Laryngeal Pathologies Analysis Using Glottal Source Features.

Fethi AMARA, Mohamed FEZARI
Department of electronic
Laboratory of Automatic and signal Annaba
Badji Mokhtar Annaba University
Faculty of engineering, bp:12, 23000 Algeria
Email: amafethi@gmail.com

Abstract—This experimental study looks at the problem of pathological disorders analysis from speech signals. Knowing that the vocal tract pathologies provide acoustic changes commonly called dysphonia. In particular, laryngeal pathologies have an effect on the larynx and vocal cords. When regarding to the speech production model, source-filter, the impact is even more apparent in the glottal source. Our study reports on: firstly, the estimation of the glottal source where the accurate estimation is considered. Secondly, parameterization of the glottal source. Many features are extracted such as normalized amplitude quotient (NAQ), difference between harmonic H1H2, peak slope. This analysis is performed around the German database which contains many recordings (normal and pathological subjects). More than 50 subjects are used where each speaker pronounce the sustained vowel [a] in normal intonation. The extracted parameters show an interesting tool for the discrimination between normal and pathological voice. In particular, the peak slope and harmonic richness factor present a significant differences. This work is developed under Matlab thanks to the project COVAREP (A Collaborative Voice Analysis Repository for speech technologies).

Index Terms—Speech signal processing, voice, laryngeal disorders, glottal source, source-filter separation, inverse filtering.

I. INTRODUCTION

Assessment voice quality for dysphonia evaluation is an active research field. The stat of the art shows that they are two main methods: first, qualitative evaluation based on the perceptual evaluation [1] GRBAS scale. Second method based on computing a measures that rating the degree of the abnormalities [2]. The major disadvantage of the first method is the variability between listeners this why a quantitative measurement is more interesting. On the other hand, it is one way in which an automatic detection system can be developed.

This experimental study focus on the analysis of the laryngeal pathologies affecting directly the larynx (muscles, nerves) and the vocal cords, such as, spasm, laryngitis and edema. The presence of those pathologies cause an organic or a functional disorder in the tract vocal which has an impact on the production of the speech signals. According to the source-filter model of speech production, laryngeal diseases effect can appear in the source of excitation named glottal source [3]. The separation of glottal source is a basic problem in speech processing.

As discussed by Thomas et al. in the overview of the current stat of the art [4], the glottal source processing have received big attention in many applications of speech processing technologies, the most important works are in speech synthesis [5], speech recognition [6] and biomedical applications [7]. The main idea behind this overview is to simplify techniques for glottal source processing starting from pitch tracking, glottal closed instant (GCI) and estimation of glottal flow.

In the literature, many features are found to characterize the glottal source, some of them are in the time domain, others are in frequency domain. Those features have proved to be effective to separate between different type of phonation [8,9].

Until recently, little attention has been given to those features to characterize laryngeal pathologies. In this study, we focus on the description and the analysis of some laryngeal pathologies using parameters related to the glottal source. Firstly, the estimation of the glottal flow waveform from speech signal is considered where the iterative adaptive inverse filtering (IAIF) algorithm is used [10]. Due to the irregularities in speech signal summarized in the disturbance of glottal cycles, analysis is carried out on a GCI (Glottal Closed Instant) synchronous basis in order to get accurate estimation. The location of those instants from speech signal is very difficult, there exist a variety of methods, the most important which present an accurate and robust methods are: SEDREAMS (Speech Event Detection using the Residual Excitation And a Mean-based Signal) [11] and SE-VQ [12] for non-modal voice.

Electroglottograph EGG gives information on the closure of vocal folds by measuring the electrical resistance between two electrodes placed around the neck. From EGG waveform, GCI can be detected accurately by locating the positive peaks in the derivate of EGG [13]. It can provide a ground truth to evaluate GCI detection algorithms.

Secondly, once glottal source is obtained a set of features describing the voice quality is extracted. It is discussed which from those parameters have the ability to separate between laryngeal disorders.

The purposes of this study were to:

- Estimate the glottal source in the case of laryngeal disorder.
- Evaluate the effect of multiple diseases on the glottal
source
• Prepare strong discriminative features for an automatic recognition purposes.

The paper is organized as follow: Section two introduce briefly the mechanism of speech production, laryngeal pathologies and therefore their effect on the speech signal. Section 3 is reserved to the proposed method. Section 4 concerns the experiment presentation. The obtained results are discussed in the section 5. The last section contains the conclusion and the future work.

II. THE EFFECTS OF LARYNGEAL DISORDERS ON SPEECH SIGNALS:

The production of the speech signal can be given by the model source /filter. The source or glottal flow is produced when the airflow coming from the lungs passing through the trachea and vibrate the vocal folds. Speech is the result of the convolved glottal flow and the impulse response representing the vocal tract cavities.

