Automatic voice pathology detection and classification using vocal tract area irregularity

Ghulam Muhammad[1], Ghadir Altuwaijri[1], Mansour Alsulaiman[1], ZulfiqarAli[1,2], Tamer A. Mesallam[3, 4, 5], MohamedFarahat[3,4], Khalid H. Malki[3,4], AhmedAl-nasheri[1]

[1]Speech Processing Group, Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia

[2]Centre for Intelligent Signal and Imaging Research, Department of Electrical and Electronic Engineering, Universiti Tekhnologi PETRONAS, Tronoh, Perak, Malaysia

[3]ENT Department, College of Medicine, King Saud University, Riyadh, Saudi Arabia

[4]Research Chair of Voice, Swallowing, and Communication Disorders, King Saud University, Riyadh, Saudi Arabia

[5]ENT Department, College of Medicine, Al-Menoufiya University, Shebin Alkoum, Egypt

Abstract

In this paper, an automatic voice pathology detection (VPD) system based on voice production theory is developed. More specifically, features are extracted from vocal tract area, which are connected to the glottis. Voice pathology is related to a vocal fold problem, and hence the vocal tract area which is connected to vocal fold or glottis should exhibit irregular patterns over frames in case of a sustained vowel for a pathological voice. This irregular pattern is quantified in the form of variance across the frames to distinguish between normal and pathological voices. The proposed VPD system is evaluated on the MEEI database with sustained vowel samples. Vocal tract irregularity features and support vector machine classifier are used in the proposed system. The proposed system achieves 99.02%±0.01 accuracy on the MEEI database. The results indicate that vocal tract irregularity measures can effectively be used in voice pathology detection.

Keywords: voice pathology detection, vocal tract area, voice disorders, support vector machine.

Introduction

Voice pathology detection refers to a detection procedure of the pathology in the vocal fold from an input voice. The detection can be subjective or objective. In subjective detection, an experienced physician hears the voice and assesses whether the voice is normal or having pathology based on his/her previous knowledge and

experience. However, this type of assessment may vary physician to physician depending on experience (Kreiman, Gerratt, & Precoda, 1990). Therefore, objective evaluation of voice pathology is gaining more attention day by day from both medical and engineering personnel.

Lieberman proposed one of the first acoustic voice parameters in pathological voice analysis in 1961 (Lieberman, 1961). Voice perturbation and quality measures, such as jitter (pitch perturbation) and shimmer (amplitude perturbation), depend on accurate extraction of fundamental frequency and the amplitude of various waveform types. However, measuring fundamental frequency (F_0) is a very difficult task; especially in the case of a pathological voice (Titze, & Liang, 1993).

Most of the work in voice pathology detection rely on the measurement of a sustained vowel, particularly, /AH/ sound, uttered by the subject. /AH/ is easy to pronounce for a patient having voice pathology; its formants are clearly distinguishable, and peaks are prominent. These three attributes of sustained /AH/ make it a proper choice for voice pathology detection. Popular acoustic features extracted from sustained vowel to detect voice pathology are shimmer and jitter, spectral centroid and spectral flatness (Moran, Reilly, Chazal, & Lacy, 2006), cepstral peak prominence (Heman-Ackah, Heuer, Michael, Ostrowski, Horman, Baroody, Hillenbrand, & Sataloff, 2003), etc. Vasilakis and Stylianou developed a new method for short-time jitter and found the AUC (area under curve) reached 94.82% (Vasilakis, & Stylianou, 2009). Though these popular features are widely used in the literature, many researchers found different relationships (poor to high) between these features and severity of voice pathology (Martin, Fitch, & Wolfe, 1995; Shrivastav, 2003). Little *et al.* found that nonlinear measures, such as recurrence period density entropy and detrended fluctuation analysis, were more stable and reliable than the classical measures on voice pathology quantification (Little, Costello, & Harries, 2009).

Markaki and Stylianou suggested a modulation spectra method to identify pathological voices and achieved an AUC of 96.26% (Markaki, & Stylianou, 2011). Arias-Londono *et al.* introduced a method, which is based on extracting eleven features using nonlinear analysis of time series, and gained a classification accuracy of 98.23% using classifier combination (Arias-Londono, Godino-Llorente, Saenz-Lechon, Osma-Ruiz, & Castellanos-Dominguez, 2011). Godino-Llorente *et al.* proposed a detection system of pathological voice by means of Gaussian mixture models and short-term mel cepstral parameters complemented by frame energy together with first and second derivatives (Godino-Llorente, Gomez-Vilda, & Blanco-Velasco, 2006). All of three methods

in (Markaki, & Stylianou, 2011; Arias-Londono, *et al*., 2011; Godino-Llorente, *et al*., 2006) were applied on a common subset of MEEI database (Kay Elemetrics Corp., 1994). Some research was conducted with continuous speech to detect voice pathology (Lowell, Colton, Kelley, & Hahn, 2011; Muhammad, Mesallam, Almalki, Farahat, Mahmood, & Alsulaiman, 2012), where the authors claim that continuous speech is more practical than sustained vowel in everyday life.

