3rd European Conference on
Speech Communication and Technology
EUROSPEECH'93
Berlin, Germany, September 19-23, 1993

ACOUSTIC MODEL AND EVALUATION OF PATHOLOGICAL VOICE PRODUCTION

Dimitar D. Deliyski

Kay Elemetrics Corp., Dept. of Research and Development
12 Maple Av., Pine Brook NJ 07058, U.S.A.

ABSTRACT·

An acoustic model of pathological voice production is presented. It describes the non-linear effects occurring in the acoustic waveform of disordered voices. The noise components such as fundamental frequency and amplitude irregularities and variations, sub-harmonic components, turbulent noise and voice breaks are formally expressed as a result of random time function influences on the excitation function and the glottal filter.

A method for quantitative evaluation of these random functions is described. The method computes their statistical characteristics which can be useful in assessing voice in clinical practice. More than 33 acoustic parameters are computed: average fundamental frequency, phonatory frequency range, several frequency and amplitude short- and long-term perturbation and variation measures, noise-to-harmonic ratio, voice turbulence and soft phonation indexes, quantitative measures of voice breaks, sub-harmonic components and vocal tremors. This set of parameters, which corresponds to the model, allows a multi-dimensional voice quality assessment. A computer system based on above model and method was developed for the CSL model 4300 (Kay Elemetrics Corp.). A group of 68 people with normal and disordered voices was analyzed using the system in order to define normative values for the acoustic voice parameters.

Keywords: acoustic voice analysis, signal processing, speech pathology, phoniatrics.

1.INTRODUCTION

The classic way to describe the acoustics of human speech is by using the *Linear Model of Speech Production* [1, 2], where the voice signal is presented as a result of a periodic impulse sequence (excitation) filtered by the glottis, the vocal tract and the lips.
However, the real voice contains irregular components which are (probably) due to the chaotic nature of the laryngeal mechanism [3]. A voice without irregularity is not perceived as human which is why the advanced speech synthesizers, based on the linear model, introduce some pitch irregularity [4, 14].

2.ACOUSTIC MODEL OF THE PATHOLOGICAL VOICE PRODUCTION

Voice pathology can cause increased noise components in the voice signal such as: fundamental frequency and amplitude irregularities and variations with different patterns, sub-harmonic frequency components, turbulent noise, voice breaks and tremors [2, 5-8]. Understanding the acoustics of these changes is the key to the development of methods for the evaluation of pathologic voices. A formal expression of these changes is given by the *Extended Acoustic Model of the Pathological Voice Production* [9] on **Fig.1**.

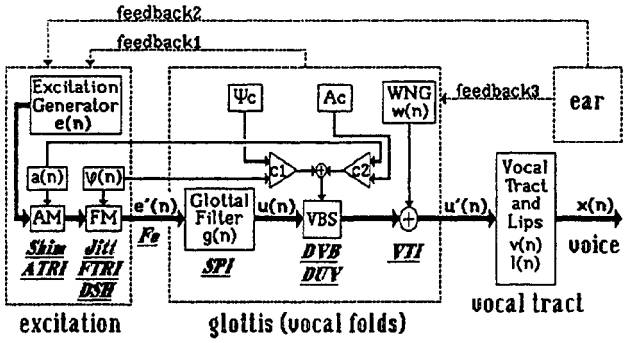

Fig.1:Extended Information Model of the Pathological Voice Production.

The discrete-time formal representation of the model describes the excitation function

$$e'(n) = a(n)\sum_{m=0}^{\infty} \delta\left[n - \lfloor mTo + \varphi(n) \rfloor\right]$$

as a modulated impulse sequence, where the frequency modulating (FM) function $\varphi(n)$ and the amplitude modulating (AM) function $a(n)$ are random time functions;

· The author is also a research associate at the Bulgarian Academy of Sciences, Institute of Informatics, Sofia, Bulgaria.

