Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2017) 25: 3201 – 3214
© TÜBİTAK
doi:10.3906/elk-1608-324

Assessment of disordered voices based on an optimized glottal source model

Mounir BOUDJERDA[1,2,*], **Abdellah KACHA**[1]

[1]Laboratory of Radiation Physics and Applications, University of Jijel, Jijel, Algeria

[2]Research Centre in Industrial Technologies CRTI, ex CSC B. P. 64, Cheraga, Algiers, Algeria

Received: 30.08.2016 • **Accepted/Published Online:** 23.12.2016 • **Final Version:** 30.07.2017

Abstract: In this paper, a method for the assessment of disordered voices is proposed. A feature named mean opening quotient (MOQ) obtained from the glottal source estimation is used as an acoustic cue to summarize the degree of severity of the voice disorder. The analysis method uses the empirical mode decomposition algorithm to estimate the glottal source excitation signal from the speech signal. The logarithm of the magnitude spectrum of the speech signal is decomposed into oscillatory modes, called intrinsic mode functions, that are clustered into two classes, the spectral envelope and the harmonic component. The exploitation of the phase information jointly with the estimated harmonic component enables the estimation of the glottal source signal. An appropriate parametric model is fitted to the estimated glottal source excitation signal. The optimal parameters of the glottal source excitation model from which the MOQ is defined are obtained by using a genetic algorithm. The presented method is tested on a corpus of natural speech including the vowel [a] uttered by 22 normophonic speakers and 229 speakers with different degrees of dysphonia. Experimental results show that the proposed method is very effective for assessing the degree of severity of the voice disorder.

Key words: Glottal source signal, voice disorders, empirical mode decomposition, genetic algorithm

1. Introduction

Voices disorders most often originate from disease or malfunction of the larynx. The cause of voice disorders can be either functional or organic. Functional dysphonia is caused by a misuse of an anatomically and physiologically intact voice system. These pathologies can be corrected by voice therapy. Laryngeal pathologies can be corrected either by voice therapy or by a surgical intervention.

A robust method able to identify and assess the degree of severity of the pathology is of great importance for monitoring the evolution of the patient's condition. Conventional methods of diagnosis and clinical evaluation of the voice are subjective [1]. Routinely, subjective assessment of pathological voices is based on listener perception of speech, i.e. the clinicians evaluate the level of perceived dysphonia according to several protocols such as the grade, roughness, breathiness, asthenia, and strain (GRBAS) scale. For this assessment method, the result is listener-dependent. A main disadvantage of perceptual rating is the intravariability as well as the intervariability of the evaluations [2,3]. For obtaining reproducible evaluations, the evaluators must have experience in voice timbre evaluation. On the other hand, objective evaluations are based on the acoustical analysis of speech. This method of evaluation of voice disorders is noninvasive and provides a severity index that allows clinicians to track the progress of patients and numerically evaluate the degree of dysphonia. Clinicians prefer acoustics markers that are correlated with perceptual characteristics. An objective method is considered powerful if it correlates strongly with the perceptual evaluation.

*Correspondence: boudjerda.m@yahoo.com

Several methods for the objective evaluation of disordered voices have been proposed in the literature. A method proposed recently in [4,5] is based on cepstral analysis. The real cepstrum of a given signal is obtained by calculating the inverse Fourier transform of the logarithm of its magnitude spectrum. An acoustic marker named cepstral peak prominence (CPP) is used to indirectly quantify the degree of dysphonia. A class of acoustic features extensively proposed to evaluate dysphonic voices reflects the departure of the voiced speech signal from the perfect periodicity. Recent methods for estimating vocal dysperiodicities can be found in [6,7].

Many voice disorders are due to a dysfunction of the vocal cords [8]. As a result, characteristics that describe the behavior of the glottal source excitation are interesting for the automatic evaluation of disordered voices. Most often, jitter and shimmer are used to quantify perturbations of the speech cycle lengths and amplitudes, respectively [7,9]. Numerous methods for quantifying the perturbations of the glottal source have been proposed. In [10], short-term jitter obtained by a spectral jitter estimator based on a mathematical description of the jitter phenomenon was used for discriminating disordered voices in continuous speech. In [11], the contribution of the glottis in the production of dysphonic voices was taken into account by using features extracted from glottal source estimation jointly with two other sets of features related to the speech and prosody. Results show that higher discrimination ability is obtained by the combination of glottal source and speech-based features. The study presented in [12] focused on a method for voice pathology detection through a biometric signature based on the speaker's glottal source power spectral density. In a recent study [13], the use of vocal source features as biomarkers for depression severity assessment was investigated. These biomarkers include jitter and shimmer of vocal fold, fundamental frequency dynamics, and level of aspiration. The investigation showed a relationship between these biomarkers and depression severity.

In this study, a vocal source feature named mean opening quotient (MOQ) is proposed as an acoustic measure to numerically quantify the degree of hoarseness in the speech signal for the assessment of voice disorders. For this purpose, the logarithm of the magnitude spectrum of the speech signal is decomposed via the empirical mode decomposition (EMD) algorithm [14] for the estimation of the glottal source signal. An appropriate parametric model is fitted to the estimated glottal source signal. The optimal parameters of the model from which the MOQ is defined are obtained through a genetic algorithm (GA).

CPP is used for comparison purpose. CPP indirectly summarizes the degree of perturbations of the glottal excitation by characterizing it in the cepstral domain. Both the cepstral-analysis–based method and GA-based method involve the separation of the speech signal components, but in different domains. The use of GA-based CPP and MOQ offers the opportunity to document the degree of perturbation of the glottal excitation in the cepstral domain and time domain, respectively.

2. Perceptual ratings and corpus

The performance of an acoustic marker is often analyzed by comparing objective measures provided by the acoustic marker to the perceptual evaluation carried out on some scale. The purpose of the perceptual rating scale is to describe the quality of a voice following qualitative and quantitative criteria. A popular scale used for subjective assessment of disordered voices is the GRBAS scale. "G" is a measure of the global quality of the voice (perceived hoarseness) and it is used in the present study to evaluate the proposed method's voice disorder assessment performance.