Voice pathology detection is still an open problem. The difficulty of detecting voice pathology is closely associated to the severity of the pathology. A patient with less severe pathology can pronounce a vowel without much difficulty; however, it is not the same with the patient having severe voice pathology; for example, a big cyst in the vocal fold prevents proper closing and opening of the fold while pronouncing a vowel sound.

In this paper, we propose a voice pathology detection (VPD) system based on the vocal tract area features. Voice pathology is related to vocal fold problem, and hence the vocal tract area which is connected to the vocal fold should exhibit irregular patterns over frames in case of a sustained vowel for a pathological voice. Typically, it is believed that there is no infralaryngeal (tracheobronchial tree) effect on the vocal tract while producing a vowel on the assumption that the voicing source has infinite impedance. However, a finely detailed analysis must recognize that the infralaryngeal structures exert an influence on the vocal tract; the articulatory configuration in the vocal tract interacts with the articulation in the vocal folds (Kent, & Kim, 2008). Therefore, additional vocal tract-related information is expected to assist in detecting the characteristics of the vocal folds, especially during phonation (Lee, *et al*., 2013). We quantify this irregularity in the form of variances along time. Several methods based on glottal noise measures, such as signal to noise ratio (SNR), harmonic to noise ratio (HNR), normalized noise energy (NNE), pitch amplitude (PA), and spectral flatness ratio (SFR), are proposed in (Parsa, & Jamieson, 2000). However, calculation of SNR, HNR, and NNE requires correct estimation of F_0. PA and SFR are based on linear prediction (LP) modeling, where PA needs computationally expensive autocorrelation, and SFR needs windowing and Fourier transform after LP. On the other hand, our proposed method is much simpler than the PA and SFR in the sense that it uses LP coefficients directly to calculate the area of different tubes in the vocal tract.

Materials and Methods

Data

To prove the proposed concept, we utilize a common database, which is MEEI database, in the experiments. Campbell and Reynolds state that "the main benefit of using standard corpora is that it allows researchers to compare performance of different techniques on common data, thus making it easier to determine which approaches are most promising to pursue" (Campbell, & Reynolds, 1999). To compare the proposed system with others, we used the same subset of the MEEI database which was used in (Markaki, & Stylianou, 2011; Arias-Londono, *et al*., 2011; Godino-Llorente, *et al*., 2006), denoted as MEEI$_{subset}$; this subset has sustained vowel /AH/ recordings for 53 normal and 173 pathological speakers. The criteria in selecting this subset took into consideration a wide variety of voice disorders and a uniformly distributed gender and age between both normal and pathological classes (Parsa, & Jamieson, 2000). A statistical description of the MEEI$_{subset}$ is shown in Table 1. In the pathological class, the voice disorders include adductor (22 speakers), vocal nodules (20 speakers), keratosis (26 speakers), vocal fold polyp (20 speakers), and paralysis (71 speakers). The rest 14 speakers have multiple disorders. The sound files in the MEEI$_{subset}$ have a sampling rate of either 50 kHz or 25 kHz. Therefore, all the files having a sampling rate of 50 kHz were down-sampled to 25 kHz to have a uniform sampling rate.

Table 1: Statistics of MEEI$_{subset}$:♂ for male, ♀ for female (Parsa, & Jamieson, 2000).

	Number of speakers		Mean age (years)		Age range (years)		Standard Deviation (years)	
	♂	♀	♂	♀	♂	♀	♂	♀
Nor mal	21	32	38.8	34.2	26 to 59	22 to 52	8.5	7.9
Pathological	70	103	41.7	37.6	26 to 58	21 to 51	9.4	8.2

