Received February 18, 2017, accepted March 6, 2017, date of publication March 9, 2017, date of current version April 24, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2680467

An Automatic Health Monitoring System for Patients Suffering From Voice Complications in Smart Cities

ZULFIQAR ALI[1], GHULAM MUHAMMAD[1], AND MOHAMMED F. ALHAMID[2]
[1]Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia
[2]Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia

Corresponding authors: Z. Ali (zuali@ksu.edu.sa), G. Muhammad (ghulam@ksu.edu.sa), and M. Alhamid (mohalhamid@ksu.edu.sa)

This work was supported by the Deanship of Scientific Research, King Saud University, Riyadh, Saudi Arabia, through the Research Group under Project RGP-1436-023.

ABSTRACT Current evolutions in the Internet of Things and cloud computing make it believable to build smart cities and homes. Smart cities provide smart technologies to residents for the improved and healthier life, where smart healthcare systems cannot be ignored due to rapidly growing elderly people around the world. Smart healthcare systems can be cost-effective and helpful in the optimal use of healthcare resources. The voice is a primary source of communication and any complication in the production of voice affects the personal as well as professional life of a person. Early screening of voice through an automatic voice disorder detection system may save life of a person. In this paper, an automatic voice disorder detection system to monitor the resident of all age group and professional backgrounds is implemented. The proposed system detects the voice disorder by determining the source signal from the speech through the linear prediction analysis. The analysis calculates the features from normal and disordered subjects. Based on these features, the spectrum is computed, which provided distribution of energy in normal and voice disordered subjects to differentiate between them. It is found that lower frequencies from 1 to 1562 Hz contributes significantly in the detection of voice disorders. The system is developed by using sustained vowel and running speech so that it can be deployed in a real world. The obtained accuracy for the detection of voice disorder with the sustained vowel is 99.94% $\pm$ 0.1, and that is for running speech is 99.75% $\pm$ 0.8.

INDEX TERMS Smart cities, health monitoring, voice disorders, liner prediction, GMM.

I. INTRODUCTION

The population of elderly people is increasing rapidly worldwide including Japan, China and Europe [1]. According to an estimate, the number of elderly citizens will rise to 10 million in the coming decade. Health issues among the old age citizen are one of the major concerns in developed as well economically growing countries like Brazil and India. Elderly citizens occupy a large portion of health-related facilities due to different health issues. Smart cities and homes can be developed to meet the needs of citizens/senior citizens in an efficient and cost-effective manner.

The recent development in internet of thing (IoT) makes the smart homes and smart cities a reality [2]. IoT sensors the data from surrounding environment and makes optimized decisions. The purpose of the smart cities is to provide the residents a quality life with basic needs of life as well as health monitoring services [1], [3], [4]. Various pillars of smart cities are depicted in Fig. 1. Smart technologies provide real time correct information to the selected users at right time. To monitor the pet animals, a system based on animal biometrics is implemented successfully in [5]. In addition, a smart healthcare is proposed in [6], the system monitors the status of patients by capturing the voice and video through sensors installed inside the smart homes. The speech is processed by using local features [7], whereas, the features from video are extracted by using interlaced derivative pattern [8].

Smart cities and homes have many benefits but security of data is still a challenging task. A smart health system with data security is suggested in [9]. This health monitoring smart system including communication technologies, combination of different apps, things (sensors and devices) and people. The components of smart system work together to monitor, track, and store the information of patients to take care of his\her health. The data is collected through mobile devices which is secured by the watermarking. A large amount of data is generated from IoT in a smart city [10], [11]. The quality

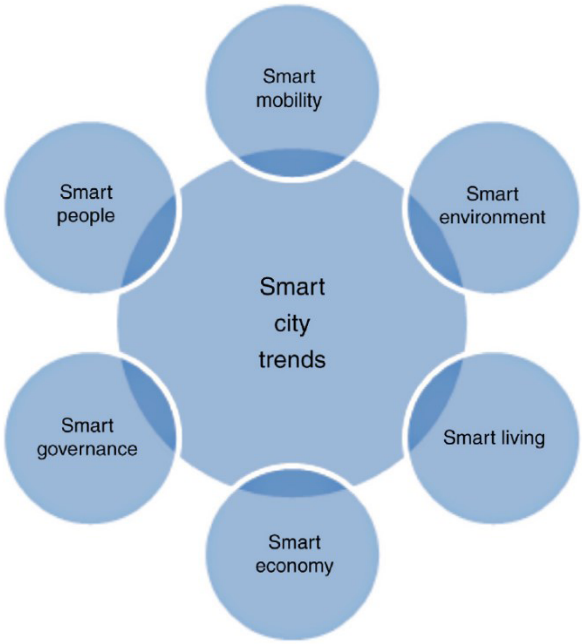

FIGURE 1. Trends of smart cities [1].

and the efficiency of data is also one of the major challenge in smart cities. However, an energy efficient cypher-physical system has been described in [12].

Automatic voice disorder detection systems in smart cities can be deployed to continuously monitor the voice of residents and inform him in case of any voice degradation. According to the medical dictionary [13], dysphonia is a difficulty in speaking, usually evidence by hoarseness. Hoarseness represents any deviation of voice quality as perceived by self or others [14]. Voice samples can be collected at different public places, i.e. school, colleges, universities, courts, and parks. Residents with the profession of high risk of prevalence of voice disorder can be evaluated regularly, and precautionary campaign could be launched to aware them.