The corpus used in this study includes natural speech comprising the vowel [a] [15]. The stimuli are uttered by 22 normophonic speakers (3 males and 19 females) aged from 19 to 48 years and 229 speakers (79 males and 150 females) with different degrees of disordered voices aged from 19 to 48 years. The voice pathologies include nodules (42), functional dysphonia (81), polyps (11), edema (29), paresis/paralysis (18), acute laryngitis (5), cysts (8), and others (34). The sampling frequency is 44,100 Hz. The perceptual assessment was carried out by five specialists in pathological voices who rated the stimuli from 0 (normophonic speaker) to 3 (highly dysphonic speaker). The global score of each stimulus was taken as the average of the five scores assigned by the judges. The corpus was recorded using a microphone (AKG C41WL, Vienna, Austria) and a recorder (Sony TCD D8, Tokyo, Japan). This corpus was evaluated at the Saint-Jan General Hospital, Bruges, Belgium.

3. Methods

The general procedure for the assessment of disordered voices is described by the flowchart in Figure 1. In the proposed method, the EMD algorithm is used in the log spectral domain for the estimation of the glottal source signal from the speech signal [16,17]. By means of the EMD algorithm, the logarithm of the magnitude spectrum of the speech signal is decomposed to oscillatory modes, called intrinsic mode functions (IMFs), that are clustered in two classes (the spectral envelope and the harmonic component) by a simple thresholding. The exploitation of the phase information jointly with the estimated harmonic component enables the estimation of the glottal source signal. A parametric model is fitted to the glottal source signal, and by using GA, the optimal parameters of the generic model of the glottal source signal are obtained. The different steps involved in the method are detailed subsequently.

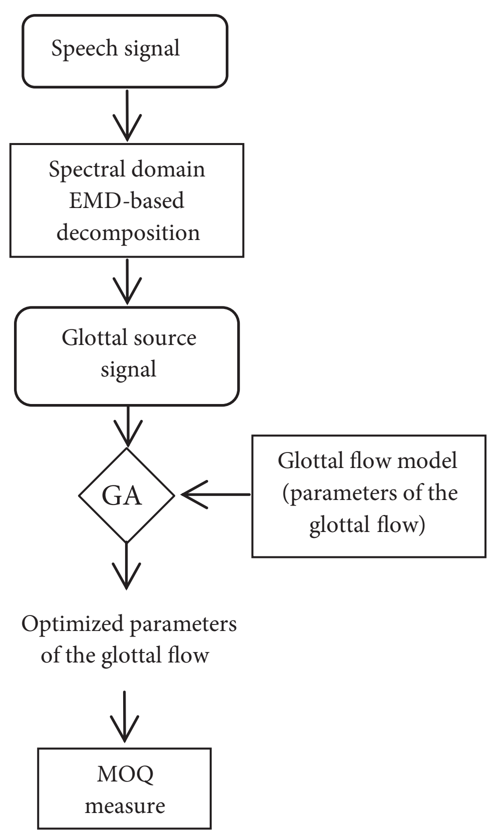

Figure 1. Procedure of the estimation of MOQ from the speech signal.

3.1. EMD-based estimation of the glottal source signal

3.1.1. Empirical mode decomposition

The EMD algorithm is a new means proposed for processing nonstationary signals. It has been proposed in [14] for the analysis of nonstationary and nonlinear processes, e.g., waves of the ocean. It has since been applied in several fields, like speech signal processing.

The application of the EMD algorithm to a given signal $s(t)$ enables the decomposition of this signal to oscillation modes, namely the IMFs. The use of an iterative sifting process allows these IMFs to be estimated. The following steps summarize the EMD algorithm [14]:

1. Initialization of the algorithm $j = 1$, $R_0(t) = S(t)$, and fixation of the threshold σ.
2. Find local maxima and minima of $R_{j-1}(t)$.
3. Interpolation between minima (maxima) to obtain the lower envelope $I_j(t)$ and upper envelope $S_j(t)$ and computation of the average envelope as $M_j(t) = (I_j(t) + S_j(t))/\,2$.
4. Calculation of the jth component: $H_j(t) = R_{j-1}(t) + M_j(t)$.
5. Processing $H_j(t)$ as $R_{j-1}(t)$. With $H_{j,0}(t) = H_j(t)$ and $M_{j,k}(t)$, $k = 0, 1, ...,$ (k is the number of sifts), and calculate $H_{j,k}(t) = H_{j,k-1}(t) - M_{j,k-1}(t)$ until:

$$SD_k = \sum_{t=0}^{T} \frac{\left|H_{j,k-1}(t) - H_{j,k}(t)\right|^2}{\left(H_{j,k-1}(t)\right)^2} < \sigma \tag{1}$$

6. Computation of the jth IMF as $IMF_j(t) = H_{j,k}(t)$.
7. Update of the residue $R_j(t) = R_{j,k-1}(t) - IMF_j(t)$ and increase the sifting index j and repeat steps 2 to 6 until the number of local extrema in $R_j(t)$ is <3.

The sum of all IMFs and the residue allows the signal $s(t)$ to be reconstructed exactly.

3.1.2. Decomposition of the speech signal via EMD algorithm

A voiced speech $s(t)$ can be expressed as the convolution of the glottal source signal (excitation) $g(t)$ and the impulse response of the vocal tract $v(t)$ [18]

$$s(t) = g(t) * v(t), \tag{2}$$

where * denotes the convolution. Applying the Fourier transformation gives

$$S(f) = G(f) \times V(f) \tag{3}$$

The application of the log to the magnitude spectrum given by Eq. (3) changes the multiplicative components to additive components.

$$log\,|S(f)| = log\,|G(f)| + log\,|V(f)| \tag{4}$$

The EMD allows the separation of the log magnitude spectrum into two components. In fact, the EMD algorithm can be interpreted like a filter bank [19]. The logarithm of the magnitude spectrum of the speech signal is decomposed via the EMD algorithm into several IMFs and a residue that can be classed into two classes assigned to the log-magnitude spectrum of the harmonic component (plus noise) and the spectral envelope. The EMD-based method for speech is presented in [16,17].

