Pakistan Journal of Science (Vol. 73 No. 1 March, 2021)

PROSPECTIVE ARABIC LANGUAGE DATA BASE FOR IDENTIFICATION AND RECOGNITION OF VOICE DISORDER

S. A. Syed[*1], M. Rashid[2] and H. Zahid[3]

[1,2]Department of Biomedical Engineering and Electrical Engineering Ziauddin University Faculty of Engineering Science Technology and Management Karachi, Pakistan.
[3]Department of Electrical Engineering and Software Engineering Ziauddin University Faculty of Engineering Science Technology and Management Karachi, Pakistan.
[*]Corresponding Author: Email: sidra.agha@zu.edu.pk

ABSTRACT: Speech is the underlying strong characteristic and voice is the vibration that the air creates as it's pushed out of the lungs and going into the vocal cords. Vocal cords are two folds of tissue within the larynx, sometimes called a speech box, the sound of these strings is what gives rise to speech.Anything that interferes with vocal cord movement or contact can cause a voice disorder. Speech production may often be compromised as social stressors contribute to chronic aphonia or dysphonia.Machine learning algorithm and non-invasive systems may play a major role in the early detection and in tracking and even development of efficient pathological speech diagnosis, based on a computerized acoustic analysis. A strong voice impairment dataset will help in recognition and classification ofincreasing number of voice disorders in and outside the Arab area. The database protocols for AVPD (Arabic voice pathology database) has been developed in a way to prevent previous data base deficiencies in widely used databases like MEEI (Massachusetts eye and ear infirmary) and SVD (Saarbrücken voice database).

Keywords: Acoustic analysis, Clinical voice pathology, Voice disorder, AVPD, machine learning technique.

(Received 05.11.2020 Accepted 30.12.2020)

INTRODUCTION

In recent years, the number of patients with speech pathology has dramatically risen, with about 17.9 million vocal problems in the U.S. alone (Bhattacharyya, 2014). Voice disorders are something in description which distinguishes from voices of similar ages, gender and social groupings in terms of quality, pitch, loudness, and / or versatility (Dejonckere, 2001). Significant indirect costs of speech-related work, short - term demands and projections of costs of primary health care approximately $5 billion a year (Bradley, 2010). Dysphonia happens when the consistency, loudness and pitch of the voice are affected. Around 10% of the population (Martin, 2016) suffer from this condition, primarily because of bad social behaviors. Hoarseness describes an abnormal, harsh, breathy, weak or strained voice quality.

Vocal production of the voice may be specified by fundamental frequency, intensity, vibration and vocal intonation according to its vocal parameters. Some indicators of social or emotive expression, indicates changes of voice (Ainsworth, 1992). Voice disorders manifest in various ways, including the presence of sensory and auditory symptoms, deviations in vocal quality and that may involve behavioral and/or organic factors (Baker, 2008). Patients with voice disorders may experience various symptoms, of which hoarseness, sore throat, vocal fatigue, and throat clearing which may be associated with intense voice use, upper respiratory tract infections, stress, and smoking (Kasama, 2007). Because manifestation of a voice disorder is multidimensional, its assessment must include a variety of factors, including perceptual voice assessment, visual laryngeal inspection, acoustic analysis, aerodynamic assessment, and vocal self-assessment (Ferreira, 2009). Voice pathology disorders can be detected using the classification tools for computer helped voice pathology, which gives promising outcomes. The development of specialist networks and decision support technologies for medical applications has led to the recent advancement in the field of artificial intelligence, which have the ability to be good medical devices. Classification systems may contribute to the increase in precision, accuracy and reliability of diagnosis (Verde, 2018). Three commonly used datasets for recognition and classification of Voice Disorder pathologies are AVPD (Arabic Voice Pathology database) , MEEI (Massachusetts Eye \& Ear Infirmary) database and SVD (Saarbrücken Voice database) and out of these three databases AVPD has been explored less by the scientists as compared to the other two (Sidra, 2020).

We shall evaluate through our study to analyse the quality,strengths and worth of AVPD database for future researchers for the purpose of classification and recognition of voice disorders.

