A software system for pathological voice acoustic analysis

B. Boyanov, B. Teston, S. Hadjitodorov, P. Mitev, D. Donskov

A SOFTWARE SYSTEM FOR PATHOLOGICAL VOICE ACOUSTIC ANALYSIS

Boyanov, B., B. Teston *, S. Hadjitodorov, P. Mitev, D. Doskov**

Center on Biomedical Engineering, Acad. G. Bonchev St, Block 105, 1113 Sofia, Bulgaria
fax: +3592 723 787 E-mail:STHADJ@BGEARN.ACAD.BG
* Lab."Parole et Langage",ESA 6057 Universite de Provence, 29, av R. Schuman, 13621, Aix en Provence, France
fax: 0033 4 42 59 50 96, E-mail: teston@lpl.univ-aix.fr
** Phoniatric Dep. Univ. Hospital "Queen Jovanna" Sofia, Bulgaria

Key words: acoustic voice analysis, pitch period, spectral and cepstral analysis

ABSTRACT

A software system for pathological voice analysis using only the resources of a personal computer with a sound card is proposed. The system is written on the basis of specific methods and algorithms for pathological voice analysis and allows evaluation of: 1) Pitch period (To); 2) Degree of unvoiceness; 3) Pitch perturbation and amplitude perturbation quotients; 4) Dissimilarity of surfaces of the pitch pulses; 5) Ratio aperiodic/periodic components in cepstra; 6) Ratio {energy in the cepstral pitch pulse}-to-{total cepstral energy}; 7) Harmonics-to-noise ratio; 8) Degree of hoarseness; 9) Ratio low-to-high frequency energies; 10) Glottal Closing Quotient. The voices of 400 persons were analyzed - 100 (50 females/50 males) normal speakers and 300 (100 females/200 males) patients. The statistical analysis shows very significant changes in PPQ, DH, DPP, DUV, APR, HNR and PECM, and significant changes in APQ and CQ.

1. INTRODUCTION

Many experimental researches on dysphonia have shown that acoustic voice analysis is an effective and non invasive tool for laryngeal pathology evaluation and detection. In the present paper a system for acoustic analysis of pathologic voice, using only the resources of a personal computer with a sound card is proposed.

2. ACOUSTIC VOICE ANALYSIS

Pitch period (To) evaluation

The To evaluation are very important for pathological voice analysis, because wrong To will produce fatal errors in the voice parameters. To minimize these errors the robust detector [4] is implemented.

Determination of pitch pulses and separates pitch periods (period per period)

The pitch peaks for each segment are detected by the geometric center of mass, using the procedure [2].

Formation of stable zones

Detection of stable zones (segments, where. To is nearly constant - deviations less than 5%).

Quantitative evaluation of the stability of pitch generation during sustained phonations

I. The following voice parameters are proposed:

1. Degree of unvoiceness (DUV) is the ratio: {Length ($Length_{unv}$) of unvoiced segments} - to - { Length ($Length_{vo}$) of voiced segments}

2. Dissimilarity of surfaces of the pitch pulses (DPP):

Step1**.** Calculation of the surfaces (SURF(i), for i=0,.., I) of the pitch pulses in the stable zone. For i-th pitch pulse SURF(i) is calculated by:

$$SURF(i) = \sum_{t=bi}^{endi} x(t) \quad (1)$$

where: x(t) - samples forming the pitch pulses;

$A_{max}(i)$ - i-th pitch pulse;

b_i - index of the first sample after the zero crossing on left of $A_{max}(i)$;

end_i - index of the first sample after the zero crossing on the right of $A_{max}(i)$.

Step 2. Calculation of DPP for the stable zone by:

$$\mathrm{DPP} = \frac{1}{I}\sum_{i=1}^{I} abs(\mathrm{SURF(i)} - \mathrm{SURF(i+1)}) \qquad (2)$$

where: I - number of pitch pulses in the zone.

II. The following widely used voice parameters are calculated by the formulae described in [7]:

1. Pitch period (PPQ) perturbation quotient.
2. Amplitude (APQ) perturbation quotient.

Spectral analysis

Formation of stable segments and spectral analysis by 2048-point - FFT.