Some laryngeal diseases such as polyp, nodule and tumors can provide changes in mass, elasticity or tension of vocal folds. Others pathologies as spasmodic dysphonia disturb the normal functioning of the larynx muscle. The major resulting symptoms are the pitch and the amplitude perturbations termed in some references by the aperiodicity in speech signal. Jitter and shimmer have been used for long time to quantify this perturbations. We find also that the asymmetric distribution of the vocal folds mass disturb their dynamics and provide a turbulence noise at the glottis, as results, the produced speech have a modified harmonic structure. Fig 1 shows speech signal and EGG of some cases: normal, laryngitis and spasmodic dysphonia. Hypothesis:

According to the description of laryngeal disorders, their effect can appear in the glottal source. We suppose that the parameters related to the glottal source are effective to characterize those pathologies.

III. PROPOSED METHOD

Fig.4 shows a block diagram describing the process set up for laryngeal disorders analysis. A short description of each step is included in the following subsections.

A. Glottal closer instant (GCI)

GCI are defined as the events describing the activity of the glottis, it occurs when the vocal folds are in the nearest position. The accurate determination of GCI is an important task to perform glottal synchronous analysis, especially in the pathological cases.

Many algorithms are dedicated to detect those instants from speech signal. In the goal to establish a comparative study two main methods are used. First one is the Detection using the Residual Excitation And a Mean based Signal (SEDREAMS) and the second is SE-VQ, this method is developed to detect those instants from non-modal like breathy, tense and harsh voice where VQ refers to voice quality. This algorithm proposed by John Kane et al to enhance the exciting algorithm SEDREAMS for a specific phonation types. This is why we are motivated to use this method in our case which concern laryngeal disorders.

We note that the determination of GCI by the two cited algorithms needs to determine the fundamental frequency and the voiced segment. The algorithm based on the Summation of Residual Harmonics (SRH) has been shown to provide a robust estimation of both F0 and voicing decision, especially in the case of pathological voice.

Electroglottography EGG is defined as the measurement of the electrical conductance of the glottis recorded at the same time with phonation operation.

EGG can provide an important information about glottal activity, it’s derivative (DEGG) contains high amplitude temporal features which corresponds to GCI and small features of opposite sign corresponds to Glottal Opening Instant (GOI).
The SIGMA algorithm (Singularity in EGG by Multiscale Analysis) developed in [14] is dedicated to detect those instants. For good analysis the EGG waveform is filtered by a low pass filter (RIF) to attenuate high frequencies which give peaks in the first derivative, those peaks can be confused with other peaks representing GCI. It is termed Lx. The obtained instants are considered as benchmark to evaluate the performances of other exciting techniques. In this study it is compared with SEDREAMS and SE-VQ.

B. Glottal source estimation

Glottal source estimation have interested many researchers in the domain of speech processing. It is well known that voiced speech is a product by an excitation signal (source) convolved with impulse response generated by the tract vocal(filter). Its production can be represented as:

\[ S = G(z) \ast V(z) \ast R(z). \]  

(1)

By removing the effect of the lip radiation R(z) and vocal tract V(z) the glottal source G(z) can be obtained. This is the principle idea of the inverse filtering termed in the literature by GIF (Glottal Inverse Filtering). Many research are concentrated, IAIF (Iterative Adaptive Inverse Filtering) developed by Alku et al. is one of interesting methods to estimate glottal source. This algorithm is based on the hypothesis that the vocal tract is considered as autoregressive filter. First approximation of the spectrum can be obtained by computing LPC analysis of order one to the speech spectrum.

C. Glottal features

As mentioned in the fig 1, after getting the glottal flow waveform we proceed to the parameterization step. Features can be divided into two categories: time domain features and frequency domain features. A brief description is in the following section.

Time domain features:

- Normalized amplitude quotient (NAQ) is proposed by Alku 2002 as a method to parameterize the glottal source. It is one of amplitude based methods, it can obtained by computing the ratio between the maximum amplitude of the glottal flow and the maximum negative amplitude of the differentiated glottal flow recognized by amplitude quotient (AQ), this quotient is normalized by T as represented in the following express:

\[ AQ = (f_{0c} \ast f_0)/d_{peak} \]  

(2)

The most previous works use NAQ was present an efficient indicator to discriminate breathy to tense voice.

- Quasi Open quotient (QOQ): this type of parameters is one variant of the standard open quotient OQ. It is defined as the ratio between the duration during which the amplitude of the glottal flow is 50% of the peak amplitude, and the pitch period.

Frequency domain features:

- H1-H2 this parameter denotes the difference between the two first harmonics (in dB) in the narrowband of the voice source spectrum.

- Harmonic richness factor (HRF) by means of this parameter the amount of harmonics in the voice source spectrum can be quantified. It can be obtained by the computation of the ratio between the sum of the amplitudes of harmonics and the amplitude at the fundamental frequency.

- Harmonic richness factor (HRF) by means of this parameter the amount of harmonics in the voice source spectrum can be quantified. It can be obtained by the computation of the ratio between the sum of the amplitudes of harmonics and the amplitude at the fundamental frequency.