People working in voice demanding professions such as teaching, singing, call centers and judiciary have high risk of suffering from voice complications [15], [16]. In addition, people working in back ground noise like stock markets also have high possibility of suffering from voice problems. Only in USA, around 17.5 million people have voice related problems [17]. Voice complications may rise due to malfunctioning of two muscular layers (called as vocal folds) residing on the top of the trachea. While, damaged nerve system (recurrent laryngeal nerve) due to head and neck injuring which controls vocal folds is also one of the reasons of voice problems [18]. Air pressure generated from lungs vibrates the vocal folds to produce voice. The vibration of vocal folds can be affected by growth of abnormal tissues on the surface of vocal folds [19]–[21]. Due to problem on the surface of vocal folds, they exhibit irregular vibrations which produced hoarseness, breathiness and harshness in a voice of a person. Voice disorders can affect people of any age, young children to senior citizens. People delay hospital visits due to busy routine of life and long waiting time. Moreover, dependency on other family members is also one of the reasons to delay in consulting a physician. In case of delay or ignorance, treatment by incision of vocal folds becomes necessary, because, some voice disorders are cancerous and life threatening, such as keratosis [14]. To avoid the risk of life, it is important to diagnose voice disorders at early stages and it is possible by the help of automatic voice disorder detection systems. By using the automatic systems, a person can evaluate his\her voice even without visiting the hospitals. Various automatic voice disorder detection system has been reported in the literature. Ali *et al.* [22] used different frequency bands in an automatic system to determine their contribution in differentiating the normal and disordered subjects. Moreover, Pouchoulin *et al.* [23] and Fraile *et al.* [24] also found significant frequency bands to detect the presence of voice disorders.

It is a comparatively easy task to develop a detection system by using sustained vowel as a speech signal remains stationary during phonation of sustained vowels. Any type of acoustic analyses, short- and long-term, can be used to analyse the sustained vowel. A number of studies has been reported those used the short-term acoustic analysis such as Mel-frequency Cepstral Coefficients (MFCC) [25]–[31] and LP coefficients, and long-term acoustic parameters [28], [32]–[35] such as shimmer, jitter, fundamental frequency, and formants [36]–[38]. However, running speech is more realistic than the artificial phonation of sustained vowel as people use running speech in daily life conversation. Running speech contains various characteristics which are crucial for evaluation of voice quality such as voice breaks, voice onset and offset information and voice termination. These vocal characteristics are not fully existed in sustained vowels [39]. In addition to these characteristics, short pauses and silence makes the analysis of running speech difficult. Long-term analysis required only voiced part of a speech, and therefore, not reliable in case of running speech. An accuracy of 86% is achieved in [40] by using long-term analysis which is not good. Short-term acoustic analysis is a preferable choice to extract the features from running speech because running speech varies quickly over time. Only a few number of studies has developed the running speech based voice disorder detection system [40]–[45]. Most of the work is done for sustained vowel.

In smart cities, the resident will be monitor by using their voices samples which will be a running speech. At the same time, most of the practitioner use sustained vowels during clinical evaluation. Therefore, the proposed system is developed in such a way that it should work for both types of speech signals, sustained vowel and running speech. In this paper, an automatic disorder detection system implemented by using Linear Prediction (LP) analysis based spectrum. The proposed system analyzed the energy variation in the spectrum to differentiate between normal and disordered subjects. LP analysis considers the vocal tract as a liner model, and it divided the vocal tract in number of tubes

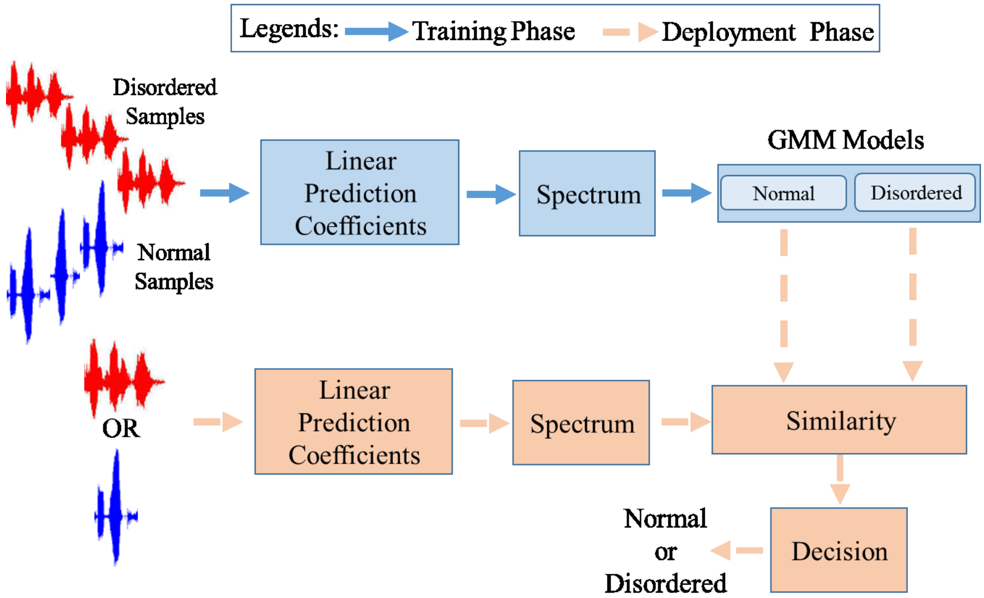

FIGURE 2. The block diagram of the proposed system.

from glottis to lips. LP analysis estimate the source signal by inverse filtering, and then, spectrum is computed by using the estimated source signal to analyse energy distribution in both types of subjects for detection of voice disorder. The proposed system provided good accuracy for detection of voice disorder for sustained vowel as well as running speech.

The rest of the paper is organized as follows: Section 2 describes the proposed automatic voice disorder detection system. The description of the voice disorder database and selection of order for LP analysis is also described in this section. Section 3 presents the experimental setup and disorder detection results by using sustained vowel and running speech. Section 4 provides the discussion, and finally, Section 5 draws some conclusions.

II. MATERIAL AND METHOD

A. THE PROPOSED SYSTEM FOR DISORDER DETECTION

The interaction of the vocal tract with the voice source is a non-linear. In case of a healthy phonation, lower frequencies are strongly depending on source than the higher frequencies. It is due to the low frequency glottal formant. Vocal folds disorder disturbs the vibration of vocal folds which exhibits irregularities in source signal. The glottis produces buzz (a low, vibrating and humming sound) which can be described by its intensity and frequency. The sound generated by voice disorder affected vocal folds contains very low intensity due to abnormal behavior of the vocal folds. To analyze the difference between intensity of normal and disorder subjects, the LP coefficients based spectrum is computed. The proposed system used the computed spectrum to differentiate between normal and disordered subjects. The block diagram of the proposed detection system is shown Fig. 2.

