Voice Pathology Detection based on the Modified Voice Contour and SVM

[1,2]Zulfiqar Ali, [1]Mansour Alsulaiman, [2]Irraivan Elamvazuthi, [1]Ghulam Muhammad, [3]Tamer A. Mesallam, [3]Mohamed Farahat, and [3]Khalid H. Malki

[1]Digital Speech Processing Group, College of Computer and Information Sciences, King Saud University, Saudi Arabia.
{zuali, msuliman, Ghulam}@ksu.edu.sa

[2]Centre for Intelligent Signal and Imaging Research (CISIR), Department of Electrical and Electronic Engineering, Universiti Tekhnologi PETRONAS, Perak, Malaysia
zulfiqar_g02579@utp.edu.my, irraivan_elamvazuthi@petronas.com.my

[3]ENT Department, College of Medicine, King Saud University, Riyadh, Saudi Arabia.
{tmesallam, mfarahat, kalmalki}@ksu.edu.sa

Abstract
In this study, a novel method based on the voice intensity of a speech signal is used for automatic pathology detection with continuous speech. The proposed method determines the peaks from the speech signal to form a voice contour. The area under the voice contour allows us to discriminate between normal and disordered subjects. In the case of disordered subjects, the calculated area under the voice contour is lower than that for a normal subject due to the malfunctioning of vocal folds, which makes the voice weaker and breathier. Some long-term features such as shimmer and jitter are based on the accurate estimation of fundamental frequency, which is itself a difficult task, especially for disordered speech signals. The proposed features do not need to estimate the pitch period or fundamental frequency during the calculation of the voice contour, and they provide a single value for the whole utterance similar to other long-term features. The voice disorder database used in this study includes 71 normal subjects, and the same number of disordered subjects. Each disordered subject has one of the following voice disorders: vocal folds cysts, laryngopharyngeal reflux disease, vocal folds polyps, unilateral vocal folds paralysis and sulcus vocalis. The accuracy of the proposed method is 100%.

Keywords: Artificial intelligence, Voice disorder detection, Modified voice contour, Continuous speech, Simpson's rule

1 Introduction

Vocal folds disorders (VFDs) occur due to the excessive use of the voice. Therefore, teachers, singers and lawyers have more risk of being affected by voice pathology. In the USA, the occurrence of voice problems in teachers is 57.7% during their lifetime and for other professions, it is 28.8% (Roy et al., 2004). The voice produced by a VFD-affected person differs from a normal person due to the malfunctioning of the vocal folds. Due to the incomplete closure and irregular vibration of the vocal folds, the speech signal of a pathological person becomes more transient. The speech signals produced by a pathological and a normal person are shown in Figure 1. A difference between the amplitudes of the two signals can be observed, and they do not have the same voice intensity. In this paper, an automatic voice pathology detection (AVPD) system based on the voice intensity of a speech signal is developed. Moreover, continuous speech samples are used for the differentiation of normal and pathological samples.

Because of its noninvasive nature, the automatic detection of voice pathology is strongly considered as a primary screening tool for the clinician. It would be of great help to an ENT specialist if an automatic system could discriminate between normal and pathological samples. The detection of VFDs with a sustained vowel has been well investigated by the research community during recent years (Markaki & Stylianou, 2011; Lee et al., 2013; Lee et al., 2014; Muhammad & Melhem, 2014), and it is a comparatively easy task than detection with continuous speech. Parsa and Jamieson (2001) concluded that the analysis of continuous speech is more challenging because of the inherent non-stationary of the signal. On the other hand, other studies emphasize the value of continuous speech over the sustained vowel in obtaining acoustic characteristics. Askenfelt and Hammarberg (1986) reported the importance of pitch and loudness variation in continuous speech as indicators of abnormal voice quality. Some of the common and most spreading voice disorders are vocal fold nodules, keratosis, vocal folds paralysis, and adductor spasmodic dysphonia (Muhammad et al., 2012).

