SPECIAL ARTICLE

Application of "Glottal Disturbogram" as a novel tool for the description of vocal disturbances

Pastiadis C[1], Papanikolaou G[2], Printza A[3]

[1] Department of Musical Studies, School of Fine Arts, Aristotle University of Thessaloniki, Thessaloniki, Greece
[2] Department of Electrical & Computer Engineering, Faculty of Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece
[3] 2nd Otolaryngology Department, Medical School, Aristotle University of Thessaloniki, Thessaloniki, Greece

Abstract
The paper introduces "Glottal Disturbogram" as a new tool for the discrimination, evaluation and representation of glottal disturbances which may be met in pathological voicing or singing. The "Glottal Disturbogram's" principles and related features also suit similar applications such as the acoustics of some families of musical instruments. Disturbogram's computational and display characteristics are presented with the use of both synthetic glottal patterns and real signals obtained from subjects with voice disorders. Results show that Disturbogram may efficiently discriminate and quantify perturbation types, offering a valuable tool in clinical or laboratory investigation of both voice disorders and normal voicing types. Hippokratia 2008; 12 (2): 122-127

Key words: vocal distrurbances, voice disorders, jitter, shimmer, noise, wave-shape variation, glottal excitation, vocal register, visualization

Corresponding author: Pastiadis C, Department of Musical Studies, School of Fine Arts, Aristotle University of Thessaloniki, Thermi Univ Campus, Thessaloniki, Greece, Tel: +30 23920 73013, e-mail: pastiadi@auth.gr

Contemporary trends in the examination, screening and diagnosis of voice and speech functionality exploit various approaches and utilize different kinds of data for qualitative and quantitative identification of fundamental parameters. Acoustic signal analysis and its associated indices are important for the determination of both stationary and dynamic attributes of the voicing mechanism, and their usefulness has been well demonstrated[1-5].

Acoustical voice examination protocols follow general guidelines proposed by various recognised scientific or technological bodies[2,5]. Much of the work done until now is based on the estimation of disturbance indices computed either from the radiated speech signal or from the underlying glottal excitation, whenever this may be available. Glottal disturbance descriptors' computation requires more elaborate processing strategies compared to direct estimation from the speech signal. Nevertheless, glottal disturbances may reflect important characteristics of the voicing mechanism. Glottal-based disturbance estimates benefit from the absence of "impurities" that manifest themselves in the speech wave as ambiguities that originate from interactions between perturbed glottal excitation and the vocal tract's response[6-8].

This paper presents applications of "Glottal Disturbogram" (hereafter called "Disturbogram"), a newly developed computational and graphical tool for the discrimination, quantification and representation of glottal disturbances[3,9]. Various types of glottal disturbances appear in different Disturbogram patterns. Its efficiency is demonstrated for various disturbances, using both synthetic glottal sequences and real pathological signals. Results are discussed under the scope of both qualitative and quantitative evaluation of the tool, aiming at the future adoption of Disturbogram as an integral part of a modern laboratory acoustic voice analysis facility, with applications in voice/speech pathology, fine arts, and other related fields.

Method's General Characteristics

Different voicing types may be characterized by different types of perturbations and various extents of perturbations. The extent, up to which these variations may be judged as normal or acceptable, seems to depend on sociological, perceptual, task and environmental factors. Variability between speech, singing, and other forms of voice activity involved in our every-day life requires special attention and analytical study, in order to bring ourselves in sound conclusions about the description of these characteristic disturbances[2,10,11].

This paper focuses on the study of voice disturbances at the acoustical level. Therefore, the acoustic analysis in the present work aims at the identification and quantification of phenomena that constitute the disturbances' space.

Various phenomena participate in the appearance of increased perturbation at the glottal flow production mechanism (e.g. laryngeal apparatus) and/or the vocal tract, especially in pathological voicing conditions. These perturbations may concern timing or intensity of glottal cycles (jitter/shimmer), the introduction of significant amount of turbulent noise due to air leakage through insufficient glottal closure, alteration of the shape of glottal pulses, or even variation of the formant structure of radiated speech[1,3,5,8,10-12].

Thus, the search for adequate and descriptive indices of jitter, shimmer and noise (which are considered as the major perturbations) must aim to the identification of proper quantities (inherent to the glottal/radiated speech signals), that tend to distinguish between various phenomena[3,13].