$n=0,1...\infty$ is discrete time (samples); *To* is the period of the sequence (samples); $\delta(n)$ is a Kronecker delta function ($\delta(n=0)=1$, $\delta(n0)=0$); and the carrier sequence is

$$e(n)=\sum_{m=0}^{\infty}\delta(n-mTo).$$

The glottal volume velocity function

$$u'(n)=\begin{vmatrix} u(n)+w(n), if\ a(n)\geq Ac\ and\ \varphi(n)\leq \Psi c \\ w(n), in\ remaining\ cases \end{vmatrix}$$

is a result of filtering of the excitation $e'(n)$ by the glottal filter, where

$$u(n)=e'(n)*g(n)=\sum_{m=0}^{\infty}e'(m)g(n-m);$$

$$g(n)=Go(n+1)e^{-cnT};c\approx 200\pi/\text{sec.};$$

The White Noise Generator (WNG) adds a component $w(n)$ which is a model of the turbulent components and the Voice Break Switch (VBS) describes the interruptions of the voice generation, where: $g(n)$ is the impulse response of the glottal filter, *Go*- scale factor, *T*- sampling period (sec.), *Ac* and Ψc- amplitude and frequency break thresholds, c1 and c2- comparators. The convolution of $u'(n)$, the impulse response of the vocal tract filter $v(n)$ and the impulse response of the lip-radiation filter $l(n)$ results into the modeled voice signal

$$x(n)=u'(n)*v(n)*l(n)$$

where $v(n)$ and $l(n)$ are considered invariable because it is assumed that the laryngeal pathology does not affect the vocal tract and the lips.

All $a(n)$, $\varphi(n)$ and $w(n)$ are random time functions. Therefore the task of acoustic evaluation of pathological voices can be regarded as the extraction of specific statistical parameters of these functions which have clinical significance. The method described below includes three separate parts: pitch extraction (demodulation), noise evaluation and long-term components (tremor) analysis.

3.PITCH EXTRACTION

The amplitude and frequency demodulation curves of the voice signal contain information about the time-domain behavior of $a(n)$ and $\varphi(n)$. The period-to-period pitch extraction [10] is the classic type of demodulation used for evaluation of voice pathology [7, 8]. However the irregularity of the disordered voice makes the pitch extraction inaccurate, often impossible.

In order to provide reliable data an adaptive time-domain pitch-synchronous method for pitch extraction was developed. It consists of the following main steps: fundamental frequency (*Fo*) estimation, *Fo* verification, period-to-period *Fo*-extraction and computation of time-domain voice parameters.

The *Fo-estimation* provides preliminary information about the pitch. It is based on short-term autocorrelation analysis with non-linear *sgn*-coding [11] of the voice signal x(n)

$$R(\tau)=\sum_{n=0}^{N-\tau-1}x'(n)x'(n+\tau),\ 0\leq\tau\leq N/2,$$

where: $x'(i)=0\ if\ Pmin<x(i)<Pmax;$

$x'(i)=1\ if\ x(i)\geq Pmax;$

$x'(i)=-1\ if\ x(i)\leq Pmin$

and $Pmax=KpAmax;$

$Pmin=KpAmin;$

Amax and *Amin* - global extremes of the current window in the voice signal $x(n)$. The length of the autocorrelation window is 30ms or 10ms depending on the *Fo*-extraction range (67-625Hz or 200-1000Hz). The sampling rate is 50kHz and every window is low-pass filtered at 1800Hz before coding. The value of the coding threshold at this stage of the analysis is $Kp=0.78$ in order to eliminate the incorrect classification of *Fo*-harmonic components as *Fo* [12]. The current window is considered to be voiced with period $To=\tau max$ if the global maximum is $Rmax(\tau max)>KdR(\tau=0)$, where the voiced/ unvoiced threshold value id $Kd=0.27$ [12].

The *Fo-verification* procedure is similar to the *Fo*-estimation. The autocorrelation function is computed again for the same windows at $Kp=0.45$ in order to suppress the influence of components sub-harmonic to *Fo*. The results are compared to the previous step and the decision about the correct *To* is made for all windows where difference is discovered.

A *period-to-period Fo-extraction* is made on the original signal x(n) using a peak-to-peak extraction measurement. It is synchronous with the verified pitch and voiced/unvoiced results computed in the previous steps. A linear 5-point interpolation is applied on the final period-to-period *Fo*-data in order to increase the resolution. This increased resolution is necessary for meaningful frequency perturbation measurements. The peak-to-peak amplitude is also extracted for every period.

