An Investigation of MDVP Parameters for Voice Pathology Detection on Three Different Databases

Ahmed Al-nasheri\textsuperscript{1}, Zulfiqar Ali\textsuperscript{1,2}, Ghulam Muhammad\textsuperscript{1}, and Mansour Alsulaiman\textsuperscript{1}

\textsuperscript{1}Digital Speech Processing Group, Department of Computer Engineering, College of Computer and Information Sciences
King Saud University, Riyadh 11543, Saudi Arabia
\textsuperscript{2}Centre for Intelligent Signal and Imaging Research (CISIR), Department of Electrical and Electronic Engineering
Universiti Teknologi PETRONAS, Tronoh 31750, Perak, Malaysia
a.alnasheri@yahoo.com, {zuali, ghulam, msuliman}@ksu.edu.sa

Abstract

In this paper, an investigation of Multi-Dimensional Voice Program (MDVP) parameters to automatically detect voice pathology in three different databases was conducted. MDVP parameters are very popular acoustic analysis among physician / clinician to detect voice pathology. The main objective of the paper is to find out the most prominent MDVP parameters irrespective to the databases used. In this study, three different databases from three distinct languages were used. The databases are Arabic voice pathology database (AVPD), Massachusetts Eye and Ear Infirmary (MEEI) (English database), and Saarbrucken Voice Database (SVD) (German database). Only the sustained vowel /a/ was used in the study. Fisher discrimination ratio (FDR) was applied to rank the parameters. Support vector machine (SVM) was used to perform the detection process. The experimental results demonstrated that there was clear difference of the performance of the MDVP parameters using these databases. The highly ranked parameters were also different from one database to another. The accuracies that achieved were varied from one database to another with the same number of MVPD parameters. The best accuracies obtained by using the three highest MDVP parameters arranged according to FDR were 99.68\%, 88.21\% and 72.53\% for SVD, MEEI and AVPD, respectively.

Index Terms: MDVP parameters, AVPD, SVD, MEEI, SVM, Fisher discrimination ratio, voice pathology detection.

1. Introduction

Voice pathologies affect the vocal folds, and these pathologies (disorder) produce irregular vibrations in the vocal folds due to the malfunctioning of the voice box. Vocal fold pathologies exhibit variations in a vibratory cycle of the vocal folds due to their incomplete closure. The voice disorder also changes the shape of the vocal tract and produces irregularities in spectral properties [1]. In addition, voice disorders affect the vocal fold vibration depending on the type of disorder and the location of the disease on the vocal folds that make them produce different basic tones.