Proposed System

The proposed VPD system, shown in Figure 1, consists of two major components, which are feature extraction and classification. Features are extracted from the vocal tract area obtained from the input voice. The vocal tract

starting from glottis and ending at lips can be divided into several tubes of uniform length. Vocal tract area function calculates area of different tubes along the vocal tract for a certain time based on the phone production system. Figure 2 shows simulated tubes of a vocal tract area. Glottis is directly connected to the vocal fold, and therefore, any abnormality in the vocal fold may be observed to the area of the tubes connected to the glottis. During vowel pronunciation, in case of voice pathology the vocal folds can have irregular opening and closing, thereby causing noise. The abnormality can be found in the tubes area across frames.
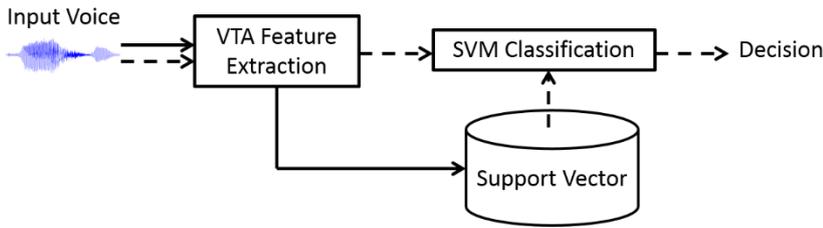

Figure 1: Block diagram of the proposed VPD system. Solid line indicates training phase, while dashed line indicates testing phase.
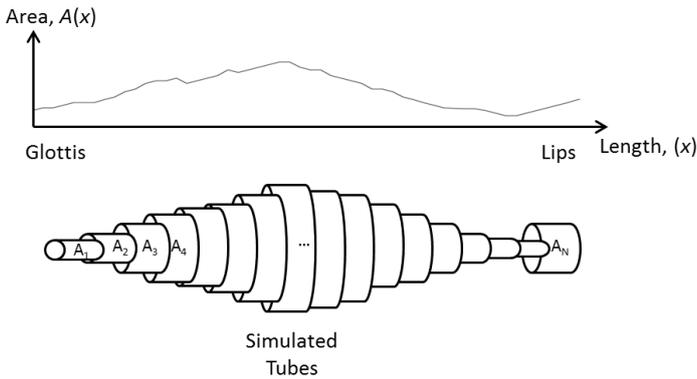

Figure 2: Simulated tubes of the vocal tract.

Vocal tract area calculation

The area of the vocal tract is calculated using LP filter. LP coding of speech is proposed early 1980's and has been used in many applications of speech (Markel, & Gray, 1982). Suppose, $x(n)$ is the current sample, and it has to be predicted by some weighting of the previous M samples. M is called the order of the filter (number of tubes), and the predicted current sample, $x'(n)$, can be found as Eq. (1).

$$x'(n) = \sum_{i=1}^{M} a_i x(n-i) \qquad (1)$$

where, a_i corresponds to i-th LP coefficients or weight. The error between the current sample and the predicted current sample is

$$e(n) = x(n) - x'(n) \tag{2}$$

The mean square error (E) is calculated by Eq. (3).

$$E = \sum_n e^2(n) = \sum_n \left(x(n) - \sum_{i=1}^{M} a_i x(n-i) \right)^2 \tag{3}$$

To find minimum error, we set the derivative of E equal zero and find the coefficients a_i.

$$\frac{\partial E}{\partial a_i} = 0 \Rightarrow \sum_n x(n-k) \sum_{i=1}^{M} a_i x(n-i) = \sum_n x(n-k)x(n) \tag{4}$$

Eq. (4) can be solved using the matrix form in Eq. (5).

$$\begin{bmatrix} r(0) & r(1) & \dots & r(M-2) & r(M-1) \\ r(1) & r(0) & \dots & r(M-3) & r(M-2) \\ \dots & \dots & \dots & \dots & \dots \\ r(M-2) & r(M-3) & \dots & r(0) & r(1) \\ r(M-1) & r(M-2) & \dots & r(1) & r(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \dots \\ a_{M-1} \\ a_M \end{bmatrix} = \begin{bmatrix} r(1) \\ r(2) \\ \dots \\ r(M-1) \\ r(M) \end{bmatrix} \tag{5}$$

where, $r(k) = \displaystyle\sum_{n=0}^{N-1-k} x(n)x(n+k)$ and N is the total number of samples in an analysis frame.

