SCIENTIFIC CULTURE
www.sci-cult.com

SCIENTIFIC CULTURE, Vol. 12, No. 5.1, (2026), pp. 843-857
Open Access. Online & Print

DOI: 10.5281/zenodo.12511068

AUTOMATIC DETECTION OF VOICE PATHOLOGIES USING GLOTTAL SOURCE CHARACTERISTICS AND BI-LSTM NETWORKS

Anil Kumar V[1]*, R Venkata Siva Reddy[2]

[1,2] School of Electronics and Communication Engineering, REVA University, Bengaluru-560064, Karnataka, India

Received: 11/12/2024
Accepted: 25/02/2025

Corresponding author: Anil Kumar V
(anilkvmr@gmail.com)

ABSTRACT

Early diagnosis intervention and objective assessment are made possible by the automatic identification of voice pathology. This work gives an efficient investigation of glottal source characteristics and looks at how well they work to identify voice pathologies. The estimation of the glottal source flow is undertaken via the quasi-closed phase (QCP) methodology, incorporating the glottal-inverse-filtering (GIF) technique to facilitate an intricate and highly refined analysis. The provided input acoustic speech signal is further processed using the IAIF (Iterative-Adaptive-Inverse-Filtering) technique to extract glottal source waveforms. Ultimately, to meticulously capture the fluctuations in the glottal source spectra associated with speech disorders. This study uses the IAIF and QCP methodologies to calculate derived MFCCs (Mel-frequency cepstral coefficients), which are suggested by glottal source waveforms. To evaluate the systems, we have used SVD (Saarbrucken Voice Disorders) database to detect the normal and pathological voices. There is information in the glottal source that can differentiate between normal and disordered voice, according to features that were analysed. Bi-LSTM or Bi-directional long short term memory is employed to identify the normal and pathological voices. From the simulation results it is observed the proposed QCP-MFCC, IAIF-MFCC and the combination feature sets provides the better results.

KEYWORDS: Quasi-closed phase (QCP), Glottal source, Iterative Adaptive Inverse Filtering (IAIF), Inverse filtering.

1. INTRODUCTION

A time-varying excitation signal stimulates lips, tongue, and jaw that constitute a time-varying vocal tract system, which produces speech. In speech communication, the primary goal is to transmit a linguistic message. In addition to language content, speech includes an extensive amount of information about the language and dialect, state of a health a person, age and gender. This study examines the difference between fit and pathological voices, to detection and categorise the pathological diseases of the voice. The voice pathology disorder arises because of infection, psychogenic and physiology causes and mainly as a result of the vocal misuse, and it has been commonly occurring in the professions like, teachers, singers, and customer services. Further, the infections, psychogenic and physiological reasons can all lead to voice disorders [1], [2]. Early diagnosis intervention and an objective assessment are made possible by the automatic detection of vocal pathology.

The provided acoustical voice signal's depiction (also known as feature extraction) and the classifier comprise the two primary steps of a standard system for detecting vocal pathology. The first stage is the primary focus of the current investigation. Three major categories may be utilized to classify feature sets used for speech pathology identification [3], [4]: (a) measurement of perturbation, (b) cepstral as well as spectral measures, and (c) complexity measures. The vocal folds' inconsistent movement and inadequate glottal closure that cause aspiration noise and aperiodicity in the acoustic voice signal are captured by perturbation measurements.

Jitter and shimmer are the most often used characteristics in this category [5]–[11]. Shimmer measures short-term amplitude disturbances, whereas jitter measures short-term fundamental frequencies (F0) perturbations [5, 12, 13]. Voice pathology detection has made use of a number of jitter and shimmer variations, which include absolute, relative, and shimmer with a amplitude perturbation of three-point quotients, as well as relative jitter five-point period perturbation quotients and jitter average perturbation [12], [13]. These attributes are estimated based on F0, although it is well known that pathological voice makes it difficult to estimate F0 accurately [14], [15]. Jitter and shimmer characteristics have been the subject of several prior research, It is significant to note that the American Speech-Language-Hearing Association excludes these elements from its recommended feature set due to insufficient clinical voice value [16]. Refer to Table 2 in [16] for further information on the suggested acoustic measurements.

The normalized-noise entropy (NNE) [17], the ratio of GNE (glottal-to-noise excitation) [18]–[20], the HNR ratio (harmonics-to-noise) [21], [22] are some well-liked perturbation techniques that measure the existence of aspiration noise. The harmonic element's energy is divided by the noise component's energy to determine HNR. The noise energy to total signal energy ratio is known as NNE. GNE computes the interrelationship of the Hilbert envelope across disparate frequency bands within the acoustic speech stream.

Because they are easy to calculate and do not need an estimation of F0, parameters derived from cepstrum and spectrum been employed for voice pathology identification [15], [23], and [24]. MFCCs [4], [15], and [25] are the most well-liked features in this category. They make use of cepstrum's decorrelating property and the mel-scale principles of human auditory perception.

Furthermore, perceptual linear prediction (PLP) [4], and LPCCs [21], [26], [27], have been used to the identification of voice pathologies. LPCCs record the vocal tract system's characteristics. The Bark scale, intensity-to-loudness and equivalent loudness-level curve conversion models of human auditory tract are the PLP features sources [28] [29]. A higher CPP value suggests that the signal's periodic structure is more pronounced. The "smoothed CPP" is a term that denotes a Cepstrum variation whose corresponding variable was smoothed to generate the CPP. [30]. Furthermore, investigations have used average spectral energies in both high- and low-frequency bands [31]. Voice pathology detection has also been studied using features obtained from the frequency-time decomposition approaches, such as empirical mode decomposition [32], modulation spectrum [33], wavelet transform [34], and iterative time and frequency transform [35], [36]. Complex measures were proposed to capture signal properties such non-linearity, aperiodicity and non-stationarity utilising non-linear dynamic analysis-based estimators. [5], [37]. It's been well recognised that natural physiological systems, including voice production, frequently experience nonlinear events. The dynamic variations in voice disorders brought on by irregular and incorrect vocal fold movement are characterised by non-linear dynamic analysis. The correlation dimension or the fractal dimension are used to calculate the popular parameters in this category [37]. Numerous studies have examined various complex measures, including entropy such as entropy of modified sample, fuzzy, HMM,

detrended, Shannon HMM and Lyapunov exponent, further, Hurst exponent [38], [39], and [40]. These characteristics record the signal's long-range correlations, regularity, predictability, and dynamic variants/invariants.

Noted that the prerequisites include selecting the appropriate window length and accurately estimating F0 for the complexity and estimation of perturbation features [14]. However, F0 is not necessary for the extraction of spectral or cepstral characteristics. It has been demonstrated in [13] and [41] that the performance of voice pathology identification using spectral properties (PLPs and MFCCs) exclusively is comparable to or superior to that provided by elements of complexity and disruption in continuous speech and prolonged vowels. More information on the research on diseased voice and the many characteristics that are used to diagnose voice pathology in latest review papers [4], [13]. In terms of classifiers, pathological voice has been treated with a number of well-known methods, including SVM, KNN, HMM, LDA, GMM, ANN, and CNN [25], [27], [40]. SVM has been determined to be the best classifier among the others for speech pathology identification [42]. Further information on the different classifiers and machine learning methods used to identify speech pathology found in the most current review that was published in [42].

Voice disorders impact the mechanism that produces speech, hence in order to analyse and diagnose voice pathology, the glottal source as well as vocal tract system must be accurately modelled and parameterized. Previous research has successfully captured the vocal tract properties scheme by obtaining cepstral or spectral features, such as PLPs and MFCCs. On the other hand, there hasn't been much focus on the effective analysis of glottal source characteristics for the assessment and diagnosis of voice issues in the past.. The authors of [43] [44] have mostly taken use of traits like HNR, GNE, and spectral energies in the glottal source's low- and high-frequency bands that capture its unique properties.