The number of dysphonic patients having different types of voice disorders has been increased significantly. In the United States, approximately 7.5 million people have vocal difficulty [2]. It has been found that 15\% of the total visitors to the King Abdul Aziz University Hospital, Riyadh complain from a voice disorder [3]. The complications caused by a voice problem in a teaching professional are significantly greater than in a non-teaching professional. Studies revealed that, in the U.S., the prevalence of voice disorders during a lifetime is 57.7\% for teachers and 28.8\% for non-teachers [4]. Approximately, 33\% of male and female teachers in the Riyadh area suffer from voice disorders [5]. At Communication and Swallowing Disorders Unit, King Abdul Aziz University Hospital, a high volume of voice disorder cases is examined (almost 760 cases per annum) in individuals with various professional and etiological backgrounds. The use of computers to detect or identify pathological problems in speech is a non-invasive method, which is advancing with time. In the last decade, much research has been done on the automatic detection of vocal fold disorders, and these tasks continue to require further investigation due to the lack of standard automatic diagnosing approaches/equipment for voice disorders. Pathology detection is the first crucial step to correctly diagnose and manage voice disorders. The use of the objective assessment that includes acoustical analysis is independent of human bias and can assess the voice quality reliably by relating certain parameters to vocal fold behavior. On the other hand, subjective measurement of voice quality is based on individual experience, which may vary. Automatic voice pathology detection can be accomplished by various types of features, which can be obtained by the long-term and short-term signal analysis. The long-term parameters can be derived by acoustic analysis [6], [7] of speech, and the short term parameters can be calculated by linear predictive coefficients (LPC) [8], [9], linear predictive cepstral coefficients (LPCC) [10], Mel-frequency cepstral coefficients (MFCC) [11], [12] etc. Different pattern matching techniques such as Gaussian mixture model (GMM) [13], [14], hidden Markov model (HMM) [15], support vector machine (SVM) [16], artificial neural networks (ANN) [17] etc. have been used to differentiate between disorder and normal samples. Multiple long-term acoustic features, namely, pitch, shimmer, jitter, APQ (amplitude perturbation quotient), PPQ (pitch perturbation quotient), HNR (harmonic to noise ratio), NNE (normalized noise energy), VTI (voice turbulence index), SPI (soft phonation index), FATR (frequency amplitude tremor), and the glottal to noise excitation ratio (GNE) are frequently used to diagnose voice pathology (referred in [14] as [2]-[12]). Furthermore, jitter and shimmer capture the vocal fold vibration characteristics for both pathological and normal people, and both parameters are widely used for clinical and scientific diagnosis [18]. Seven acoustic parameters, including shimmer...
and jitter, are extracted by means of an iterative residual signal estimator in Rosa et al. [19], and jitter provided 54.8% pathology detection accuracy between 21 pathologies. Thirty-three different long-term acoustic parameters with their definitions, derived from Multi-Dimensional Voice Program (MDVP), are listed in Arjmandi et al. [20]. Twenty-two acoustic parameters are selected from the list, and they were extracted from the voice samples of the Massachusetts Eye and Ear Infirmary (MEEI) database. In this study, 50 dysphonic patients and 50 normal persons were used for the detection. The 22 parameters are calculated for each sample and fed to six different classifiers to compare their accuracies. Two feature reduction techniques are also used before applying classification methods. Binary classifier SVM has shown the best results compared to other classifiers. Fishers discriminative ratio (FDR) was used in [21], MFCC and six acoustic parameters jitter, shimmer, NHR, SPI, APQ and RAP (Relative Average Perturbation) are extracted, and the results are compared with the NN-based voice pathology detection system [22]. Sáenz-Lechón et al compared between their proposed parameters based on wavelet transform and some of the MDVP parameters to discriminate between pathological and normal voices [23]. To ensure the reliability of the acoustic MDVP parameters, some of these parameters were compared to the same parameters which are extracted using Praat and the result showed that there was no significant different between the two computer software [24]. Recently, MPEG-7 audio descriptors and multi-directional regression based features are used in voice pathology detection and found to have good accuracy [26, 27].

Another recent study investigates the most discriminative frequency region for voice pathology detection [28]. In general, MDVP parameters have a good ability to discriminate between normal and pathological voices like other tools that used to extract acoustic parameters such as WPCVox [25]. In this paper, the well-known MDVP parameters are used in three different databases, which are AVPD, MEEI, and SVD, to detect voice pathology. MDVP parameters are commonly used by the physician / clinician to assess the voice pathology; however, MDVP is commercial software, and it includes parameters of voice quality measurement. The rest of the paper is organized as follows: section 2 provides overview of the speech databases. Section 3 presents the methodology and tools using in this study; Section 4 gives experimental setup; Section 5 presents results and discussion, and finally, Section 6 draws some conclusions.

2. Databases

2.1. Arabic voice pathology database (AVPD)

The speech samples collected in different sessions at Communication and Swallowing Disorders Unit [4], King Abdul Aziz University Hospital, Riyadh, Saudi Arabia, by experienced phoneticians in a soundproof room using a standardized recording protocol. The database collection is one of the major tasks of the ongoing project funded by National Plans for Science and Technology (NPSST), Saudi Arabia, for the duration of two years. The protocol of the database designed such that it should avoid different shortcomings of MEEI database [24]. The project records the sustained vowels as well as the speech of patients having vocal fold disorders and the same of normal persons after clinically assessing the persons. The recording has different types of texts; three vowels with onset and offset information, isolated words containing Arabic digits and some common words, and continuous speech. The selected text covers all Arabic phonemes. All speakers record three utterances of each vowel /a/, /u/ and /i/, while, isolated words and continuous speech are recorded once to avoid making a burden on patients. The sampling frequency of the database is 50 kHz, and the speech is recorded by using Kay Pentax computerized speech lab (CSL Model 4300).

2.2. Massachusetts Eye and Ear Infirmary (MEEI) voice disorder database

This database developed by MEEI Voice and Speech Lab. It includes more than 1,400 voiced samples of sustained vowel /a/ and first 12 seconds of Rainbow Passage. It is openly commercialized by Kay Elemetrics [29]. It was recorded in two different environments. The sampling frequency for normal samples is 50 kHz while that of the pathological samples is 25 kHz or 50 kHz. It is used in most of the studies of voice pathology detection and classification.

2.3. Saarbruecken voice database (SVD)

The SVD is a freely downloadable database [30]. It has been recorded by the Institute of Phonetics of Saarland University. This database contains sustained vowels /a/, /i/ and /u/ with different intonations: normal, low, high and low-high-low and a spoken sentence in German “Guten Morgen, wie geht es Ihnen?” that mean in English “Good morning, how are you?” is included. Very few studies of voice pathology detection have been done on this database.