The following *time-domain voice parameters* are computed from the extracted pitch data:

Fundamental frequency information measurements: *Average Fundamental Frequency Fo* /Hz/ [2], *Average Pitch Period To* /ms/, *Highest Fundamental Frequency Fhi* /Hz/, *Lowest Fundamental Frequency Flo* /Hz/, *Standard Deviation of the Fundamental Frequency STD* /Hz/ [5], *Phonatory Fundamental Frequency Range PFR* /semi-tones/, *Length of Analyzed Data Sample Tsam* /sec/ and *Number of Pitch Periods PER.*

Short and long-term frequency perturbation measurements: *Absolute Jitter Jita* /us/ [13], *Jitter Percent Jitt* /%/ [13], *Relative Average Perturbation RAP* /%/ [7], *Pitch Period Perturbation Quotient PPQ* /%/ [8], *Smoothed*

Pitch Period Perturbation Quotient sPPQ /%/ and *Fundamental Frequency Coefficient Variation vFo /%/* [5].

Short and long-term amplitude perturbation measurements: *Shimmer in dB ShdB* /dB/ [13], *Shimmer Percent Shim /%/* [13], *Amplitude Perturbation Quotient APQ /%/* [8], *Smoothed Amplitude Perturbation Quotient sAPQ /%/* and *Peak-to-Peak Amplitude Coefficient of Variation vAm /%/* [5].

Voice break related measurements: *Degree of Voice Breaks DVB /%/* [15] - the ratio of the total length of areas representing voice breaks to the time of the complete voiced sample; and *Number of Voice Breaks NVB.* The criteria for voice break area can be a missing impulse for the current period or an extreme irregularity of the pitch period.

Sub-harmonic components related measurements: *Degree of sub-harmonics DSH /%/-* the ratio of the number of auto-correlation windows with incorrect sub-harmonic period classification to the total number of autocorrelation windows; and *Number of Sub-Harmonic Segments NSH*

Voice irregularity related measurements: *Degree of Irregular Vocalization DUV /%/* [15]- the ratio of the number of auto-correlation windows classified as unvoiced to the total number of autocorrelation windows; and *Number of Unvoiced Segments NUV.*

4.NOISE EVALUATION

The analysis of the voice signal in the frequency domain provides another approach to the evaluation of its irregularity (noise). The amount of in-harmonic spectral components correlates to the perception of hoarseness of the pathological voice [16]. *To* evaluate the level of noise components and separate the turbulent noise correlating to the intensity of the function *w(n)*, a pitch-synchronous frequency-domain method was developed. The following parameters are extracted: *Noise to Harmonic Ratio NHR-* a general evaluation of the noise presence in the analyzed signal (including amplitude and frequency variations, turbulence noise, sub-harmonic components and/or voice breaks); *Voice Turbulence Index VTI-* mostly correlating with the turbulence components caused by incomplete or loose adduction of the vocal folds; and *Soft Phonation Index SPI-* an evaluation of the poorness of high-frequency harmonic components that may be an indication of loosely adducted vocal folds during phonation.

The algorithm consists of the following general procedures:

1. Election of two groups of windows of 81.92 ms (4096 points) of the voice signal. The first group includes a sequense of windows of the voiced areas in the analyzed signal with a half window overlap. The second group includes four non-contiguous windows, where the frequency and amplitude perturbations are the lowest for the signal.
2. For every window in both groups the following steps apply: low-pass filtering (cutoff 6000Hz, order 22, Hamming window), downsampling to 12.5kHz and conversion of the real signal into analytical one using Hilbert filtering; computation of the power spectrum of the window using a 1024-points Complex Fast Fourier Transform (FFT) on the analytical signal; calculation of the average fundamental frequency within the current window from the time-domain analysis data and synchronous harmonic/in-harmonic separation; computation of the current window's *NHR*, *SPI* and *VTI*. *NHR* is a ratio of the in-harmonic energy in the range 1500-4500Hz to the harmonic spectral energy (70-4500 Hz) and *SPI* is a ratio of the lower-frequency (70-1600Hz) to the higher-frequency (1600-4500Hz) harmonic energy for the first group of windows. *VTI* is a ratio of the spectral in-harmonic high-frequency energy (2800-5800Hz) to the spectral harmonic energy (70-4500Hz) for the second group of windows.
3. Computation of the average values of NHR, SPI and VTI.