Eq. (5) can be solved using Levinson-Durbin recursive algorithm as follows (Markel, & Gray, 1982).

$$\textit{Initialization}: E^0 = r(0)$$
$$\textit{For } m \geq 1,$$
$$a_m{}^m = k_m = \left[r(m) - \sum_{i=1}^{m-1} a_i{}^{m-1} r(m-i) \right] / E^{m-1}$$
$$a_i{}^m = a_i{}^{m-1} - k_m a_{(m-i)}{}^{m-1}, \quad \textit{for } i = 1,2,...,m-1$$
$$E^m = E^{m-1} \left[1 - k_m{}^2 \right]$$
$$\textit{if } m = M, \textit{ then Stop.}$$

In the above algorithm, E^m corresponds to the m-th order prediction error, and k is called reflection coefficient. The vocal tract area of the m-th tube can found using Eq. (6).

$$A_{M+1} = 1$$
$$A_m = A_{m+1} \times \frac{1+k_m}{1-k_m}, \quad for\ m = M,...,2,1 \qquad (6)$$

Feature extraction based on vocal tract area

The vocal tract areas of m number of tubes are calculated frame by frame for each input voice file. If the input voice is a sustained vowel, for a normal speaker without any voice pathology, the area of the m-th tube across the frames will not vary a significant amount. On the other hand, the area will differ for a pathological voice because of irregular opening and closing of the vocal fold. Figure 3 shows an example of such case. In the top row of the figure, the vocal tract areas of two different frames for a normal voice (file AXH1NAL in MEEI database) are shown. We can find that the areas of the tubes close to the glottis are almost similar for the two frames. In the bottom row of the figure, the same are shown for a pathological voice (file ALB18AN). We see that the areas close to the glottis are not similar in the two frames. From the above observation, we calculate three types of features as described below.

*Average*: The average area (Avg_m) of each tube (A_m) is calculated along the frames.

Variance: The variance (Var_m) of each tube is calculated along the frames.

Ratio_Variance: The ratio ($R_{m,\ m-1}$) of the area of the two successive tubes is calculated along the frames. We can formulate it as
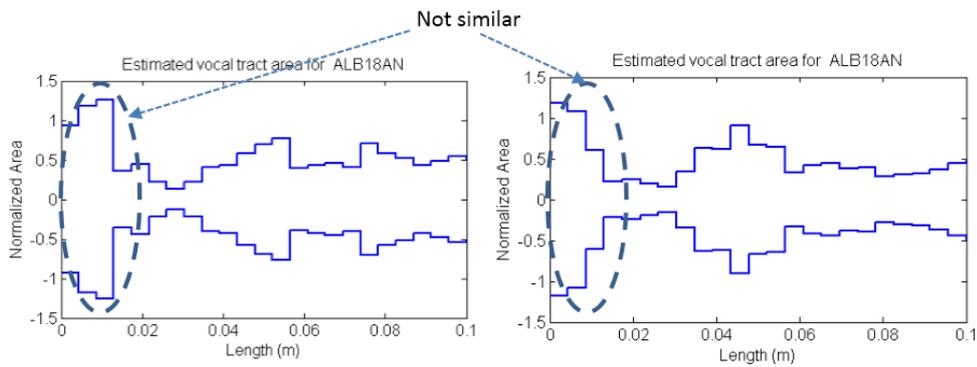
$$R_{m,m-1} = \frac{A_{m-1}}{A_m}. \qquad (7)$$

So, there will be m-1 number of ratio. We take the variance of the ratio along the frames and denote it as ($R\text{-}var_{m,m-1}$). The reason behind this feature is to capture the relative shape differences between the tubes (as evident in Figure 3).

The final feature vector set is the concatenation of these three types of features as [$Avg_m\ Var_m\ R\text{-}var_{m,m-1}$].
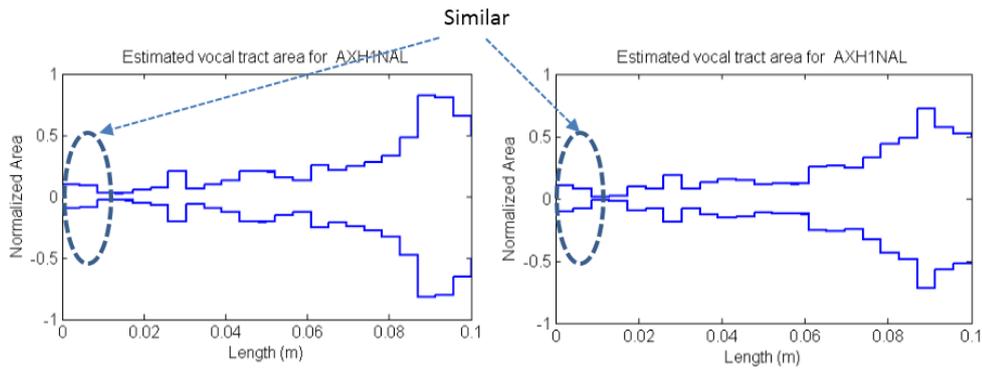
(a) Normal voice.

(b) Pathological voice.

Figure 3: Vocal tract area for (a) normal and (b) pathological voices for two different frames. For the normal voice, area close to glottis is similar between frames, while those for pathological voice are different.