Evaluation of the ratio low-to-high energy using the equation described in [3].

Cepstral analysis

The cepstra (c(t)) for the segment is calculated pitch synchronously.

I. The following voice parameter is proposed:

The parameter Ratio aperiodic (noisy) - to - periodic (harmonic) (APR) components is calculated by means of the following algorithm:

Step 1. Detection of the pitch peak (c (To)) in the cepstra.

Step 2. Calculation of a threshold:

$$THR_{cep} = qc\,(To). \qquad (3)$$

where: q = 0.6.

Step 3. Detection of cepstral pitch peaks (cPP(k), for k=1,.. K) by amplitude selection of cepstral maximas at quefrencies kTo - (c(kTo):

if c(kTo) > THR_{cep}

then cPP(k) = c(kTo)

else cPP(k) = 0.

Step 4. Calculation of the energy ($E_{cp}(k)$) for every cPP(k) by:

$$E_{cp}(k) = \sum_{t=\mathrm{beg(k)}}^{end(k)} c(t) \qquad (4)$$

where: c(t) - samples forming the pitch pulses;

beg(k) - index of the first sample after the zero crossing on left of cPP(k)

end(k) - index of the first sample after the zero crossing on the right of cPP(k).

Step 5. Calculation of periodic components energy (E_{per}):

$$E_{per} = \sum_{k=1}^{K} E_{cp}(k) \qquad (5)$$

where: K=3, because our experiments for visual analysis of the cepstrum for 50 normal speakers and 80 patients, show that:

a) for normal voices three cepstral pitch peaks are present for 98% of the phonations;

b) for pathologic voices three cepstral pitch peaks are present for 95% of the phonations;

Step 6. Liftering to eliminate cepstral components with lower quefrency than the pitch period (c(t)=0 if t>0.9To).

Step 7. Calculation of the energy (E_{clif}) of the liftered cepstra. It is assumed that E_{clif} represents the energy of the noisy components

Step 8. The APR is evaluated by:

$$APR = E_{clif}/E_{per} \qquad (6)$$

II. The following, proposed by the authors in [2] voice parameter is calculated:

The parameter ratio {energy concentrated in the cepstral pitch impulse}-to-{total cepstral energy}-PECM.

Evaluation of the quantity of noisy components in the voice

Most of the laryngeal and many neurological diseases are correlated with the presence of noisy components in the voice. In order to obtain more reliable results the quantity of these components is evaluated both in spectral and time domains.

1. Analysis in spectral domain.

The degree of hoarseness (DH) is calculated as the noise-to-harmonics ratio by means of the approach [9]. The calculation of the DH is realized over the high resolution spectra of the stable segments in order to:

a) Minimize distortions in the spectrum, because segments (here zones) containing integer number of periods with constant values are analyzed;

b) Eliminate distortions in the spectrum caused by the jitter.

2. Analysis in time domain. The harmonics-to-noise ratio (HNR) is evaluated by means of a modification of the method [8]. This method is implemented, because it is used in several systems for voice analysis. However, in this method is observed an artificial increase of the

energy of the noise components due to the differences in the values of the pitch periods rather
than to the presence of noisy components. In order to solve this problem the following modification of the method is proposed:

1. The evaluation of HNR is carried out only over the stable zones in order to analyze portions of the signal, where To is with constant values. Only zones longer than 50 To are analyzed in order to obtain stable value for HNR.
2. The mean To for the zone is used as the reference period for HNR calculation, rather than the longest period.

Glottal analysis

The glottal waveforms are obtained by pitch synchronous iterative adaptive inverse filtering [1] and the Closing Quotient (CQ) is calculated.

Averaged parameter's values

Weighted mean values are calculated using the procedure described in [3].

Hardware requirements of the system

Only the resources of a personal computer with a high quality sound card (like "Sound Blaster") and an electret microphone are used. Electret microphones are recommended, because they have good frequency response curves.