MATERIALS AND METHODS

We have tried to study in depth the parameters and contents of AVPD in comparison of MEEI and SVD. Formulation of AVPD has overcame the deficiencies of previous databases of voice disorders (Mesallam 2017). The Arab voice pathology (AVPD) database may have a likely effect on voice disorders diagnoses in the Arab region. It was found that 15% of all visitors to King Abdulaziz University Hospital in Saudi Arabia have been identified to be complaining of speech impairment. This AVPD database that is used in this review is Arabic voice pathology database (AVPD) (Al-Nasheri, 2017). Samples of words and voices were recorded at various sessions in King Abdul Aziz University Hospital in Riyadh, Saudi Arabia, Communication & Swallowing Disorders Unit. In a sound treatment room, a standard recording protocol was used to collect voices of the patient by experienced phoneticists. The database protocol has been developed to prevent specific MEEI data base deficiencies. The AVPD contains recordings of long - lasting vowels and vocal folding diseases, together with the same records of persons who have normal speech. After the scientifically tested use of a laryngeal stroboscope, normal and pathological vocal folds were established. In the case of pathology, a 1–3 scale was issued to measure the perceptual extent of voice disorders, 3 being the most severe. Each sample was associated with a gravity rating based on a panel of three medical experts. The text types are different: (1) three long - lasting vowels with information in the beginning and the offset; (2) single, Arabic, and some common words; and (3) continuous speech. In all Arabic phonemes, the chosen text was carefully selected. Three utterances of each vowels have been recorded by most speakers: /a/, /u/ and /i/. Only once have been reported single words and constant speech to avoid overloading patients. The test frequency is 50 kHz for all regular and disease samples collected in AVPD (Sáenz-Lechón, 2006).

During the development of the AVPD, different shortcomings of the SVD (Saarbruecken Voice Database, n.d) and MEEI (Al-Nasheri, 2017) databases were avoided. For eg, the SVD and MEEI databases only include the phonation portion, whereas De Krom (Krom, 1944) proposed that full documentation of a vowel, including onset and offset sections, would offer more acoustic knowledge than only long term phonation. Furthermore, lasting seriousness plays an essential part in the pathology evaluation and can never be contained in the SVD or MEEI databases. An uncertainty matrix offers details on real and wrongly categorized topics in an automated disorder identification framework. This matrix will evaluate the cause for error in terms of perceptual magnitude. Often computer systems cannot distinguish between usual or highly serious pathological topics. This is also why in the AVPD, perceptual magnitude is considered and graded across the 1 to 3 spectrum, where 3 is nextremely serious speech disruption. In addition, after the clinical assessment the usual subjects in the AVPD are reported under the same situation as those in pathology subjects. Standard participants have not been clinically tested in the MEEI database, but they have no history of speech problem (Parsa & Jamieson 2000). No such details are given in the SVD database. The AVPD balance the number of ordinary and disease-related people. 51 percent of the overall AVPD topics are ordinary topics. Instead, in the MEEI and SVD datasets, the number of average people was 7% and 33%, respectively. Compared with abnormal topics, the amount of regular topics in the MEEI collection is troubling. The MEEI archive comprises 7% and 93% of usual and pathological samples respectively. An automated condition identification device focused on the MEEI database may also be biased and cannot produce accurate results.