3. Method

Various samples of normal and three distinct voice disorders were taken from three different databases as shown in Table 1. We chose these disorders because only these disorders are common in the three databases. The numbers of male and female speakers are shown, respectively, inside the parenthesis in Table 1.

<table>
<thead>
<tr>
<th>Database</th>
<th>Normal</th>
<th>Cysts</th>
<th>Pathological</th>
<th>Polyp</th>
<th>Σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>AVPD</td>
<td>118</td>
<td>13</td>
<td>32</td>
<td>30</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>(93-25)</td>
<td>(7-6)</td>
<td>(16-16)</td>
<td>(14-16)</td>
<td></td>
</tr>
<tr>
<td>MEEI</td>
<td>53</td>
<td>10</td>
<td>70</td>
<td>15</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>(21, 32)</td>
<td>(6-4)</td>
<td>(38-32)</td>
<td>(8-7)</td>
<td></td>
</tr>
<tr>
<td>SVD</td>
<td>262</td>
<td>6</td>
<td>194</td>
<td>44</td>
<td>244</td>
</tr>
<tr>
<td></td>
<td>(137-125)</td>
<td>(5-1)</td>
<td>(121-73)</td>
<td>(19-25)</td>
<td></td>
</tr>
</tbody>
</table>

In the experiments, we only used samples for sustained vowel /a/. Each voice sample has a total of 33 MDVP acoustic parameters. The MDVP parameters of these samples were extracted using MDVP [31] kayPENTAX. This program extracts 33 parameters, but we only selected 22 parameters out of 33 parameters because the rest of the parameters do not reflect voice quality, or they are not reported for some voices. In this study, fisher discriminative ratio (FDR) was used between two classes, one from normal and the other from pathological, for the whole parameters in each database individually. The purpose of using FDR is to find which features contribute higher to detection of pathologies, and it can be calculated as shown in (1).

\[ FDR_i = \frac{(\mu_i - \mu_N)^2}{\sigma_i^2 + \sigma_N^2} \quad \text{where, } i = 1, 2, 3, ..., 22 \] (1)
where $\mu_N$ and $\mu_P$ represent the mean for classes of normal and pathological samples, respectively, while $\sigma_N$ and $\sigma_P$ represent the same but as variances, and where $j$ represents the feature number. In the experiments, the features were fed to support vector machine for the decision making whether the subject is normal or pathological. To ensure from the accuracy we got, every experiment was repeated ten times and then we took the average to report the results.

4. Experiment

For each database, four different experiments were performed with various numbers of parameters. First, we selected the 22 parameters that were used and defined in [31]. We performed the experiments with the 22 parameters from these databases individually. After that, we used FDR and sorted these parameters in descending order according to the fisher ratio score. We chose the first top 10 parameters, and performed the experiments with these parameters. Next, we chose the three top parameters from each database and performed the experiments with them. To develop a general system independent of the databases, we chose the four common parameters from the first 10 top parameters and performed the experiment with them from each database individually.

5. Results and Discussion

The results of the performed experiments for pathology detection are expressed in terms of accuracy (ACC), sensitivity (SN: the proportion of pathological samples that positively identified), specificity (SP: the proportion of normal samples that negatively identified) and area under the Receiver Operating Characteristic (ROC) curve called AUC, which are shown in Table 2. As we can see from this table, the accuracies are varied from one database to another with the same number of MDVP parameters.

The variation in the accuracies from one database to another may refer to different reasons such as (i) the severity levels of voice disorders, which are not the same in the three databases as we can notice from Table 2 how much sensitivity (that belongs to pathological samples) is varied from one database to another, (ii) the recording environment and the regulation of the recording are not the same in the three databases, (iii) in the same database the recoding environments for pathological and normal samples are not the same as in MEEI database, and (iv) the number of the taken samples from these databases in this study are not the same. In case of 22-parameters, the highest accuracy is 76.36% with MEEI database, which is comparable to the result in [32] where Arjmandi uses the same database and the same classifier, but different pathological samples. The accuracies with the two other databases, SVD and AVPD, are comparable (little less) to the accuracy obtained by MEEI. To the best of our knowledge, this is the first instance of evaluating the MDVP parameters in these two databases, and it needs more investigation. In case of 10 top parameters, the accuracy is increased by 11% with the MEEI samples, and the accuracy remained almost constant in case of SVD and AVPD. The reason behind that may be due to the top 10 parameters are not the same in all databases, and their FDR values are different too. In case of 3 top MDVP parameters, the accuracies are dramatically increased especially with the samples that were taken from SVD database. The best accuracies are 99.68%, 88.21%, and 72.53% for SVD, MEEI, and AVPD, respectively. The top three features that gave high accuracy with SVD are $vAm$ (peak amplitude variation in period-to-period), $APQ$ (amplitude perturbation quotient), and $PFR$ (phonatory fundamental frequency). $vAm$ has FDR value of 1.470 which is far greater than FDR values of 0.538 and 0.510 that belong to PFR and APQ, respectively.