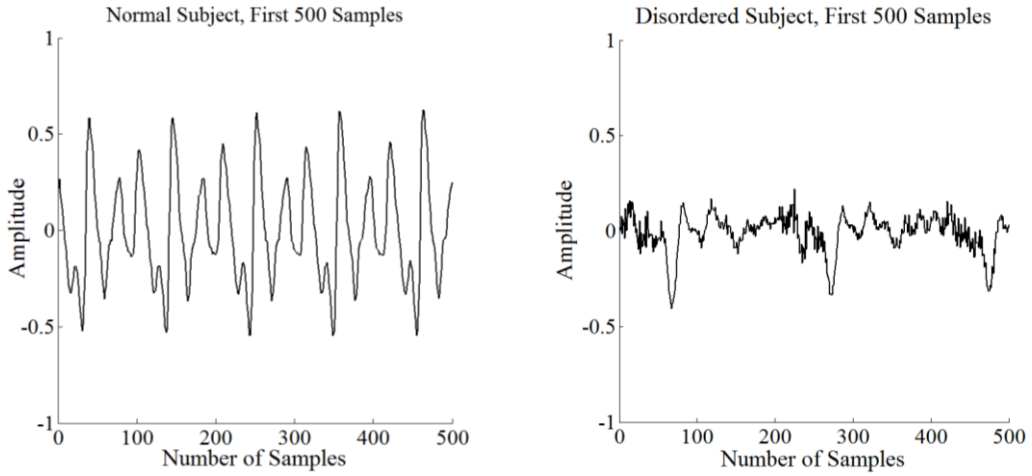
Artificial intelligence (AI) in biomedical engineering is as vital as in other fields of science for providing solutions to the problems that are computationally expensive or hard to solve by following traditional methods. Computational AI deals with pattern recognition, and different studies have used fuzzy logic (Aghazadeh & Heris, 2009), artificial neural network (Paulraj et al., 2009) and support vector machine (SVM) (Al Mojaly et al., 2014) to successfully implement AVPD systems. In this paper, SVM is used as a binary classifier to make a decision about a speech sample.

Like other long-term features such as shimmer, jitter (Arjmandi et al., 2011) and cepstral peak prominence (CPP) (Heman-Ackah et al., 2003), the proposed features also provide a single value for a whole speech signal. The measurement of long-term acoustic features normally involves the accurate estimation of the pitch period, which is a very difficult task, especially in pathological samples. The proposed features do not need to estimate the pitch period or fundamental frequency during the calculation of the modified voice contour (MVC). Arjmandi et al. (2011) used 22 long-term features including shimmer and jitter and the obtained accuracy was 91.5%. Watts and Awan (2011) reported that an accuracy of 91% with CPP was obtained for an AVPD system. The objective of this study is to propose new features that may provide better results than existing acoustic features.

Short-term features are also used in different studies to develop different AVPD systems. The most commonly used short-term features for pathology detection are LPC (Linear Prediction Coefficients), LPCC (Linear Prediction Cepstral Coefficients) and MFCC (Mel-frequency Cepstral Coefficients). LPC and LPCC have been used in many studies (Anusuya & Katti, 2010; Neto et al., 2007; Gelzinis et al., 2008; Childers and Sung-Bae, 1992; Marinaki et al., 2004) to develop voice pathology assessment systems. The correct acceptance rate of 73% with LPC and 73% with LPCC was obtained in Neto et al. (2007), when edema was detected from normal samples and other pathologies, cysts, nodules, paralysis and polyps. The efficiencies of LPC and LPCC were 85% and 80%, respectively. To conduct the study, 120 subjects, including 67 patients and 53 normal persons from the Massachusetts Eye & Ear Infirmary (MEEI) database (Massachusetts Eye & Ear Infirmary Voice & Speech Lab, 1994), were considered

and experiments were performed by using the sustained vowel /a/. MFCC coefficients were also calculated to make a comparison with LPC and LPCC and they achieved an efficiency of 52%, very low compared to LPC and LPCC. The high false acceptance rate of 74% showed that MFCC was unable to detect edema from other pathologies as well as other features. For the second case, when all normal persons were grouped into one class and all pathologies were combined in the second class, the results for MFCC were much better than the other features, which shows that MFCC performed well for the detection of disorders but less well for the discrimination between the types of disorders. As for Fourier transformation-based MFCC (Bou-Ghazale & Hansen, 2000; Murphy & Akande, 2007), non-parametric features simulate the behavior of the human ear, and the study also concluded that MFCC behaves like a clinician, because for a clinician, it is also easier to detect a disorder by auditory perception than to discriminate between disorders. By contrast, LPC and LPCC, as parametric features, represent the human vocal tract system, and hence may perform better in the classification of VFDs. In Gelzinis et al. (2008), MFCC and LPC fed into SVM and k-nearest neighbours for the classification of three classes: healthy, diffuse and nodular. The database used in the study contained sustained vowels only as recorded at the Department of Medicine, Lithuania, The classification rate obtained for MFCC was 73.08% and that for LPC was 67.31%.

In Ali et al. (2013), an MFCC-based system for disordered detection by using text-dependent running speech was developed. The running speech of a limited number of normal and disordered subjects (12 and 26, respectively) was used to evaluate the developed system. An accuracy of 91.66% was reported. The Gaussian mixture model was implemented as the classification technique with a varying number of Gaussian mixtures. A limited number of samples were used for the experiments; therefore, no reliable conclusions could be made.



(a) (b)

Fig. 1: Amplitudes in voice samples of (a) normal subjects and (b) disordered subjects

In this paper, features based on the voice intensity of a speech signal are proposed. An AVPD system by means of connected speech is developed, and SVM is used for the classification of normal and pathological speech samples.

