International Journal of Nanotechnology

ISSN online: 1741-8151 - ISSN print: 1475-7435
https://www.inderscience.com/ijnt

Unsupervised voice activity detection with improved signal-to-noise ratio in noisy environment

Shilpa Sharma, Rahul Malhotra, Anurag Sharma, Jeevan Bala, Punam Rattan, Sheveta Vashisht

DOI: 10.1504/IJNT.2023.10056477

Article History:

Received:	04 January 2022
Last revised:	24 March 2022
Accepted:	24 March 2022
Published online:	31 May 2023

Unsupervised voice activity detection with improved signal-to-noise ratio in noisy environment

Shilpa Sharma*

Computer Science and Engineering,
CT Group of Institutions,
Jalandhar, India

and

Lovely Professional University,
144411, India
Email: shilpa13891@gmail.com

Rahul Malhotra

Department of Electronics Communication Engineering,
CT Group of Institutions,
Jalandhar, 144020, India
Email: blessurahul@gmail.com

Anurag Sharma*

Faculty of Engineering, Design and Automation,
GNA University,
Phagwara, 144401, India
Email: er.anurags@gmail.com
*Corresponding authors

Jeevan Bala

Department of Computer Science and Engineering,
Lovely Professional University,
Phagwara, 144411, India
Email: jeevan.26699@lpu.co.in

Punam Rattan

Department of Computer Application,
Lovely Professional University,
Phagwara, 144411, India
Email: punamrattan@gmail.com

Sheveta Vashisht

Department of Computer Science and Engineering,
Lovely Professional University,
Phagwara, 144411, India
Email: sheveta.16856@lpu.co.in

Abstract: To identify voiced and unvoiced signals, this research provides an extended voice characteristic detection strategy for noisy settings that uses feature extraction and unvoiced feature normalisation. In a high signal to noise ratio environment, the proposed method develops a recognition model by recovering characteristics for categorisation of spoken and unvoiced signals. The novelty of this method is that it uses feature extraction to classify voiced and unvoiced signals with a higher signal-to-noise ratio (SNR). Furthermore, by combining two classifiers in a hybrid model, the model is less affected by noise for speech features, and identification performance improves. The model was tested for its ability to increase recognition accuracy. The proposed method produces better results than existing methods, with an accuracy of 99.73% and SNR of 25.61 dB. The proposed model LFV-KANN also handles increases in noise power efficiently through the hybridisation of two classifiers: artificial neural network (ANN) and K-means clustering.

Keywords: TIMIT dataset; support vector machine; voice activity detector; unsupervised learning.

Reference to this paper should be made as follows: Sharma, S., Malhotra, R., Sharma, A., Bala, J., Rattan, P. and Vashisht, S. (2023) 'Unsupervised voice activity detection with improved signal-to-noise ratio in noisy environment', *Int. J. Nanotechnol.*, Vol. 20, Nos. 1/2/3/4, pp.421–432.

Biographical notes: Shilpa Sharma received her BTech in Computer Science from Lovely Institute Technology in 2009 and MTech in computer Science Engineering from DAVIT, Punjab, India. She has 12 years of experience as an Assistant Professor. Her area of interests includes data mining, soft computing, computer networking and artificial intelligence. She has published various articles in national/international conferences and journals. She has guided more than 10 MTech students. She published around 20 research papers in national, international journals and, conferences.

Rahul Malhotra obtained his UG and PG degree in Electronics and Communication Engineering and received PhD degree from IKG PTU, Jalandhar, Punjab. He is currently working as Professor in CT Group of Institutions, Jalandhar, India. His research areas include image processing, IoT, signal processing, machine learning, text mining. He has more than 18 years of teaching experience. He has published various articles in national/international conferences and journals. He has guided more than 95 MTech students and seven research scholars. He published around 100 research papers in national, international journals and conferences.

Anurag Sharma obtained his UG and PG degree in Electronics and Communication Engineering and received PhD degree from National Institute of Technology, Jalandhar, India in 2019. He is currently working as Professor in the Faculty of Engineering and Design Automation, GNA University, Phagwara, India. He has more than 18 years of teaching experience. He has published various articles in national/international conferences and journals. He has vast research and industrial experience in the fields of biomedical engineering, data sciences, image processing, machine learning. He has guided

more than 90 MTech students and more than five Research Scholars. He published around 100 research papers in national, international journals and conferences.