2. RELATED WORKS

In order to extract the glottal source waveforms, two distinct signal processing methods were employed in this section: QCP and IAIF.

2.1. Quasi-Closed Phase (QCP) Methodology

A newly suggested GIF approach for automated estimation of a glottal source waveform obtained using the QCP approach from the voice [45]. On the basis of closed-phase (CP) concepts [46] assessment, the approach uses linear prediction (LP) study to determine the vocal tract modeling from a small number of voice found in the glottal cycle's CP. When creating the vocal tract model, QCP utilizes every voice sample from the analyzed frame, in contrast to the CP approach. Weighted linear prediction (WLP) analysis is used in this, and AME [47] waveform is used as the weighted function. The GCI's are used in the construction of the AME function, which reduces the open phase samples' contribution to the covariance or auto-correlation function. This approach yields better estimatation of V(z). The provided acoustic speech signal is inversely filtered using the function V(z) to finally estimate the waveform of glottal flow. In the calculation of the flow of a glottal source non-modal and modal kinds of phonation, furthermore, The QCP algorithm outperformed four inverse filtering methods [45]. This explains why QCP was used in this study's glottal inverse filtering procedure. Fig. 1 the block diagram that outlines the procedures in the QCP technique.

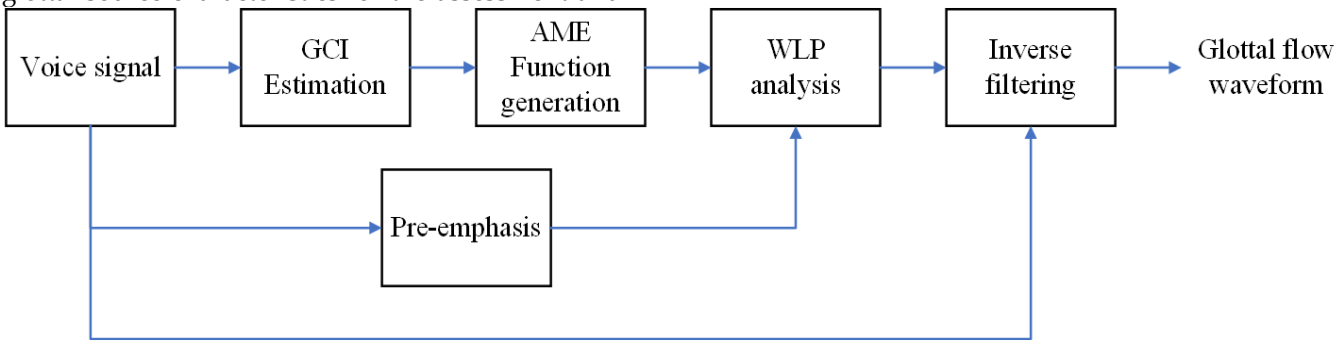


Figure 1: Block diagram of QCP

2.2. IAIF Method

IAIF [48] is a GIF method based on autocorrelation LP that depends on prior understanding of the overall structure of vocal tract and glottal source transfer functions. Using a twice-repeated iterative analytic process, IAIF attempts to eliminate the glottal source's tilting impact from the voice spectrum. The first iteration uses a first-order IIR

filter to represent the glottal source's component, calculates the vocal chord transfer coefficients, and then inversely filters the frame using the results. The glottal source envelope is approximated in second iteration using LP analysis of order. The IAIF technique has the advantage of being simply calculated from the voice waveform input without the requirement for further constraint estimate.

3. METHODOLOGY

3.1. Feature Extraction: Glottal Source – QCP features

Various parameterization approach been developed to describe glottal flow waveform in a compact manner. These techniques may be classified into two groups: glottal features in time and frequency-domain, which are also referred to as glottal parameters.

3.1.1. Glottal Features : Time-domain

Characteristics based on time and amplitude can be used to parameterize glottal flow in time-domain data [49], [50]. To define the measurements, the glottal source waveform is used to extract key temporal instants like glottal closure, the principal and subsidiary glottal opening, and the lowest and maximum glottal flow. The most popular technique for time-dependent attributes is to calculate the time-duration proportions of opening and closing phases glottal sources waveform inside a glottal cycle. The glottal flow's amplitude and derivative are used in amplitude-based features [51, 52]. Voice quality analysis has made considerable use of the standardised amplitude quotient due to its strong correlation with the closure quotient [52]. Since critical time instants can be hard to find, time-based characteristics are sometimes computed by substituting the time instants at which the glottal flow crosses a level for the actual closure and opening instants. The value that falls between the glottal flow's highest and minimum amplitude during a glottal cycle is known as this level [50].

3.1.2. Glottal Features: Frequency domain

Although calculating time-domain characteristics from the waveform of the glottal source is simple, formant ripple and other distortions result from the inverse filter's insufficient cancellation of formants [50]. Deriving frequency domain properties for the waveform of glottal source is helpful in these situations. The glottal source's spectrum is used to compute frequency-domain characteristics, which basically quantify the spectrum's slope. Using the level of F0 and its harmonics, the glottal source spectral slope has been measured in a number of experiments. The PSP [53], the HRF [54], and the amplitude difference of F0 and H1-H2 first harmonic [55] are the most often utilised characteristics. The ratio of amplitudes of all harmonics beyond F0 to the amplitude of F0 is known as the HRF. The glottal flow spectrum's low frequencies are fitted with a parabola to get PSP [53]. QCP spectra for healthy and infected voice samples are shown in Fig 2.

The waveforms of glottal flow were calculated using the QCP-GIF approach are characterised in this work using a over-all of 12 glottal parameters, 3 frequency-domain and 9 time-domain features, as described in [49]. These characteristics are shown in Table 1 and were retrieved using the APARAT Toolbox [49].

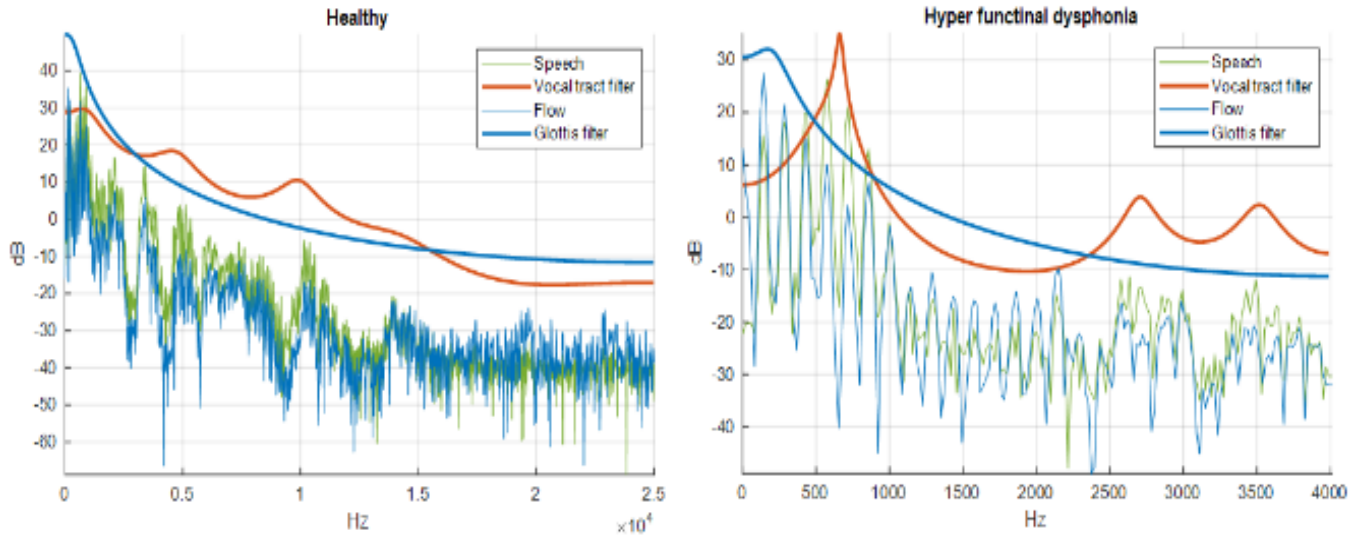


Figure 2: exhibits healthy and path QCP spectra.

