An Investigation of Multidimensional Voice Program Parameters in Three Different Databases for Voice Pathology Detection and Classification

*Ahmed Al-nasheri, *Ghulam Muhammad, *Mansour Alsulaiman, *,†Zulfiqar Ali, ‡,§,¶Tamer A. Mesallam, ‡,§Mohamed Farahat, ‡,§Khalid H. Malki, and *Mohamed A. Bencherif, *‡§*Riyadh, Saudi Arabia,* †*Tronoh, Perak, Malaysia, and* ¶*Shebin Alkoum, Egypt*

Summary: Background and Objective. Automatic voice-pathology detection and classification systems may help clinicians to detect the existence of any voice pathologies and the type of pathology from which patients suffer in the early stages. The main aim of this paper is to investigate Multidimensional Voice Program (MDVP) parameters to automatically detect and classify the voice pathologies in multiple databases, and then to find out which parameters performed well in these two processes.

Materials and Methods. Samples of the sustained vowel /a/ of normal and pathological voices were extracted from three different databases, which have three voice pathologies in common. The selected databases in this study represent three distinct languages: (1) the Arabic voice pathology database; (2) the Massachusetts Eye and Ear Infirmary database (English database); and (3) the Saarbruecken Voice Database (German database). A computerized speech lab program was used to extract MDVP parameters as features, and an acoustical analysis was performed. The Fisher discrimination ratio was applied to rank the parameters. A *t* test was performed to highlight any significant differences in the means of the normal and pathological samples.

Results. The experimental results demonstrate a clear difference in the performance of the MDVP parameters using these databases. The highly ranked parameters also differed from one database to another. The best accuracies were obtained by using the three highest ranked MDVP parameters arranged according to the Fisher discrimination ratio: these accuracies were 99.68%, 88.21%, and 72.53% for the Saarbruecken Voice Database, the Massachusetts Eye and Ear Infirmary database, and the Arabic voice pathology database, respectively.

Key Words: MDVP parameters–AVPD–SVD–MEEI–SVM.

INTRODUCTION

Voice pathologies affect the vocal folds, producing irregular vibrations due to the malfunctioning of many factors contributing to vocal vibrations.[1] In addition, voice pathologies affect vocal-fold vibration differently depending on the type of disorder and the location of the disease in the vocal folds, making them produce different basic tones.

The number of dysphonic patients has increased significantly. In the United States, approximately 7.5 million people have vocal difficulties.[2] It has been found that 15% of all visitors to King Abdul Aziz University Hospital, Riyadh, complained of a voice disorder.[3] The impact of voice problems on teaching professionals is significantly greater than for nonteaching professionals. Studies revealed that, in the United States, the prevalence of voice pathologies during a lifetime is 57.7% for teachers and 28.8% for nonteachers.[4] Approximately 33% of male and female teachers in the Riyadh area suffer from voice pathologies.[5] The Communication and Swallowing Disorders Unit, King Abdul Aziz University Hospital, examines a high volume of voice disorder cases (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, a noninvasive method, is advancing over time. In the last decade, much research has been done on the automatic detection of vocal-fold pathologies, which continues to require further investigation due to the lack of standard automatic diagnostic approaches/equipment for voice pathologies. Detection of pathology is the first crucial step to correctly diagnose and manage voice pathologies. Objective assessment, including acoustical analysis, is independent of human bias and can assist clinicians in making decisions. We firmly believe that clinicians have the final decision regarding medical diagnosis; objective assessment can only be used as an assistive tool. 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 long-term and short-term signal analysis. Long-term parameters can be derived from acoustic analysis[6,7] of speech, and short-term parameters can be calculated using linear predictive coefficients,[8,9] linear predictive cepstral coefficients (LPCC),[10] Mel-frequency cepstral coefficients,[11,12] and so on.[13] Different pattern-matching techniques, such as a Gaussian mixture model,[14,15] hidden Markov model,[16] support vector machine,[17] artificial neural networks,[18] and so on, have been used to differentiate between disordered and normal samples. Multiple long-term acoustic features, namely, pitch, shimmer,

Accepted for publication March 31, 2016.

From the *Digital Speech Processing Group, Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia; †Centre for Intelligent Signal and Imaging Research (CISIR), Department of Electrical and Electronic Engineering, Universiti Tekhnologi PETRONAS, Tronoh, Perak 31750, Malaysia; ‡ENT Department, College of Medicine, King Saud University, Riyadh, Saudi Arabia; §Research Chair of Voice, Swallowing, and Communication Disorders, King Saud University, Riyadh, Saudi Arabia; and the ¶ENT Department, College of Medicine, Al-Menoufiya University, Shebin Alkoum, Egypt.