The vocal tract, form glottis to lips, gives rise to the formants. LP analysis [46], [47] determines the formant structure and remove its effect from a speech signal to estimate the source signal. LP analysis considers the vocal tract as a linear tube. The vocal tract can be partitioned into number tubes depending on the order of LP analysis. In case of LP analysis of order R, the number of tubes will be R and the current samples $s'(n)$ will be estimated by R previous samples, as given by Eq. (1)

$$s'(n) = \sum_{r=1}^{R} p_r \times s(n-r) \tag{1}$$

where, p_r refers to LP coefficients. The inverse filtering removes effects of formants from speech signal, the estimated signal after filtering also referred as residue. In order to calculate an accurate LP model, the error between the current and the estimated sample should by minimum. It can be done by making the first order derivative ξ (mean square error) equal to zero in Eq. (2). The obtained set of equations can be solved by applying Levinson-Durbin recursive algorithm [48]. The solution of the equations by the algorithm provided LP coefficients.

$$\xi = \sum\nolimits_{n} e^2(n) \tag{2}$$

where, e is error between the current sample and the estimated sample and can be defined by Eq. (3) as

$$e = s'(n) - s(n) \tag{3}$$

To compute the spectrum based on the obtained LP coefficients, the frequency response of the transfer function, given

TABLE 1. Distribution of normal and pathological samples [49].

Subjects	Gender	Number of Samples	Mean Age (Years)	Age Range (Years)	Standard Deviation (Years)
Pathological	Male	70	41.7	26-58	9.4
	Female	103	37.6	21-51	8.2
Normal	Male	21	38.8	26-59	8.5
	Female	32	34.2	22-52	7.9

in Eq. (4), is determined.

$$H\left(e^{jw}\right) = \frac{B\left(e^{jw}\right)}{A\left(e^{jw}\right)} = \frac{\sqrt{\sigma^2}}{p_r(1)+p_r(2)e^{-jw}+p_r(3)e^{-j2w}+\ldots+p_r(R)e^{-j(R-1)w}} \tag{4}$$

where σ^2 denotes the gain during LP analysis, and p_r's are the LP coefficients. The coefficient $p_r\,(1)$ is near to the glottis and $p_r(R)$ is closer to the lips.

To develop the proposed system, two phases are implemented for automatic classification of normal and disordered subjects. The first phase is used to train the system, while, the second phase is used to test the system. In the training phase, the proposed system takes the calculated spectrum as an input and generate acoustic models for normal and disordered subjects. The Gaussian Mixture Models (GMM) are generated by using different number of mixtures. K-means algorithm is used to initial the parameter in GMM models $\Theta = (\mu_i, \Sigma_i)$, where, μ_i and Σ_i are the mean vector and covariance matrix of the i^{th} Gaussian component, respectively. Moreover, the parameters are tuned by using Expectation-Maximization (EM) algorithm to converge to a model giving a maximum log-likelihood value.

The testing phase is used to evaluate the performance of the system. In this phase, the computed spectrum of an unknown speech samples is compared with GMM models of normal and disordered subjects. If the unknown sample has more similarity with disordered subject than normal subject, then, then unknown sample belongs to a disordered patient. Otherwise, the unknown sample belongs to a normal person. The database used to perform the experiments in this study is described in the following subsection.

B. MEEI DATABASE AND PARAMETER SELECTION

The proposed system is evaluated by using sustained vowel /a/ as well as running speech. The normal and disordered samples of sustained vowels and running speech are taken from Massachusetts Eye & Ear Infirmary (MEEI) database [50]. To compare the results with existing studies, the same subset of the samples is used which has been used in [31], [49], and [51]–[54]. The subset contains 173 disordered subjects and 53 normal subjects, and distribution of these samples is given in Table 1.

In disordered subjects, number of male and female speakers are 70 and 103, respectively. Moreover, in normal subjects, number of males and females are 21 and 32. The age range of both genders for normal and disordered subjects is almost similar. The subset is selected in a way that age and gender of both types of subjects are evenly distributed, and various disordered are considered in it [49]. In MEEI database, samples are recorded at two different sampling frequencies, i.e. 25 kHz or 50 kHz. To have a unique sampling frequency, f_s, all samples of the subset are down-sampled at 16 kHz.

One of main parameter in the LP analysis is order of the filter (number of LP coefficients) and it can affect the estimation of source signal. In fact, LP analysis perform inverse filtering to get the voice source from a speech signal. The order of the filter need to be adjusted in a way that LP analysis should be able to determine the formant peaks. Because, removing the effect of the formants from the speech signal will provide accurate estimation of the source signal.

In case of sampling frequency of 10 kHz, the analysis bandwidth is 0 to 5 kHz. If the number of coefficients are adjusted to 10 then it means that there are two coefficients for each 1 kHz of the computed frequency domain spectrum. Markel and Gray [55] recommended, an extra pair of coefficients to model the spectral slope and they do not affect the analysis in negative way. The spectral slope is result of both the shape of a single glottal pulse and the effects of lip radiation. Markel and Gray suggested that couple of extra coefficients are also helpful in analyzing the closely space formants and they model the speaker variations in voice quality.

For a vowel recorded at 10 kHz with comfortable pitch, at least 12 coefficients are required. However, these number of coefficients are not sufficient for the higher sampling rates. At least one coefficient is required for each 1 kHz for the higher sampling rates to avoid missing peaks (formants). In this study, all samples of both types of subjects are down-samples to 16 kHz and the number of LP coefficients (order of filter) are determined by using the relation given by Eq. (5)

$$R = \frac{f_s}{1000} + 2 \tag{5}$$

III. EXPERIMENTAL SETUP AND DETECTION RESULTS

All samples of sustained vowel and running speech in MEEI are down-sampled to 16 kHz. According to Eq. (5), the number of calculated LP coefficient becomes 18. The LP coefficients are computed from each frame of each normal and disordered subjects. By using these LP coefficients, the spectrum is calculated at different frequency resolution to observe the energy distribution.