Sl. No.	Parameters:Time Domain F
1	OQ1: Open-Quotient, derived from primary glottal opening
2	SQ1: Speed-Quotient, derived from primary glottal opening
3	OQ2: Open-Quotient, derived from the opening of the secondary glottal
4	SQ2: Speed-Quotient, derived from secondary glottal opening
5	AQ: Amplitude-Quotient
6	NAQ: Normalized-Amplitude-Quotient
7	OQa: Open-Quotient, calculated from LF-model
8	CIQ: Closing-Quotient
9	QoQ: Quasi-Open-Quotient
Frequency Domain Features	
1	HRF: Harmonic-Richness-Factor
2	H1-H2: The first and second glottal harmonics' amplitude differences
3	PSP: Parabolic-Spectral-Parameters

Table 1: Frequnecy and Time domain features obtained by QCP

3.2. Feature Extraction: Glottal Source – IAIF features

A technique called IA-IF automatically separates an audio voice sample into the vocal tract transfer function and the glottal source. The IA-IF speech model covers vocal tract, lip, and glottal radiation effects. The IA-IF method uses adaptive LPC (all-pole) for vocal tract transfer. A lower-order LP filter that is adaptive first determines the speech sample signal the spectrum, and it is then averaged to account for the glottal source impact. This results in the glottis being excited. The IAIF technique has two versions. To provide an initial estimate of the glottal contribution, the initial iteration constructs a first-order LPC framework. To provide a more accurate representation for the glottal impact the next cycle computes the greater-order LPC model.

Fig. 3 depicts the Iterative IA-IF block diagram. The blocks labelled (a) through (f) correspond to the first iteration, while the remaining blocks labelled (g) through (k) correspond to the second iteration. The following steps make up the IA-IF approach for glottal flow estimation. The provided voice sample is filtered using a high-pass filter in the first block (Block a) to eliminate lower frequency background noise that is caused by the microphone. By calculating an LPC order of one, the following block (Block b) predicts the initial combined effects of lip radiation and glottal flowing$Fge-1$ in Figure 3 represents the LPC inverse filter transfer. Since the spectral system has just one adaptive pole on the real axis in the Z-domain during LPC order one, preliminary estimations of glottal flow and lip radiation effects are presented. If lip radiation and glottal flows are estimated, a higher-order LPC analysis may simulate speech sound resonant structure. This effect should be deliberately avoided. In third block (Block c), the $X(n)$ has its glottal flow and lip radiation cancelled by the use of inverse filtering. The primary vocal tract filter approximation $Fvt-1$ is calculated by analysing the output of Block C utilising LPC order-N (Block D). Block E filters the speech stream using inverse filtering to remove the estimated vocal tract transfer function. The lip radiation effect, which is regarded as a fixed differentiator, is eliminated using the integration block (Block F). The resulting signal is used to generate a first estimate of the glottal flow. Additionally, the spectral envelope parametric model of the anticipated glottal flow is generated by estimating LPC order two (Block G). The goal will be to compute the greater-order LPC assessment more precisely compared to block (b) while yielding models that show the false formant-like peaks, as block (e) yielded the estimated flow by extracting the vocal tract from the filtered speech signal. In addition, inverse filtering (block h) is used to remove the glottal component. Block (i) calculates the final estimate of vocal tract $Fvt-2$ using LPC analysis of order-N. By reducing the lip radiation and vocal tract impact using inverse filtering (block j) and integration block (block k), the glottal flow signal g(n) finally produces.

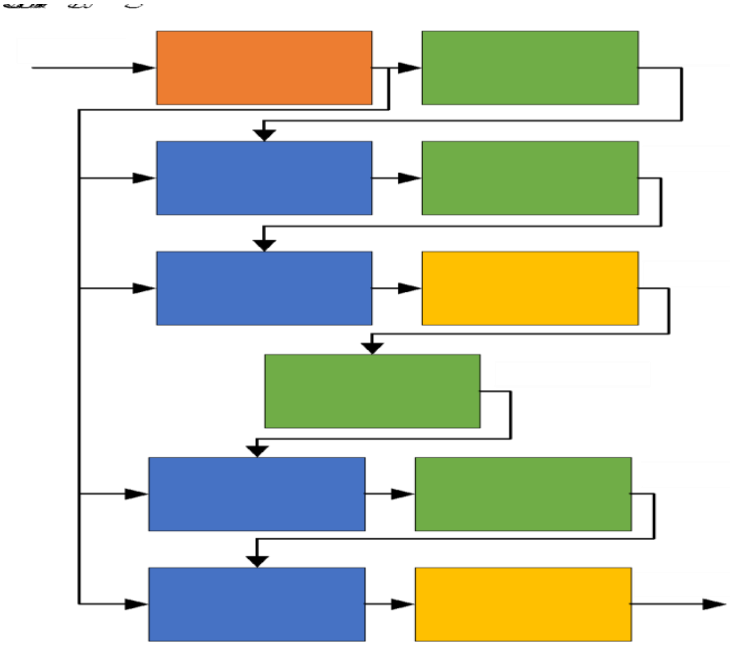


Figure 3: Schematic of IA-IF method

$$F_{ge-1}(z) = 1 + pz^{-1} \quad (1)$$

$$F_{vt-1}(z) = 1 + \sum_{l=1}^{N} p(l)z^{-1} \quad (2)$$

$$F_{ge-2}(z) = 1 + qz^{-1} + rz^{-2} \quad (3)$$

$$F_{vt-2}(z) = 1 + \sum_{l=1}^{N} q(l)z^{-1} \quad (4)$$

3.3. Proposed MFCC Feature Extraction the Derived from Glottal Flow Waveforms

When the glottal source features are examined, it becomes evident that the frequency-domain features which are obtained from the glottal source spectrum are more discriminable than the features of time-domain. This encourages us to employ the whole glottal source spectrum for vocal pathology identification rather than just a few select characteristics. The spectrograms of the pathological and normal speech glottal flow waveforms, as assessed by the QCP approach, are displayed in Fig.4. The sick and normal voices differ significantly in glottal flow spectrum harmonic structure.

To compactly represent and capture these fluctuations, we propose extracting MFCCs from glottal source waveforms spectral content [56]. Despite using glottal source waveform rather than an Acoustic voice signal, the suggested feature extraction method for MFCC is equivalent to typical MFCC feature computation.

Figure 4 depicts the MFCCs extraction method from QCP and IAIF glottal source waveforms. By dividing the glottal source waveform into overlapping time frames, the method utilizes short-term spectral evaluation, with each frame's spectrum determined through the application of the DFT. By employing a 1024-point DFT, the spectrum is estimated via Hamming windowing in a frame 25 ms alongside a 5 ms shift. The DCT and logarithm are utilized to process the mel-cepstrum, which is obtained from mel-scale investigation of the glottal source's spectrum. For each frame, the first 13 coefficients including the 0th coefficients from the complete mel-cepstrum are considered. The cepstral coefficients that are generated as a result of the computation of glottal source waveforms using ZFF and QCP are regarded to as MFCC ZFF and MFCC-QCP, accordingly. Static cepstral coefficients are also used to compute delta and double-delta coefficients.