5.TREMOR ANALYSIS

The pitch extraction process yields the amplitude and frequency demodulation curves of the voice signal. These curves contain information about the long-term amplitude and frequency variability (tremor) of the voice signal [17]. Methods for frequency and amplitude tremor analysis are developed. The algorithm for frequency tremor analysis includes the following steps:

1. Division of the *Fo*-data resulting from pitch extraction into windows of 2 sec. length with 1 sec. step overlap.
2. Apply the following procedures to every window: low-pass filtering of the *Fo*- data (cutoff 30Hz) and downsampling to 400Hz; calculation of the total energy of the resulting signals; subtraction of the DC-component and computation of the autocorrelation function on the residual signal; division of the autocorrelation data by the total energy and accumulation of the results from every window. The maxima of the resulting autocorrelation curve show the intensity and frequency of the long-term (up to 30Hz) frequency-modulating components.
3. Calculation of the *Fo-Tremor Intensity Index FTRI /%/-* the value of the global maximum of the average autocorrelation curve and the corresponding position *Fo-Tremor Frequency Fftr /Hz/*

The same method applies for computation of the *Amplitude Tremor Intensity Index ATRI /%/* and the *Amplitude-Tremor Frequency Fatr /Hz/* from the peak-to-peak amplitude data resulting from pitch extraction.

6.APPLICATION

Based on the model and the methods described above a *Multi-Dimensional Voice Program MDVP* was developed utilizing the *Computerized Speech Lab* (CSL) model 4300 (Kay Elemetrics Corp.). CSL, a hardware/software system which uses an MS-DOS based computer as host, includes signal conditioning, 16-bit A/D converters, dual digital signal processors (DSP16A & TMS32025) and support peripherals. The MDVP system computes a set of 33 acoustic

voice parameters in about 16 seconds and provides flexible routines for graphical representation of the results[Fig.2-3]. Also a user-upgradable voice database allows automatic comparison of the current results with different nosological groups.

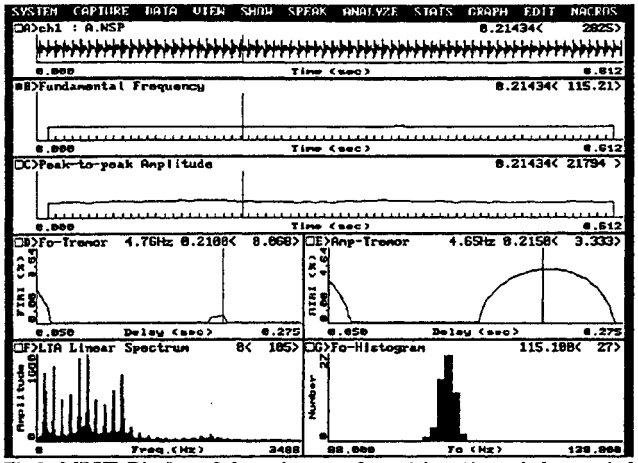

Fig.2: MDVP-Display of the voice waveform (view A), period-to-period fundamental frequency (B), peak-to-peak amplitude (C), *Fo*-tremor (D) and amplitude tremor (E) autocorrelation curves, long-term average linear spectrum of the voiced areas of the signal (F) and histogram of the distribution of *Fo* (G).

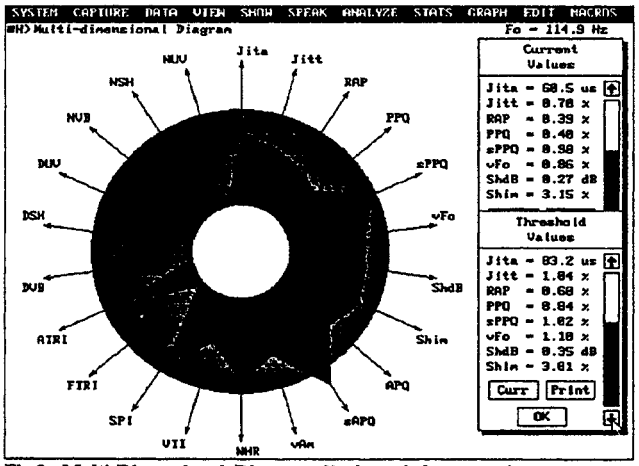

Fig.3: Multi-Dimensional Diagram display of the acoustic parameters. The area within the circle shows the normative threshold range and the polygon - the currently computed values.