The rest of the paper is organized as follows. Section 2 explains the proposed method. Section 3 describes the speech database. Section 4 describes the information regarding the experimental setup, and explains the baseline results as well as the results of the proposed method. Section 5 presents a discussion and provides a comparison with existing methods. Finally, Section 6 draws the conclusion.

2 Method

The proposed method is based on the idea that normal and dysphonic patients can be classified by measuring the voice intensity of a speech signal. The method is never used for voice pathology detection. The method measures the voice intensity of an utterance by calculating the area under the MVC. The contour is obtained by normalizing the cubic polynomial fitted through the peaks. These peaks are found from each frame of the utterance, and Simpson's rule is used to calculate the area under the contour.

The method is divided into five major components, and these are grouped into three steps: (1) frame blocking and peak calculation, (2) fitting and adjustment of a polynomial and (3) calculation of the area under the MVC by using Simpson's rule. To make the decision between normal and disordered speech, a binary classifier SVM is used. A block diagram for the proposed AVPD system is shown in Figure 2.

To determine the MVC, peaks are found after blocking the whole speech signal into frames. The length of each frame is 32 milliseconds, and it contains 512 samples. The peaks higher than a certain threshold value *thresh* are determined in each frame. A frame showing the calculated peaks, with *thresh* = 0.05, is depicted in Figure 3(a).

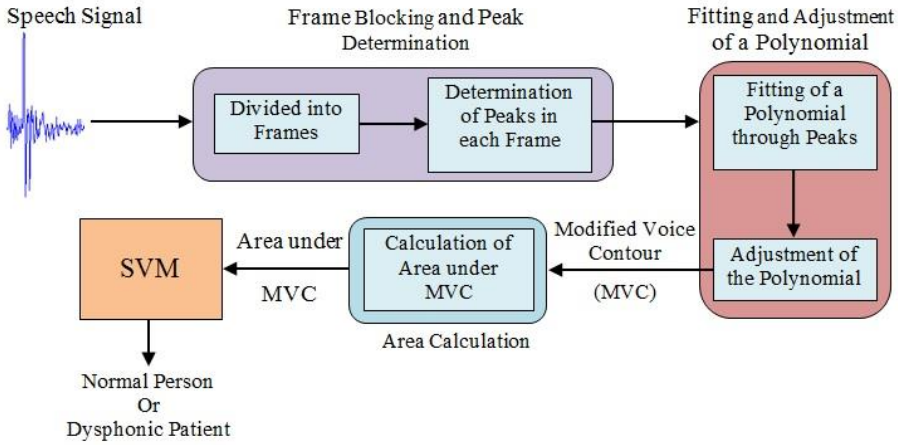


Fig. 2. Block diagram of the proposed method for the automatic voice disorder detection system

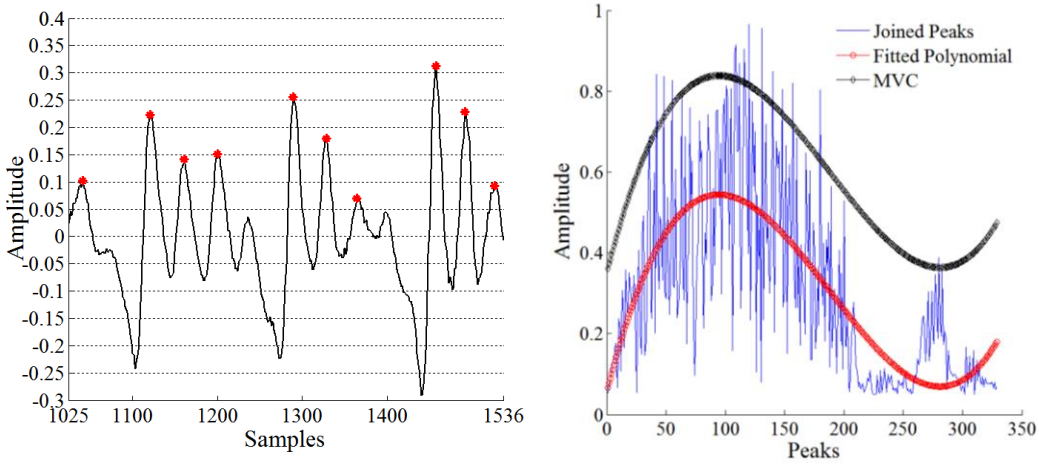
After the calculation of the peaks in each frame, they are joined together. Then, a polynomial of degree three, g(x), is fitted through these peaks to form a curve as shown in Figure 3(b). The curve is composed by circles. As observed in Figure 3(b), the fitted polynomial passes through the peak points, and hence does not make an envelope over the joined peak points. Therefore, to get an MVC, a factor is added in the polynomial g(x). The factor is given by Eq. (1) as:

$$factor = \left(\max(peaks) - \max(g)\right) * 0.70 \tag{1}$$

where *peaks* is a vector containing the peaks of all frames and *g* is a vector containing all the points on the fitted polynomial. After adding the factor into the fitted polynomial g(x), the MVC composed by 'diamonds', shown in Figure 3(b), is obtained as:

$$MVC = g(x) + factor \tag{2}$$

The area under the MVC is calculated with Simpson's rule of numerical integration. SVM takes the area under the MVC to make the decision about the presence of voice pathology. Dysphonic patients are represented by +1 and normal persons are represented by -1. SVM is implemented by using LIBSVM (Chang & Lin, 2011) with a radial basis function as a kernel, given by Eq. (3).