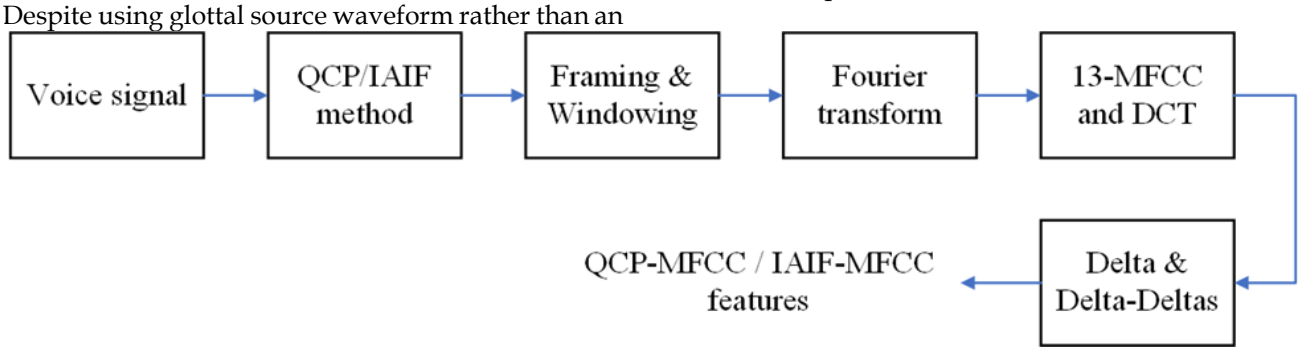


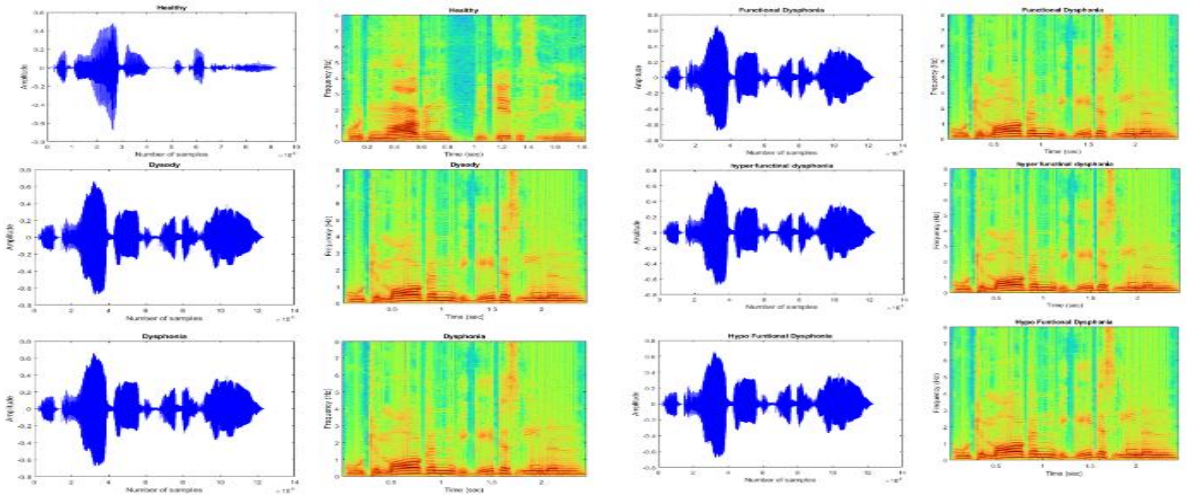
Figure 4: Block diagram of QCP/IAIF based MFCC feature extraction method

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

The databases, feature sets, classifier, and assessment metrics utilized in pathological and normal voices detection and are described in this section.

4.1. Database description

The Saarbruecken Speech Database (SVD) is used to evaluate the proposed scheme. 2000 speakers participated in the creation of the Saarbruecken speech database (SVD), which includes voice and electroglottography (EGG) signal groups [57] [58]. The medical records of 1356 patients with diverse vocal disorders and 687 healthy people (428 women and 259 men) are included. During the entire process of recording, the vowels "i," "a," and "u," as they seem in natural voice, Vowels with falling-raising pitches, as well as high and low pitches, are all employed. Furthermore, "Guten Morgen, wiegeht es Ihnen?" is also spoken in German. "Hello, how are you today?" every recorded SVD voice was processed with a 16-bit resolution at 50 kHz. Fig.5 shows the time and frequency domain of healthy and different pathological voice samples.



Figur 5: Time domain and the frequency spectrum representation of healthy and different pathological voice samples.

4.2. The Proposed feature sets obtained from glottal source and variables utilized for extraction of feature

In this work, glottal source characteristics of four sets are examined in total, as follows:

- Characteristics in the frequency (HRF, H1–H2, PSP) and time domains (OQ1, SQ1, OQ2, SQ2, AQ, NAQ, OQa, CIQ, QoQ) obtained from the glottal source waveforms produced by the QCP approach. Every glottal cycle's worth of these characteristics are extracted, and QCP analysis is done using Hamming window with the 25-ms frame size along with a 5-ms frame-shift.
- The IAIF approach was utilised to obtain the glottal source features. Further, in this work the IAIF method was applied in this investigation with orders p=10 and q=4. Every step of the LP analysis was carried out utilizing the Hamming window.
- The waveforms of the glottal source acquired by QCP are utilized to compute QCP-MFCC, and a hamming window is employed as the window function with a 5 ms frame-shift and 25 ms frame size. The coefficients double-delta and delta of the first 13 static cepstral coefficients are computed to create feature vector of 39-dimensional.
- The glottal source waveforms obtained by IAIF are used to compute IAIF-MFCC, and a hamming window is employed as the window function with a 5 ms frame-shift and 25 ms frame size. The first 13 static cepstral coefficients' delta and double-delta coefficients are computed to create the 39-dimensional feature vector.

4.3. Classifier

The most current literature study used a variety of classifiers to detect the out-of-the-ordinary noises, including ANN, SVM, GMM, and RF Random Forests [59] [60]. In this works, to classify the normal and pathological voices, Bi-directional LSTM, or Bi-LSTM, is what we have chosen. The input size of 39, the hidden layer's 256 neurons, using 64 as the mini-batch size, and the dropout value of 0.5 are the many parameters that are employed in the network. A recurrent neural network (RNN) architecture known as Bi-LSTM is an effective method for capturing

dependencies in sequential data. A "Bi-LSTM" is a recurrent neural network, which can efficiently capture the dependencies in a sequential data. Through forward and backward processing of input sequences, the Bi-LSTM model improves upon the normal LSTM architecture, allowing it to simultaneously collect context from the past and the future.

4.4. Simulation results

Fig.6 and Fig.7 shows the time and frequency domain of healthy and different pathological voice samples glottal derivatives of the iterative adaptive inverse filter (IAIF) and Quasi closed phase (QCP) respectively. From the spectrogram of IAIF and QCP can observe that the IAIF estimates the glottal flow and vocal tract filter using an iterative deconvolution approach and produces cleaner vocal tract formant structure but may blur source characteristics. Further, the glottal flow estimation is done via inverse filtering (linear prediction + iteration). Whereas, QCP models the speech signal during the glottal closed phase to isolate glottal flow and reduce vocal tract influence and enhances spectral features related to source (glottal) pathology, improving detection of abnormal phonation. Further, the glottal flow estimation implicitly separated using the closed-phase advantage.

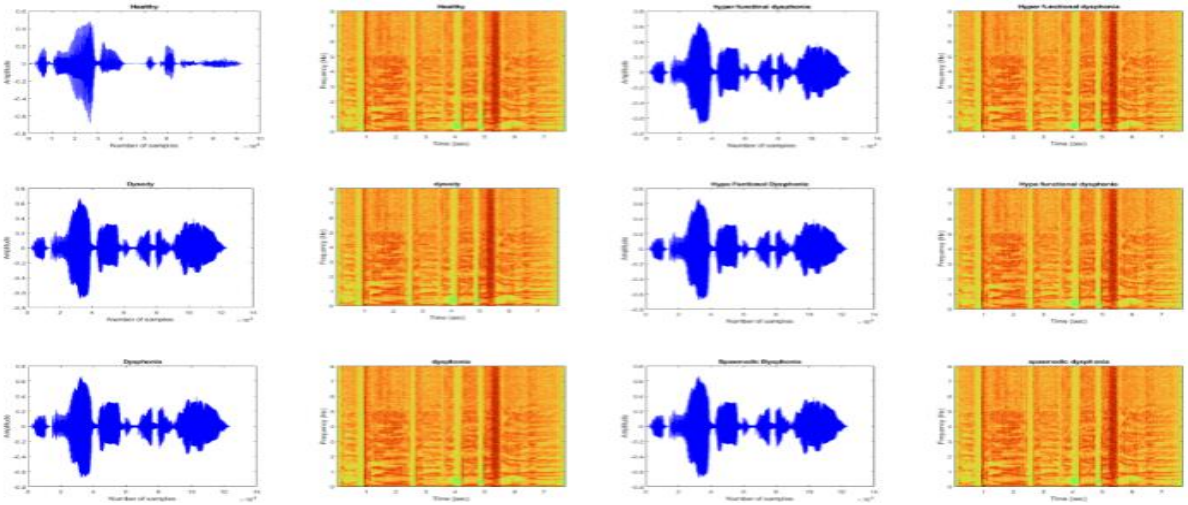


Figure 6: Time domain and the iterative adaptive inverse filter (IAIF) based frequency spectrum representation of healthy and different pathological voice samples.

