OBJECTIVE DIAGNOSIS OF LARYNGEAL PATHOLOGY USING THE WIGNER-VILLE DISTRIBUTION

Dimitar DELIYSKI Tomasz ZIELINSKI*

Bulgarian Academy of Sciences, Institute of Industrial Cybernetics and Robotics, Acad.G.Bonchev str., bl.2, 1113 Sofia, BULGARIA
**Institute of Electrical Engineering IMISUE, University of Mining and Metallurgy, Al.Mickiewicza str. 30, 30-059 Krakow, POLAND*

RESUME

Cette conférence présente une nouvelle possibilité d'application de la *distribution de Wigner-Ville* (WVD), comme un moyen de traitement numérique du signal vocal. La méthode décrite est assignée de donner des évaluations quantitatives de certains paramètres, utilisés dans la pratique médicale. Elle comprend deux niveaux de traitement principaux: analyse en *domaine temporel* et analyse *WVD temporelle-fréquencielle.* L'analyse en domaine temporel représente une suite adaptive de mesurages de la fréquence principale momentanée $Fo^{(i)}$, précédée d'une estimation autocorrélative des limites de la fréquence principale de la voix et de sa régularité. Les paramètres temporels suivants sont déterminés: degré de nonvocalisation, degré de subharmoniques, degré d'interruptions et degré de perturbations de la fréquence principale et de l'amplitude des impulsions du signal. La version de la WVD appliquée (OTSWVD) est optimisée par un polissage temporel, synchronic aux variations temporelles de $Fo^{(i)}$. La valeur moyenne du *degré de raucité* (DH_{wv}) du spectre d'OTSWVD est évalue. Elle diffère de celle du degré de raucité (DH_{sp}), extracté de la *spectrogramme* conventionelle, par une sensibilité augmentée en pathologie, grace aux *interférences* nonpolissées en cas de nonstationarité de la voix. Finalement, des résultats comparatifs des DH_{sp} et DH_{wv} sont présentés.

SUMMARY

This paper presents a new possibility for application of the *Wigner-Ville distribution* (WVD) - as a tool for voice signal processing, used in objective diagnostics of laryngeal pathology. The described method for quantitative evaluation of some acoustic parameters, used in medical practice, covers two main levels of processing: *time-domain* and *time-frequency WVD* analysis. The time-domain analysis includes autocorrelation pitch frequency estimation, voiced/unvoiced decision and adaptive peak-to-peak measure of momentum frequencies ($Fo^{(i)}$) in time direction. A synchronous with $Fo^{(i)}$ *optimal time-smoothed WVD* (OTSWVD) analysis is discussed and an algorithm for the evaluation of the *degree of hoarseness* (DH_{wv}) from the OTSWVD spectrum, where the cross-terms are used, is proposed. Finally, the compared experimental results of DH_{wv} and DH_{sp} extracted from the conventional *spectrogram* are presented.

1. INTRODUCTION

The traditional methods of laryngeal diseases diagnosing are subjective (perceptual) or semi-objective. Hearing evaluation, indirect and direct laryngoscopy and stroboscopy fall under the first group. The spectrographic (sonagraphic) analysis of voice is the general semi-objective method permitting visualization of the time-frequency representation of signals [1].

Lately a tremendous effort is made at working out digital methods for *objective* (quantitative) evaluation of acoustic voice parameters with the purpose of reliable disease diagnosis. Some parameters as pitch frequency and amplitude perturbation quotients (FPQ, APQ) [2] and degree of hoarseness (DH) [3], computed by different time-domain or frequency-domain techniques, are fundamental in the pathological voice researches [4,5,6].

