7. Conclusions

The present framework for reliable voice activity detection is focusing on sustained vowels and continuous speech with babble noise. The framework is supported by a two-step noise reduction method of the Wiener and SVM classification. The implementation of the entire framework using an open-source Linux platform and GNU Octave is cost-effective. Experimental investigations are carried out using NOIZEUS continuous speech corpus with babble noise. The results revealed that the methods involving linear and nonlinear features, m_1, m_5, and m_6 appear to be more robust in VAD using SVM with an accuracy around 93%. For sustained vowels, the laryngeal pathologies data from SVD and laryngeal cancer data depicted that, the time-domain based VAD methods having accuracy of around 80% with better TP rate, specificity, precision, AUC and F-score. SVM prediction rate is found to be better in the methods m6 (ideal 63.9% and predicted 67.9%) along with much reduced FP-rate of 11.59% and a bit higher FP-rate in method 7. In summary, the combination of biologically inspired features like MFCC and nonlinear features has proved to be the emphasis in the reliable VAD and further assessment of signal to noise ratio, which is useful in speaker verification. Because the dynamic behaviour of vocal folds is more prominently revealed by the consideration of phase in nonlinear features in the VAD methods m_6 and m_7. This VAD forms an integral part of the pre-processing stage of speech involving continuous and sustained vowels affected with babble noise in case of pathological voice analysis and detection. There is more scope in the development of VAD by an in-depth analysis of nonlinear methods and higher-order machine learning methods with the involvement of phase and magnitude.

References

[1] R. Amir Hossein Poorjam, Max A. Little, Jesper Rindom Jensen and Mads Græsboll Christensen, (2018),"A Supervised Approach to Global Signal-to-Noise Ratio Estimation for Whispered and Pathological Voices", 978-1-5386-4658-8/18/2018 IEEE, ICASSP.Vol. 1, No. 1, pp. 123-456, 2009.

[2] Cyril Plapous, Claude Marro, and Pascal Scalart, "Improved Signal-to-Noise Ratio Estimation for Speech Enhancement", *IEEE Transactions on Audio, Speech and Language Processing.*2006.

[3] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin, "*A Practical Guide to Support Vector Classification*, Department of Computer Science", National Taiwan University, Taipei 106, Taiwan, May 2006.

[4] Carlos M. Travieso, Jesús B. Alonso, J.R. Orozco-Arroyave, J.F. Vargas-Bonilla, E. Nöth, Antonio G. Ravelo-García, "Detection of different voice diseases based on the nonlinear Characterization of speech signals", *Expert Systems with Applications* (pp. 184–195), 2017.

[5] Dang Minh Cong, (2015), "*Noise Reduction In Speech Enhancement By Spectral Subtraction With Scalar Kalman Filter"*, Ha Noi – 2015.

[6] Broder, Gerard Chollet, Anna Esposito, Marcos Faundez-Zanuy, Maria Marinaro, "*Non Linear Speech Modeling and Applications: Advanced Lectures, and revised selected papers"*, Springer, May, 2005.

[7] Chakrabarti, Hamzeh Ghasemzadeh, Mehdi Tajik Khass, Meisam Khalil Arjmandi, Mohammad Pooyan, "Detection of vocal disorders based on phase space parameters and Lyapunov spectrum", *Biomedical Signal Processing and Control*, *Elsevier*-2015.

[8] Lasonas Kokkinos, Student Member, Petros Maragos, "Nonlinear Speech Analysis Using Models for Chaotic Systems", *IEEE Transactions On Speech and Audio Processing*, Vol. 13, No. 6, Nov, 2005.

[9] In-Chul Yoo, Hyeontaek Lim, Dongsuk Yook, "Formant-Based Robust Voice Activity Detection", *IEEE/ACM Transactions On Audio, Speech, and Language Processing*, Vol. 23, No. 12, Dec, 2015.

[10] Jack J. Jiang, Yu Zhang, and Clancy McGilligan, "Chaos in Voice, From Modeling to Measurement", *Journal of Voice*, Vol. 20, No. 1, pp. 2–17, 2006.

[11] Jun Tae Kim, Sung Hoon Jung, Kwang-Hyun Cho, "Efficient harmonic peak detection of Vowel sounds for enhanced voice activity detection", *IET Signal Process*, Vol. 12 Iss. 8, pp. 975- 982, 2018

[12] Lavneet Singh, Girija Chetty and Savleen Singh, "A Novel Algorithm Using MFCC and ERB Gammatone Filters in Speech Recognition", *Journal of Information Systems and Communication*, ISSN: 0976-8742 & E-ISSN: 0976-8750, Volume 3, Issue 1, pp. -358-364, 2012.

[13] M. H. Moattar and M. M. Homayounpour,"A Simple But Efficient Real-Time Voice Activity Detection Algorithm", *17th European Signal Processing Conference* (EUSIPCO 2009) Glasgow, Scotland, Aug, 2009.

[14] Md. Moinuddin, Arunkumar N. Kanthi, "Speaker Identification based on GFCC using GMM"", *International Journal of Innovative Research in Advanced Engineering* ISSN: 2349-2163, Volume 1, Issue 8, 2014.

[15] Nitin N Lokhande, Navnath S Nehe, Navnath S Nehe, "Voice Activity Detection Algorithm for Speech Recognition Applications", *International Conference on Computational Intelligence* (ICCIA), 2011.

[16] Pham Chau Khoa, "*Noise Robust Voice Activity Detection"*, School of Computer Engineering, 2012.

[17] Rainer Hegger, Holger Kantz, Thomas Schreiber, *"Practical implementation of nonlinear time series methods: The TISEAN package"*.

[18] Simon Graf, Tobias Herbig, Markus Buck and Gerhard Schmidt, "Features for voice activity detection: a comparative analysis", *EURASIP Journal on Advances in Signal Processing* 2015:91, DOI 10.1186/s13634-015-0277-z.

[19] W. J. Barry and M. Putzer, "*Saarbrucken Voice Database",* Institute of Phonetics, Univ. of Saarland, http://www.stimmdatenbank.coli.uni-saarland.de/.

[20] Wilson Burgos, "*Gammatone and MFCC Features In Speaker Recognition"*, Melbourne, Florida-2014.

[21] W. Q. Ong and A. W. C. Tan," Robust Voice Activity Detection Using Gammatone Filtering and Entropy", *International Conference on Robotics, Automation and Sciences* (ICORAS)-2016.

[22] Yan Zhang, Zhen-min Tang,Yan-ping Li,Yang Luo, "A Hierarchical Framework Approach for Voice Activity Detection and Speech Enhancement, Hindawi Publishing Corporation, *Scientific World Journal*, Volume 2014.

[23] Youngmoon Jung, Younggwan Kim, Hyungjun Lim, Hoirin Kim,"Linear-Scale Filterbank for Deep Neural Network-Based Voice Activity Detection", *Conference of The Oriental Chapter of International Committee for Coordination and Standardization of Speech Databases and Assessment Technique* (O-COCOSDA) Seoul, Korea, 2017.

[24] Zulfiqar Ali, Muhammad Talha, "Innovative Method for Unsupervised Voice Activity Detection and Classification of Audio Segments", *IEEE Access*, Volume 6, 2018.