Classical acoustical models of voiced speech production, seen from a systemic point of view, suggest that the major disturbances' generation consists of some kind of operations performed on the glottal sequence level (such as addition of noise or modulation in original glottal sequence). The vocal tract mechanism alters both the quality and the extent at which glottal perturbations appear in radiated speech, since voicing includes and consists of interlinked physical processes both at the glottal level and the flow modulation process obtained through the oral and pharyngeal tracts. Consequently, discrimination and magnitude estimation of vocal disturbances is theoretically more feasible in the glottal source since vocal tract effects may be efficiently removed from analysed speech signals[3,13]. In the present study, inverse filtering of the lip radiated pressure wave during steady vowel phonation is used as a means of obtaining glottal excitation[1,3,7]. A period of the flow derivative sequence, called the glottal wavelet, is considered the major describing constituent of the process[1,9]. This approach, namely the micro-analytical study at glottal period level, may also be followed at other similar biomedical modalities and musical acoustics[9,14].

Perturbation analysis of consecutive glottal pulses' pairs affords introduction of a new graphical representation tool for the description of vocal pathologies as well as normal voicing conditions. Such an attempt has a twofold aiming; first, to estimate characteristic disturbance indices that reflect glottal dysfunction, and second, to graphically demonstrate and discriminate between various types of disturbances. In this way, the tool may be seen from two perspectives: either as a unified method for discrimination, estimation and graphical representation of characteristic disturbance indices that reflect glottal dysfunction, or as a stand-alone graphical visualization tool that may be fed with various disturbance indices' data. Moreover, the Disturbogram's graphical similarities with classical spectrographic analysis allow incorporation and implementation in existing analysis framework[15-17].

Supported Indices and Glottal function parameters

The Disturbogram's computational approach is based on pitch-synchronous correlation analysis[1,3,9]. Such complex functional forms imply more general waveform fits and offer some major advantages over direct point estimates, especially in cases of glottal signals contaminated with uncorrelated noise[12,13].

Disturbogram treats major disturbance types, which are:

• Sequence's amplitude variation (referred to as Shimmer)

• Sequence's period variation (referred to as Jitter)

• Addition of stochastic multi-frequency spectral components (referred to as Noise)

Another important form of signal aperiodicity, which may be also shown, is the waveform type diversity (or what is called wave-shape disturbance[3,10]), which manifests itself as an intra-speaker variation and may characterize different phonation types.

The method offers the possibility of obtaining and representing local (short-term) disturbance estimates, as well as global (long-term) ones, together with additional spectral and numerical information regarding glottal operation. First, noise components and short-time Signal-to-Noise-Ratios (SNR) are treated using signal denoising algorithms. Jitter and shimmer are estimated in short-time two-period long windows, while a "global" shimmer estimate is also provided. Moreover, average glottal pulse spectral characteristics are represented in the shimmer's spectrographic representation, thus enhancing voice pathology characterisation.

At the moment, the representations refer to computed local and global mean rectified values of the 1st order Perturbation Function (MR_1) for shimmer/jitter and noise's local power spectrum density (psd) estimates[5,10,12,13] (Table 1). However, various other types of perturbation

Table 1. Supported disturbances and characteristics with respective indices.

Disturbance/Characteristic	Index			
Shimmer	Local Shimmer:	$Sh(i)$	(%)	†
	Global Shimmer:	MR_1	(%)	†
		$mRAP$	(%)	*
		$RMS-s$	(%)	*
Jitter	Local Jitter:	$Jitt(i)$	(%)	†
	Global Jitter:	MR_1	(%)	*
		$mFPQ$	(%)	*
		$RMS-j$	(%)	*
Noise	PSD**:	$N(\omega,t)$		†
Average Glottal Wavelet's Spectrum	ESD**:	$G(\omega,t)$		†
Average F0				†

† Visual
* Numerical
** PSD – Power Spectral Density, ESD – Energy Spectral Density

indices may also be considered in case of deploying Disturbogram as a visualization-only tool[5,10,12-13,18]. Additionally, Disturbogram offers computation and indication of mean value of glottal fundamental frequency (F0), average glottal wavelet's energy spectral density (esd), and short-time glottal noise psd. Table 1 summarizes supported types of disturbances, together with descriptive indices and parameters. Definitions of these indices are presented in Table 2. Their properties have been extensively discussed in the relevant literature[3,9,12,13,18].