The DH is usually defined as the ratio of the energies of inharmonic to harmonic components within the power spectrum of the signal [3,7]. The *spectrogram* enables the determination of the average degree of hoarseness DH_{sp} for all the spectra of the time-frequency representation. The spectrogram in its turn can be shown to be a smoothed version of another time-frequency representation - the *Wigner-Ville distribution* (WVD) [8]. Due to the smoothing process involved, many details of the signal time-frequency structure depicted by WVD in the spectrogram are blurred or altogether invisible. WVD is thus the basis for time-frequency signal analysis with significantly improved resolution [9]. The application of WVD, however, is complicated by the interference terms [10] which have to be partially suppressed.

A method for assessing acoustic parameters, necessary for voice disorders diagnosis is described here.

The experimental DH results extracted from the spectrogram and WVD for normal and pathological voices are compared.

Only a voiced speech - vowel "A" is analysed because of its strong harmonic structure and high first formant (F_1). The preliminary processing of the signal includes two-channel analogue filtration and synchronous analogue-to-digital conversion with a frequency band of 70-700Hz, 60dB/oct for the first channel and of 70-6000Hz, 60dB/oct for the second one. The sampling rate is 16KHz.

The method includes two general levels of signal processing: *time-domain* analysis and *time-frequency WVD* analysis.

2. TIME-DOMAIN ANALYSIS

The time-domain processing of normal and pathological voice is discussed in details in [11]. It insures relevant information about the momentum pitch frequency ($Fo^{(i)}$) variations and interruptions in time and the evaluation of some time-domain informative parameters. It is used then for a synchronous with $Fo^{(i)}$ WVD-analysis. The sampled signal $x_1(t)$ from the first channel is used here, because it is necessary to eliminate the influence of F_1 and to provide the required accuracy [12].

2.1. Pitch frequency estimation

First the *pitch frequency range* is extracted making use of the short-time autocorrelation function with nonlinear sign($x_1(t)$) coding [13]

$$R^{(i)}(\tau)=\sum_{t=0}^{N^{(i)}-\tau-1} x_1(t)x_1(t+\tau), \quad 0 \le \tau \le \frac{N^{(i)}}{2},$$

where:

$$x_1(t)x_1(t+\tau)=\begin{cases} 0, \text{if } P_{min} < x_1(t) < P_{max}, \\ \quad \text{or } P_{min} < x_1(t+\tau) < P_{max}, \\ 1, \text{if } x_1(t)\ge P_{max} \text{ and } x_1(t+\tau)\ge P_{max}, \\ \quad \text{or } x_1(t)\le P_{min} \text{ and } x_1(t+\tau)\le P_{min}, \\ -1, \text{if } x_1(t)\ge P_{max} \text{ and } x_1(t+\tau)\le P_{min}, \\ \quad \text{or } x_1(t)\le P_{min} \text{ and } x_1(t+\tau)\ge P_{max}, \end{cases}$$

and

$$P_{max}=K_pA_{max},$$
$$P_{min}=K_pA_{min}.$$

$R^{(1)}(\tau)$ is computed first with a time window $N^{(1)}= 2To_{max}=$ 30ms at the begining of $x_1(t)$, where: To_{max} is the maximal human pitch period To ($2.5ms\le To\le 15ms$ [14]), A_{max} and A_{min} are the global extrema of $x_1(t)$ in the analysed segment and K_p=0.78 [12].

The global maximum $R^{(i)}_{max}(\tau_{max})=\max\{R(\tau)\}$ for $2.5ms\le\tau\le 15ms$ is found. If:

$$R^{(i)}_{max}(\tau_{max}) \ge K_dR(\tau=0) ,$$

where K_d=0.27 [12], the analysed segment $N^{(i)}$ is classified as voiced, otherwise - as unvoiced. In the case of unvoiced decision the next $R^{(i)}(\tau)$ with time window $N^{(i)}=N^{(1)}$ is called. If a voiced segment is found the computed pitch period is $To_c^{(i)}=\tau_{max}$.

