Investigation of Voice Pathology Detection and Classification on Different Frequency Regions Using Correlation Functions

Ahmed Al-nasheri, Ghulam Muhammad, Mansour Alsulaiman, and Zulfiqar Ali, *Riyadh, Saudi Arabia*

Summary: Objectives and Background. Automatic voice pathology detection and classification systems effectively contribute to the assessment of voice disorders, which helps clinicians to detect the existence of any voice pathologies and the type of pathology from which patients suffer in the early stages. This work concentrates on developing an accurate and robust feature extraction for detecting and classifying voice pathologies by investigating different frequency bands using correlation functions. In this paper, we extracted maximum peak values and their corresponding lag values from each frame of a voiced signal by using correlation functions as features to detect and classify pathological samples. These features are investigated in different frequency bands to see the contribution of each band on the detection and classification processes.

Material and Methods. Various samples of sustained vowel /a/ of normal and pathological voices were extracted from three different databases: English, German, and Arabic. A support vector machine was used as a classifier. We also performed a *t* test to investigate the significant differences in mean of normal and pathological samples.

Results. The best achieved accuracies in both detection and classification were varied depending on the band, the correlation function, and the database. The most contributive bands in both detection and classification were between 1000 and 8000 Hz. In detection, the highest acquired accuracies when using cross-correlation were 99.809%, 90.979%, and 91.168% in the Massachusetts Eye and Ear Infirmary, Saarbruecken Voice Database, and Arabic Voice Pathology Database databases, respectively. However, in classification, the highest acquired accuracies when using cross-correlation were 99.255%, 98.941%, and 95.188% in the three databases, respectively.

Key Words: Voice pathology detection and classification–Frequency investigation–Arabic Voice Pathology Database (AVPD)–Saarbruecken Voice Database (SVD)–Massachusetts Eye and Ear Infirmary (MEEI).

INTRODUCTION

Recently, lifestyle comes with an increased risk of pathological voice problems. About 25% of the population are engaged in works that are "vocally demanding." For these individuals, either their jobs require excessive vocalization or their work environments force them to speak above a high noise level. Examples of professionals with heavy vocal demands include teachers, lawyers, auctioneers, aerobics instructors, singers, actors, and manufacturing supervisors. As a consequence, working on digital processing of speech signals was found to provide a noninvasive analytical technique that is considered to be an effective assisting tool to medical doctors when identifying voice disorders specifically in their early stages. Voice pathologies affect the vocal folds, producing irregular vibrations due to the malfunctioning of many factors contributing to vocal vibrations. Vocal fold pathologies exhibit variations in the vibratory cycle of the vocal folds due to their incomplete closure. Voice disorders also affect the shape of the vocal tract (supraglottal) and produce irregularities in spectral properties.[1] It is well known that there is no infralaryngeal (tracheobronchial tree) effect on the vocal tract during the production of a vowel if we consider that the voicing source has an infinite resistance. However, in an accurate detailed analysis, we must realize that the infralaryngeal structures do influence the vocal tract; the articulatory configuration in the vocal tract interacts with the articulation in the vocal folds.[2] Because of this interaction, supplemental vocal tract-related information is predictable to help in detecting the characteristics of the vocal folds, especially during phonation.[3] In addition, voice disorders 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. Vocal folds' vibration depends on several factors such as mucus present on the vocal folds tissue, stiffness, tension, muscles in the larynx, closing and opening of the folds, etc. These factors are affected differently for various voice pathologies. Due to the position and the size of the pathologies, vocal folds closing behaves differently during the vibration. Therefore, the vibration varies from one type of pathology to another. This vibration produces glottal source excitation frequencies, as well as affects the supraglottal (the bottom part of the vocal tract) area, which in turn contributes to the frequency of the output voice signal.

The number of patients with dysphonia has increased significantly, and in the United States alone, approximately 7.5 million people have vocal difficulties.[4] It has been found that 15% of all visitors to King Abdul Aziz University Hospital, in Riyadh, Saudi Arabia, complain of a voice disorder.[5] 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 disorders during a lifetime is 57.7% for teachers and 28.8% for nonteachers.[6] Approximately 33%

Accepted for publication January 27, 2016.

From the Digital Speech Processing Group, Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia.