- Parabolic spectral parameter (PSP) this parameter describes how the spectral decay of glottal flow behaves with the respect of theoretical limit corresponding to maximal spectral decay.

- Peak slop: the peak slope parameter is derived by using wavelet analysis to decompose the glottal source signal into octave bands. A sliding window is then used for measuring time-domain maxima in the different frequency bands. A regression line is fit to log10 of these peaks, and the peak slope coefficient is used as the peak slop parameter. For the wavelet analysis, different scaled versions of a cosine modulated gaussian pulse are convolved with the input signal.

IV. Experiment presentation

A. Database description

The database presents an essential factor to develop a detector, according to the overview of Nicolas Saenz et al [14], the use of standard speech corpora might be necessary to compare the obtained results with those that exist, it allows researchers to test the effectiveness and the reliability of the used methods. This work is built around Saarbrucken Voice Database (SVD), it is a free database developed by Putzer et al at the Institute of Phonetics, University of Saarland (Germany) [15]. It contains healthy and pathological recordings as follow:

- Sustained vowels /a/, /i/, /u/ pronounced at different intonations (low, normal, high and low-high-low) during 1-3s.
- Sentence "Guten Morgen, wie geht es Ihnen?", it means (Good morning, how are you?).
- Electroglottogram EGG.

All files are digitized at 50 KHz and downsampled to 16 KHz for this study. The table below presents the used speech corpora:
B. GCI detection

As mentioned above two algorithms are used to locate GCI from speech signal. Fig 5 shows the obtained GCI via SEDREAMS (red impulse) and GCI using SE-VQ (green impulse) positioned on the speech signal (A). Comparing with GCI obtained from DEGG by SIGMA algorithm (B and C), it appears clearly that GCI obtained by the algorithm SE-VQ is more accurate than those obtained by SEDREAMS. The difference will be more seriously with pathological voice as shown in fig 6. The algorithm SE-VQ is more preferment with laryngeal pathologies.

C. Glottal flow estimation

As mentioned above the glottal flow is estimated using the IAIF algorithm based on GCI synchronous. The location of the window of analysis is very important. We have used hamming window of two period of length. It is centered on GCI as mentioned in fig 7.

V. RESULTS AND DISCUSSION

The results are presented by the boxplot.

TABLE I

<table>
<thead>
<tr>
<th></th>
<th>Number</th>
<th>Age</th>
<th>Gender</th>
<th>Vowel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laryngitis</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leukoplakie</td>
<td>10</td>
<td>20-60</td>
<td>male</td>
<td>/a/</td>
</tr>
<tr>
<td>Edema</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spasm</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The analysis is established into two tasks:

Task 1: Discrimination between normal and pathological voice
The first lecture shows that the HRF and Peak Slope are more efficient for the binary separation (normal/pathological). HRF values are remarkably increased for all pathological voices compared with the normal voice values in particular, with the edema which presents the maximum values. It confirms our hypothesis that the harmonics structure of the speech signal in some pathological cases is modified. Unlike HRF the peak slope values are decreased significantly where laryngitis and leukoplakia presents the maximum value. In the second range, the QOQ and PSP provide less discriminative features compared with the two first parameters but it remains always legible. For both parameters values are decreased in relation to the normal value. From those results it is remarkable that those parameters are more efficient with the organic diseases (laryngitis, leukoplakia and edema) compared with the functional diseases presented in the study by the spasmodic dysphonia. For the remain parameters (H1H2 and NAQ) it fails to provide a clear discrimination excepting with leukoplakia which presents a different significance in NAQ and edema with important value of H1H2.

Task2: Discrimination between pathological voice:
The obtained results are very close to each other for the four diseases, it is really difficult to separate between them using those parameters, notably when separation concerns organic diseases.

VI. CONCLUSION AND FUTURE WORK
This work focus on the use of the parameters describing the glottal source to characterize the laryngeal pathologies. Both organic (laryngitis, leukoplakia and edema) and functional (spasmodic dysphonia) diseases are considered. A set of parameters include features in time and frequency domain is used. This study shows that the binary discrimination (normal/pathological voice) using the vowel /a/ at a normal intonation is easier than the discrimination between pathological voice. In the top of the list, peak slop and HRF provide a better and clear separation. In the second range, QOQ and PSP are less discriminative. While H1H2 and NAQ fail to offer a significant information.

In future work we are firstly, motivated to extract the same features from different vowels (/a/, /i/ and /u/) with different intonations (low high and low-high-low). Where the main objective is to find a best separation in particular with pathological voice. Secondly, prepare the obtained features using features selection algorithms to develop an automatic recognition system.

ACKNOWLEDGMENT
This work is developed thanks to the project COVAREP.

REFERENCES
[9] John Kane, Christer Gobl wavlet maxima dispersionfor breathy to tense voice IEEE transaction on audio, speech and language processing 21 (6) june 2013.