Excepting the cases of abrupt changes in the pitch or amplitude within the analysed segment, classified as unvoiced (pathologic), errors of the type $To_c^{(i)}$=eTo are possible, where: To- correct pitch period; e=1/3,1/2- harmonic errors, due to the signal structure; e=2,3- subharmonic errors, due to the nonstationary character of the pathologic voice and e=1- no errors. Another sixteen $R^{(i)}(\tau)$ with *adaptive time window* $N^{(i)}=nTo_c^{(i-1)}$ are called, where: n=2 if $To_c^{(i-1)}$>7.5ms, n=4 if $5ms\le To_c^{(i-1)}\le 7.5ms$ and n=6 if $To_c^{(i-1)}$<5ms. These adaptive autocorrelation functions are computed only for estimated values of τ. The resulting sequence of $To_c^{(i)}$ periods insures statistic data for pitch period range corrective decision [11].

The advantages of the described algorithm for pitch frequency range estimation are increased accuracy, particulary for pathologic voices and high speed [12], both due to the use of autocorrelation analysis, the adaptive windows, including exact number of pitch periods, the nonlinear coding, the voiced/unvoiced decision and the preliminary analogue filtration of $x_1(t)$.

2.2. Momentum pitch frequencies tracing

Proceeding from the aproximate information in 2.1. about the pitch period at the begining of $x_1(t)$, synchronous with To_c time-domain peak-to-peak distances measurement becomes possible. All values of extrema $A^{(i)}_{min}$ and $A^{(i)}_{max}$, corresponding to the glottal impulses [2,4], their time positions $T^{(i)}_{min}$, $T^{(i)}_{max}$ and the distances between $T^{(i)}_{min,max}$ and $T^{(i-1)}_{min,max}$, corresponding to the *momentum pitch periods* $To^{(i)}$ are determinated. This procedure continues throughout the analysed voice signal. If the amplitude or frequecy and phase conditions [11]:

$$A^{(i)}_{min,max} \le \frac{0.18}{i-1}\sum_{n=1}^{i-1} A^{(n)}_{min,max} ,$$

$$\frac{\left|T^{(i)}_{O} - \frac{1}{i-1}\sum_{n=1}^{i-1} T^{(n)}_{O}\right|}{\frac{1}{i-1}\sum_{n=1}^{i-1} T^{(n)}_{O}} \ge 0.25$$

are accomplished during analysis in (i)-position, it is considered as a point of *interruption of the pitch generation*. The time point $T^{(i-1)}_{min,max}$ is accepted to be the begining of an interruption zone. The restoration point of pitch generation is found again using the procedure described above in 2.1.

2.3. Time-domain parameters evaluation

Some results from the time-domain analysis of $x_1(t)$ can be used for quantitative evaluation of voice pathology [15]. The following informative parameters [11] are determined:

(1) *Degree of unvoiceness* (DUV,%), as a ratio of the number of unvoiced decisions in 2.1. to the total number of autocorrelation functions (N_{ac}).

(2) *Degree of subharmonics* (DSH,%), as a ratio of the number of segments, with $To_c^{(i)}$=eTo, e=2,3 (see 2.1.) to N_{ac}. This parameter correlates with the dyplophonic voice effect.

(3) *Degree of interruptions* (DI,%), as a ratio of the summary time duration of all interruption zones (see 2.2.) to the total duration of the analysed signal $x_1(t)$.

(4) *Pitch frequency perturbation quotient* (FPQ,%) [2], as

$$FPQ = \frac{\frac{1}{M-k+1}\sum_{n=1}^{M-k+1}\left|\frac{1}{k}\sum_{r=1}^{k} F_O^{(n+r-1)} - F_O^{(n+m)}\right|}{\frac{1}{M}\sum_{n=1}^{M} F_O^{(n)}} ,$$

where: $Fo^{(i)}=1/To^{(i)}$ - momentum pitch frequencies, m=1/2(k-1), k=7 and M- total number of $To^{(i)}$ values, measured in $x_1(t)$.