(a) (b)

Fig. 3. (a) Peaks determination in a frame and (b) MVC

$$K(x, x') = \exp\left(-\gamma \left\| x - x' \right\|^2\right) \tag{3}$$

where x is the training sample, x' is the testing sample and γ is a free parameter. SVM is a linear classifier; however, in most cases, the data are not linearly separable. Therefore, the kernel function is implemented to map the original input space to the higher dimensional space, whose features are lineally separable.

3 Data

The speech database used for the experiments in this paper has been considered in many studies of automatic pathology detection and classification systems and automatic speech recognition systems (Muhammad et al., 2012; Ali et al., 2013; Alsulaiman, 2014). The speech database was recorded at the Research Chair of Voice, Swallowing, and Communication Disorders, King Abdul-Aziz Hospital, Riyadh in a soundproof room. Standardized recording equipment, KayPentax Computerized Speech Lab (CSL Model 4300), was used for the recording. The database contains speech samples of both normal and pathological speakers. The total number of subjects is 142; 50% of them are patients and 50% are normal speakers, as listed in Table 1. The normal speakers did not have any previous or current history of voice disorders. Each speaker recorded one utterance of each Arabic digit from one to nine, as mentioned in Table 2. All speakers were Arab natives. The total number of available samples for the investigation is 1278 (= 142 x 9).

Table 1: Statistics of the speech database

Speakers	**Male**	**Female**	**Total**
Normal	53	18	71
Patients	50	21	71
Total:	103	39	142

Pathological speakers have five different types of VFDs, namely cysts, laryngopharyngeal reflux disease, unilateral vocal fold paralysis, polyps and sulcus vocalis. The classification of these voice disorders was based on the available clinical data, including video-laryngostroboscopic examination. In

some cases such as polyps and cysts, the clinical diagnosis was confirmed by histopathology after the excision of the lesion.

Table 2: List of Arabic digits

Digits		
Symbol	*In Roman English*	*In Arabic*
1	Wahed	واحد
2	Athnayn	أثنين
3	Thalathah	ثلاثه
4	Arbaah	أربعه
5	Khamsah	خمسه
6	Setah	سته
7	Sabaah	سبعه
8	Thamanyah	ثمانية
9	Tesaah	تسعه

4 Experimental Setup and Results

To ensure the performance of the proposed system, all results were obtained by using five-fold cross validation. The automatic systems were carried out using the following parameters:

(a) *Sensitivity (SE):* The likelihood that the system detects a pathological voice when the input is also pathological.

$$SE = \frac{TP}{TP + FN} \times 100$$

(b) *Specificity (SP):* The likelihood that the system detects a normal voice when the input is also normal.

$$SP = \frac{TN}{TN + FP} \times 100$$

(b) *Accuracy (ACC):* The ratio between correctly detected files and the total number of files.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \times 100$$

where true negative (TN) means that the system detects a normal subject as a normal subject, true positive (TP) means that the system detects a pathological subject as a pathological subject, false negative (FN) means that the system detects a pathological subject as a normal subject and false positive (FP) means that the system detects a normal subject as a pathological subject.

4.1 Results of the Proposed Method for Pathology Detection

The area under the MVC of normal speakers and dysphonic patients was calculated by following the steps mentioned in Figure 2. The calculated area was fed into SVM to differentiate between normal and dysphonic people. Different values of the variable *thresh*, i.e. 0.03 and 0.04, were considered in the peak determination process. The experimental results provided in Table 3 show that the values of *thresh* affect the detection rate. In Table 3, 95% C. I. represents the 95% confidence interval and AUC stands for the area under the ROC (receiver operating characteristic) curves. An accuracy for each fold is

determined, and their averaged accuracy with standard deviation (STD) is presented in Table 3.The developed system was verified for male and female speakers separately. Seventy percent of male speakers were used for training and the remaining 30% used for testing the system. Similarly, female speakers were also divided by the same ratio.