Setup

The minimum value of M (order of the LP filter) is set to 16, because the typical vocal tract length is around 17 cm. We varied M between 16, 20, 24, and 28. Frame length is set to 30 ms with no overlapping.

The classification process was done using a ten-fold cross-validation SVM. RBF kernel was chosen for SVM. In ten-fold cross-validation, the normal files and the pathological files are randomly divided into ten equal groups each. In each iteration, nine groups each from the normal and the pathological are used for training, while the remaining are for testing. Therefore, at the end of ten iterations, all the ten groups are tested. There is no overlapping of the samples between the training set and the testing set. The optimization of the SVM parameters was done with the training set. Loose grid search and fine grid search were conducted to obtain the optimal parameters values for 'c' and 'gamma'. The whole program ran 20 times and the final accuracy was

obtained by averaging accuracies of 20 runs. We used LIBSVM, which is a library for SVM (Chih-Chung, & Chih-Jen, 2011).

Results

Figure 4 shows the effect of the order of LP on the proposed system. All the three types of irregularity measures are calculated on all the tubes. From the figure, we see that 24-order performs the best and *Var* has the highest detection capability with 95% accuracy. From this outcome, we find that tube lengths of around 7 mm (17 cm / 24 tubes) each can effectively detect the irregularity in the voice production. If the tube length is greater than 7 mm, the irregularity is smoothed, and if the tube length is less than 7 mm, noise is emphasized. All the subsequent experiments are done with 24-order LP.

Figure 5 shows the accuracies of the system using various numbers of tubes (starting from the first tube). For example, considering the average area of only the first tube (Avg_1), the accuracy is 32%, while the variance of the first tube (Var_1) gives 77% accuracy. The variance of the ratio between the first and the second tubes ($R\text{-}var_{2,1}$) shows 51% accuracy. Using the first three tubes, the accuracies obtained by *Avg*, *Var*, and *R-var* are 41%, 89%, and 75%, respectively. The last bar of each type corresponds to the result shown in Figure 4 with 24-LP. These results demonstrate that variances of the vocal tract areas of the tubes have very good discrimination capability between normal and pathological voices. *Var* of only first three tubes has accuracy of 89%, which is interesting. Shape ratio of the successive tubes also has a good contribution towards the discrimination.

The performance of combining all the three irregularity measures is given in Table 2. The best accuracy (99.02% with a standard deviation of 0.01) is obtained using all the tubes. In this case, the number of features is (24 *Avg* + 24 *Var* +23 *R-var* =) 71. The sensitivity (*Sn*) and specificity (*Sp*) is 99.8% and 97.5%, respectively, and the AUC = 0.9915 (see Figure 6). The ROC curve with AUC is shown in Figure 6 (with 95% confidence interval). Figure 7 shows the discriminative capability of the first three variances in the form of a scatter plot.
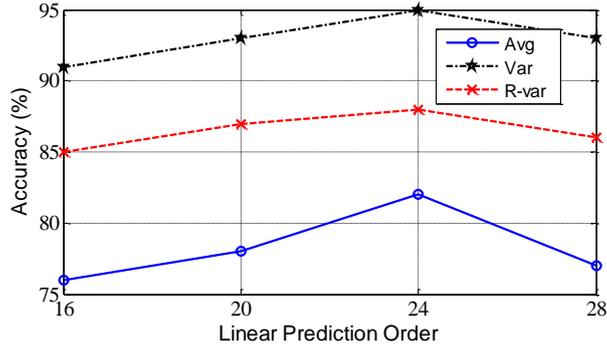
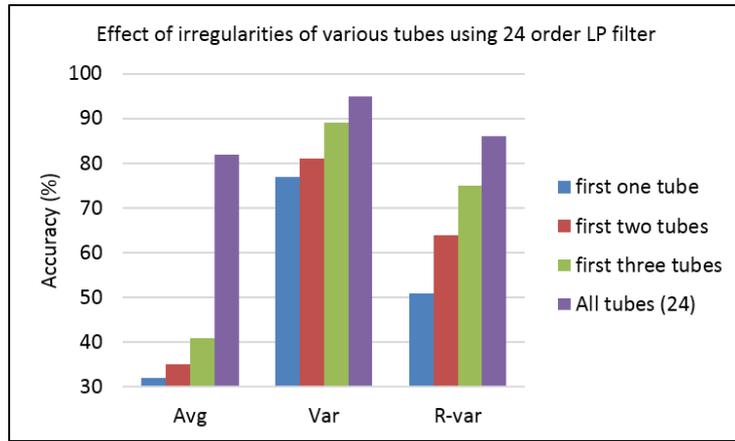Figure 4: *Accuracy (%) of the system using various LP orders with three different measurement types.*

Figure 5: *Accuracy (%) of the system using various vocal tract area irregularity measurements with 24-*

order LP filter.