Various experiments are performed to evaluate the proposed system for detection of voice disorders. Both sustained vowel and running speech are used to conduct the experiments and the results of the experiments are presents by using different performance measures: sensitivity, specificity, and accuracy. Sensitivity (SEN) is a ratio between truly detected disordered samples and total number of disordered samples. Specificity (SPE) is a ratio between truly classified normal samples and total number of normal samples, and accuracy (ACC) is a ratio between all truly identified samples and total number of samples. The measures are calculated by using following relations.

$$SEN = \frac{true\ Abnorm}{true\ Abnorm + false\ Health} \times 100 \tag{6}$$

$$SPE = \frac{true\ Health}{true\ Health + false\ Abnorm} \times 100 \tag{7}$$

$$ACC = \frac{trure\ Abnorm + true\ Health}{total\ Abnorm + total\ Health} \times 100 \tag{8}$$

where *true Abnorm* means when a disordered sample is detected as a disordered sample by the system, *false Abnorm* means that when a normal sample is detected as a disordered sample, *true Health* means a normal sample is detected as a normal sample by the system, *false Health* means a disordered samples is detected as a normal sample, *total Health* represents the total number of normal samples, and *total Abnorm* stands for the total number of disordered samples.

To avoid the biasness of the training and testing samples during the evaluation of the proposed system, 5-folds cross validation approach is implemented for disorder detection with sustained vowel and running speech. The samples are divided into five disjoint testing subset. Each time one of the subsets is used to train the system, and the reaming four subsets are used to test the system. The results obtained in different scenarios are provided in the following subsections.

A. DISORDER DETECTION WITH SUSTAINED VOWEL AT DIFFERENT RESOLUTIONS

An inverse relation exists between the frequency and time resolution. To increase the frequency resolution, the time resolution should be decreased. In other words, the length of analysis window will be increased. The relation for frequency resolution is given by Eq. (9)

$$\Delta f = \frac{1}{T} = \frac{f_s}{N} \tag{9}$$

According to Eq. (9), the frequency resolution will be increase by decreasing Δf. It can be achieved either by decreasing sampling frequency f_s or increasing the number of samples N in the analysis window. In the experiments, we gradually increase the number of samples to increase the frequency resolution. The experiments are performed by using N equal to 128, 256 and 512, one by one. The results for

TABLE 2. Disorder detection results with N = 128 by using sustained vowel /a/.

Number of Gaussians	%SPE±STD	%SEN±STD	%ACC±STD
4	95 ± 6.1	85 ± 3.2	90 ± 4.5
8	95 ± 5.6	85 ± 2.8	90 ± 2.8
16	95 ± 4.8	85 ± 1.9	90 ± 2.2
32	85 ± 3.5	90 ± 0.9	87.50 ± 2.0
50	90 ± 2.1	95 ± 0.7	92.50 ± 1.1

TABLE 3. Disorder detection results with N = 512 by using sustained vowel /a/.

Number of Gaussians	%SPE±STD	%SEN±STD	%ACC±STD
4	95 ± 1.8	99.9± 2.5	97.50 ± 1.3
8	95 ± 0.9	95 ± 2.9	95 ± 2.1
16	95 ± 0.6	99.9 ± 1.8	97.50 ± 1.2
32	95 ± 0.6	99.9 ± 0.8	97.50 ± 0.7
50	99.9 ± 0.4	100 ± 0	99.94 ± 0.1

N = 128 and N = 512 are provided in Table 2 and Table 3, respectively. The results with N = 256 are almost same as for N = 128, therefore, they are not listed. All results are obtained by using sustained vowel.

The averaged results of 5-folds are presented in the Table 2 and 3 with standard devastation (STD) among the results of 5-folds. The maximum detection accuracy obtained with N = 128 to classify the normal and disordered subjects is 92.50% ± 1.1. The results in Table 3 suggest that by increasing the frequency resolution the detection accuracy is increased to 99.94% ± 0.1. The accuracy of 99.94% ± 0.1 is achieved with N = 512.

By increasing the frequency resolution, the computed LP based spectrum differentiates between the normal and disordered subjects significantly. In addition, a comparison between the accuracies for N = 128 and N = 512 is depicted in Fig. 3. It can be seen in Fig. 3 that accuracy is increased by increasing the number of Gaussian mixtures. It means that high number of Gaussian generates perfect acoustic model for normal and disordered subjects. Moreover, it can be observed from Table 2 that an accuracy of 90% is obtained with 4 mixtures. There is clear difference between SPE (95%) and SEN (85%). By increasing the number of mixtures, not only the accuracy is improved by 2% but also the difference between SPE (90%) and SEN (95%) is reduced.

B. DISORDER DETECTION WITH RUNNING SPEECH

A system based on running speech is more realistic than the artificial utterance of sustained vowel. Hence, the proposed system is also evaluated by using the running speech. As for sustained vowel, the best results are obtained by using high-frequency resolution, i.e. N = 512. Therefore, the experiments for running speech are performed only by using N = 512, and the results are presented in Table 4. The best obtained accuracy by using running speech is 99.75% ± 0.8 with 50 Gaussian mixtures. The same kind of

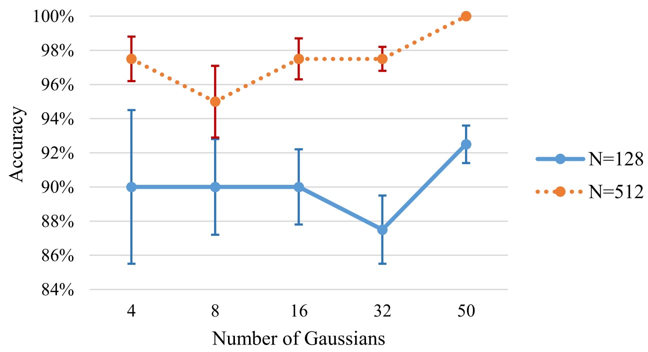

FIGURE 3. A comparison of accuracy between N=128 and N=512.