Table 2. Definitions of indices of Table 1.

$Sh(i)$	$\frac{1}{N-1}\left\|A_i - A_{i+1}\right\| \Big/ \overline{A}$
	where $\overline{A} = \frac{1}{N}\sum_i A_i$ A_i is the amplitude of the i-th glottal period N is the number of glottal periods
MR_1	$\sum_i Sh(i)$
$mRAP$	The Relative Average Perturbation (RAP) will be $\leq mRAP$
$RMS-s$	$\sqrt{\frac{1}{N}\sum_i \left(A_i - \overline{A}\right)^2} \Big/ \overline{A}$
$Jitt(i)$	Similar to $Sh(i)$ but using period length instead of period amplitude
MR_1	$\sum_i Jitt(i)$
$mFPQ$	The Frequency Perturbation Quotient (FPQ) will be $\leq mFPQ$
$RMS-j$	Similar to $RMS-s$ but referring to period length

Visualization of Disturbances and Glottal parameters

The Disturbogram's graphical form is presented in Figure 1. It consists of four major axes pairs where different types of phenomena are displayed, and a text region for numerical presentation of parameters. Except for the first axes pair, the remaining three are spectrographic displays, combining spectral amplitude information along time.

In the first axes diagram, the glottal signal is presented for reviewing purposes. In the second axes, short-term shimmer and spectral characteristics of average glottal pulse are represented within a spectrographic display. As already mentioned, analysis is performed in a pitch-synchronous manner using consecutive pulses pairs with glottal period-long sliding. In the third axes, short-term jitter is computed similarly, and average F0 is also provided. Finally, the fourth axes pair displays short-term noise spectrum.

Disturbogram's visualizations incorporate special design of colour graphics that facilitate identification and quantification of displayed characteristics. Accompanying colorbars enhance intelligibility of represented patterns. Colour tone selection is made from within a 3-d additive colour space (RGB), using nullification of one colour component at a time. More specifically, in the second axes pair, each time slice represents the average glottal pulse's spectrum. The slice's colour grade reflects spectral amplitude variation, following the first colorbar's scaling. Long-term (global) average shimmer estimate is identified from colour shade family in correspondence to the second colorbar's scaling. Colour grading differences between time slices represent short-term (local) shimmer variation. In the third axes pair, short-term jitter values, at each time slice, may be identified by "visual summation" (or integration) of colour grade values along that slice. Similarly, coloration variations reflect local jitter values, around the mean F0 which is represented as a yellow line crossing the y-axis at the specific value. The fourth axes pair is a purely spectrographic display of glottal noise time evolution. As it can be seen, simultaneous representation of short-term and long-term amplitude, timing and noise disturbance values together with average glottal pulse spectrum and noise spectrum is possible, following their respective estimates. For example, referring to the synthetic signals' Disturbograms in Figure 1 and Figure 2, one can observe higher shimmer values in Figure 1, both locally and globally. Average shimmer may be estimated from the red shading of the shimmer pattern and the respective second colorbar (actual value ~18.4%), while local shimmer at each time slice may be computed by visual integration of colour values along the slice, according to the first colorbar. As it can be observed higher shimmer is localized at time slices past 20ms, each slice corresponding to a pair of glottal periods. On the other hand, Figure 2 shows much lower global shimmer values, and also much lower local shimmer estimates. Noise contamination is also obvious in Figure 2 from the high frequency noise components present in the fourth axes pair. However, the Disturbogram's underlying computational framework efficiently removes noise effects both on shimmer and jitter estimates and reveals absence of any other disturbance except for noise contamination. No jitter is present in both signals, while their average F0 is correctly identified at 110 Hz. The average glottal pulse's spectral profile is also displayed in the second axes pair of both Disturbograms, and reveals the glottal pulse's low-pass characteristics with energy concentration around 500 Hz.

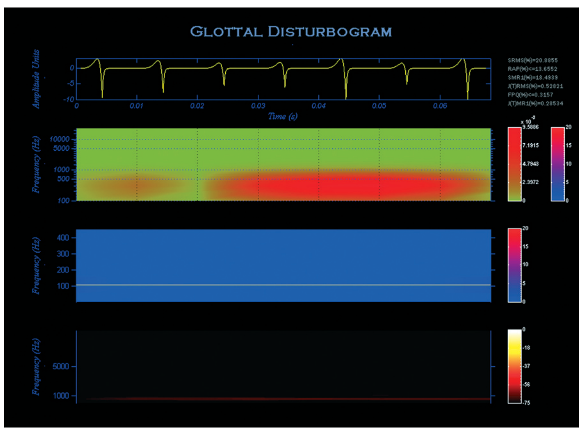

Figure 1. Typical Disturbogram representation of a synthetic signal with significant shimmer and absence of jitter and noise disturbances.