Table 2: *Performance of the combined irregularity measures. The numbers inside the parenthesis*

represent feature dimension.

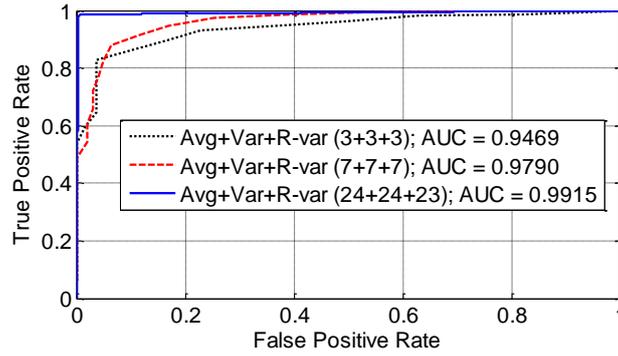| Features	%Acc	%Sn	%Sp
Avg+Var+R-var (3+3+3)	95.14±1.13	97.1±0.03	93.1±0.87
Avg+Var+R-var (7+7+7)	97.84±0.07	98.4±0.12	96.5±1.3
Avg+Var+R-var (24+24+23)	99.02±0.01	99.8±0.02	97.5±0.04

Figure 6. *ROC curves of the proposed system with different dimensions of irregularity measures.*
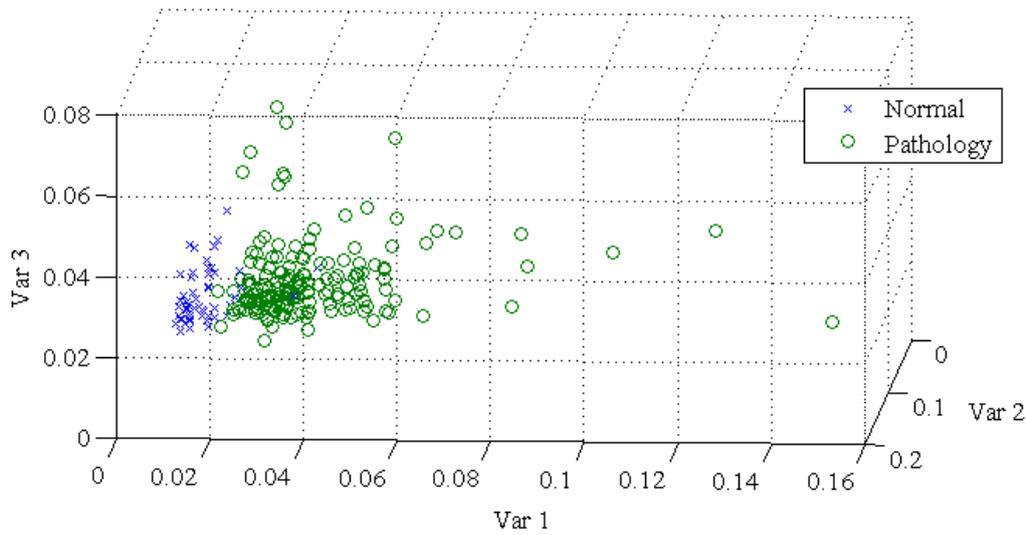

Figure 7. *3-D scatter plot to show the discriminative capability of the three features (Var 1, Var 2, and Var 3).*

Comparison with other methods of VPD system

We compare the performance of different methods of the VPD system. The accuracies together with the feature and classification names are given in Table 3. The proposed vocal tract area irregularity measures and SVM classification based VPD system performs better than the methods described in (Markaki, & Stylianou, 2011; Arias-Londono, *et al.*, 2011; Godino-Llorente, *et al.*, 2006). The accuracies of these methods are taken from their papers, because all the methods in these papers are applied on the same subset of the MEEI database. From the table we find that the best accuracy (99.02%) is achieved with the proposed method, and the second best accuracy (98.23%) is obtained with the method in (Arias-Londono, *et al.*, 2011).

Table 3. Comparison of different approaches on voice pathology detection.