TABLE 4. Disorder detection results with $N = 512$ by using running speech.

Number of Gaussians	%SPE±STD	%SEN±STD	%ACC±STD
4	95.5 ± 2.9	97.5 ± 2.5	96.50 ± 2.2
8	96 ± 1.5	98 ± 1.9	97 ±1.9
16	97 ± 1.1	99.5 ± 0.9	98.25 ± 1.5
32	99 ± 0.9	99.5 ± 0.8	99.25 ± 1.2
50	99.25 ± 0.7	99.9 ± 0.2	99.75 ± 0.8

trend is noticed in case of running speech that accuracy of the system increased by increasing the number of mixtures. In addition, the standard deviation among the results of 5-folds decreased by increasing the mixtures, which makes the results more reliable.

IV. DISCUSSION

A health monitoring system for smart cities is proposed in this study. The proposed system computed spectrum by using LP coefficients to differentiate between normal and disordered patients. It can be deployed in homes to monitor the elderly citizen, and at different public places to observe the residents of various professional background. Therefore, the system is evaluated by using the running speech so that it could process the daily conversion of people to determine the presence of voice disorders. However, the system is also tested by using the sustained vowels. In this way, it can be implemented in clinics to provide the complementary information to the medical practitioners about the existence of voice disorders.

The irregular vibrations disturb the vocal folds which effect the source signal and causes to produce harsh, strained, weaker and breathier sound. LP analysis apply inverse filtering to determine the source signal. Then, the spectrum is calculated to observe the strength of source signal, and it founds the presence of voice disorder. Due to weaker and breathiness, the voice affected by disorders contains low energy than normal people. It can be observed from Fig. 4 that spectrum of disordered subject does not contain high energies and exhibits irregular pattern. While, Fig. 5 shows that high energies are presented in a regular pattern for a normal sample. In both figures, the spectrum of a sustained vowel of equal length is depicted, and it can be noticed that frequencies from 1 Hz to 1562 Hz are more significant in detection of voice disorders. That range of frequency has different energy pattern for normal and disordered subjects.

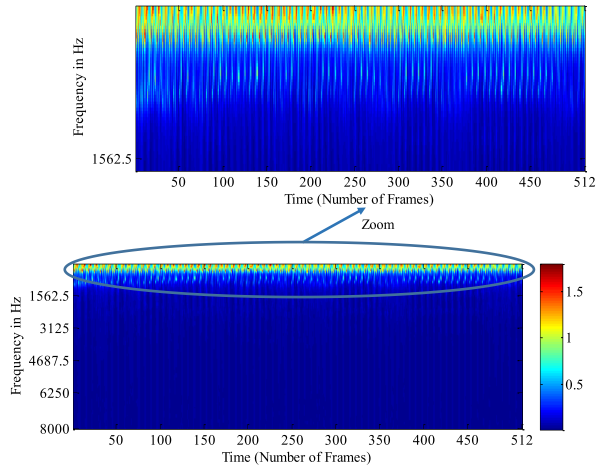

FIGURE 4. LP analysis based spectrum of a voice disordered subject.

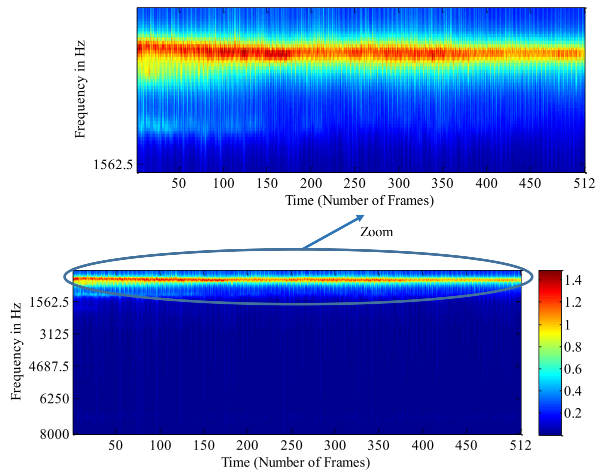

FIGURE 5. LP analysis based spectrum of a normal subjects.

The fact is also supported by the study [22]. In the study, normal and disordered subjects are decomposed into different frequency bands to determine the contribution of each frequency band in differentiation of normal and disordered subjects. The study concluded that the lower frequency band of 1 Hz to 1562 Hz contains lower energy for disorder subjects as compared to normal subjects. For this band, there is significant difference in the energy pattern of normal and disordered subjects. The obtained detection accuracy of the band, 1 Hz to 1562 Hz, was 91.28% ± 0.4.

In our proposed automatic system, we did not implement voice activity detection module. Activity detection is itself a difficult task [44]. Some existing system based on running speech implemented the voice activity detection module [41]–[43]. Such systems are not using running speech truly, because they are extracting only voiced part of running speech to develop the systems. Running speech contains silence, voiced and unvoiced parts; therefore, it is challenging task to develop the system with running speech.