Address correspondence and reprint requests to Ahmed Al-nasheri, Digital Speech Processing Group, Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia. E-mail: a.alnashari@yahoo.com

Journal of Voice, Vol. 31, No. 1, pp. 3–15
0892-1997

http://dx.doi.org/10.1016/j.jvoice.2016.01.014

of male and female teachers in the Riyadh area suffer from voice disorders.[7] The Communication and Swallowing Disorders Unit at 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 disorders, which continues to require further investigation due to the lack of standard automatic diagnostic approaches/equipment for voice disorders. Detection of pathology is the first crucial step to diagnose and manage voice disorders correctly. 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, and an 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 analyses. Long-term parameters can be derived from the acoustic analysis[8,9] of speech, and short-term parameters can be calculated using linear predictive coefficients,[10,11] linear predictive cepstral coefficients,[12] Mel-frequency cepstral coefficients,[13,14] and so on. Different pattern-matching techniques, such as a Gaussian mixture model,[15,16] hidden Markov model,[17] support vector machine (SVM),[18] artificial neural networks,[19] and so on, have been used to differentiate between disordered and normal samples. Multiple long-term acoustic features, namely pitch, shimmer, jitter, amplitude perturbation quotient, pitch perturbation quotient, harmonic-to-noise ratio, 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 (referenced in Godino-Llorente et al[16] as in other studies[4–14]). 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.[20] Seven acoustic parameters, including shimmer and jitter, are extracted by means of an iterative residual-signal estimator in Rosa et al,[21] and jitter provided 54.8% accuracy of detection for 21 pathologies. A total of 33 different long-term acoustic parameters with their definitions, derived from the *Multi-Dimensional Voice Program* (*MDVP*),[22] are listed in Arjmandi et al.[23] There were 22 acoustic parameters selected from the list extracted from voice samples in the Massachusetts Eye and Ear Infirmary (MEEI) database. A total of 50 patients with dysphonia and 50 normal persons participated for detection. The 22 parameters were calculated for each sample and fed to 6 different classifiers to compare their accuracies. Two feature-reduction techniques were also used before applying classification methods. Binary classifier SVM showed the best results compared with other classifiers, with a recognition rate of 94.26%. In Wang and Jo,[24] Mel-frequency cepstral coefficients and six acoustic parameters (jitter, shimmer, NHR, soft-phonation index, amplitude perturbation quotient, and Relative Average Perturbation) were extracted, with the results being compared with those of the Neural Network (NN)-based voice pathology detection system.[25] 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.[26] To ensure the reliability of the acoustic *MDVP* parameters, some of them were compared with the same parameters extracted using *Praat*; results showed no significant difference between the two computer software approaches.[27] Recently, MPEG-7 audio descriptors and multidirectional, regression-based features have been used in voice pathology detection, with good accuracy.[28,29] Another recent study investigated the most discriminative frequency region for voice pathology detection.[30]

Correlation functions are considered as one of the common methods for extracting various characteristics from speech signals. They are known as a domain that has certain good properties that can be used as features. The methods based on correlation function applied on a short section of voice signal can provide substantial information that enables us to estimate the vocal tract transfer function. For example, these methods result in many peak values with periodicity as the same periodicity as of the input signal. Therefore, to examine the periodicity of the signal, it is common to examine its autocorrelation function. This indicates that the correlation function of a periodic signal is also periodic. Consequently, finding pitch, fundamental frequency, etc of the signal will be possible by using these methods. In many researches, it is observed that the normal voice has more periodicity than the pathological one, and therefore performing correlation functions on these types of classes will provide an excellent allusion that can be used to discriminate between normal and pathological voices. For instance, Von Leden et al observed that the pathological samples have a strong tendency for frequent and rapid changes in the regularity.[31] In addition, Lieberman found that pathological voices tend to show unusually large cycle-to-cycle fluctuations in the fundamental period.[32] In this work, we performed different forms of correlation functions such as autocorrelation on the signal itself frame by frame, cross-correlation between two successive frames in the same signal, and cross-correlation between two successive filters frame by frame. It is preferable to use a short segment of the voice signal instead of the whole signal because the noise tends to be cancelled out in the autocorrelation process in this short segment.[33]

As we observe, every voice disorder produces different frequencies depending on the type of voice disorder and its location on the vocal fold, as we described before. Consequently, observing the frequency band is very important to see which frequency band contributes more to the detection and classification of voice disorders. For instance, in Pouchoulin et al,[34] the authors found that the lower frequencies between 0 and 3000 Hz are more suitable for discriminating dysphonic voices than the higher frequencies. In addition, Fraile et al[35] found that the power in bands between 2000 and 6400 Hz is significantly less stable in dysphonic voices.

In this paper, we mainly focus on developing a computationally less expensive method for voice pathology detection. Specifically, we concentrate on extracting a feature set having low dimension. In the proposed method, the input voice is passed through a bank of band pass filters, and each filter output is