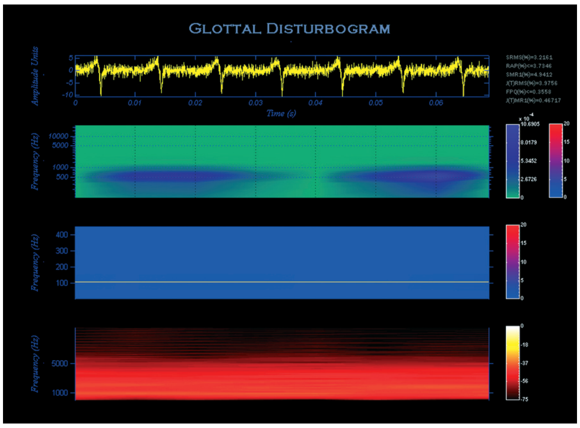

Figure 2. Disturbogram of a synthetic signal with noise contamination and absence of jitter and shimmer. Robustness of Disturbogram analysis precludes high values of shimmer/jitter artifacts.

Results from application onto real signals

In this section, we briefly present examples of application of Disturbogram analysis onto real glottal signals obtained from pathological subjects after inverse filtering of steady vowel phonations.

Figure 3 presents the Disturbogram of a female patient with a polyp of the vocal fold before operation. A mild amount of shimmer and jitter is present (Shimmer MR_1=3.2%, Jitter MR_1=3.17%). Mean F0 is below 200Hz (~180Hz). The glottal sequence does not show high-frequency noise contamination, as we can observe both from the waveform display and the Disturbogram's noise axes pair. However, a low-frequency noisy spectral component may be observed. This low-frequency effect wouldn't be easily removed using high-pass filtering without affecting the waveform shape, indicating a relation of this effect to the signal itself. Actually, it may be attributed to wave-shape variation that can be observed during the evolution of glottal excitation. Classical denoising performs best for noise components unrelated to the signal, thereby offering greater SNR improvement mainly for higher noise frequencies. However, noise components that are highly related to signal's spectral components are hard to discern. A heuristic approach of fitting an average period waveform to each one noisy signal's period may yield better estimates by reducing lower frequency stochastic components as well. On the other hand, this strategy would not remove waveshape variability, which, from one point of view, could be considered as a benefit, especially for discrimination and representation of phenomena such as those that, besides vocal pathology, take place in occasions of register type alteration in the vicinity of singers' F0 limit points.

Figure 4 shows Disturbogram analysis for the same female speaker as above, after the polyp's surgical removal. A clear improvement of shimmer and jitter patterns may be observed (Shimmer MR_1=0.87%, Jitter MR_1=1.28%). Mean F0 is raised at expected values for a female speaker (~270Hz). However, waveshape variability still remains.

Figure 5 presents the results of Disturbogram analysis for a patient with leukoplakia (Shimmer MR_1=4.84%, Jitter MR_1=2.78%). Although a relatively consistent waveshape is observed at macro-scale, lower frequency noise components are present, due to gradual variation at waveshape details after 80ms. There are also patterns of mild jitter and high shimmer.

Discussion and further research

This paper presents a novel vocal disturbances' description tool which may be used either autonomously for computational and graphical purposes or, providing it with appropriate disturbance indices' estimates, as an independent visualization tool. Disturbogram combines the legacy of classical spectrographic representation forms with modern computational and visualization

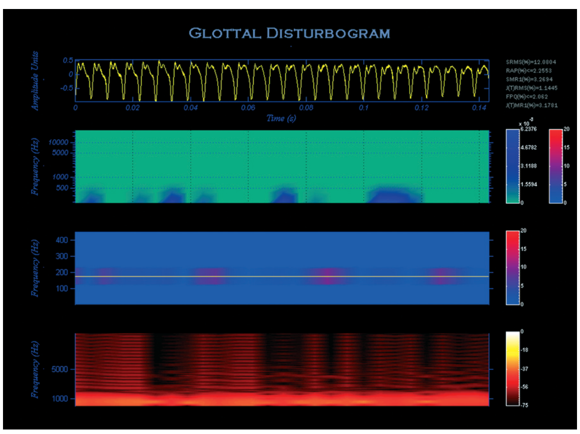

Figure 3a. Disturbogram from a female patient with vocal fold polyp before surgical removal. Moderate shimmer and jitter is present, together with F0 lowering. Wave-shape variation manifests itself as low-frequency noisy spectral components.