Table 3: Performance of the proposed method for voice pathology detection

Performance Measures	**MVC**			
	thresh = 0.03		*thresh = 0.04*	
	Male	***Female***	***Male***	***Female***
SE	70%	100%	100%	100%
SP	92%	100%	98%	100%
ACC (%) ± STD	81 ±2.07	100 ± 0	99 ± 0.83	100 ± 0
AUC	83.6 %	100%	100%	100%
95% C.I.	75.6- 91.6	100-100	96-99	100-100

With *thresh* = 0.03, an accuracy of 81% with STD of 2.07 is achieved for male speakers. The developed system recognizes normal persons well as the acceptance of TNs is 92% but does not perform well for dysphonic patients as the acceptance of TPs is 70%. FPs are only 8%, but FNs are up to 30%, which is very high. The long range of the C.I., from 75.6 to 91.6, reflects the unreliability of the system. The performance parameters for *thresh* lower than 0.03 also show the same kind of behavior. For *thresh* = 0.04, the accuracy of the pathology detection system is 99% with STD = 0.83. TPs and TNs are 100% and 98%, respectively. The system responded well for both normal and dysphonic patients. The 95% C.I. is [96 99] for males and [100 100] for females, which shows the reliability of the system. The AUC for male and female subjects are 83.6% and 100%, respectively, as shown in Figure 4(a). In Figure 4(b), the ROC curves for both genders with *thresh* = 0.04 are shown, and the AUC is 100% for each. The performance of the developed system with *thresh* higher than 0.04 did not show any improvement in the results. Therefore, they are not included in the Table 3.

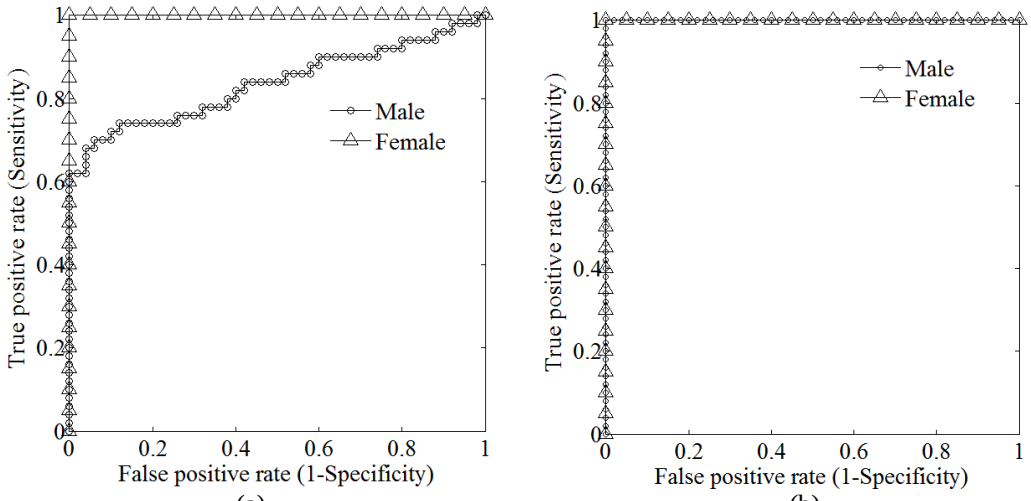


Fig. 4. ROC curve for the proposed method with (a) *thresh*=0.03 and *thresh*=0.04

4.2 Baseline Results with LPCC and MFCC for Pathology Detection

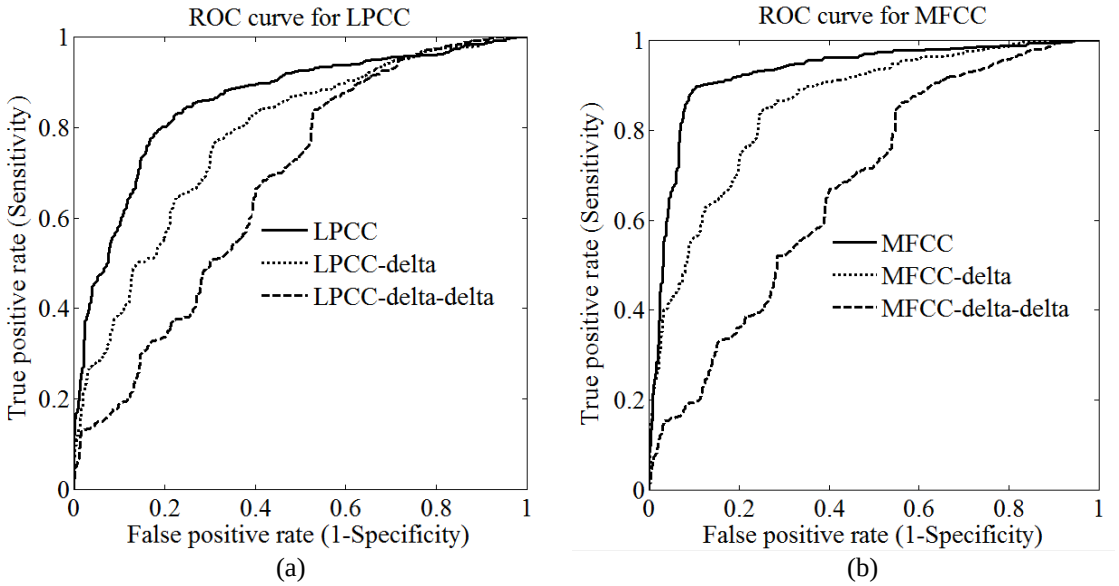
The results of the pathology detection system with the different numbers of LPCC and MFCC are presented in Table 4. Twenty-four coefficients contain 12 static coefficients and 12 delta coefficients. Similarly, 36 coefficients contain 12 static, 12 delta and 12 delta-delta coefficients.