Table 2. Result of different MDVP Parameters from three databases

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Databases</th>
<th>ACC%</th>
<th>SN%</th>
<th>SP%</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>22-parameters</td>
<td>AVPD</td>
<td>71.63</td>
<td>52.69</td>
<td>84.23</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>MEEI</td>
<td>76.36</td>
<td>93.22</td>
<td>45.2</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>SVD</td>
<td>72.58</td>
<td>62.17</td>
<td>82.4</td>
<td>0.72</td>
</tr>
<tr>
<td>10-parameters</td>
<td>AVPD</td>
<td>70.42</td>
<td>51.34</td>
<td>83.34</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>MEEI</td>
<td>89.71</td>
<td>92.17</td>
<td>86.27</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>SVD</td>
<td>68.52</td>
<td>51.77</td>
<td>84.3</td>
<td>0.69</td>
</tr>
<tr>
<td>3-parameters</td>
<td>AVPD</td>
<td>72.53</td>
<td>49.89</td>
<td>86.83</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>MEEI</td>
<td>88.21</td>
<td>90.63</td>
<td>84.83</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>SVD</td>
<td>99.68</td>
<td>99.75</td>
<td>99.63</td>
<td>0.99</td>
</tr>
<tr>
<td>Common</td>
<td>AVPD</td>
<td>71.16</td>
<td>43.09</td>
<td>88.85</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>MEEI</td>
<td>81.71</td>
<td>80.52</td>
<td>84.97</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>SVD</td>
<td>67.86</td>
<td>46.91</td>
<td>87.96</td>
<td>0.67</td>
</tr>
</tbody>
</table>

Figure 1: Scatter plot for the three to MDVP Parameters from SVD database.

From the scatter plot in Figure 1, we can see $vAm$ parameter has the best ability to discriminate the pathological samples more than the others two parameters, and it can be illustrated in Figure 2 that the probability density functions (pdf) of normal and pathological samples using the $vAm$ parameter have almost no overlapping. Figure 3 illustrates the ROC curve of the 3 top parameters from the three databases. It shows that the best performance is obtained with the features extracted from SVD database. One of the reasons behind that is in SVD database, the pathological samples have high severity, so it has a clear difference between normal and pathological samples. The 95% confidence interval (C.I.) is [0.9449 0.9870], and 1-tail p-value is zero (<0.05) describing the significance of the data in the two classes.
In general, the result we got is comparable to previous study in case of using MEEI database and it represents a good result if we compare our initial work on AVPD database with the initial use of SVD in the field of voice pathology detection as in [33].

From the results of all the experiments, we find the followings.

- Training with MEEI does not make the system robust, because the system is confused whether it is classifying the environments or normal / pathology.
- Training with SVD makes the system more robust than the systems trained with other database, because SVD samples are clearly distinguishable between normal and pathology.
- Training with two databases offsets some shortcomings of training with one database.
- There is a need to develop more robust features that can successfully differentiate between normal and pathological samples irrespective of the databases. Moreover, the features should be able to classify low severe pathological samples as in the case of AVPD.

### 6. Conclusion

In this work, we evaluate MDVP parameters on three different databases AVPD, MEEI, and SVD with four different experiments individually. The accuracies of the detection are varied from one database to another with the same number of MDVP parameters. The best accuracies we got are 99.68 %, 88.21%, and 72.53% for the samples that were taken from SVD, MEEI, and AVPD respectively. Some of the MDVP parameters show high ability of discrimination between normal and pathological subjects such as vAm, APQ, and PFR. In a future work, we will try to find out some features that can correctly detect voice pathologies independent of the databases.

### 7. Acknowledgment

This project was funded by the National Plan for Science, Technology and Innovation (MAARIFAH), King Abdulaziz City for Science and Technology, Kingdom of Saudi Arabia, Award Number (12-MED-2474-02).
8. References


[31] Kay Elemetrics, Multi-Dimensional Voice Program (MDVP) [Computer Program], 2012.