The process used by the EMD-based method to separate the two components of the speech signal (harmonic and spectral envelope) is performed frame by frame. As an illustration of the effectiveness of the decomposition algorithm, Figure 2 displays the different components. In this example, the frame length is 200 ms extracted from a natural vowel [a] uttered by a dysphonic speaker.

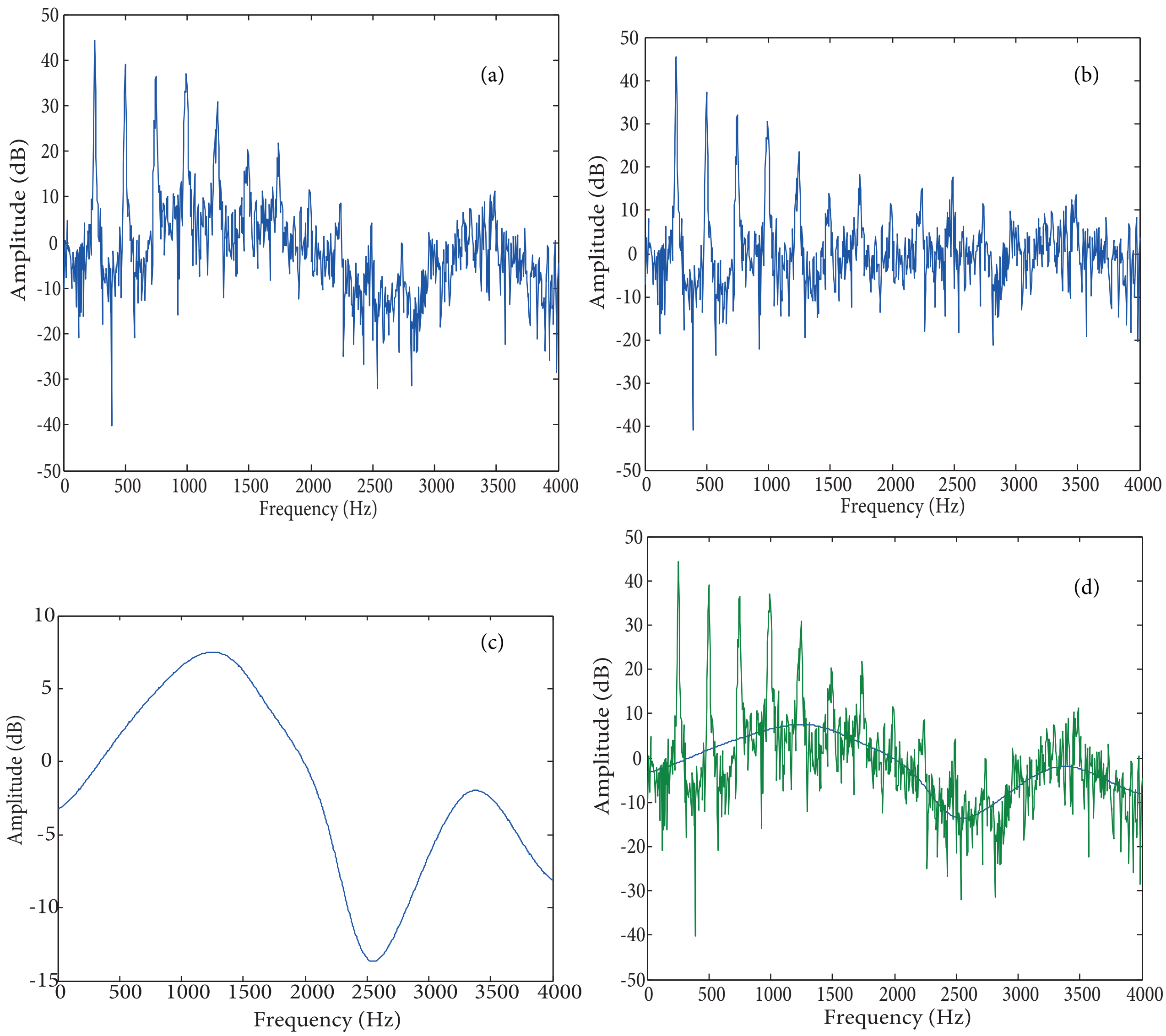

Figure 2. Decomposition of a 200 ms speech frame of sustained [a] into two components via the EMD algorithm. (a) Logarithm magnitude spectrum. (b) Harmonic component. (c) Spectral envelope component. (d) Estimated envelope superposed to the sum of the two components.

3.1.3. Estimation of the glottal source signal

The estimation of the glottal source signal combines the logarithm of the harmonic component and the phase information estimated from the complex spectrum of the speech signal. The flowchart given in Figure 3 illustrates the different steps of the method.

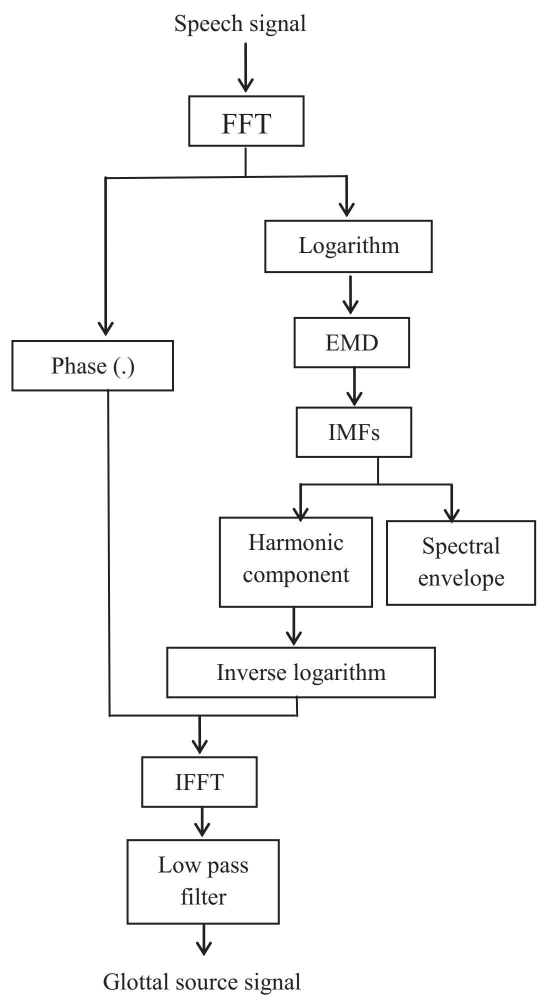

Figure 3. Estimation of the glottal source signal.