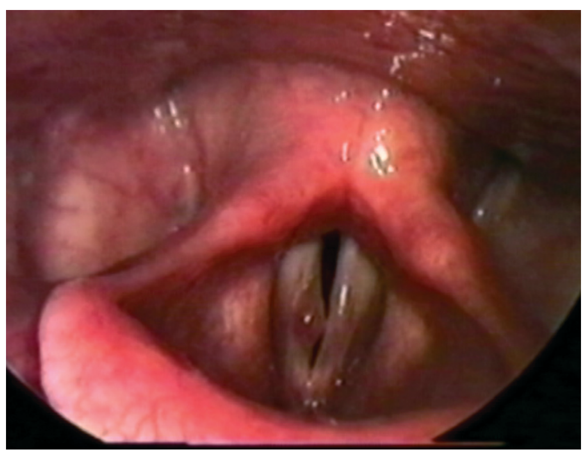

Figure 3b. Laryngoscopic view: female patient with vocal fold polyp.

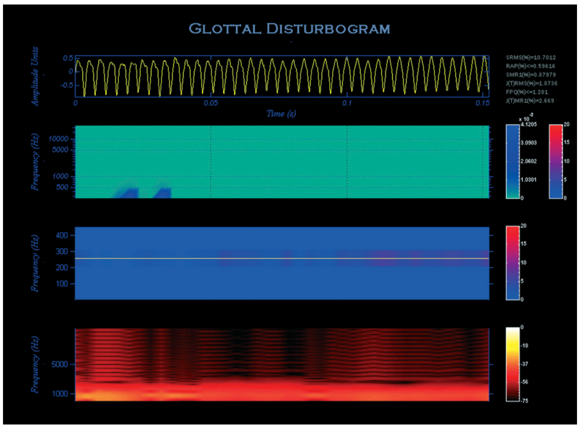

Figure 4. Disturbogram from the female patient in fig.3 after surgical removal of the polyp. Both shimmer and jitter are significantly reduced, and F0 increased. However, remaining wave-shape variation is also evident as low-frequency noise.

strategies. These characteristics allow incorporation of Disturbogram in existing frameworks and easier adaptation to current analysis, screening and diagnosis systems[15].

The validity of Disturbogram's use is consolidated both by simulations using synthetic signals with various types of disturbances at various degrees, and application on real glottal signals obtained using speech inverse filtering. The tool and its underlying computational method are capable of discriminating major types of disturbances, and may reveal hidden patterns in perturbation time evolution. It also supports important spectral information regarding the glottal wavelet or any possible stochastic components. Quantification of parameters is obtained by the use of comprehensive colorbars and the accompanying numerical output.

Further research includes addressing the problem of optimal separation of interrelated perturbation types and especially the effect of waveshape change on jitter, shimmer and noise. An intriguing perspective would be the possibility of more efficient signal modelling with inclusion of un-modelled dynamics or possible non-linear phenomena using pattern matching techniques and non-linear models[19].

A major objective of combined application of Disturbogram analysis on synthetic and real voice signals is the construction of "Disturbogramic" patterns of vocal register types and pathologies for exploitation in intelligent characterisation and screening of examined vocal function (e.g. singing, speech, general voice pathology). Currently, a pilot deployment of Disturbogram analysis is taking place in a laboratory voice signals' acquisition, processing and analysis framework[16,17].