Method	%Acc	Feature	Classification
Proposed	99.02	Vocal tract area irregularity	SVM
Method in (Markaki, & Stylianou, 2011)	94.1	Modulation Spectra	SVM
Method in (Godino-Llorente, *et al.*, 2006)	94.07	Cepstral	GMM
Method in (Arias-Londono, *et al.*, 2011)	98.23	MFCC + others	GMM+SVM

Discussion

The LP coefficients correspond to the excitation signal of the vocal tract area function. In case of normal voice, the LP coefficients may be modeled as a deterministic signal (all-pole), consisting of the sinusoids with equal amplitudes. In normal production of a sustained vowel, the vocal fold vibrates with periodic and stable structure, and the vocal tract is approximately fixed. In case of some pathologies, there might be vocal fold tension irregularity, for example, the vocal folds can have irregular opening and closing. The LP coefficients cannot be estimated on only the closed phase of the cycle, the interpretation of the area estimates is not straightforward. We can generally say that, in this case, the LP coefficients are modeled as a sum of incoherent sinusoidal with random phases. This irregularity may produce frequency jitter, turbulent noise, subharmonics, two independent fundamental frequencies, and other effects. These measurements are used before in the literature to detect voice pathology (Moran, *et al.*, 2006; Heman-Ackah, *et al.*, 2003; Vasilakis, & Stylianou, 2009; Parsa, & Jamieson, 2000). However, in this paper, we propose a simple mechanism to detect voice pathology. In this mechanism, features are extracted directly from the vocal tract area tubes across frames, thereby reducing the extra calculation needed to measure jitter, noise, subharmonics, fundamental frequencies, etc. Voice is the production of a source (vocal fold) and a filter (vocal tract), and therefore, voice quality should depend on the both. Nunes *et al.* studied that supraglottic tract has a clear effect on the voice pathology; however, they could not quantify the relation (Nunes, de Souza, Duprat, Silva, Costa, & Paulino, 2009). This can be quantified by a measure of irregularity in the vocal tract.

To quantify the irregularity of the vocal tract area, we calculate the average, variance, and ratio between two successive tubes in the vocal tract. In the experiments, the variance is proved to be more significant than the other two measures. The features from the first three tubes achieve accuracy as high as 95%, indicating their significant contribution to detect the irregularity. Supraglottic tract can be represented by the tubes close to the glottis (for example, first three to seven tubes), and these tubes contribute more to the detection performance (Table 2). This study is the first kind of quantifying the vocal tract irregularity to detect voice pathology.

Markakai *et al*. reported 94.1% accuracy using modulation spectra features and SVM classification (Markaki, & Stylianou, 2011). The total number of features in their case was 91, and this number of optimal features wa found using a high order singular value decomposition. Godino-Llorente *et al*. achieved 94.07% accuracy using 12 MFCC and 12 delta-MFCC with GMM (Godino-Llorente, *et al*., 2006). However, our proposed method achieved 95.14% accuracy using only nine features (Table 2, first row), without any optimal feature selection.

Arias-Londono *et al*. used different entropy and noise measures together with MFCC features (Arias-Londono, *et al*., 2011). They fused two classifiers, generative and discriminative, in the form of SVM to achieve 98.23% accuracy. Fusing classifiers is always a time consuming process. In our proposed method, we got 97.84% accuracy with only 21 features and 99.02% accuracy with 71 features. These accuracies were obtained without any classifier combination.

To the best of our knowledge, 99.02% is one of the highest accuracies obtained in MEEI database reported so far in the literature. To investigate the proposed features capability of detecting voice pathologies independent of the classifiers, we performed another set of experiments using GMM-based classifiers with number of Gaussian mixture = 2, 4, 8, 16, and 32. The number of features was fixed to 71. The accuracies obtained were 93.2%, 94.6%, 97.1%, 98.8%, and 96.8%, respectively. The accuracy of 98.8% closely matches with 99.02% obtained with the SVM classifier indicating the robustness of the proposed features.

Conclusions

A VPD system based on vocal tract area irregularity was proposed. Three measurements of the irregularity, namely, average, variance, and ratio of the successive tubes, were investigated. The proposed system achieved 99.02% accuracy using these three irregularity measurements on sustained vowels. We can summarize the findings to detect voice pathology as follows:

- The irregularity measure of the vocal tract area is useful.

- Supraglottic tract has more contribution than the remaining part of the vocal tract towards detecting voice pathology.

- Variance of vocal tract tubes across the utterance is more important than mean of those tubes.

- A reasonable length of each vocal tract tube is around 7 mm for this specific task.

How to improve the performance of the proposed system in continuous speech will be a future study.

Acknowledgements

References

Kreiman, J., Gerratt, B. R., & Precoda, K. (1990). Listener experience and perception of voice quality. *Journal of Speech and Hearing Research*, 33, 103–115.