Figure 4a shows the original excitation signal superimposed to the estimated excitation signal for a frame of 50 ms extracted from a synthetic sustained vowel [a] uttered by a normophonic speaker. As observed, a high-frequency component interpreted as an artifact related to the decomposition method is present in the estimated excitation signal. This component can be attenuated by using a low-pass filter. Figures 4b and 4c show the estimated excitation signal low-pass filtered by a second-order Butterworth filter with a cut-off frequency of 2000 Hz and 1000 Hz, respectively. In this study, a second-order low pass Butterworth filter with a cut-off frequency of 1000 Hz has been applied.

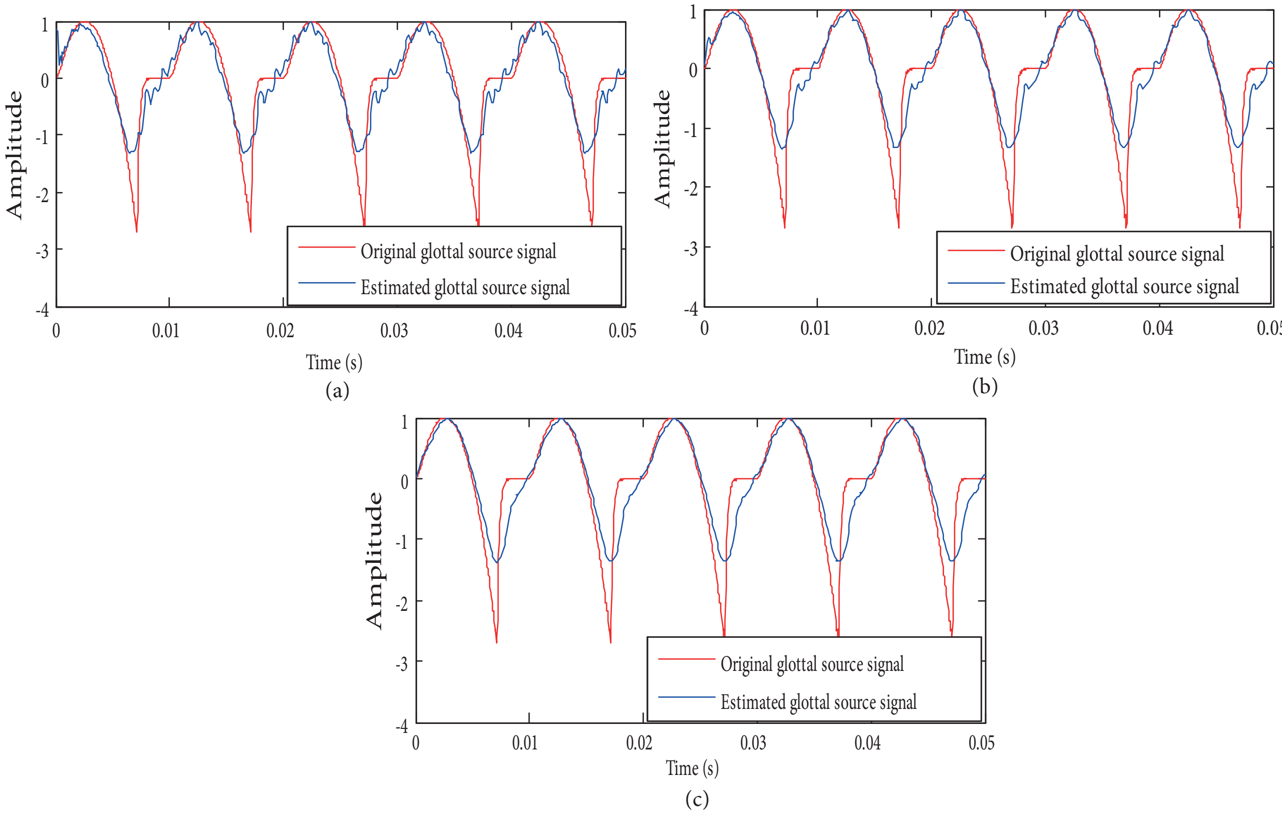

Figure 4. Original excitation signal superimposed to the estimated glottal source signal for a frame of 50 ms extracted from a sustained vowel [a] uttered by a normophonic speaker (a) without filtering, (b) with low-pass filtering at 2000 Hz, and (c) with low-pass filtering at 1000 Hz.

3.2. GA-based optimization of the glottal flow model

3.2.1. Description of the glottal flow models

Many glottal source models are available in the literature. The model proposed by Fant et al., known as the LF model [20], is widely used in speech synthesis and modeling [21,22]. Phases and parameters of the glottal source with its derivative are presented in Figure 5 according to the LF model [20]. The glottal flow model is described in terms of two phases, the open phase and the closed phase.

The open phase itself is divided into the opening and the closing phases. The opening phase is defined by the increase of the glottal flow from the initial state to its maximum amplitude A_v at time T_p defining the opening instant, as depicted in Figure 5. By convention, this phase represents the closure of vocal cords [23]. The closing phase is described via the decrease of the glottal source from A_v to a point at time T_e defining the closing instant where the derivative achieved its negative minimum E (E: maximum excitation). Into the closed phase, the glottal flow returns to the initial state after a brutal closure T_a, where the glottal flow derivative achieves 0 after maximum excitation. In practice it is appropriate to put $T_c = T_0$, that is the totality of the fundamental period [20].