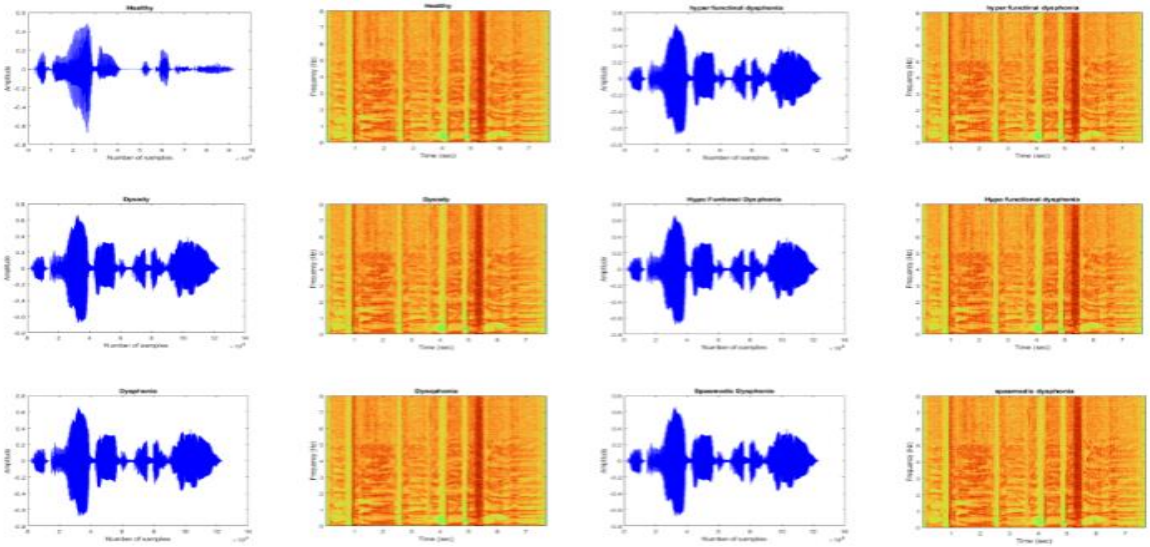


Figure 7: Time domain and the Quasi closed phase (QCP) based frequency spectrum representation of healthy and different pathological voice samples.

Distributions of the glottal source features obtained for both normal and diseased voice using the QCP approach are displayed in Fig. 8. Three frequency-domain features and nine time-domain features are displayed in the fig.8 (a)-(i). It is evident that when compared to time-domain features, frequency-domain features produce a higher ability to distinguish between normal and disordered voice. NAQ is more effective than the other features at differentiating between normal and diseased voice in

the time-domain features. When compared to normal speech, the open quotients OQ1 and OQ2 exhibit greater variances in pathological voice, while QoQ shows less discriminability. Conversely, the open quotient (OQa) based on the LF model exhibits strong discriminability. There are general minor variations in the distributions of AQ, ClQ, SQ1, and SQ2 between normal and diseased voice. The inability to recognize crucial glottal temporal instants (primary and secondary glottal opening, and the instant of glottal closure) could be the cause of this.

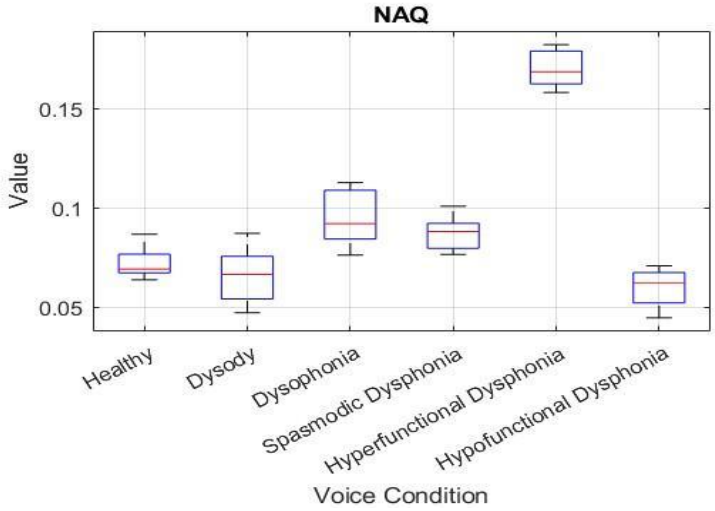


Figure 8 (a): NAQ glottal source parameter distribution derived from QCP model for both healthy and different pathological voices

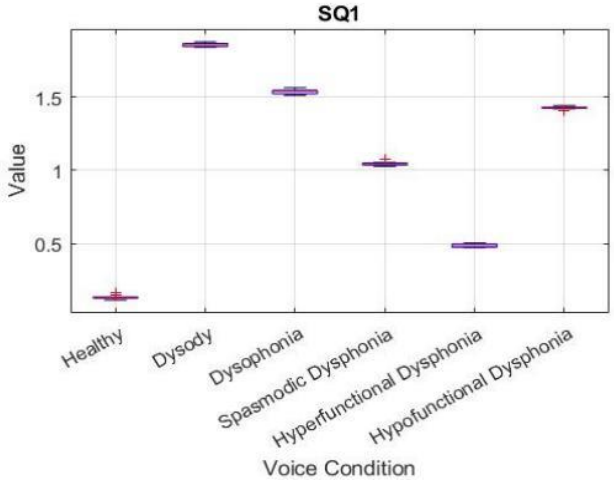


Figure 8 (b): SQ1 glottal source parameter distribution derived from QCP model for both healthy and different pathological voices

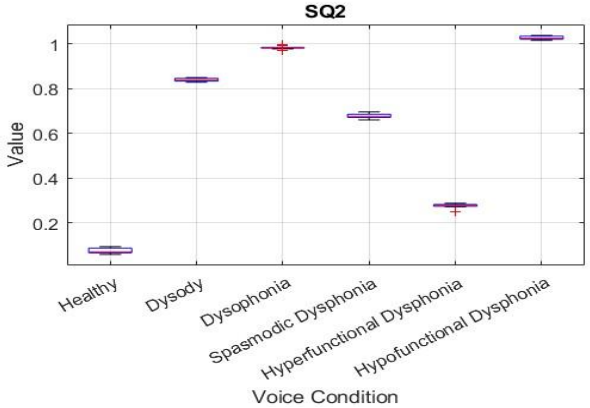


Figure 8 (c): SQ2 glottal source parameter distribution derived from QCP model for both healthy and different pathological voices