Lieberman, P. (1961). Perturbation in vocal pitch. *J. Acoustic America*, 33, 597-603.

Titze, I.R., & Liang, H. (1993). Comparison of F0 extraction methods for high-precision voice perturbation measurements. *J. Speech Hear Res*, 36, 1120-1133.

Moran, R. J., Reilly, R. B., Chazal, P., & Lacy, P. D. (2006). Telephony-based voice pathology assessment using automated speech analysis. *IEEE Trans. Biomedical Engineering*, 53 (3), pp. 468-477.

Heman-Ackah, Y. D., Heuer, R.J., Michael, D.D., Ostrowski, R., Horman, M., Baroody, M.M., Hillenbrand, J., & Sataloff, R.T. (2003). Cepstral peak prominence: a more reliable measure of dysphonia. *Ann Otol Rhinol Laryngol.*, 112 (4), pp. 324-333.

Vasilakis, M., & Stylianou, Y. (2009). Spectral jitter modeling and estimation. *Biomed. Signal Process. Control*, pp. 183-193.

Martin, D., Fitch, J., & Wolfe, V. (1995). Pathologic voice type and the acoustic prediction of severity. *Journal of Speech and Hearing Research*, 38, 765–771.

Shrivastav, R. (2003). The use of an auditory model in predicting perceptual ratings of breathy voice quality. *Journal of Voice*, 17 (4), pp. 502–512.

Little, M.A., Costello, D.A.E., & Harries, M.L. (2009). Objective dysphonia quantification in vocal fold paralysis: comparing nonlinear with classical measures. *Journal of Voice*, 25 (1), pp. 21-31.

Markaki, M., & Stylianou, Y. (2011). Voice Pathology Detection and Discrimination Based on Modulation Spectral Features. *IEEE Trans. Speech Audio Processing*, 19 (7), pp. 1938-1948.

Arias-Londono, J.D., Godino-Llorente, J.I., Saenz-Lechon, N., Osma-Ruiz, V., & Castellanos-Dominguez, G. (2011). Automatic Detection of Pathological Voices using Complexity Measures, Noise Parameters and Mel-Cepstral Coefficients. *IEEE Trans. Biomed. Eng.*, 58 (2), pp. 370-379.

Godino-Llorente, J., Gomez-Vilda, P., & Blanco-Velasco, M. (2006). Dimensionally reduction of a pathological voice quality assessment system based on Gaussian mixture models and short-term cepstral parameters. *IEEE Trans. Biomed. Eng.*, 53 (10), pp. 1943-1953.

Kay Elemetrics Corp. (1994). *Disordered Voice Database (CD-ROM), Version 1.03*. Boston, MA: Massachusetts Eye and Ear Infirmary (MEEI), Voice and Speech Lab.

Lowell, S. Y., Colton, R. H., Kelley, R. T., & Hahn, Y. C. (2011). Spectral- and cepstral-based measures during continuous speech: capacity to distinguish dysphonia and consistency within a speaker. *Journal of Voice*, 25 (5), pp. e223 - e232.

Muhammad, G., Mesallam, T.A., Almalki, K.H., Farahat, M., Mahmood, A., & Alsulaiman, M. (2012). Multi Directional Regression (MDR) Based Features for Automatic Voice Disorder Detection. *Journal of Voice*, 26 (6), pp. 817.e19-817.e27.

Parsa, V. & Jamieson, D. (2000). Identification of pathological voices using glottal noise measures. *J. Speech, Lang., Hear. Res.*, 43 (2), pp.469-485.

Markel, J.E., & Gray, A. H. (1982). *Linear Prediction of Speech*, Springer-Verlag New York, Inc. Secaucus, NJ.

Campbell, J.P., & Reynolds, D.A. (1999). Corpora for the evaluation of speaker recognition systems. *Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* '99, 2, pp. 829–832, Phoenix, USA.

Chih-Chung, C., & Chih-Jen, L. (2011). LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2 (27), 1-27. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

Kent, R. D. and Kim, Y. (2008) Acoustic Analysis of Speech, in The Handbook of Clinical Linguistics (eds M. J. Ball, M. R. Perkins, N. Müller and S. Howard), Blackwell Publishing Ltd., Oxford, UK. doi: 10.1002/9781444301007.ch22, p. 364-365.

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*, Vol. 2013, Article ID 758731. DOI: 10.1155/2013/758731

Nunes, R.B., de Souza, A., Duprat, A., Silva, M., Costa, R., & Paulino, J. (2009). Vocal tract analysis in patients with vocal fold nodules, clefts and cysts. *Braz J Otorhinolaryngo*. 75(2), pp. 188-192.