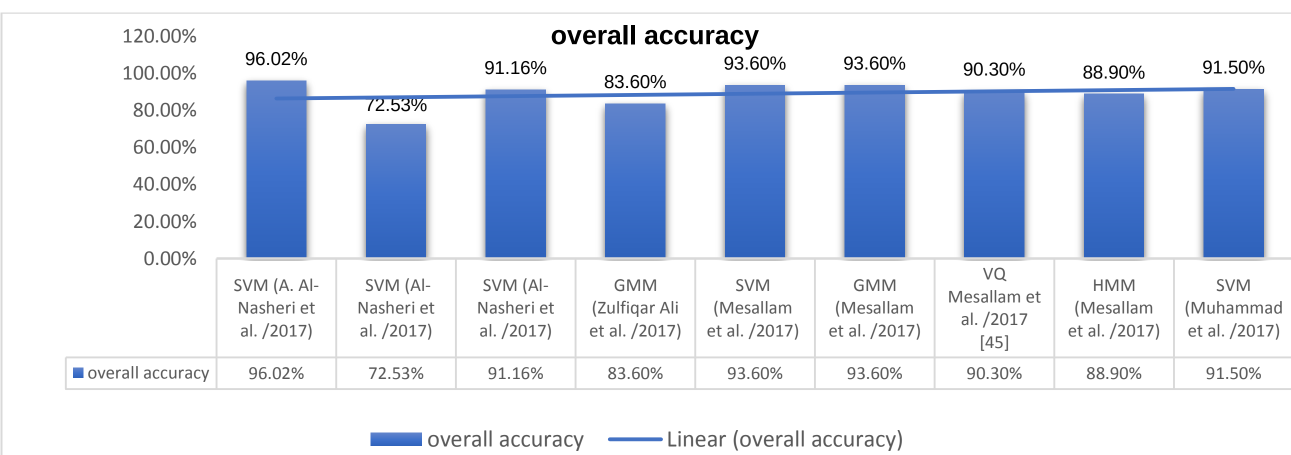

Figure 1. SVM = Support Vector Machine, GMM = Gaussian Mixture Model, VQ = Vector Quantizer, HMM = Hidden Markov Model. Graph of accuracies calculated by different researchers by applied various machine learning algorithm on AVPD dataset.

RESULTS

Figure 1 exhibits that only few researchs have utilized AVPD and from this figure the trend in accuracies cannot be denied that the researchers obtained good accuracies using different machine learning approaches which shows the potential of this database due to its better pre signal processing environment. Al-Nasheri in his all three studies used SVM for as a classifier but with different features for recognition of voice disorder patterns and except MVDP (Multi dimensional Voice Programming) parameters both peak, log entropy and eight frequency band with correlational features showed good accuracies. Whereas Mesallam *et al*. in 2017 used MFCC (Mel Frequency Cepstral Coefficient) feature with four different classifiers in his study and all of them has accuracies above 85% which is considered as good reported outcome in machine learning terms. Since SVM is a world reknown classifier and prvides high seperability of data therefore it is mostly utilized by the researchers.

Zulfiqar in his study have used GMM (which is a probabilistic model that learns the subpopulation in an automatic manner) at AVPD for recognition and classification of voice pathologies which has resulted in 83.6% accuracy.The results of this study proved that common speech features does not necessarily correlate with the voice therefore they are unreliable for the purpose of voice pathology detection. Messalam also used GMM at AVPD and SVD for comparative analysis and reported an accuracy of 93.3%.

Messalam also practiced Vector Quantization method at AVPD and reported 90.3% accuracy through the lossy compression of voice data and presiction of voice pathology.

DISCUSSION

A normal voice is essentially in quality and allows sufficient communication and unnecessary effort or inconvenience It is not deniable that voice problems are linked to negative effects on quality of life. In what is perceived as a' normal voice' there is a huge variation. It is problematic to determine its essential properties because a continuum exists between a normal and a disordered voice. The perceptional correlates of frequency in voice are known as pitch or subjective level sensations that are appropriate for age and sex and are known as loudness or subjective noise sensations that are suitable for the environment. There are however no universal criteria to determine the characteristics and limits of a normal voice but these few parametrs are key checks to determine voice quality. In voice pathology evaluation primarily databases may be used for testing and developing state of art algorithms for recognition and classification of voice disorders, The significance of using data base for this purpose increases as database once made can be used in multiple studies without the hassle involvement of of ethical issues since it is recorderd. In voice databases perceptual severity has a major role to play, which either in SVD or MEEI repositories is not accessible. A confusion matrix may provide information on honestly and categorized topics in an automated disturbance detection system on the basis of outliers and how many times the system failed to detect the correct voice pathology. The cause for misclassification can be calculated by the perceptual severity of this structure. Automatic systems can at times not differentiate between typical abnormal subjects and relatively severe ones. This is why the perceptive severity in the AVPD is also taken into account in grades 1 - 3, in which 3 is a highly severe speech disorders. In comparison the typical AVPD participants are reported in the same state as those used for the pathological subjects following the clinical assessment (Tamer *et al*. 2017). A clinical examination of standard MEEI topics is not conducted although the history of the speech problem is incomplete (Parsa, 2000). No such information is provided in the SVD database. In AVPD, according to the MEEI database, all normal and pathological specimens are recorded at a single AVPD sampling frequency. Deliyski *et al*. concluded that the precision and the efficiency of the acoustic analysis is affected by the frequency of the sampling (Deliyski *et al*. 2005). . However, there is a vowel in the MEEI database and three vowels are registered in the AVPD. While three vows are also recorded in the SVD, they are only reported once. In the AVPD, three vowels are repeatedly reported, as some studies have suggested to model the intraspeaker variability for more than one single sample of the same vowel. The total length of the reported study, that is 60 seconds, is another important feature of the AVPD. By regular as well as disordered individuals any text reported in an AVPD is of the same duration. Between normal and pathologic topics, the recording times in the MEEI database vary. In comparison, the connected language (sentence) duration in the SVD database is only 2 seconds, which is not enough to build an automatic speech detection system. In addition, the SVD database cannot be used for a text - independent system. The AVPD is 18 seconds long on average and comprises seven sentences. The length of Al-Fateha speech is 18 seconds and it is segmented into two components to develop text - independent structures (Tamer *et al*. 2017). The capturing of independent words is another special feature of the AVPD, other than its perceptual momentum. There are 14 isolated terms in the AVPD: Arabic (zero to ten) and three common ones. In the missing Arabic letters typical words were registered. Thus, each Arabic letter is included in the AVPD. Inclusion of all Arabic letters in this database not only itself increases the acceptability of this dataset for voice pathology detection as well opens a wide room for Artificial Inteleegence techniques to be applied for achievement of promising results.