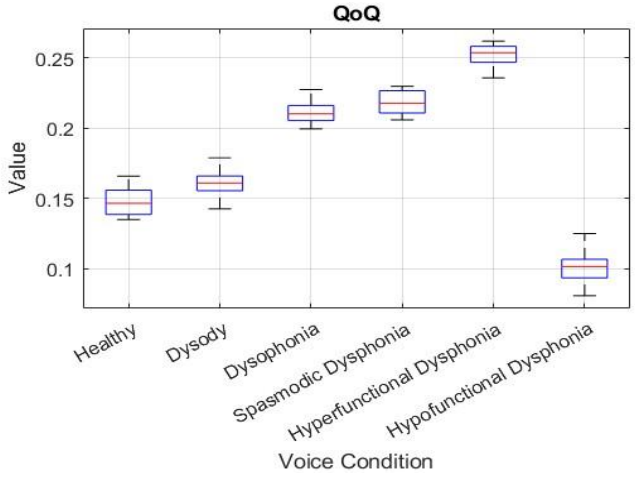


Figure 8 (d): QoQ glottal source parameter distribution derived from QCP model for both healthy and different pathological voices

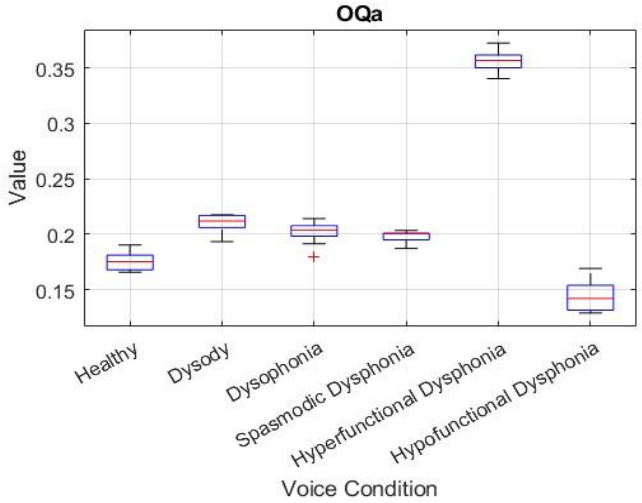


Figure 8 (e): OQa glottal source parameter distribution derived from QCP model for both healthy and different pathological voices

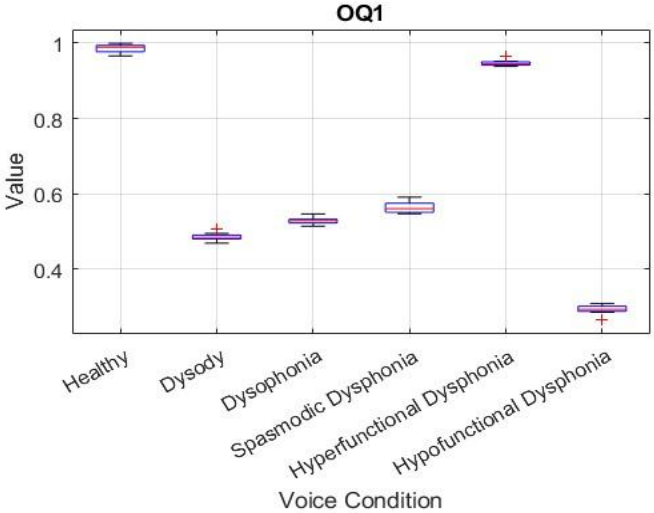


Figure 8 (f): OQ1 glottal source parameter distribution derived from QCP model for both healthy and different pathological voices

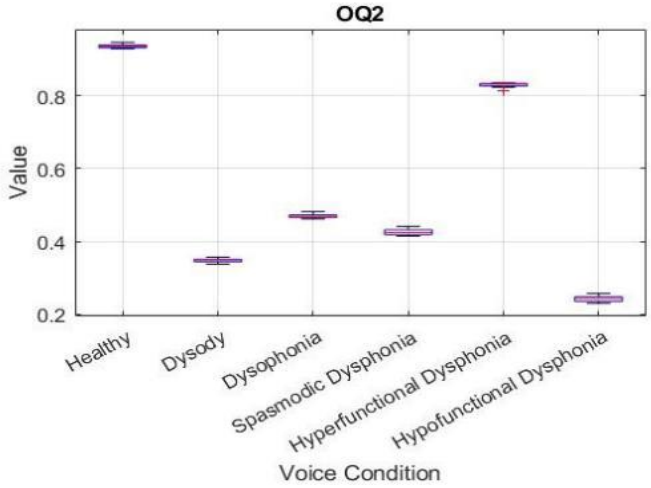


Figure 8 (g): OQ2 glottal source parameter distribution derived from QCP model for both healthy and different pathological voices

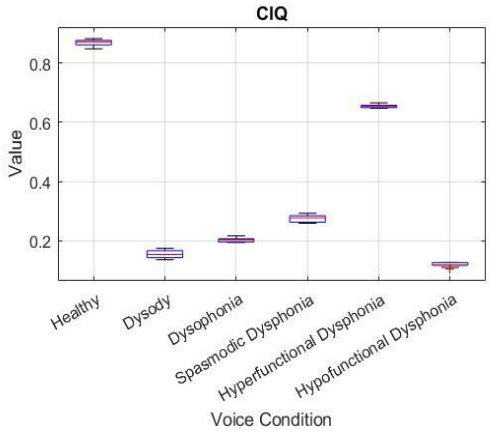


Figure 8 (h): CIQ glottal source parameter distribution derived from QCP model for both healthy and different pathological voices

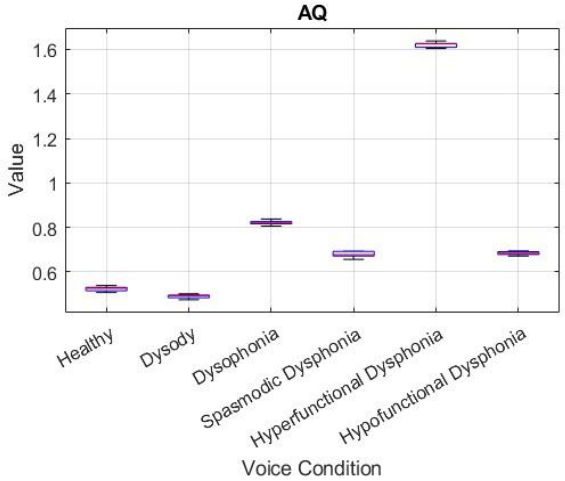


Figure 8 (i): AQ glottal source parameter distribution derived from QCP model for both healthy and different pathological voices

This study proposes normal and abnormal voice identification using Bi-LSTM classifier with different feature sets like, QCP features, IAIF features, QCP-MFCC, IAIF-MFCC, along with these features sets we have combined glottal features sets like QCP+IAIF features and QCP-MFCC+ IAIF-MFCC to evaluate the system. In total, 6 feature sets were utilized to detect the voices of pathological and normal. QCP is the first feature set, IAIF is the second, QCP-MFCC is the third, IAIF-MFCC is the fourth, QCP+IAIF is the fifth, and QCP-MFCC+IAIF-MFCC is the sixth.

Table 2: Results of proposed system and other state-of-the-art methods

Feature-set	Classifier	Accuracy (in %)	Sensitivity	Specificity
IDP [59]	SVM	93.2	0.943	0.923
MFCC [61]	Bi-LSTM	85.2	0.862	0.882
CQCC [61]	Bi-LSTM	88.9	0.872	0.891
APGDF [61]	Bi-LSTM	87.3	0.868	0.884
Proposed methods				
QCP	Bi-LSTM	90.1	0.912	0.898
IAIF	Bi-LSTM	91.2	0.926	0.914
QCP-MFCC	Bi-LSTM	91.6	0.931	0.926
IAIF-MFCC	Bi-LSTM	**93.7**	0.946	0.938
QCP + IAIF	Bi-LSTM	92.8	0.939	0.932
QCP-MFCC + IAIF-MFCC	Bi-LSTM	**94.6**	0.967	0.958

From the table 2, its shows that the QCP, IAIF, QCP-MFCC, IAIF-MFCC, and the combination of QCP + IAIF, and QCP-MFCC + IAIF-MFCC, performs better than other methods. The previous algorithm IDP-SVM [59] also performs better with a 93.2% accuracy. The glottal source waveforms obtained by QCP which contains time-domain an frequency domain features with MFCC derivatives and IAIF features with MFCC derivatives performs better than other methods.