(5) *Peak-amplitude perturbation quotient* (APQ,%)[2], similar to FPQ, defined as a mean value of perturbation quotients for $A^{(i)}_{min}$ and $A^{(i)}_{max}$, where M is the total number of $A^{(i)}_{min,max}$ values.

The borders of variations of these parameters and their diagnostic potential are discussed in [15,16].

3. TIME-FREQUENCY WVD ANALYSIS

The *Wigner-Ville distribution*

$$WVD_x(t,f)=\int_\tau x(t+\tfrac{\tau}{2})x^*(t-\tfrac{\tau}{2})e^{-j2\pi f\tau}d\tau \quad \in \mathbb{R}$$

where: x(t)- signal, t- time, f- frequency, has the advantage of *high resolution* and the shortcomming of *interference (cross) terms* in the time-frequency spectrum.

3.1. Voice implementation of the WVD

The voiced speech has a strong harmonic structure. The ideal voice can be shown as a superposition of complex sinusoids

$$x(t)=\sum_{k=1}^{H} A_k \exp(j2\pi f_k t),$$

where: f_k=kFo, Fo- pitch frequency, A_k- amplitude of k-harmonic, H- number of Fo-harmonics in the spectrum. The interference terms in frequency-domain are situated in each point $f_{ij}=(f_i+f_j)/2$, i,j=1,2,...H, oscillating in the time direction with a period of $td_{ij}=1/fd_{ij}$, where $fd_{ij}=|f_i-f_j|$.

These cross-terms can be suppressed by time-frequency smoothing. The *spectrogram* is a time-frequency smoothed version of the WVD [8,9]. An advantage of the *smoothed pseudo WVD* [17] is the possibility of independent time and/or frequency smoothing. Because of the time-frequency structure of the WVD spectrum the *time-smoothed WVD* (TSWVD)

$$TSWVD_x(t,f)= WVD_x(t,f) \underset{t}{*} u(t) \;,$$

where u(t)- time-smoothing window, suppresses the time oscillations in the f_{ij}-points, reducing the time resolution but preserving the frequency resolution.

The cross-terms suppression degree depends on the lenght t_u of u(t). Fortunately the voice pitch period To=1/Fo is divisible by all time-oscillating periods td_{ij}. Then, if t_u=To, an *optimal time-smoothing* of voice is attainable.

In the ideal-voice (stationary) case the cross-terms are removed by optimal time-smoothing. The normal real voice is accepted as being pseudo-stationary for short-time representations [1,14]. The pathologic voice characteristics differ from the normal ones by an increased nonstationarity [5]. Because of the high frequency resolution of the TSWVD, all nonstationary factors (pitch variations, noise, interruptions) bring about an increased degree of interference terms into the TSWVD-spectrum, which can be used as a quantitative measure of voice pathology.

The optimal TSWVD's *time resolution* is $\Delta t=t_u$=To. Then the TSWVD momentum spectrum can be called *"pitch spectrum"*.

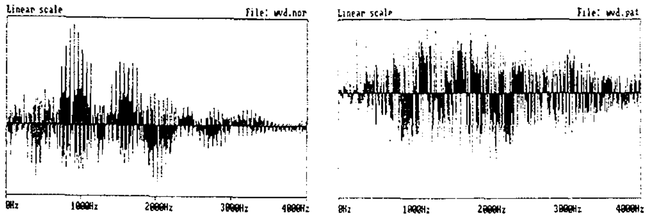
Examples for optimal TSWVD of normal and pathologic voices and their comparision with WVD and the spectrogram are shown in Fig.1-5. The experiments are realized using the *Computer Development System for WVD-Analysis* ellaborated by [18].