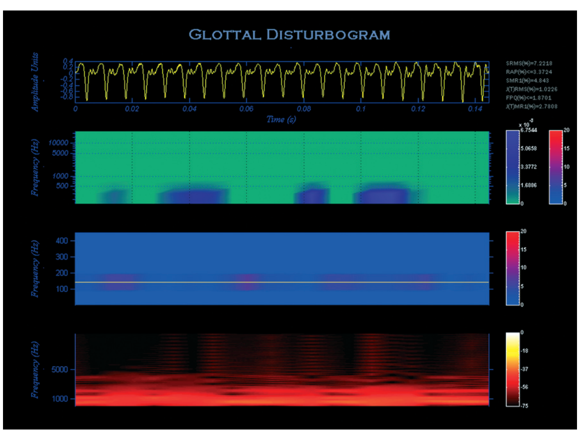

Figure 5. Disturbogram of a patient with leukoplakia.

References

1. Childers D. Speech Processing And Synthesis Toolboxes. New York, John Wiley & Sons, 2000
2. Dejonckere PH, Bradley P, Clemente P, et al. A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. Guideline elaborated by the Committee on Phoniatrics of the European Laryngological Society (ELS). Eur Arch Otorhinolaryngol 2001; 258: 77-82
3. Pastiadis C. Contemporary Analysis Techniques for the Speech Impaired (PhD Dissertation). Thessaloniki: Aristotle University of Thessaloniki; 2002
4. Pastiadis C, Metaxas S, Papanikolaou G, Printza A. Acoustic Voice Analysis in the Evaluation of Vocal Fold Polyps. Hel ORL H-N Surg 2003; 1: 38-47
5. Titze I. Workshop on Acoustic Voice Analysis - Summary Statement. 1994 February 17-18; Denver, Colorado; National Center for Voice and Speech; 1994
6. Bless D. Measurement of vocal function. Otolaryngol Clin North Am 1991; 24: 1023-1033
7. Pastiadis C, Papanikolaou G. Higher-order statistics based inverse filtering for analysis of esophageal voice production. Preprints of the 104th Audio Engineering Society Convention; 1998 May 16-19; Amsterdam; 1998; Preprint No. 4678
8. Sihvo M, Vilkman E. Towards objective measurement of normal and disturbed voices. In: Aulanko R, Leiwo M. (eds) Studies in Logopedics and Phonetics 2; Publications of the Department of Phonetics, University of Helsinki; Series B: Phonetics, Logopedics and Speech Communication 3; 1991; 157-171
9. Pastiadis C, Papanikolaou G. Analysis and visualization of musical instruments' and voice's acoustic disturbances. Proceedings of the 2nd HELINA Conference; 2004 September 27-28; Thessaloniki; 2004; 207-212
10. Titze I. Principles of Voice Production. Iowa City, Prentice-Hall Inc.; 1994
11. Koufman J, Isaacson G. The spectrum of vocal dysfunction. Otolaryngol Clin North Am 1991; 24: 985-988
12. Michaelis D, Fröhlich M, Strube HW, Kruse E, Story B, Titze I. Some simulations concerning jitter and shimmer measurement. In: Lehmann T, et al. Proceedings of the 3rd International Workshop on Advances in Quantitative Laryngoscopy, Voice and Speech Research; 1998 June 19-20; Aachen: Universitδtsklinikum, RWTH, 71-80
13. Murphy P. Perturbation-free measurement of the Harmonics-to-Noise Ratio in voice signals using pitch synchronous harmonic analysis. J Acoust Soc Am 1999; 105: 2866-2881
14. Fletcher N, Rossing D. The physics of musical instruments. 2nd ed. Berlin, Springer-Verlag; 1998
15. Gopal H. Technical Issues Underlying the Development and use of a Speech Research Laboratory. In: Syrdal A, Bennett R, Greenspan S, (eds). Applied Speech Technology. Boca Raton, CRC Press; 1995, 315-342

16. Pastiadis C, Printza A, Psyllidou G, Papanikolaou G, Triaridis S, Kekes G. Development of an integrated and interactive voice signals acquisition, analysis and database system, for the examination of dysphonic patients. Presented at the 13th Convention of the Hellenic ENT Society; 2005; Athens
17. Pastiadis C, Psyllidou G, Papanikolaou G. A new integrated system for laboratory environments of speech/voice examination. Preprints of the 120th Audio Engineering Society Convention; 2006 May 20-23; Paris; 2006; Preprint No. 6828
18. Michaelis D, Frolich M, Strube HW. Selection and combination of acoustic features for the description of pathologic voices. J Acoust Soc Am 1998; 103: 1628-1639
19. Haykin S. Neural Networks. A comprehensive foundation. New York, Macmillan College Publishing Company; 1994