5. CONCLUSION

In this work, glottal source characteristics were analyzed in both diseased and normal voices, and these features were then used to the identification of voice pathologies. The glottal flows computed using the IAIF methodology and the QCP inverse filtering method are the two signals processing techniques that have been employed to identify the glottal source features. The glottal source waveform, which was obtained via the QCP and IAIF approaches, was then used to extract the resulting MFCCs. Glottal source traits are useful in differentiating between normal and diseased voice, according to feature analysis. Further, the combination of QCP-IAIF and QCP-MFCC with IAIF-MFCC provides the 92.8% and 94.6 % of accuracy. These combination of feature set provides the best result compare to other state-of-the-art methods.

AUTHOR CONTRIBUTIONS

Conceptualization, methodology, software - Anilkumar V; validation - R Venkata Siva Reddy ; formal analysis, investigation, Anilkumar V; original draft preparation, Anilkumar V; writing—review and editing, Anilkumar V; visualization, Anilkumar V; supervision, R Venkata Siva Reddy;

REFERENCES

[1] A. Aronson, Clinical Voice Disorders; An Interdisciplinary Approach. Thieme Inc, 1985.

[2] N. R. Williams, "Occupational groups at risk of voice disorders: a review of the literature," Occupational Medicine, vol. 53, no. 7, pp. 456–460, 2003.

[3] J. D. Arias-Londo˜ no, J. I. Godino-Llorente, N. S´ aenz-Lech´on, V. Osma Ruiz, and G. Castellanos-Dom´ ınguez, "Automatic detection of patho logical voices using complexity measures, noise parameters, and mel cepstral coefficients," IEEE Transactions on Biomedical Engineering, vol. 58, no. 2, pp. 370–379, 2011.

[4] J. A. G. Garc´ ıa, L. Moro-Vel´ azquez, and J. I. Godino-Llorente, "On the design of automatic voice condition analysis systems. part I: review of concepts and an insight to the state of the art," Biomedical Signal Processing and Control, vol. 51, pp. 181– 199, 2019.

[5] M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of parkinson's disease," IEEE Transactions on Biomedical Engineering, vol. 56, no. 4, p. 1015, 2009.

[6] D. G. Silva, L. C. Oliveira, and M. Andrea, "Jitter estimation algorithms for detection of pathological voices," EURASIP Journal on Advances in Signal Processing, pp. 9:1–9:9, Jan. 2009.

[7] M. Vasilakis and Y. Stylianou, "Voice pathology detection based on short-term jitter estimations in running speech," Folia Phoniatrica Logopedica, vol. 61, no. 3, pp. 153–170, 2009.

[8] Y. Zhang, J. J. Jiang, L. Biazzo, and M. Jorgensen, "Perturbation and nonlinear dynamic analyses of voices from patients with unilateral laryngeal paralysis," Journal of Voice, vol. 19, no. 4, pp. 519– 528, 2005.

[9] V. Parsa and D. G. Jamieson, "Acoustic discrimination of pathological voice," Journal of Speech, Language, and Hearing Research, vol. 44, no. 2, pp. 327–339, 2001.

[10] J. R. Orozco-Arroyave, E. A. Belalcazar-Bolaos, J. D. Arias-Londoo, J. F. Vargas-Bonilla, S. Skodda, J. Rusz, K. Daqrouq, F. Hnig, and E. Nth, "Characterization methods for the detection of multiple voice disorders: Neurological, functional, and laryngeal diseases," IEEE Journal of Biomedical and Health Informatics, vol. 19, no. 6, pp. 1820 1828, Nov 2015.

[11] J. Mekyska, E. Janousova, P. Gomez-Vilda, Z. Smekal, I. Rektorova, I. Eliasova, M. Kostalova, M. Mrackova, J. B. Alonso-Hernandez, M. Faundez-Zanuy, and K. L. de Ipia, "Robust and complex approach of pathological speech signal analysis," Neurocomputing, vol. 167, pp. 94– 111, 2015.

[12] N. Roy, J. Barkmeier-Kraemer, T. Eadie, M. P. Sivasankar, D. Mehta, D. Paul, and R. Hillman, "Evidence-based clinical voice assessment: A systematic review," American Journal of Speech-Language Pathology, 2013.

[13] J. A. G. Garc´ıa, L. Moro-Vel´azquez, and J. I. Godino-Llorente, "On the design of automatic voice condition analysis systems. part II: review of speaker recognition techniques and study on the effects of different variability factors," Biomedical Signal Processing and Control, vol. 48, pp. 128–143, 2019.

[14] C. Manfredi, M. D'Aniello, P. Bruscaglioni, and A. Ismaelli, "A com parative analysis of fundamental frequency estimation methods with application to pathological voices," Medical Engineering & Physics, vol. 22, no. 2, pp. 135– 147, 2000. 11

[15] C. R. Watts and S. N. Awan, "Use of spectral/cepstral analyses for differentiating normal from hypofunctional voices in sustained vowel and continuous speech contexts," Journal of Speech, Language, and Hearing Research, vol. 54, no. 6, pp. 1525–1537, 2011.

[16] R. R. Patel, S. N. Awan, J. Barkmeier-Kraemer, M. Courey, D. Deliyski, T. Eadie, D. Paul, J. G. ˇ Svec, and R. Hillman, "Recommended proto cols for instrumental assessment of voice: American speech-language hearing association expert panel to develop a protocol for instrumental assessment of vocal function," American Journal of Speech-Language Pathology, vol. 27, no. 3, pp. 887–905, 2018.

[17] H. Kasuya, S. Ogawa, K. Mashima, and S. Ebihara, "Normalized noise energy as an acoustic measure to evaluate pathologic voice," The Journal of the Acoustical Society of America, vol. 80, no. 5, pp. 1329–1334, 1986.

[18] J. I. Godino-Llorente, V. Osma-Ruiz, N. S´aenz-Lech´ on, P. G´omez Vilda, M. Blanco-Velasco, and F. Cruz-Rold´an, "The effectiveness of the glottal to noise excitation ratio for the screening of voice disorders," Journal of Voice, vol. 24, no. 1, pp. 47–56, 2010.

[19] V. Parsa and D. G. Jamieson, "Identification of pathological voices using glottal noise measures," Journal of Speech, Language, and Hearing Research, vol. 43, no. 2, pp. 469–485, 2000.

[20] D. Michaelis, T. Gramss, and H. W. Strube, "Glottal-to-noise excitation ratio a new measure for describing pathological voices," Acta Acustica united with Acustica, vol. 83, no. 4, pp. 700–706, 1997.

[21] Y. Qi and R. E. Hillman, "Temporal and spectral estimations of harmonics-to-noise ratio in human voice signals," The Journal of the Acoustical Society of America, vol. 102, no. 1, pp. 537–543, 1997.

[22] J. Lee, S. Kim, and H. Kang, "Detecting pathological speech using contour modeling of harmonic-to-noise ratio," in ICASSP, May 2014, pp. 5969–5973.

[23] R. Fraile, J. I. Godino-Llorente, N. S´ aenz-Lech´on, J. M. Guti´errez Arriola, and V. Osma-Ruiz, "Spectral analysis of pathological voices: sustained vowels vs running speech," in MAVEBA, 2011, pp. 67–70.

[24] A. Benba, A. Jilbab, and A. Hammouch, "Discriminating between patients with parkinsons and neurological diseases using cepstral analysis," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 24, no. 10, pp. 1100–1108, Oct 2016.

[25] J. I. Godino-Llorente and P. G. Vilda, "Automatic detection of voice impairments by means of short-term cepstral parameters and neural net work based detectors," IEEE Transactions on Biomedical Engineering, vol. 51, no. 2, pp. 380–384, 2004.