For one fundamental period T_0 and at a sampling frequency F_s, the sample length of one cycle is $N = \text{int}(T_0 \times F_s)$ [24]. Therefore, the corresponding parameters in discrete domain are: $N_e = \text{int}((T_e \times N)/T_0)$, $N_p = \text{int}((T_p \times N)\ /T_0)$, $N_a = \text{int}((T_a \times N)/\ T_0)$ and $N_c = \text{int}((T_c \times N)/T_0)$. The glottal source model according

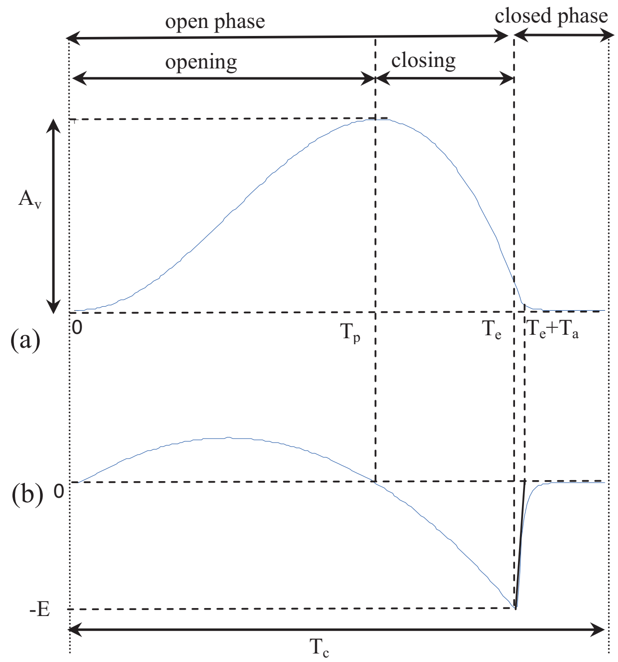

Figure 5. Glottal flow model (a) and its derivative (b).

to the LF model in the discrete domain is defined by [24]:

$$e_g(n) = \begin{cases} E_0 \times e^{\alpha n} \sin(\omega_g n), & 0 \leq n < N_e \\ -\frac{E}{\varepsilon N_a}\left[e^{-\varepsilon(n-N_e)} - e^{-\varepsilon(N_c-N_e)}\right], & N_c \leq n < N_0 - 1. \end{cases} \tag{5}$$

The parameters E_0, α, ω_g, and ε are obtained through the following constraints [24]:

$$\begin{cases} \int_0^{T_0} e_g(t)\, dt = 0 \\ \omega_g = \frac{\pi}{N_p} \\ \varepsilon N_a = 1 - e^{-\varepsilon(N_c - N_e)} \\ E_0 = -\frac{E}{e^{-\alpha N_e}\, sin(\omega_g N_e)} \end{cases} \tag{6}$$

Noise emerges due to turbulence generated during the open phase of phonation, especially for dysphonic speakers [25].

3.2.2. GA-based model parameters optimization

GAs are inspired by the theory of evolution. The goal of the GA is to find the extrema of a function defined in space of data. In order to solve a problem by GA, an evolutionary process is used, where possible solutions (chromosomes) will be used for creating new solutions. Such a group of possible solutions will be named a population. One specific population will survive and will be used in the next generation of populations. Solutions used to create new solutions (offspring) are chosen according to their fitness. The chromosome that has more chances to reproduce is the more suitable [26]. The flowchart of the GA-based method for the optimization of the model parameters is presented in Figure 6. The use of the GA involves the following steps [26]:

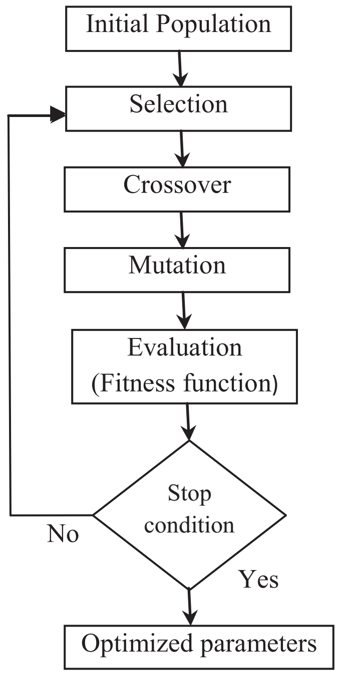

Figure 6. GA-based estimation of the glottal flow parameters.

- Generation of the initial population: Each chromosome is represented as a binary string, e.g. chromosome 1: 10000100 and chromosome 2: 11111111. The initial population is generated by a mechanism that should be able to produce a nonhomogeneous population of individuals that will serve as a base for the future generations. Generally, the number of generations is between 50 and 100 [26]. In this work, the number of generations is fixed to 100. The population size (Ps) and the number of individuals affect directly the convergence of the genetic algorithm. A good population size is between 20 and 30 [26]. In this study, Ps = 20 and the number of individuals = 2 because we want to find the optimized values of T_p and T_e.

- Selection: Populations are evaluated and classified according to the fitness function. Populations that have the best fitness values are chosen for the next generation.

- Crossover: Genes are recombined from parents to form a new chromosome, e.g., chromosome 1: 10000100 and chromosome 2: 11111111 could be crossed over after the third locus to create two new offspring chromosomes, 10011111 and 11100100. The crossover is applied to the population with a constant probability (Pc). In general, the choice of Pc is very large and depends on the application. Usually, Pc is between 0.5 and 0.95 [27]. For this application, Pc = 0.75.

- Mutation: For creating a new offspring, some of the bits in a chromosome changes. In binary encoding, a few randomly selected bits can be changed from 0 to 1 or from 1 to 0, e.g., the chromosome 00000100 could be changed in the second position to create the chromosome 01000100. Generally, for the mutation operator, it is preferable to use a low probability of mutation (Pm). Usually, Pm is between 0.01 and 0.3 [27,28]. For this application, Pm = 0.1.

- Evaluation function: This turns over a value of R+ named fitness of the individual. In this study, each individual of the population is evaluated via a fitness function δ defined as the sum of the absolute differences between the model of the glottal source $e_g(t)$ and the estimated glottal source signal $e(t)$.

$$\delta = \sum_{n=1}^{N} |e_g(n) - e(n)| \tag{7}$$

with N denoting the frame length of the glottal source signal.