Few disorder detection systems [40], [44], [45] has been implemented without voice activity detection module; however, their accuracy is not good. A comparison of the

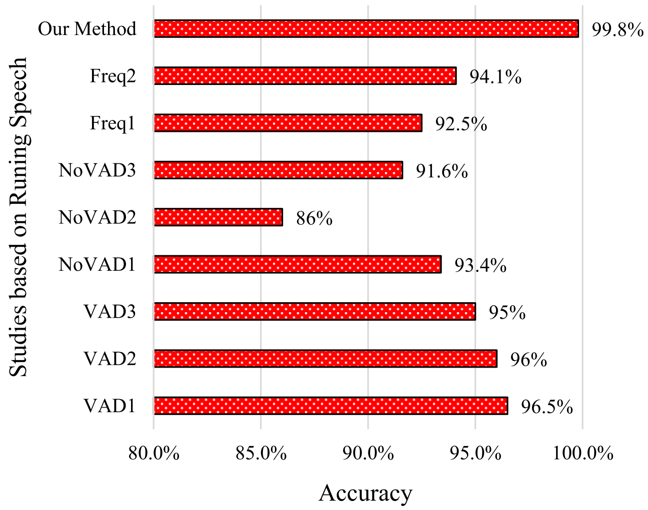

FIGURE 6. Comparison for accuracy of our proposed system with existing system.

proposed system with the existing running speech based disorder detection systems is provided in Fig. 6. The comparison of accuracy is done with three types of studies. The studies used voice activity detection are represented by "VAD", the studies which did not implement voice activity detection module are denoted by "NoVAD", and those studies which have used the frequency bands for detection of voice disordered are mentioned as "Freq" in Fig. 6. Our proposed method in this study has outperformed existing systems based on running speech, and an accuracy of 99.8% is achieved. Moreover, our system is also evaluated with the sustained vowel so that medical practitioners can use it in the clinics.

V. CONCLUSIONS

A number of residents in smart cities may belong to the professions which have high risk of prevalence of voice disorders. Voice complications have the negative impacts over the life of individuals. They disturb not only the daily routine of a personal but also affect the professional life. In addition, an increase in the population of senior citizens in developing as well as economically growing countries is alarming. A smart health care system can assist the residents of different professions and of different age groups. Moreover, the early diagnosis of voice disorder is very crucial. Often, people ignore it due to busy and hectic routine of life. Consequently, they risk their lives and undergo the incision process to cure the voice disorders. It is not financially viable, and also wastage of resources at health centers. To avoid these circumstances, a smart healthcare system is proposed and developed. The system monitor the voices of specific residents such as lawyers, teachers, singers, and who are working in background noise like stock markets. Furthermore, the proposed system can monitor the health of all residents specially who have strong family history of voice complications. The proposed system investigates the energy distribution across the spectrum computed by using LP coefficients to differentiate between normal and voice disordered subjects. The subjects suffering from voice disorders exhibit irregular vocal folds vibration which makes the voice weaker, whisper and breathier. Therefore, the voice of a disorder subject contains lower energy as compared to a normal subject. The proposed system is also implemented by using running speech which is a challenging task due to quick variations in it over time. Due to the running speech, the proposed system can be used in the real world as people used running speech in daily life conversion. The performance of the proposed system is promising as the obtained disorder detection accuracy for running speech is 99.75 ± 0.8. The accuracy of the proposed system with the sustained vowel is also good, i.e. 99.94 ± 0.1.

REFERENCES

[1] T. Guelzim, M. S. Obaidat, and B. Sadoun, "Introduction and overview of key enabling technologies for smart cities and homes," in *Smart Cities Homes*. Cambridge, MA, USA: Morgan Kaufmann, 2016, pp. 1–16, ch. 1.

[2] T. K. L. Hui, R. S. Sherratt, and D. D. Sánchez, "Major requirements for building smart homes in smart cities based on Internet of Things technologies," *Future Generat. Comput. Syst.*, to be published, doi: http://dx.doi.org/10.1016/j.future.2016.10.026.

[3] M. S. Hossain, "Patient state recognition system for healthcare using speech and facial expressions," *J. Med. Syst.*, vol. 40, p. 272, Dec. 2016.

[4] M. S. Hossain and A. El Saddik, "A biologically inspired multimedia content repurposing system in heterogeneous environments," *Multimedia Syst.*, vol. 14, no. 3, pp. 135–143, Aug. 2008.

[5] S. Kumar and S. K. Singh, "Monitoring of pet animal in smart cities using animal biometrics," *Future Generat. Comput. Syst.*, to be published. [Online]. Available: http://dx.doi.org/10.1016/j.future.2016.12.006.

[6] M. S. Hossain, "Patient status monitoring for smart home healthcare," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Jul. 2016, pp. 1–6.

[7] A. Mahmood, M. Alsulaiman, and G. Muhammad, "Automatic speaker recognition using multi directional local features (MDLF)," *Arabian J. Sci. Eng.*, vol. 39, no. 5, pp. 3799–3811, 2014.

[8] G. Muhammad, "Automatic speech recognition using interlaced derivative pattern for cloud based healthcare system," *Cluster Comput.*, vol. 18, no. 2, pp. 795–802, Jun. 2015.

[9] M. S. Hossain and G. Muhammad, "Cloud-assisted industrial Internet of Things (IIoT)—Enabled framework for health monitoring," *Comput. Netw.*, vol. 101, pp. 192–202, Jun. 2016.

[10] H. B. Sta, "Quality and the efficiency of data in 'smart-cities,'" *Future Generat. Comput. Syst.*, to be published. [Online]. Available: http://dx.doi.org/10.1016/j.future.2016.12.021.

[11] U. Aguilera, O. Peña, O. Belmonte, and D. López-de-Ipiña, "Citizen-centric data services for smarter cities," *Future Generat. Comput. Syst.*, to be published.

[12] M. S. Hossain, M. A. Rahman, and G. Muhammad, "Cyber-physical cloud-oriented multi-sensory smart home framework for elderly people: An energy efficiency perspective," *J. Parallel Distrib. Comput.*, vol. 103, pp. 11–21, May 2017.

[13] *The American Heritage Stedman's Medical Dictionary*, aceessed on Jan. 24, 2016. [Online]. Available: http://dictionary.reference.com/browse/dysphonia

[14] T. Mau, "Diagnostic evaluation and management of hoarseness," *Med. Clin. North Amer.*, vol. 94, no. 5, pp. 945–960, Sep. 2010.

[15] N. Roy, R. M. Merrill, S. D. Gray, and E. M. Smith, "Voice disorders in the general population: Prevalence, risk factors, and occupational impact," *Laryngoscope*, vol. 115, no. 11, pp. 1988–1995, Nov. 2005.