Conclusion: Natural voice is the pulmonary air pulse acoustic substance communicating with the larynx, which adduces actual vocal folds to create recurrent and / or seasonal vibrations and it is a challenging task to produce a set of computer processed voices (data base), particularly of disordered voices, due to special requirements of sampling frequency and pre processing techniques. From available voice disorder databases as discussed above in detail the potential and worth of AVPD in determination of voice disorders , Different speech recognition software can be created by AVPD for dysphonic patients espacially due to the isolated terminology and cannot be generated through means of SVD and MEEI databases.

It is recommended for future research to utilize this database for application of multiple machine learning and artificial intelligence techniques so that user friendly softwares may be developed for recognition and classification of voice disorders to aid the patient especially children with voice disorder, their families, Speech Language pathologists and Supporting hospital staff. Developing a speech recognition device can determine how exact the vocabulary of a voice-disordered patient is or has changed during therapy may have a major prognostic benefit in voice-disorder management.

REFERENCES

Ainsworth, W. A., (1992), Clinical Voice Disorders: An Interdisciplinary Approach, 3rd edn. By Arnold E. Aronson.Pp. 394. Thieme, 1990. DM 78.00 hardback. ISBN 3 13 598803 1. Experimental Physiology, 77 doi: 10.1113/expphysiol.1998.sp004249.

Al-Nasheri *et al*. (2017), "An Investigation of Multidimensional Voice Program Parameters in Three Different Databases for Voice Pathology Detection and Classification," (in eng), J Voice, vol. 31, no. 1, pp. 113.e9-113.e18, doi: 10.1016/j.jvoice.2016.03.019.

Al-Nasheri, A., Muhammad, G., Alsulaiman, M., Ali, Z., Malki, K. H., Mesallam, T. A., & Farahat Ibrahim, M. (2017). Voice Pathology Detection and Classification Using Auto-Correlation and Entropy Features in Different Frequency Regions. IEEE Access, 6, 6961-6974. https://doi.org/10.1109/ACCESS.2017.2696056

Al-Nasheri, G. Muhammad, M. Alsulaiman, and Z. Ali (2017), "Investigation of Voice Pathology Detection and Classification on Different Frequency Regions Using Correlation Functions," (in eng), J Voice, vol. 31, no. 1, pp. 3-15, Jan 2017, doi: 10.1016/j.jvoice.2016.01.014.

Baker J. (2008), The role of psychogenic and psychosocial factors in the development of functional voice disorders. J Speech Lang Pathol; 10(4):210-30. http://dx.doi.org/ 10.1080/17549500701879661. PMid:20840038.