3.2.3. Estimation of the optimized opening and open phases

The GA has been used to obtain the opening phase T_p and open phase T_e of the glottal flow model defined by Fant et al. [20]. The different parameters of the GA have been set to the values given in 3.2.2.

The acoustic marker used to summarize the amount of aperiodicities within an utterance is the MOQ, defined as the mean of the opening phase length to one fundamental period. For a given signal, the interval of this signal is divided into L frames. The MOQ is calculated as

$$MOQ = \frac{1}{L}\left[\frac{\sum_{i=1}^{L}(T_p)_i}{T_0}\right] \tag{8}$$

The MOQ reflects the increase in the glottal flow from the initial state to its maximum amplitude, which corresponds to the closure of the vocal cords. If the closure of the vocal cords is altered due to a malfunction, this affects the MOQ. The MOQ provides an indication of vocal cord closure and may be used as an acoustic descriptor for the assessment of disordered voices. It is expected that normophononic and dysphonic speakers exhibit different MOQ measures that reflect the degree of severity of the voice disorder.

4. CPP

The real cepstrum of a given signal is defined as the inverse Fourier transform of the logarithm of its magnitude spectrum [5]. CPP measures the log-amplitude of the first harmonic of the speech cepstrum. CPP is computed as follows [4]:

- Compute the cepstrum of each analysis frame of speech. In this work, the analysis frame is 2048 samples.

- Fit a linear regression line to logarithm cepstrum between the maximum quefrency and 1 ms.

- The local (by frame) CPP is the height with regard to the regression line of the most prominent cepstral peak between the maximum and the minimum expected vocal quefrencies.

- The global CPP is the average of all local CPPs.

5. Results and discussion

Previous studies have shown that the best frame length that provides an accurate decomposition of the speech signal into its components (harmonic and spectral envelope) is 200 ms [17]. The frame length has been set to this value.

Figure 7 shows the estimated glottal source signals of a frame of 200 ms derived from a sustained vowel [a] uttered by a normophonic speaker assigned an average score of 0 as well as by a dysphonic speaker assigned an average score of 3. The values of the mean fundamental frequency F_0 of both frames are 215 Hz and 155 Hz, respectively. The estimated glottal source signal corresponding to the normophonic speaker (Figure 7a) is quasiperiodic whereas the estimated glottal source signal corresponding to the dysphonic speaker is irregular. These results are in good agreement with those published in the literature [29].

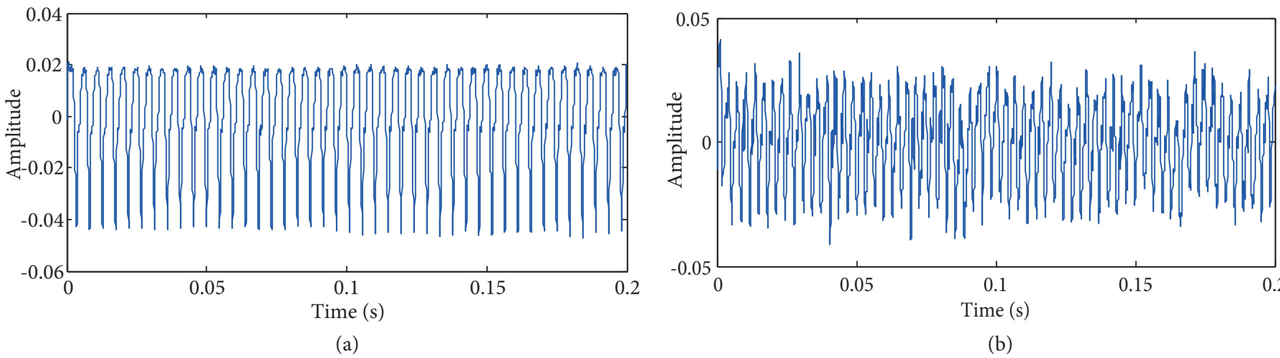

Figure 7. Glottal source signal estimates corresponding to two frames of 200 ms extracted from sustained vowels [a] uttered by a normophonic speaker (a) and a dysphonic speaker (b).

The glottal source signal is modeled by a LF model, the optimal parameters of which are obtained by a GA algorithm. For the optimization procedure, the frame length is fixed to about three times the possible maximal fundamental period. For a maximal fundamental period of 12.5 ms (i.e. a minimal possible fundamental frequency of 80 Hz), this corresponds to 37.5 ms. Accordingly, the frame length in the optimization procedure is set to 40 ms. The optimization is focused on the opening and closing phases. The brutal closure is fixed to $N\mathrm{a} = 10$ and 3 pitch periods are used to construct it.

Figure 8 shows the estimated glottal source signals via the EMD-based method superposed to the optimized glottal source via the LF model and GA for the first three pitch periods extracted from a sustained vowel [a] uttered by a normophonic speaker and a dysphonic speaker. The optimized LF model via GA provides an accurate estimate of the glottal source. Table 1 gives the mean and standard deviation of the fundamental frequency and period as well as the values of T_p,T_e, MOQ, and the fitness function for the normophonic and dysphonic speakers. The amount of the dysperiodicity in both glottal flows is quantified by estimating the MOQ via the EMD-GA–based method, which is 0.56 for the normophonic speaker and 0.24 for the dysphonic speaker. The MOQ of the pathological voice is small compared to the MOQ of normal voice.

In order to show the effectiveness of the EMD-GA method for the assessment of disordered voices, the MOQ has been tested on the corpus of natural speech. Figure 9 displays the MOQ estimates versus the average perceived grade scores for vowel [a]. The values of the estimated MOQ range from 0.18 to 0.70. The MOQ values of normal voices range from 0.56 to 0.68 and the MOQ values of pathological voices decreases as the degree of severity of voice increases.