[16] N. Roy, R. M. Merrill, S. Thibeault, R. A. Parsa, S. D. Gray, and E. M. Smith, "Prevalence of voice disorders in teachers and the general population," *J. Speech Lang. Hear Res.*, vol. 47, pp. 281–293, Apr. 2004.

[17] N. Bhattacharyya, "The prevalence of voice problems among adults in the United States," *Laryngoscope*, vol. 124, no. 10, pp. 2359–2362, Oct. 2014.

[18] L. H. S. Rosenthal, M. S. Benninger, and R. H. Deeb, "Vocal fold immobility: A longitudinal analysis of etiology over 20 years," *Laryngoscope*, vol. 117, no. 10, pp. 1864–1870, Oct. 2007.
[19] A. I. R. Fontes, P. T. V. Souza, A. D. D. Neto, A. de M. Martins, and L. F. Q. Silveira, "Classification system of pathological voices using correntropy," *Math. Problems Eng.*, vol. 2014, 2014, Art. no. 924786.
[20] S. L. Thibeault, R. M. Merrill, N. Roy, S. D. Gray, and E. M. Smith, "Occupational risk factors associated with voice disorders among teachers," *Ann. Epidemiol.*, vol. 14, no. 10, pp. 786–792, Nov. 2004.
[21] J. Bohlender, "Diagnostic and therapeutic pitfalls in benign vocal fold diseases," *GMS Current Topics Otorhinolaryngol. Head Neck Surgery*, vol. 12, pp. 1–19, Dec. 2013.
[22] Z. Ali, I. Elamvazuthi, M. Alsulaiman, and G. Muhammad, "Detection of voice pathology using fractal dimension in a multiresolution analysis of normal and disordered speech signals," *J. Med. Syst.*, vol. 40, no. 1, pp. 1–20, Jan. 2015.
[23] G. Pouchoulin, C. Fredouille, J.-F. Bonastre, A. Ghio, and J. Révis, "Characterization of the Pathological Voices (Dysphonia) in the frequency space," in *Proc. Int. Congr. Phonet. Sci. (ICPhS)*, 2007, pp. 1993–1996.
[24] R. Fraile, J. I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruiz, and J. M. Gutiérrez-Arriola, "Characterization of dysphonic voices by means of a filterbank-based spectral analysis: Sustained vowels and running speech," *J. Voice*, vol. 27, no. 1, pp. 11–23, Jan. 2013.
[25] S. C. Costa, B. G. A. Neto, and J. M. Fechine, "Pathological voice discrimination using cepstral analysis, vector quantization and hidden Markov models," in *Proc. 8th IEEE Int. Conf. BioInformat. Bioeng. (BIBE)*, Oct. 2008, pp. 1–5.
[26] J. I. Godino-Llorente and P. Gomez-Vilda, "Automatic detection of voice impairments by means of short-term cepstral parameters and neural network based detectors," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 2, pp. 380–384, Feb. 2004.
[27] W. Jianglin and J. Cheolwoo, "Vocal folds disorder detection using pattern recognition methods," in *Proc. 29th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBS)*, Aug. 2007, pp. 3253–3256.
[28] A. Gelzinis, A. Verikas, and M. Bacauskiene, "Automated speech analysis applied to laryngeal disease categorization," *Comput. Methods Programs Biomed.*, vol. 91, no. 1, pp. 36–47, Jul. 2008.
[29] A. A. Dibazar, S. Narayanan, and T. W. Berger, "Feature analysis for automatic detection of pathological speech," in *Proc. 2nd Conf. Joint Eng. Med. Biol., 24th Annu. Conf. Annu. Fall Meeting Biomed. Eng. Soc. (EMBS/BMES)*, vol. 1, Oct. 2002, pp. 182–183.
[30] A. A. Dibazar, T. W. Berger, and S. S. Narayanan, "Pathological Voice Assessment," in *Proc. 28th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBS)*, Aug. 2006, pp. 1669–1673.
[31] J. D. Arias-Londoño, J. I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruiz, and G. Castellanos-Domínguez, "An improved method for voice pathology detection by means of a HMM-based feature space transformation," *Pattern Recognit.*, vol. 43, no. 9, pp. 3100–3112, Sep. 2010.
[32] B. G. A. Neto, J. M. Fechine, S. C. Costa, and M. Muppa, "Feature estimation for vocal fold edema detection using short-term cepstral analysis," in *Proc. 7th IEEE Int. Conf. Bioinformat. Bioeng. (BIBE)*, Oct. 2007, pp. 1158–1162.
[33] D. G. Childers and K. SungBae, "Detection of laryngeal function using speech and electroglottographic data," *IEEE Trans. Biomed. Eng.*, vol. 39, no. 1, pp. 19–25, Jan. 1992.
[34] M. Marinaki, C. Kotropoulos, I. Pitas, and N. Maglaveras, "Automatic detection of vocal fold paralysis and edema," in *Proc. 8th Int. Conf. Spoken Lang. Process.*, 2004, pp. 1–4.
[35] G. Muhammad, T. A. Mesallam, K. H. Malki, M. Farahat, M. Alsulaiman, and M. Bukhari, "Formant analysis in dysphonic patients and automatic Arabic digit speech recognition," *Biomed. Eng. Online*, vol. 10, no. 1, p. 41, 2011.
[36] A. Al-Nasheri *et al.*, "Voice pathology detection with MDVP parameters using Arabic voice pathology database," in *Proc. 5th Nat. Symp. Inf. Technol., Towards New Smart World (NSITNSW)*, Feb. 2015, pp. 1–5.
[37] A. Al-Nasheri *et al.*, "An investigation of multidimensional voice program parameters in three different databases for voice pathology detection and classification," *J. Voice*, vol. 31, no. 1, pp. 113.e9–113.e18, Jan. 2016. [Online]. Available: http://dx.doi.org/10.1016/j.jvoice.2016.03.019.
[38] M. K. Arjmandi, M. Pooyan, M. Mikaili, M. Vali, and A. Moqarehzadeh, "Identification of voice disorders using long-time features and support vector machine with different feature reduction methods," *J. Voice*, vol. 25, no. 6, pp. e275–e289, Nov. 2011.
[39] B. Hammarberg, B. Fritzell, J. Gaufin, J. Sundberg, and L. Wedin, "Perceptual and acoustic correlates of abnormal voice qualities," *Acta Otolaryngol.*, vol. 90, nos. 1–6, pp. 441–451, Nov./Dec. 1980.
[40] V. Klára, I. Viktor, and M. Krisztina, "Voice Disorder Detection on the Basis of Continuous Speech," in *Proc. 5th Eur. Conf. Int. Federation Med. Biol. Eng.*, 2012, pp. 86–89.
[41] V. Parsa and D. G. Jamieson, "Acoustic discrimination of pathological voice: Sustained vowels versus continuous speech," *J. Speech Lang. Hear Res.*, vol. 44, pp. 327–339, Apr. 2001.
[42] J. I. Godino-Llorente, R. Fraile, N. Sáenz-Lechón, V. Osma-Ruiz, and P. Gómez-Vilda, "Automatic detection of voice impairments from text-dependent running speech," *Biomed. Signal Process. Control*, vol. 4, no. 3, pp. 176–182, 2009.
[43] J. R. Orozco, J. F. Vargas, J. B. Alonso, M. A. Ferrer, C. M. Travieso, and P. Henriquez, "Voice pathology detection in continuous speech using nonlinear dynamics," in *Proc. 11th Int. Conf. Inf. Sci., Signal Process. Appl. (ISSPA)*, Jul. 2012, pp. 1030–1033.
[44] K. Umapathy, S. Krishnan, V. Parsa, and D. G. Jamieson, "Discrimination of pathological voices using a time-frequency approach," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 3, pp. 421–430, Mar. 2005.
[45] Z. Ali, M. Alsulaiman, G. Muhammad, I. Elamvazuthi, and T. A. Mesallam, "Vocal fold disorder detection based on continuous speech by using MFCC and GMM," in *Proc. 7th IEEE GCC Conf. Exhibit. (GCC)*, Nov. 2013, pp. 292–297.
[46] X. Lu and J. Dang, "An investigation of dependencies between frequency components and speaker characteristics for text-independent speaker identification," *Speech Commun.*, vol. 50, no. 4, pp. 312–322, Apr. 2008.
[47] B. S. Atal, "Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification," *J. Acoust. Soc. Amer.*, vol. 55, no. 6, pp. 1304–1312, Jun. 1974.
[48] J. Benesty and T. Gansler, "Computation of the condition number of a non-singular symmetric toeplitz matrix with the Levinson-durbin algorithm," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2362–2364, Jun. 2006.
[49] V. Parsa and D. G. Jamieson, "Identification of pathological voices using glottal noise measures," *J. Speech, Language, Hearing Res.*, vol. 43, pp. 469–485, 2000.
[50] Massachusette Eye & Ear Infirmry Voice & Speech LAB, "Disordered voice database model 4337 (ver. 1.03)," Kay Elemetrics Corp., Boston, MA, USA, Tech. Rep., 1994.
[51] T. Villa-Cañas *et al.*, "Automatic detection of laryngeal pathologies using cepstral analysis in Mel and Bark scales," in *Proc. Symp. Image, Signal Process., Artif. Vis. (STSIVA)*, Sep. 2012, pp. 116–121.
[52] G. Muhammad and M. Melhem, "Pathological voice detection and binary classification using MPEG-7 audio features," *Biomed. Signal Process. Control*, vol. 11, pp. 1–9, May 2014.
[53] M. Markaki and Y. Stylianou, "Voice pathology detection and discrimination based on modulation spectral features," *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, no. 7, pp. 1938–1948, Sep. 2011.
[54] J. I. Godino-Llorente, P. Gomez-Vilda, and M. Blanco-Velasco, "Dimensionality reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 10, pp. 1943–1953, Oct. 2006.
[55] J. D. Markel and A. H. J. Gray, *Linear Prediction of Speech*. Berlin, Germany: Springer-Verlag, 1976.