Address correspondence and reprint requests to Ghulam Muhammad, Digital Speech Processing Group, Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia. E-mail: ghulam@ksu.edu.sa

Journal of Voice, Vol. 31, No. 1, pp. 113.e9–113.e18
0892-1997

http://dx.doi.org/10.1016/j.jvoice.2016.03.019

jitter, amplitude perturbation quotient (APQ), pitch perturbation quotient (PPQ), noise-to-harmonic ratio (NHR), normalized noise energy, voice-turbulence index, soft-phonation index, frequency amplitude tremor, and glottal-to-noise excitation ratio are frequently used to diagnose voice pathology (cited in Reference 15 as References 2–12). Furthermore, jitter and shimmer capture vocal-fold vibratory characteristics for both pathological and normal people, and both parameters are widely used for clinical research purposes.[19] Seven acoustical parameters, including shimmer and jitter, are extracted by means of an iterative residual-signal estimator in Rosa et al,[20] and jitter provided 54.8% accuracy of detection for 21 pathologies. Thirty-three different long-term acoustic parameters with their definitions, derived from the Multidimensional Voice Program (MDVP),[21] are listed in Arjmandi et al.[22] Twenty-two acoustic parameters were selected from the list extracted from voice samples in the Massachusetts Eye and Ear Infirmary (MEEI) database. Fifty dysphonic patients and 50 normal persons were used for detection. The 22 parameters were calculated for each sample and fed to six different classifiers to compare their accuracies. Two feature-reduction techniques were also used before applying classification methods. The binary classifier support vector machine showed the best results compared to other classifiers, with a recognition rate of 94.26%. In Reference 23, Mel-frequency cepstral coefficient and six acoustic parameters—jitter, shimmer, NHR, soft-phonation index, APQ, and relative average perturbation (RAP)—were extracted, with the results compared to the Neural Network (NN)-based voice pathology detection system.[24] Sáenz-Lechón et al compared their proposed parameters based on wavelet transform with some of the MDVP parameters to discriminate between pathological and normal voices.[25] To ensure the reliability of the acoustic MDVP parameters, some of them were compared to the same parameters extracted using *Praat*; results showed no significant difference between the two computer software approaches.[26] Recently, MPEG-7 audio descriptors and multidirectional, regression-based features have been used in voice-pathology detection, with good accuracy.[27,28] Another recent study investigated the most discriminative frequency region for voice-pathology detection.[29] In general, MDVP parameters are well able to discriminate between normal and pathological voices, as are other tools that are used to extract acoustic parameters, such as WPCVox.[30]

In this paper, the well-known MDVP parameters are investigated in three different databases—(1) Arabic voice pathology database (AVPD); (2) MEEI;[31] and (3) Saarbruecken Voice Database (SVD)[32]—to detect and classify voice pathology. MDVP parameters are commonly used by physicians or clinicians to assess voice pathology;[33,34] however, MDVP is a commercial software. The objective of this study is to investigate the capability of MDVP parameters to detect and classify voice pathologies in a cross-database scenario and to find out which of these parameters perform best in each individual database.

MATERIAL AND METHODS

Data

In this study, we used three different databases: (1) MEEI; (2) SVD; and (3) AVPD. We chose only three types of pathological voices—(1) vocal-folds cyst; (2) unilateral vocal-fold paralysis; and (3) vocal-fold polyps—because only these pathologies are common in all three databases. We selected these three databases in our study due to the following reasons:

- MEEI database is one of the most popular databases in the field of voice pathology. It is considered as the basis of many studies with voice pathology assessment; however, it has some limitations as mentioned in subsection of MEEI database description. Therefore, for comparison purposes, we used it, but we did not solely rely on it.
- SVD is a German database that is free and downloadable with rich variation of samples. This variation of samples makes it possible to carry several types of experiments in different research purposes. It has very little use in voice pathology.
- AVPD is our Arabic developed database, and this is the first time that it is involved in research.
- Other used databases in most research are private and not available on the net.

Voiced signals can be seen in three types as qualitatively classified by Titze in Reference 35. Type 1 signals are nearly periodic, type 2 signals contain strong modulations or bifurcations, whereas type 3 signals are irregular and aperiodic. It has been suggested that traditional perturbation methods of voice signal analysis, such as jitter and shimmer, are appropriate only for type 1 or type 2 signals. For the MEEI database, some experiments are performed by excluding the type 3 signals.

The number of samples in each database is shown in Table 1, where the number of male and female speakers are shown, respectively, inside parentheses. The three used databases are each described below.

TABLE 1.
Normal and Pathological Samples From Three Different Databases

		Pathological			
Database	Normal	Cysts	Paralysis	Polyp	Total
AVPD	118 (93, 25)	13 (7, 6)	32 (16, 16)	30 (14, 16)	75
MEEI	53 (19, 34)	10 (6, 4)	66 (34, 32)	19 (8, 11)	95
SVD	262 (100, 162)	6 (1, 5)	195 (64, 131)	43 (25, 18)	244