For comparison's sake, CPP has been computed. Figure 10 shows the CPP estimates versus the average

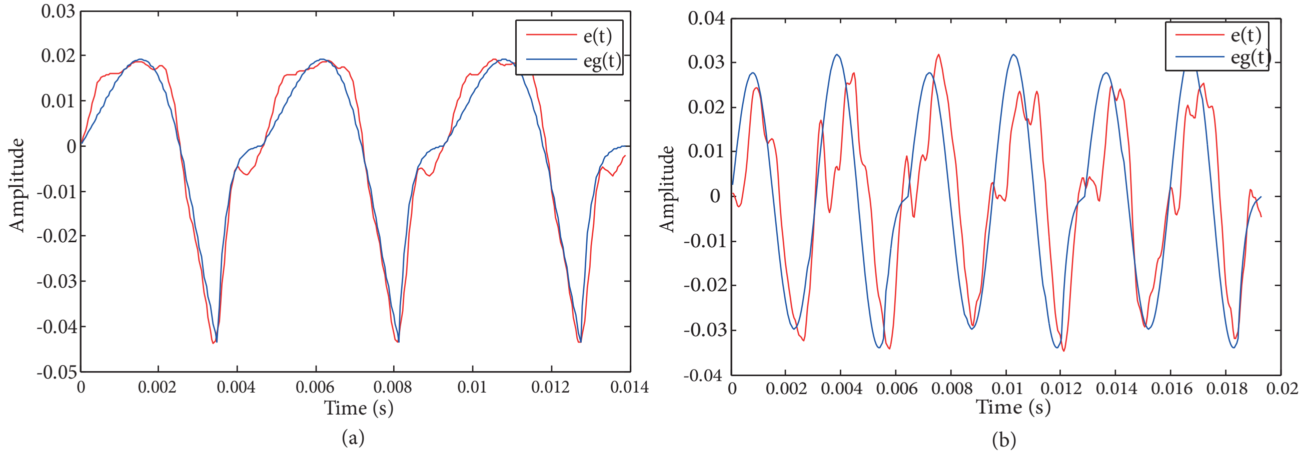

Figure 8. Three pitch periods of the glottal source signal estimates and optimized glottal source using the LF model and GA extracted from sustained vowel [a] uttered by a normophonic speaker (a) and a dysphonic speaker (b).

Table 1. Mean and standard deviation of the fundamental frequency and period, values of the optimized parameters for glottal model, acoustic marker, and fitness function for normophonic and dysphonic speakers.

	F_0 (Hz)		T_0 (ms)		T_p (ms)	T_e (ms)	MOQ	$\sigma(\%)$
	Mean	Std	Mean	Std				
Normophonic	215	0.9	4.6	0.02	2.57	3.49	0.56	0.30
Dysphonic	155	6.6	6.4	0.39	1.57	5.59	0.24	1.07

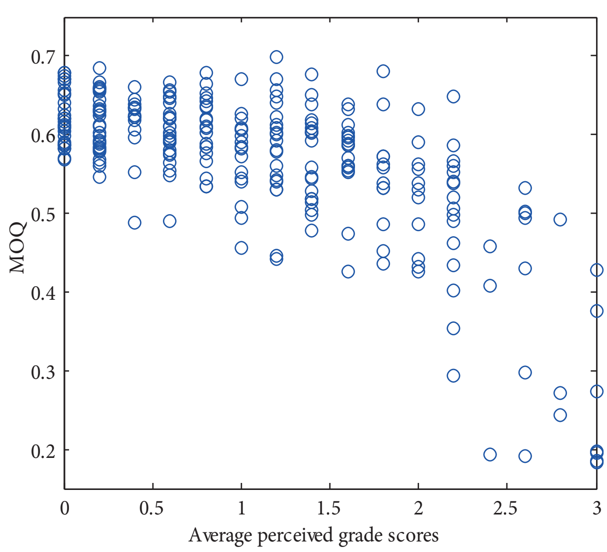

Figure 9. MOQ estimates versus average perceived grade scores for vowel [a].

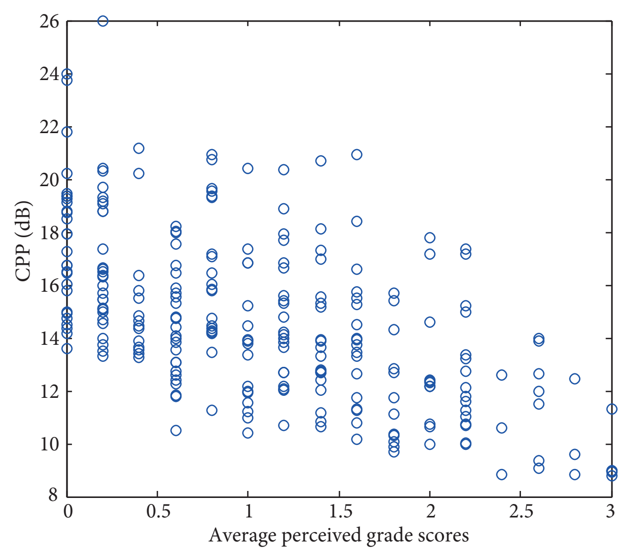

Figure 10. CPP estimates versus average perceived grade scores for vowel [a].

perceived grade scores for vowel [a]. The CPP values of disordered voices decreases as the degree of vocal severity increases. Table 2 gives the correlations of MOQ values based on the proposed method with mean "G" scores, and the last column of Table 2 presents the correlations of CPP values with mean "G" scores. The correlation between the mean "G" scores and the values of the acoustic marker obtained by means of the MOQ estimated via the EMD algorithm and GA is stronger than that obtained by the method based on cepstral analysis.

Table 2. Correlation coefficients between the acoustic markers MOQ/CPP and average "G" scores.

	MOQ	CPP
Correlation	–0.68	–0.56

6. Conclusion

A new method for the assessment of disordered voices is presented. MOQ obtained from the glottal source estimation is used as an acoustic cue to evaluate the overall quality of the pathological voices obtained from speakers with disordered voices.

The EMD algorithm is applied as an alternative for the estimation of the glottal source signal from the speech signal. A GA is applied to estimate the optimal parameters of a generic model of the glottal flow.

The proposed method was tested on a corpus of natural speech including vowel [a] uttered by 251 speakers (normophonic and dysphonic). Experimental results show that the MOQ is highly correlated with the mean "G" scores. The correlation between the mean "G" scores and the values of the MOQ acoustic marker estimated via the empirical mode decomposition algorithm and genetic algorithm is stronger than that obtained by the method based on CPP analysis.