3.2. DH evaluation using WVD

The *average degree of hoarseness using optimal TSVWD* (DH_{wv}) is introduced as a quantitative measure of the interference terms degree in the TSWVD spectrum. It determines the nonstationarity of the voice signal, discussed in 3.1. The sampled signal $x_2(t)$ from the second channel (see Introduction) is used in this processing. The algorithm for DH_{wv} extraction includes the following procedures:

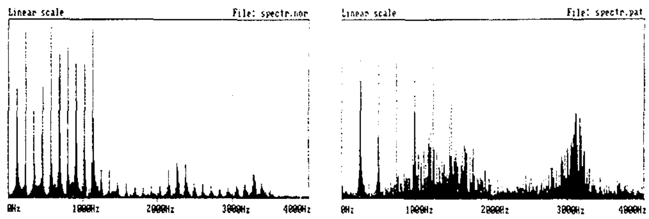
(1)The *complex analytic signal* $\hat{x}_2(t)$ from $x_2(t)$ is determined using Hilbert filtration.

(2)The *optimal TSWVD time-frequency spectrum* from $\hat{x}_2(t)$ is computed. The *time step* between TSWVD's "pitch spectrums" is Δt=1/Fo, where Fo- average pitch frequency value for $x_1(t)$, computed from 2.2. The interruption zones are not analysed. The WVD algorithmic implementation is described in [19]. A rectangular windowed 1024-point complex fast Fourier transform (FFT) is used. Then the *frequency resolution* is Δf=7.81Hz at a sampling rate of 16KHz.



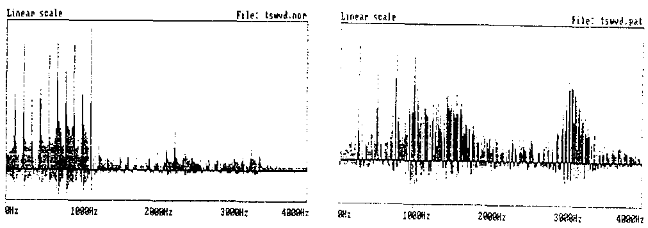
a) b)

Fig.1.WVD momentum spectra of a)normal and b)pathological voices.



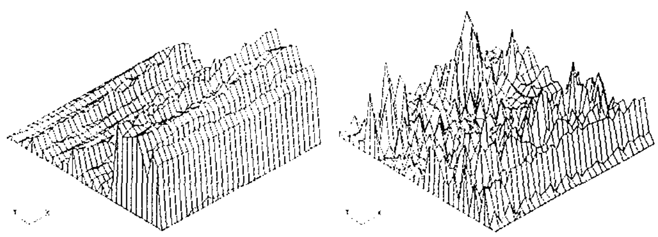
a) b)

Fig.2.Spectrograms of a)normal and b)pathological voices.



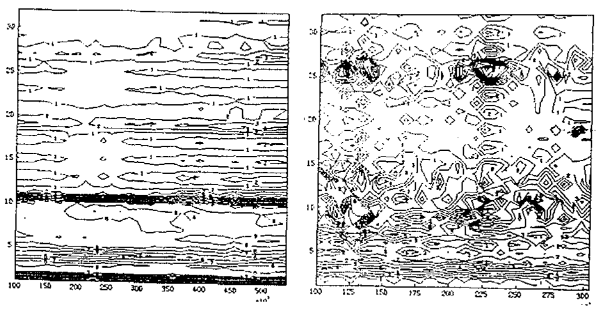
a) b)

Fig.3.Optimal TSWVD "pitch spectra" of a)normal and b)pathological voices.



a) b)

Fig.4.Three-dimensional plots of optimal TSWVD spectra (Δt=To) for a)normal and b)pathological voices.



a) b)

Fig.5.Contour plots of optimal TSWVD spectra (Δt=To) for a)normal and b)pathological voices.

(3)The *mean values* $F_0^{(n)}$ *of the momentum pitch frequencies* for each rectangular FFT window $w^{(n)}(t)$ are computed synchronously with 2.2., where n- number of "pitch spectra" in the TSWVD. Then the *optimal time-smoothing window lenghts* are $t_w^{(n)}=1/F_0^{(n)}$. The sampling rate errors of $F_0^{(n)}$ for 1024-point $w^{(n)}(t)$ is equal to ±0.0488%.