Table 4: Performance of the LPCC and MFCC for voice pathology detection

Performance Measures	**Number of LPCC Coefficients**			**Number of MFCC Coefficients**		
	12	*24*	*36*	*12*	*24*	*36*
SE	73%	93%	97%	90%	96%	97%
SP	85%	33%	22%	89%	39%	18%
ACC (%) ± STD	79.1±4.8	63.1±6.6	59.6±6.0	89.6±2.7	67.7±3.9	57.4±6.7
AUC	85.8%	77.8%	67.4%	93%	85.6%	67.5%
95% C.I.	83.8 - 87.8	75.3 - 80.4	64.4 - 70.3	91.5 - 94.5	83.5 - 87.7	64.6 - 70.0

Accuracy decreases for both LPCC and MFCC as the number of coefficients increases, as depicted in Figure 6. The best obtained accuracy for LPCC is 79.13%, where TPs and TNs are 73.23% and 85.02%, respectively. The maximum obtained accuracy for the MFCC-based system is 89.69%, where TP is 89.92% and TN is 89%. The detection rate of MFCC is better than that of LPCC. The FNs and FPs of MFCC are also lower than those of LPCC. The performance of the proposed system is better than both LPCC- and MFCC-based AVPD systems.

The ROC curves for the different numbers of LPCC and MFCC are shown in Figure 5. It can be observed that the AUC for both LPCC- and MFCC-based detection systems decreases as the number of coefficients increases. The maximum AUC is for MFCC (93%) with 12 coefficients, and the 95% C.I. is [91.5 – 94.5].



(a) (b)

Fig 5. ROC curve for different numbers of (a) LPCC and (b) MFCC

5 Discussion

The speech production system of a subject suffering from voice pathology differs from a normal subject. Phonation, resonance and articulation are three important phases to produce speech. Vocal folds directly affect phonation and resonance in the speech production system. The abnormal behavior of the vocal folds makes the voice weaker, whispering and breathier due to too far apart vocal folds. This malfunctioning of the vocal folds makes a speech signal more transient and noisy. In this paper, a novel method based on the voice intensity of the speech signal is proposed to develop an AVPD system.

The method calculates one feature for a whole utterance produced by a subject. The peaks from the speech signal are determined to make a voice contour. Breathiness in the voice of a patient makes the voice weaker; therefore, the amplitude in the dysphonic speech signal is lower than the amplitude in a normal speech signal. The area under the voice contour for a disordered subject is also less than that of a normal subject due to the lower amplitude in dysphonic subjects. Moreover, the proposed method did not depend on the language of the database. In case of any language, speech signal of a disordered subject will contain lower amplitude than a speech signal of a normal subject.

To explain the working of the voice intensity-based AVPD system, the area under the MVC is plotted in Figure 6. As can be observed from Figure 1, the amplitude of normal speakers is higher than that of pathological patients. Figure 5 also shows that normal speakers have higher voice intensity than pathological speech samples. To observe the significance of the area under the MVC for normal and disordered subjects, a two-tailed t-test is performed. For male subjects, the *p*-value is 0.00004 (< 0.05), which shows that the area under the MVC for normal and disordered subjects is statistically significant. Similarly, the *p*-value = 0.000003 for female subjects shows the significance of the two classes.

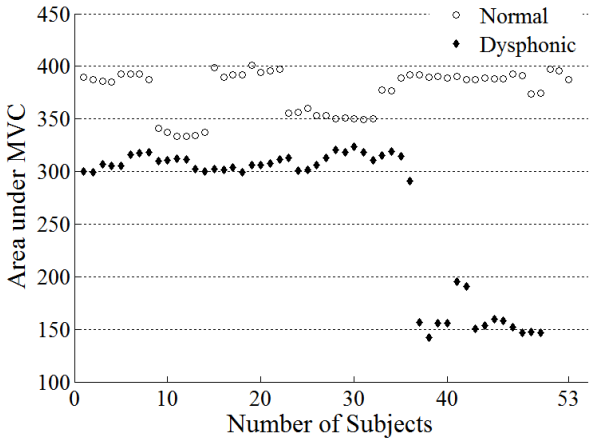


Fig. 6. Area under the MVC for male speakers for Digit 6 with *thresh* = 0.04

Similar to other long-term features such as shimmer, jitter (Arjmandi et al., 2011) and CPP (Heman-Ackah et al., 2003), the proposed features also provide a single value for a whole utterance. The measurement of long-term acoustic features normally involves the accurate estimation of the pitch period, which is a very difficult task, especially in pathological samples. The proposed features do not need to estimate the pitch period or fundamental frequency during the calculation of the voice contour, and this is a positive aspect of the proposed features.