References

[1] Kondo K. Subjective Quality Measurement of Speech: Its Evaluation, Estimation and Applications. Berlin, Germany: Springer, 2012.

[2] Labuschagne IB, Ciocca V. The perception of breathiness: acoustic correlates and the influence of methodological factors. Acoust Sci Tech 2016; 37: 191-201.

[3] Kreiman J, Vanlancker-Sidtis D, Gerratt BR. Defining and measuring voice quality. In: Proceedings of the Workshop on Voice Quality; 27–29 August 2003; Geneva, Switzerland, pp. 115-120.

[4] Alpan A, Grenez F, Schoentgen J. Cepstral analysis of perceptually rated synthetic disordered speech stimuli. In: Proceedings of the International Workshop on Models and Analysis of Vocal Emission for Biomedical Applications; 25–27 August 2011; Florence, Italy, pp. 131-134.

[5] Herman-Ackah YD, Michael DD, Goding GS. The relationship between cepstral peak prominence and selected parameters of dysphonia. J Voice 2002; 16: 20-27.

[6] Kacha A, Grenez F, Schoentgen J. Estimation of dysperiodicities in disordered speech. Speech Commun 2006; 48: 1365-1378.

[7] Alpan A, Maryn Y, Kacha A, Grenez F, Schoentgen J. Multi-band dysperiodicity analyses of disordered connected speech. Speech Commun 2011; 53: 131-141.

[8] Zhonga Z, Jiang T, Zhang W, Yao H, Xiao S. Analyzing speech of patients with vocal polyps based on channel parameters and fuzzy logic systems. Comput Math Appl 2011; 62: 2834-2842.

[9] Murphy PJ, Akande OO. Noise estimation in voice signals using short-term cepstral analysis. J Acoust Soc Am 2007; 121: 1679-1690.

[10] Vasilakis M, Stylianou Y. Voice pathology detection based on short-term jitter estimations in running speech. Folia Phoniatr Logop 2009; 61: 153-170.

[11] Drugman T, Dubuisson T, Dutoit T. On the mutual information between source and jitter contributions for voice pathology detection. In: Proceedings of Interspeech 2009; 6–10 September 2009; Brighton, UK. pp. 1463-1466.

[12] Gomez-Vilda P, Fernandez-Baillo R, Rodellar-Biarge R, Lluis VN, Alvarez-Marquina A, Mazaira-Fernandez LM, Martinez-Olalla R, Godino-Llorente JI. Glottal source biometrical signature for voice pathology detection. Speech Commun 2009; 51: 759-781.

[13] Quatieri TF, Malyska N. Vocal-source biomarkers for depression: link to psychomotor activity. In: Proceedings of Interspeech 2012; 9–13 September 2012; Portland, OR, USA. pp. 1059-1062.

[14] Huang NE, Shen Z, Long S, Wu M, Shih H, Zheng Q, Yen N, Tung C, Liu H. The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis. Proc R Soc London Ser A 1998; 454: 903-995.

[15] Maryn Y, Corthals P, Van Cauwenberge P, Roy N, De Bodt M. Toward improved ecological validity in the acoustic measurement of overall voice quality: combining continuous speech and sustained vowels. J Voice 2010; 24: 540-555.

[16] Kacha A, Grenez F, Schoentgen J. Assessment of disordered voices using empirical mode decomposition in the log-spectral domain, In: Proceedings of Interspeech 2012; 9–13 September 2012; Portland, OR, USA. pp. 66-69.

[17] Kacha A, Grenez F, Schoentgen J. Multiband vocal dysperiodicities analysis using empirical mode decomposition in the log-spectral. Biomed Signal Process Control 2015; 17: 11-20.

[18] De Krom G. A cepstrum-based technique for determining a harmonic-to-noise ratio in speech signals. J Speech Hear Res 1993; 36: 254-266.

[19] Flandrin P, Rilling G, Conçalvès P. Empirical mode decomposition as a filter bank. IEEE Signal Proc Let 2004; 11: 112-114.

[20] Fant G, Liljencrants J, Lin Q. A four-parameter model of glottal flow. STL-QPSR 1985; 4: 1-13.

[21] Sahoo S, Routray A. A novel method of glottal inverse filtering. IEEE/ACM T Audio Speech Lang Process 2016; 24: 1230-1241.

[22] Alonso JB, Ferrer MA, Henríquez P, López-de-Ipina K, Cabrera J, Travieso CM. A study of glottal excitation synthesizers for different voice qualities. Neurocomputing 2015; 150: 367-376.

[23] de Corbiere S, Fresnel É, Freche C. La voix: la corde vocale et sa pathologie. Technical report, Collège International de Médecine et Chirurgie de l'Hôpital Américain de Paris, 2001 (in French).

[24] Fu Q, Murphy P. Robust glottal source estimation based on joint source-filter model optimization. IEEE T Audio Speech Lang Process 2006; 14: 492-501.

[25] Parsa V, Jamieson DG. Identification of pathological voices using glottal noise measures. J Speech Lang Hear R 2000; 43: 469-485.

[26] Lowen R, Verschoren A. Foundations of Generic Optimization: Volume 2: Applications of Fuzzy Control, Genetic Algorithms and Neural Networks. Berlin, Germany: Springer Science & Business Media, 2008.

[27] Chiroma H, Abdulkareem S, Abubakar A, Zeki A, Gital AYU, Usman MJ. Correlation study of genetic algorithm operators: crossover and mutation probabilities. In: Proceedings of the International Symposium on Mathematical Sciences and Computing Research 2013; 6–7 December 2013; Perak, Malaysia. pp. 39-43.

[28] Patil VP, Pawar DD. The optimal crossover or mutation rates in genetic algorithm: a review. International Journal of Applied Engineering and Technology 2015; 5: 38-41.

[29] Drugman T, Dubuisson T, Dutoit T. Phase-based information for voice pathology detection. In: Proceedings of the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing; 22–27 May 2011; Prague, Czech Republic. pp. 4612-4615.