ZULFIQAR ALI received the master's degrees in computational mathematics from the University of the Punjab, Lahore, and in computer science from the University of Engineering and Technology, Lahore, with the specialization in system engineering. Since 2010, he has been a full-time Researcher with the Digital Speech Processing Group, Department of Computer Engineering, King Saud University, Saudi Arabia. He is also a member of the Center for Intelligent Signal and Imaging Research, Universiti Teknologi PETRONAS, Malaysia. His research interests include speech and language processing, medical signal processing, privacy and security in healthcare, multimedia forensics, and computer-aided pronunciation training systems.

GHULAM MUHAMMAD received the B.S. degree in computer science and engineering from the Bangladesh University of Engineering and Technology in 1997 and the M.S. and Ph.D. degrees in electrical and computer engineering from Toyohashi University and Technology, Japan, in 2003 and 2006, respectively. He is currently a Full Professor with the Department of Computer Engineering, CCIS, King Saud University, Riyadh, Saudi Arabia. He supervised over ten Ph.D. and master's theses. He is involved in many research projects as a Principal Investigator and a Co-Principal Investigator. He has authored or co-authored over 120 publications, including IEEE/ACM/Springer/Elsevier journals, and flagship conference papers. He has a U.S. patent on audio processing. His research interests include image and speech processing, cloud and multimedia for healthcare, serious games, resource provisioning for big data processing on media clouds, and biologically inspired approach for multimedia and software system. He was a recipient of the Japan Society for Promotion and Science Fellowship from the Ministry of Education, Culture, Sports, Science and Technology, Japan. He received the best faculty award of Computer Engineering Department at KSU during 2014 and 2015.

MOHAMMED F. ALHAMID received the Ph.D. degree in computer science from the University of Ottawa, Canada. He is currently an Assistant Professor with the Software Engineering Department, King Saud University, Riyadh, Saudi Arabia, where he is also serving as the Chairman of the department. His research interests include recommender systems, social media mining, big data, and ambient intelligent environment.