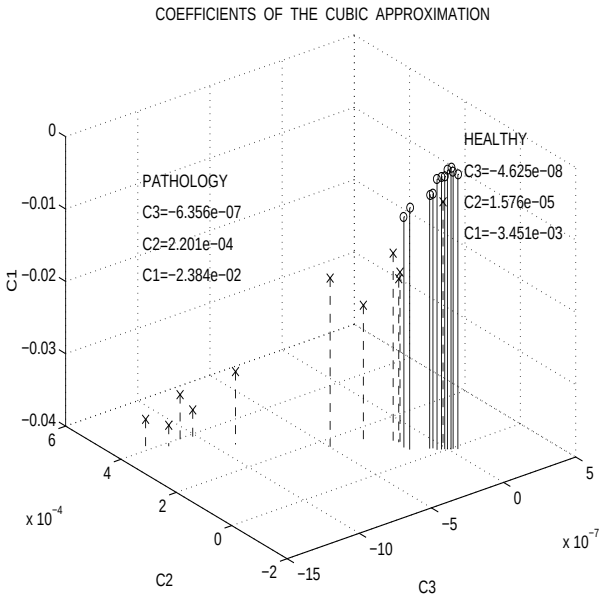

Figure 3: Scatter plot for healthy and pathology conditions in the parametric C_1, C_2, C_3 space.

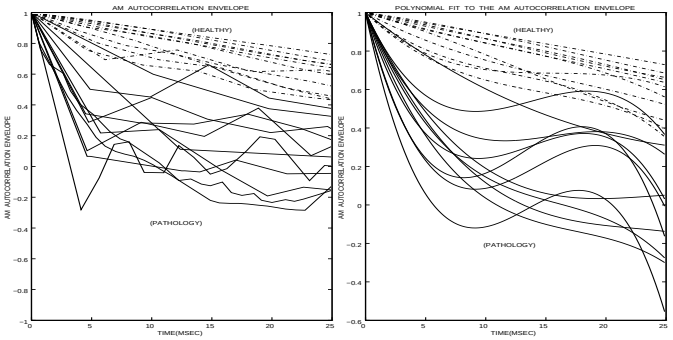

Figure 4: Piece-wise linear approximation and cubic polynomial fit to the autocorrelation envelope of first formant AM modulation for healthy and pathology conditions on the eleven subjects used in this study.

[3] Hansen J. H. L., Gavidia-Ceballos L., Kaiser J. F. "A Nonlinear Operator based Speech Feature Analysis Method with Application to Vocal Fold Pathology Assessment", submitted to *IEEE Trans. Biomedical Engineering*, 29 pgs. Oct. 1995.

[4] Kaiser J. F. "On a simple algorithm to calculate the 'energy' of a signal. *Proceedings of the IEEE Inter. Conf. on Acoustics, Speech, Signal Proc. (ICASSP)*, pp. 381-384, 1990.

[5] Maragos P., Kaiser J. F., Quatieri T. F. "Energy Separation in Signal Modulations with Application to Speech Analysis", *IEEE Trans. on Signal Proc.*, 41(10):3024-3051, Oct. 1993.

[6] Strum R. D. and Kirk D. E. "*First Principles of Discrete Systems and Digital Signal Processing*", Addison-Wesley Publishing Company, 1988.

[7] Teager H. M. and Teager S. M. "A phenomenological model for vowel production in the vocal tract". In Daniloff R. G., editor, *Speech Science. Recent Advances.* College-Hill Press, San Diego, California, 1985.

[8] Tompkins W. J. and Webster J. G., editors "*Design of Microcomputer-based Medical Instrumentation*", Prentice-Hall, Inc. , Englewood Cliffs, New Jersey 07632, 1981.