Bhattacharyya, N. (2014), the prevalence of voice problems among adults in the United States. The Laryngoscope, 124: 2359-2362. Doi: 10.1002/lary.24740

Bradley P. (2010) Voice Disorders: Classification. In: Annika M., Bernal-Sprekelsen M., Bonkowsky V., Bradley P., Iurato S. (Eds) Otorhinolaryngology, Head and Neck Surgery. European Manual of Medicine. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-68940-9_60

Dejonckere, P., Bradley, P., Clemente, P. *et al*. (2001). A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques. European Archives of Oto-Rhino-Laryngology 258, 77–82 . https://doi.org/10.1007/s004050000299

Dimitar D. Deliyski, Heather S. Shaw & Maegan K. Evans (2005) Influence of sampling rate on accuracy and reliability of acoustic voice analysis, Logopedics Phoniatrics Vocology, 30:2, 55-62, DOI: 10.1080/1401543051006721

Ferreira, Léslie Piccolotto; SANTOS, Janine Galvão dos and LIMA, Maria Fabiana Bonfim de (2009). Sintoma vocal e sua provável causa: levantamento de dados em uma população. Rev. CEFAC [online], vol.11, n.1, pp.110-118. ISSN 1982-0216. https://doi.org/10.1590/S1516-18462009000100015.

G. Muhammad *et al*. (2017), "Voice pathology detection using interlaced derivative pattern on glottal source excitation," Biomedical Signal Processing and Control, vol. 31, pp. 156-164, doi: https://doi.org/10.1016/j.bspc.2016.08.002.

Kasama, S. T., & Brasolotto, A. G. (2007). Percepção vocal e qualidade de vida [Vocal perception and life quality]. Pro-fono: revista de atualizacao cientifica, 19(1), 19–28. https://doi.org/10.1590/ s0104-56872007000100003

Krom, G. (1994). Consistency and Reliability of Voice Quality Ratings for Different Types of Speech Fragments. Journal of Speech, Language, and Hearing Research, 37(5), 985-1000. https://doi.org/10.1044/ jshr.3705.985

L. Verde, G. De Pietro and G. Sannino, "Voice Disorder Identification by Using Machine Learning Techniques," in IEEE Access, vol. 6, pp. 16246-16255, 2018, doi: 10.1109/ACCESS.2018.2816338.

Martins, R. H., do Amaral, H. A., Tavares, E. L., Martins, M. G., Gonçalves, T. M., & Dias, N. H. (2016). Voice Disorders: Etiology and Diagnosis. Journal of voice: official journal of the Voice Foundation, 30(6), 761.e1–761.e9. https://doi.org/10.1016/j.jvoice.2015.09.017

N. Sáenz-Lechón, J. I. Godino-Llorente, V. Osma-Ruiz, and P. Gómez-Vilda (2006), "Methodological issues in the development of automatic systems for voice pathology detection," Biomedical Signal Processing and Control, vol. 1, no. 2, pp. 120-128, doi: https://doi.org/10.1016/j.bspc.2006.06.003.

Parsa, V., & Jamieson, D. (2000). Identification of Pathological Voices Using Glottal Noise Measures. Journal of Speech, Language, and Hearing Research, 43(2), 469-485. https://doi.org/10.1044/jslhr.4302.469

Saarbruecken Voice Database - Handbook", Stimmdatenbank.coli.uni-saarland.de. [Online]. Available: http://www.stimmdatenbank.coli.uni-saarland.de/help_en.php4. [Accessed: 18- Jul-2020].

Sidra Abid Syed, Munaf Rashid, Samreen Hussain. Meta-analysis of voice disorders databases and applied machine learning techniques[J]. Mathematical Biosciences and Engineering, 2020, 17(6): 7958-7979. doi: 10.3934/mbe.2020404

Tamer A. Mesallam, Mohamed Farahat, Khalid H. Malki, Mansour Alsulaiman, Zulfiqar Ali, Ahmed Al-nasheri, Ghulam Muhammad, (2017), "Development of the Arabic Voice Pathology Database and Its Evaluation by Using Speech Features and Machine Learning Algorithms", Journal of Healthcare Engineering, vol. 2017, Article ID 8783751, 13 pages. https://doi.org/10.1155/2017/8783751

Z. Ali *et al*., (2017). "Intra- and Inter-database Study for Arabic, English, and German Databases: Do Conventional Speech Features Detect Voice Pathology?" Journal of Voice, vol. 31, no. 3, pp. 386.e1-386.e8, doi: https://doi.org/10.1016/j.jvoice.2016.09.009.