[26] J. I. Godino-Llorente, S. Aguilera-Navarro, and P. G. Vilda, "Lpc, LPCC and MFCC parameterisation applied to the detection of voice impairments," in INTERSPEECH, 2000, pp. 965–968.

[27] J. C. Saldanha, T. Ananthakrishna, and R. Pinto, "Vocal fold pathology assessment using mel-frequency cepstral coefficients and linear predic tive cepstral coefficients features," Journal of Medical Imaging and Health Informatics, vol. 4, no. 2, pp. 168–173, 2014.

[28] M. A. Little, D. A. Costello, and M. L. Harries, "Objective dysphonia quantification in vocal fold paralysis: Comparing nonlinear with clas sical measures," Journal of Voice, vol. 25, no. 1, pp. 21– 31, 2011.

[29] H. Hermansky, "Perceptual linear predictive (plp) analysis of speech," The Journal of the Acoustical Society of America, vol. 87, no. 4, pp. 1738–1752, 1990.
[30] R. Fraile and J. I. Godino-Llorente, "Cepstral peak prominence: A comprehensive analysis," Biomedical Signal Processing and Control, vol. 14, pp. 42–54, 2014.
[31] T. Drugman, T. Dubuisson, and T. Dutoit, "On the mutual information between source and filter contributions for voice pathology detection," in INTERSPEECH, 2009.
[32] M. Kaleem, B. Ghoraani, A. Guergachi, and S. Krishnan, "Pathological speech signal analysis and classification using empirical mode decom position," Medical & Biological Engineering & Computing, vol. 51, no. 7, pp. 811–821, Jul 2013.
[33] M. Markaki and Y. Stylianou, "Using modulation spectra for voice pathology detection and classification," in EMBC, Sep. 2009, pp. 2514 2517.
[34] R. Behroozmand and F. Almasganj, "Optimal selection of wavelet packet-based features using genetic algorithm in pathological assess ment of patients speech signal with unilateral vocal fold paralysis," Computers in Biology and Medicine, vol. 37, no. 4, pp. 474– 485, 2007.
[35] K. Umapathy, S. Krishnan, V. Parsa, and D. G. Jamieson, "Discrimi nation of pathological voices using a time-frequency approach," IEEE Transactions on Biomedical Engineering, vol. 52, no. 3, pp. 421–430, March 2005.
[36] B. Ghoraani and S. Krishnan, "A joint time-frequency and matrix decomposition feature extraction methodology for pathological voic classification," EURASIP Journal on Advances in Signal Processing, no. 1, Sep 2009.
[37] M. A. Little, P. E. McSharry, S. J. Roberts, D. A. Costello, and I. M. Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," Biomedical engineering online, vol. 6, no. 1, p. 23, 2007.
[38] C. D. P. Crovato and A. Schuck, "The use of wavelet packet transform and artificial neural networks in analysis and classification of dysphonic voices," IEEE Transactions on Biomedical Engineering, vol. 54, no. 10, pp. 1898–1900, Oct 2007.
[39] E. S. Fonseca, R. C. Guido, P. R. Scalassara, C. D. Maciel, and J. C. Pereira, "Wavelet time-frequency analysis and least squares support vector machines for the identification of voice disorders," Computers in Biology and Medicine, vol. 37, no. 4, pp. 571–578, 2007.
[40] J. D. Arias-Londoo and J. I. Godino-Llorente, "Entropies from markov models as complexity measures of embedded attractors," Entropy, vol. 17, no. 6, pp. 3595–3620, 2015.
[41] J. A. G. Garc´ ıa, "Contributions to the design of automatic voice quality analysis systems using speech technologies," January 2018. [Online]. Available: http://oa.upm.es/49565/
[42] S. Hegde, S. Shetty, S. Rai, and T. Dodderi, "A survey on machine learning approaches for automatic detection of voice disorders," Journal of Voice, 2018. [Online]. Available: http: //www.sciencedirect.com/science/article/pii/S0892199718301437
[43] L. A. Forero, M. Kohler, M. M. Vellasco, and E. Cataldo, "Analysis and classification of voice pathologies using glottal signal parameters," Journal of Voice, vol. 30, no. 5, pp. 549– 556, 2016.
[44] G. Muhammad, M. Alsulaiman, Z. Ali, T. A. Mesallam, M. Farahat, K. H. Malki, A. Al-nasheri, and M. A. Bencherif, "Voice pathology detection using interlaced derivative pattern on glottal source excita tion," Biomedical Signal Processing and Control, vol. 31, pp. 156 164, 2017.
[45] M. Airaksinen, T. Raitio, B. Story, and P. Alku, "Quasi closed phase glottal inverse filtering analysis with weighted linear prediction," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 22, no. 3, pp. 596–607, Mar. 2014.
[46] D. Wong, J. Markel, and A. Gray, "Least squares glottal inverse filtering from the acoustic speech waveform," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 27, pp. 350–355, 1979.
[47] P. Alku, J. Pohjalainen, M. Vainio, A.-M. Laukkanen, and B. H. Story, "Formant frequency estimation of high-pitched vowels using weighted linear prediction," The Journal of the Acoustical Society of America, vol. 134, no. 2, pp. 1295–1313, 2013.
[48] P. Alku, "Glottal wave analysis with pitch synchronous iterative adap tive inverse filtering," Speech Commun., vol. 11, no. 23, pp. 109–118, 1992
[49] M. Airas, "Tkk aparat: An environment for voice inverse filtering and parameterization," Logopedics Phoniatrics Vocology, vol. 33, no. 1, pp. 49–64, 2008.

[50] P. Alku, "Glottal inverse filtering analysis of human voice production a review of estimation and parameterization methods of the glottal excitation and their applications," Sadhana, vol. 36, no. 5, pp. 623 650, 2011.

[51] P. Alku and E. Vilkman, "Amplitude domain quotient for characteri zation of the glottal volume velocity waveform estimated by inverse filtering," Speech Communication, vol. 18, pp. 131–138, 1996.

[52] P. Alku, T. Backstrom, and E. Vilkman, "Normalized amplitude quo tient for parameterization of the glottal flow," The Journal of the Acoustical Society of America, vol. 112, pp. 701–710, 2002.

[53] P. Alku, H. Strik, and E. Vilkman, "Parabolic spectral parameter- A new method for quantification of the glottal flow," Speech Communi cation, vol. 22, no. 1, pp. 67–79, 1997.

[54] J. Hillenbrand and R. A. Houde, "Acoustic correlates of breathy vocal quality: Dysphonic voices and continuous speech," Journal of Speech, Language, and Hearing Research, vol. 39, no. 2, pp. 311–321, 1996.

[55] I. Titze and J. Sundberg, "Vocal intensity in speakers and singers," The Journal of the Acoustical Society of America, vol. 91, pp. 2936–2946, 1992.

[56] S. R. Kadiri and P. Alku, "Mel-Frequency Cepstral Coefficients of Voice Source Waveforms for Classification of Phonation Types in Speech," in Proc. Interspeech 2019, 2019, pp. 2508–2512.

[57] M. P¨ utzer and W. J. Barry, "Instrumental dimensioning of normal and pathological phonation using acoustic measurements," Clinical Linguistics & Phonetics, vol. 22, no. 6, pp. 407–420, 2008.

[58] "Saarbr¨ ucken voice database, institute of phonetics, univ. of saarland," 2010, http://www.stimmdatenbank.coli.uni-saarland.de/ (Last viewed April 20, 2019).

[59] Muhammad, Ghulam, et al. "Voice pathology detection using interlaced derivative pattern on glottal source excitation." Biomedical signal processing and control 31 (2017): 156-164.

[60] Al-Dhief, Fahad Taha, et al. "A survey of voice pathology surveillance systems based on internet of things and machine learning algorithms." IEEE Access 8 (2020): 64514-64533.

[61] AnilKumar, V., and R. Venkata Siva Reddy. "Classification of voice pathology using different features and bi-lstm." 2023 International Conference on Smart Systems for applications in Electrical Sciences (ICSSES). IEEE, 2023.