(4)*Central harmonic frequencies* $F_k^{(n)}=kF_0^{(n)}$, k=1,..H are precisely computed. *Estimated bands*

$$\Delta F_k^{(n)} = 0.000976kF_0^{(n)} + 2\Delta f$$

of the harmonic frequencies $F_k^{(n)}$ are defined taking into consideration the sampled signal and spectrum structure errors [7].

(5)*Harmonic energies* $P_H^{(n)}$ of TSWVD in frequency band 1-6KHz are defined as a sum of the energies of all the estimated harmonic bands $\Delta F_k^{(n)}$, k=H',H'+1,...H", where H'- first harmonic after 1KHz and H"- last harmonic before 6KHz. *Summary energies* $P_S^{(n)}$ in the same band are computed. Then the *inharmonic energies* are $P_I^{(n)}=P_S^{(n)}-P_H^{(n)}$.

(6)The *average degree of hoarseness* DH_{WV} is finally defined as a mean value of all "pitch spectrum" degrees of hoarseness $DH_{WV}^{(n)}=P_I^{(n)}/P_H^{(n)}$.

(7)*Degree of hoarseness perturbation quotient* (DHPQ) is additionally defined, as described in 2.3.4.

4. EXPERIMENTAL RESEARCH

The voice of 32 persons with normal phonation and 80 patients with laryngeal diseases, classified in 5 groups [15] is analysed using the described method. Two types of average degrees of hoarseness are extracted - from the spectrogram (DH_{SP}) as in 3.2. and from the optimal TSWVD (DH_{WV}). The comparision of DH_{SP} and DH_{WV} shows an increased sensibility of DH_{WV} in case of pathology. The experimental results for DH_{SP}'s and DH_{WV}'s values distribution are illustrated in Fig.6.

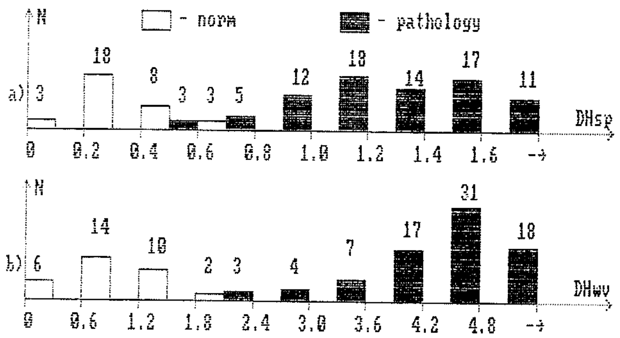


Fig.6. Histograms of the values distribution of a)DH_{SP} and b)DH_{WV}, where N - number of patients.

A *screening analysis* for another control group of 60 persons (30 - normal, 30 - with pathology) using the method of the 7-th nextdoor neighbour on 7 informative parameters (DUV, DSH, DI, FPQ, APQ, DH_{WV} and $DHPQ_{WV}$) is carried out. The results from the screening show that the discriminational decision, when DH_{WV} and $DHPQ_{WV}$ are included into the screening procedure, is more exact with respect to DH_{SP} and $DHPQ_{SP}$ inclusions. The diagnosis of some early stages of laryngeal diseases (some cases of Noduli cantatorii incipiens and Polypus chordae vocalis), "inperceptible" for the time-domain parameters and DH_{SP} [16] becomes possible by the use of DH_{WV}.

5. CONCLUSION

The Wigner-Ville distribution is the basis for high resolution time-frequency analysis of voice signals. We compared the *optimal time-smoothed Wigner-Ville distribution* (TSWVD) with conventional *spectrogram* and found TSWVD to allow the creation of more precise methods for evaluation of the degree of hoarseness of the voice, used for objective diagnosis of laryngeal diseases. The TSWVD time-frequency structure gives more pictorial visual presentation about the pathological divergences of the voice.