3. EXPERIMENTAL RESEARCH

During a period of one year in the Phoniatric Department the voices of 400 persons were recorded - 100 (50 females/50 males) normal speakers and 300 (100 females/200 males) patients. All sustain the vowel "a" every 4 months three times.

The signals were quantified directly into the computer's memory at 22 kHz and resolution 16 bit using computer's sound card "Sound Blaster" and its "SONY" electret microphone.

The evaluation of statistical significance of the calculated parameters was done by ANOVA and t-test (described in [5]).

RESULTS

The statistical analysis shows (tables 1 and 2): a) Very significant changes in DH, DPP, DUV, APR and PECM. b)Significant changes in PPQ, APQ, HNR and CQ. c)Non significant in RLH for female speakers.

REFERENCES

[1] Alku P., Vilkman E: Amplitude domain quotient for characterization of the glottal volume velocity waveforms estimated by IF. Speech Com., v. 18, 131-138, 1996.

[2] Boyanov B., S. Hadjitodorov. "Acoustic analysis of pathological voices and screening of laryngeal diseases", IEEE Eng. in Medecine & Biology, v. 16, 74 - 82, 1997.

[3] Boyanov B, Hadjitodorov S, Teston B., Doskov D. Software system for pathological voice analysis. In: Proc. Larynx'97, June 16-18,1997, 139-142.

[4] Boyanov B, Hadjitodorov S, Doskov D and Teston B. Robust hybrid pitch detector for pathological voice analysis,. In: Proc. Larynx'97, Marseille, June 16-18, 1997, pp. 55-58.

[5] Dixon W: BMDP-81 Biomedical Computer Programs, Los Angeles, University of California Press, 1981.

[6] Hess W: Pitch determination of speech signals. Springer Verlag. N.Y. 1983.

[7] Koike Y: Application of some acoustic measures for the evaluation of laryngeal dysfunction. Studia Phon. v. 5, 717-723, 1973.

[8] Yumoto E: The quantitative evaluation of hoarseness. A new harmonics to noise ratio method. Archives of Otolaryngology, vol. 109, pp. 48-52, 1983.

[9] Yunik M., Boyanov B: Method for evaluation of the noise-to-harmonic-component ratios in pathological and normal voices. Acustica, 70, 89-91, 1990.

	NORMAL		PATHOLOGY		Values of
Parameter	mean	std. dev.	mean	std. dev.	T-test
PPQ[%]	0.281	0.114	0.815	0.406	12.22
APQ[%]	0.92	0.59	1.4	0.73	4.33
DUV [%]	99.8	1.41	94.06	4.23	12.27
DPP[abs]	0.021	0.003	0.032	0.01	10.12
APR [%]	11.7	1.1	28.02	4.6	33.6
PECM[abs]	0.372	0.02	0.302	0.033	16.1
HNR[dB]	14.52	6.73	11.75	2.381	2.8
DH[abs]	0.28	0.091	0.35	0.17	3.28
RLH [abs]	1.03	0.59	0.92	0.63	1.05
CQ [abs]	6.12	2.14	9.22	3.15	7

Table 1. Mean values and standard deviations of the evaluated parameters and values of the T-test for normal and pathological voices of female persons.

	NORMAL		PATHOLOGY		Values of
Parameter	mean	std.dev.	mean	std. dev.	The T-test
PPQ[%]	0.631	0.467	1.264	0.959	6.68
APQ[%]	3.027	1.23	4.811	3.27	6.16
DUV [%]	97.2	3.25	90.58	7.32	9.56
DPP[abs]	0.043	0.02	0.11	0.073	11.38
APR [%]	18.2	2.3	32.15	5.1	28.72
PECM[abs]	0.346	0.038	0.253	0.041	15.23
HNR[dB]	10.45	9.828	2.34	1.876	5.81
DH[abs]	0.22	0.027	0.30	0.07	12.79
RLH [abs]	1.72	0.41	1.15	0.31	9.19
CQ [abs]	10.32	8.12	19.35	16.45	5.52

Table 2. Mean values and standard deviations of the evaluated parameters and values of the T-test for normal and pathological voices of male persons.