In this paper, the developed AVPD system used running speech, which is a comparatively difficult task than sustained vowel-based systems. Most automatic pathology assessment systems, pathology detection and pathology classification presented in the literature are developed by using a sustained vowel /ah/ (Markaki & Stylianou, 2011; Lee et al., 2013; Lee et al., 2014; Muhammad & Melhem,

2014). A speech signal remains stationary in the case of a sustained vowel, but it varies over time in the case of continuous speech. This is the reason why pathology assessment systems that use continuous speech are challenging and require more investigation. Moreover, these systems are more realistic because people use continuous speech in their conversations in daily life. Running speech contains fluctuations in vocal characteristics in relation to voice onset, voice termination and voice breaks, which are considered as crucial in quality voice evaluation. These characteristics are not fully represented in short signals of phonation such as a sustained vowel (Hammarberg et al., 1980).

The results of the proposed system are also compared with the results of existing systems in the literature. A comparison of the systems using running/continuous speech is listed in Table 5. The results of the existing systems are taken from the studies mentioned in the first column. From Table 5, it can be concluded that the proposed system achieved higher accuracy than the existing systems.

Table 5: Comparison of the proposed system with existing connected speech-based systems

Reference	Database	No. of Samples (N + P)	Features	Accuracy
Godino-Llorente et al. (2009)	MEEI	23 + 117	MFCC	96%
Klára et al. (2012)	Private	26 + 33	MFCC, shimmer, jitter, and HNR	86%
Ali et al. (2013)	Private	12 + 26	MFCC	91.66%
Proposed Method	Private	71 + 71	MVC	100%

6 Conclusion

In this study, new features are proposed to develop an AVPD system by using continuous speech. The proposed features are based on the voice intensity of the speech signal. Voice intensity is measured by calculating the area under the MVC. VFD makes the voice weaker and whispering; therefore, a speech signal comprises a low amplitude in the case of a dysphonic patient. Due to the malfunctioning of the vocal folds, the voice intensity in a signal produced by a patient is always lower than that in a normal person. The proposed method calculates one feature for an utterance, and the obtained accuracy is 100%. In this study, the proposed method uses the connected speech to develop the AVPD system, which is a comparatively difficult task than sustained vowel-based systems. Moreover, the extracted words from a sentence can also be used with the proposed method. Based on the obtained result, it can be concluded that the proposed features outperform both types of cepstral coefficients, namely MFCC and LPCC. Contrary to other existing long-term features, the proposed features do not need the accurate estimation of pitch period or fundamental frequency, which is itself a difficult task.

In future work, we will modify the proposed method to make it robust when a normal and pathological person varies the mouth-to-microphone distance and pitch level while recording.

Acknowledgment

This project was funded by the National Plan for Sciences, Technology and Innovation (MAARIFAH), King Abdulaziz City for Science and Technology, Kingdom of Saudi Arabia, Award Number (12-MED2474-02).

References

Aghazadeh, B. S., & Heris H. K. (2009). Fuzzy logic based classification and assessment of pathological voice signals. *International conference of the IEEE on Engineering in Medicine and Biology Society*, 328-331.

Al Mojaly, M., Muhammad, G. & Alsulaiman, M. (2014). Detection and classification of voice pathology using feature selection. *11th International Conference on Computer Systems and Applications (AICCSA)*, 571-577.

Ali, Z., Alsulaiman, M., Muhammad, G., Elamvazuthi, I., & Mesallam, T. A. (2013). Vocal fold disorder detection based on continuous speech by using MFCC and GMM. *7th IEEE GCC Conference and Exhibition (GCC)*, 292-297.

Alsulaiman, M. (2014). Voice pathology assessment systems for dysphonic patients: Detection, classification, and speech recognition. *IETE Journal of Research, 60*(2), 156-167.

Anusuya, M. A., & Katti, S. K. (2010). Front end analysis of speech recognition: A review. *International Journal of Speech Technology*, *14*, 99-145.

Arjmandi, M. K., Pooyan, M., Mikaili, M., Vali, M., & Moqarehzadeh, A. (2011). Identification of voice disorders using long-time features and support vector machine with different feature reduction methods. *Journal of Voice*, *25*(6), 275-e289.