6. ACKNOWLEDGMENTS

The autors wish to express their appreciation to Dr. Luba Drumeva, phoniatrician at the Higher Military-Medecine Institute, Sofia for her collaboration in providing clinical data and physiological interpretation of the acoustic parameters, used in this study.

REFERENCES

[1].Rabiner,L., R.Schafer. *Digital Processing of Speech Signals.* Prentice-Hall. N.J. 1978.

[2].Koike,Y., T.Calcatera. *Acoustic Mesures for Detecting Laryngeal Pathology.* Acta Otolaryngologica. v.84, n.1, 1977.

[3].Yumoto,E., W.Gould, T.Baer. *Harmonic-to-Noise Ratio as an Index of the Degree of Hoarseness.* J. Acoust. Soc. Amer. v.82, n.1, 1982.

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

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

[6].Schultz-Coulon,H.J., F.Klingholz. *Objektive und Semiobjektive Untesuchungsmethoden der Stimme.* in Proc: XV-th Congress of European Phoniatricians UEP'88. Erlangen. W.Germany. 1988. pp.1-90.

[7].Boyanov,B., D.Deliyski. *Method for Quantitative Evaluation of the Inharmonic Components in the Voice Signal.* Bioautomation. Sofia. n.7, [in bulgarian], 1988.

[8].Claasen,T., W.Mecklenbrauker. *The Wigner Distribution - a Tool for Time-Frequency Signal Analysis.* Philips J. Res. 35, 1980. pp.217.

[9].Wokurek,W., F.Hlawatsch, G.Kubin. *Wigner Distribution Analysis of Speech Signals.* in Proc: Int. Conf. of Digital Signal Processing. Florence. Italy. 1987. pp.294-298.

[10].Hlawatsh,F. *Interference Terms in the Wigner Distribution.* in Proc: Int. Conf. of Digital Signal Processing. Florence. Italy. 1984. pp.363-367.

[11].Deliyski,D. *Method for Time-Domain Analysis of the Voice Signal.* Electroindustry and Instrumentation. Sofia. n./in press/, [in bulgarian], 1989.

[12].Deliyski,D. *Investigation of the Autocorrelation Function Characteristics in Pathologic Voice Signal Analysis.* in Proc: Int. Conf. on Statistical Theory of Communications STS'88. Varna. Bulgaria. [in russian], 1988.

[13].Rabiner,L. *On the Use of Autocorrelation Analysis for Pitch Detection.* IEEE Trans. ASSP. v.ASSP-22, n.3, 1974.

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

[15].Drumeva,L., D.Doskov, B.Boyanov, D.Deliyski. *Attempt at Dysphonia Objective Analysis via Personal Computer.* in Proc: XV-th Congress of the Union of European Phoniatricians UEP'88. Erlangen. W.Germany. 1988. pp.25-28.

[16].Nikolov,Z., D.Deliyski, L.Drumeva, B.Boyanov. *Computer System for Diagnostics of Pathological Voices.* in Proc: XXI-st World Congress of the International Assosiation of Logopedics and Phoniatrics. Prague. Czechoslovakia. /in press/ 1989.

[17].Flandrin,P., W.Martin. *Pseudo-Wigner Estimators for the Analysis of Nonstationary Processes.* in Proc: IEEE ASSP Spectr. Est. Workshop II. Tampa. Florida. 1983. pp.181-185.

[18].Zielinski,T. *A Personal Computer Development System for the Wigner-Ville Transform and its Applications.* in Proc: European Signal Processing Conference EUSIPCO'88. ed.J.Lacoume, North-Holland. 1988, 23.3IMP.

[19].Zielinski,T. *On a Software Implementation of the Wigner-Ville Transform.* Computer Physics Comm., North-Holland. 1988.