In order to extract the normative threshold values of the acoustic parameters sustained phonation of the vowel 'a' of 15 persons (7m,8f) with normal voice production and of 53 patients (25m,28f) with laryngeal diseases were analyzed using the MDVP system. The following nosological groups were included in the study: laryngeal cancer, benign neoplasms, chronic laryngitis, functional dysphonia and paralysis of a recurrent nerve. The computed normative threshold values for this database are:

Frequency perturbation measurements:

Jita	Jitt	RAP	PPQ	sPPQ(55p)	vFo
83.2 us	1.04 %	0.68 %	0.84 %	1.02 %	1.10%

Amplitude perturbation measurements:

ShdB	Shim	APQ	sAPQ(55p)	vAm
0.35 dB	3.81 %	3.07 %	4.23 %	8.20 %

Voice break, sub-harmonic and voice irregularity measurements:

DVB	DSH	DUV	NVB	NSH	NUV
0 %	0 %	0 %	0	0	0

Noise and tremor evaluation measurements:

NHR	VTI	SPI	FTRI	ATRI
0.19	0.061	14.12	0.95 %	4.37 %

The normative values may vary depending on the nosological groups included in the specific study. A separate database is recommended to be selected or created for different applications.

REFERENCES

[1]. Fant, C.G.M. *Acoustic theory of Speech Production.* The Hague: Mouton 1960.

[2]. Davis, S. *Acoustic Characteristics of Normal and Pathological Voices. Speech and Language Research and Theory.* Academic Press. N.J. 1979.

[3]. Titze, I., Baken, R., Herzel, H. Evidence of chaos in vocal fold vibration. *Vocal Fold Physiology.* Edited by Ingo Titze. Singular Publishing, USA. 1993.

[4]. Klatt, D.H., Klatt, L.C. Analysis, synthesis, and perception of voice quality variations among female and male talkers. J.Acoust.Soc.Am. **87**, (2), 820-836, 1990.

[5]. Hirano, M. *Clinical Examination of Voice.* Springer Verlag. Vienna. 1981.

[6]. Kent, R. Vocal Tract Acoustics. Journal of Voice, Vol.7, No.2, 97-117, 1993.

[7]. Koike, Y. Application of some acoustic measures for the evaluation of laryngeal dysfunction. Stud.Phonol.**VII**,17-23,1973.

[8]. Koike,Y, Takahashi,H.,Calcaterra,T. Acoustic measures for de-tecting laryngeal pathology. Acta Laryngol. **84**, 105-117, 1977.

[9]. Deliyski, D. *Digital Processing of Voice Signals in the Diagnosis of Laryngeal Diseases.* Doctoral Dissertation, Bulgarian Academy of Sciences, Institute of Industrial Cybernetics and Robotics, Sofia, Bulgaria /in Bulgarian/ 1990.

[10]. Hess, W. *Pitch Determination of Speech Signals. Springer Verlag.* N.Y. 1983.

[11]. Rabiner, L. On the use of autocorrelation analysis for pitch detection. IEEE Trans. ASSP. Vol.ASSP-22, No.3, 1974.

[12]. Deliyski, D. Investigation of the autocorrelation function characteristics in pathologic voice signal analysis. 3-th International Conf. on Statistical Theory of Communications STS'88, Varna, Bulgaria, pp.17 /in Russian/, 1988.

[13]. Pinto, N., Titze, I. Unification of perturbation measures in speech signals. J.Acoust.Soc.Amer. **87**, (3), 1278-1289, 1990.

[14]. Hillenbrand,J. Perception of aperiodicities in synthetically generated voices. J.Acoust.Soc.Amer. **83**(6),2361-2371,1988.

[15]. Nikolov, Z., Deliyski D., Drumeva L., Boyanov B. Computer system for diagnostics of pathological voices. in Proc: XXI-st Congress International Association of Logopedics and Phoniatrics. Prague, Czechoslovakia, Vol.1, 973-976, 1989.

[16]. Kasuya, H., Ogawa Sh., K.Mashima K., Ehibara S. Normalized noise energy as an acoustic measure to evaluate pathologic voices. J.Acoust.Soc.Amer. **80**, (5), 1986.

[17]. Winholtz W., Ramig L. Vocal tremor analysis with the vocal demodulator. J. Speech Hearing Res., Vol.35, 562-573, 1992.