Askenfelt. A. G., & Hammarberg, B. (1986). Speech wave form perturbation analysis: A perceptual acoustic analysis of seven measures. *Journal of Speech, Language, and Hearing Research*, *29*, 50-64.

Bou-Ghazale, S. E., & Hansen, J. H. L. (2000). A comparative study of traditional and newly proposed features for recognition of speech under stress. *IEEE Transactions on Speech & Audio Processing*, *8*(4), 429-442.

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, *2*(3), 1-27.

Childers, D. G., & Sung-Bae K. (1992). Detection of laryngeal function using speech and electroglottographic data. *IEEE Trans. Biomed. Eng.*, *39*(1), 19-25.

Gelzinis, A., Verikas, A., & Bacauskiene, M. (2008). Automated speech analysis applied to laryngeal disease categorization. *Journal of Computer Methods and Programs in Biomedicine, 91*(1), 36-47.

Godino-Llorente, J. I., Fraile, R., Sáenz-Lechón, N., Osma-Ruiz, V., & Gómez-Vilda, P. (2009). Automatic detection of voice impairments from text-dependent running speech. *Biomedical Signal Processing and Control, 4*(2009), 176-182.

Hammarberg, B., Fritzell, B., Gauffin, J., Sundberg, J., & Wedin, L. (1980). Perceptual and acoustic correlates of abnormal voice qualities. *Acta Otolaryngol*, *90*, 441-451.

Heman-Ackah, Y. D., Heuer, R. J., Michael, D. D., Ostrowski, R., Horman, M., Baroody, M. M., *et al.* (2003). Cepstral peak prominence: A more reliable measure of dysphonia. *Ann Otol Rhinol Laryngol, 112*, 324-333.

Klára, V., Viktor, I., & Krisztina, M. (2012). Voice disorder detection on the basis of continuous speech. *5th European Conference of the International Federation for Medical and Biological Engineering*, *37*, 86-89.

Lee, J. W., Kang, H. G., Choi, J. Y., & Son, Y. I. (2013). An investigation of vocal tract characteristics for acoustic discrimination of pathological voices. *Biomed Research International*, *2013*, 1-12.

Lee, J. W., Kim, S., & Kang, H. G. (2014). Detecting pathological speech using contour modeling of harmonic-to-noise ratio. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5969-5973.

Marinaki, M., Kotropoulos, C., Pitas, I., & Maglaveras, N. (2004). Automatic detection of vocal fold paralysis and edema. *ICSLP '04*, 1-4.

Markaki, M., & Stylianou, Y. (2011). Voice pathology detection and discrimination based on modulation spectral features. *IEEE Transactions on Audio, Speech, and Language Processing, 19*(7), 1938-1948.

Massachusetts Eye & Ear Infirmary Voice & Speech Lab (1994). Disordered Voice Database Model 4337 (Ver. 1.03), *Kay Elemetrics Corp*, Boston, MA.

Muhammad, G., Mesallam, T. A., Malki, K. H., Farahat, M., Alsulaiman, M., & Bukhari, M. (2011). Formant analysis in dysphonic patients and automatic Arabic digit speech recognition. *Biomed Eng Online*, *10*(41), 1-12.

Muhammad, G., & Melhem, M. (2014). Pathological voice detection and binary classification using MPEG-7 audio features. *Biomedical Signal Processing and Control, 11*(2014), 1-9.

Muhammad, G., Mesallam, T. A., Malki, K. H., Farahat, M., Mahmood, A., & Alsulaiman, M. (2012). Multidirectional regression (MDR)-based features for automatic voice disorder detection. *J Voice*, *26*(6), 817 e19-27.

Murphy, P. J., & Akande, O. O. (2007). Noise estimation in voice signals using short-term cepstral analysis. *Journal of the Acoustical Society of America*, *121*(3), 1679-1690.

Neto, B. G. A., Costa, S. C., Fechine J. M., & Muppah, M. (2007). Feature estimation for vocal fold edema detection using short-term cepstral analysis. *7th International Conference on Bio-Informatics and Bio-Engineering*, 1158-1162.

Parsa, V., & Jamieson, D. G. (2001). Acoustic discrimination of pathological voice: Sustained vowels versus continuous speech. *Journal of Speech, Language, and Hearing Research, 44*, 327-339.

Paulraj, M. P., Yaacob, S., & Hariharan, M. (2009). Diagnosis of vocal fold pathology using time-domain features and systole activated neural network. *5th International Colloquium on Signal Processing & Its Applications*, 29-32.

Roy, N., Merrill, R. M., Thibeault, S., Parsa, R. A., Gray, S. D., & Smith, E. M (2004). Prevalence of voice disorders in teachers and the general population. *J Speech Lang Hear Res.*, *47*(2), 281-93.

Watts, C. R., & Awan, S. N. (2011). 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, 54*, 1525-1537.