Jeevan Bala is an Assistant Professor at Lovely Professional University. She teaches graduate, postgraduate and doctoral students Computer Science and Engineering subjects and guide them on research projects, academic matters and career decisions. She holds a PhD in meta-heuristic techniques, Machine Learning. She did her MTech in Bioinformatics. She hold research experience including international exposure, research publications in reputed (WoS/Scopus) journals with high impact factors, and experience of working with research writing and technical tools like LaTex and MATLAB. Her research interest includes digital image processing, machine learning, meta-heuristic techniques, and bioinformatics.

Punam Rattan is an experienced Associate Professor with a demonstrated history of working in the higher education industry. Strong administrative professional with a PhD focused in Computer Application from IKG Punjab Technical University. Head of School of Engineering and Technology and School of Computer Applications & Information technology. Her research interest includes database, datamining.

Sheveta Vashisht received her BTech in Computer Science from Lovely Institute Technology in 2009 and MTech in computer Science Engineering from LPU, Punjab, India. She has 10 years of experience as an Assistant Professor. Her area of interests includes data mining, soft computing, cyber crime. She has published various articles in national/international conferences and journals. She has guided more than 10 MTech students.

1 Introduction

A voice framework aids in identifying key factors that affect system performance and is used in conjunction with a variety of application areas in the cell phone voice chat environment. It is applied to a speech machine (such as a voice activation or noise – cancelling algorithm) and is designed with key factors affecting system performance. The efficacy of a noise-cancelling algorithm, which is a method for reducing noise by estimating noise signals, is dependent on the success of a voice recognition system. The proposed model is able to determine the optimum window length and window overlapping size. To improve the results of the recommended model, four feature extraction approaches were used to evaluate the audio segment. The main disadvantage of the supervised voice activity detection (VAD) approach is that it necessitates a large number of training datasets, which might result in mismatches between the trained and untrained training dataset. The detecting of the voice section is intimately related to recognition performance in the domains of speech coding, voice improvement, and voice recognition. In a voice recognition enhancement system, detecting the voice segment is crucial since it affects the accurate noise section estimation. The most frequent classical noise cancellation method in a speech improvement approach is frequency reduction, which deducts the noise signal paired with the speech signal using a noise signal estimates in the spectral domain. Gelly and and Gauvain [1] developed frequency reduction, which is simple to use, has benefits, and has a high recognition rate.

Based on speech activity recognition, there are many methods to obtain the mean of the noise in various sections where the speech segment is absent for general noise estimate approaches. The estimation of a noise signal has a significant impact on the identification accuracy of voice improvement systems, and when the prediction of the noise signal is inadequate, an unwanted voice is heard due to the background noise. However, if the noisy signals estimation is large, the speech signal's intelligibility suffers as a result of the loss of frames containing primary sounds, because the prediction does not really distinguish among low-energy initial low energy speech and noises. A voice recognition algorithm can distinguish between speech and non-speech signals using a decision rule that applies an appropriate signal to the feature parameter to identify the speech and noises signals [2]. Non-speech feature normalisation is a technique for detecting noise and voiced signals, as well as feature parameters in a variety of formats, including zero crossing rate, spectral energy, likelihood ratio based on the spatial model, linear prediction coefficient, and so on [3]. The existing technique suffers from degradation of identification capacity because classification accuracy for noisy and voiced signals falls when the energy level of the silent segment grows with a low signal to noise ratio. As a result, this research provides an effective voice features recognition strategy for noisy situations by extracting features for classification relative to noisy and voiced signals. The suggested method constructs a detection model in a high signal-to-noise ratio (SNR) environment by extracting characteristics for the categorisation of noisy and voiced segments.

Speech is a historical means of communication that is today used in a variety of applications, notably biometric reconstruction, machine communication etc. New ways are being developed to differentiate speech signals from audio signals that are a mixture of noise and speech as the use of speech in more and more applications rises at an exponential rate. The term VAD refers to the system that distinguishes between voice and background noise. The suggested technique constructs a recognition model by retrieving attributes for classification of the spoken and unvoiced data in a high SNR environment. This approach is gaining popularity because it effectively reduces background noise and is a viable option for speech learning, auditory surveillance, and speaking language recognition, among many other uses.

Various concerns connected to VAD were addressed in prior work. Feature extractions and an unique model are commonly used to produce VAD. Speech detection can be classified as either supervised or unsupervised by the VAD. Because it requires specialised data that is not labelled, the supervised VAD approach is less difficult but more accurate than unsupervised alternatives. But even so, numerous strategies for both approaches have been developed, such as a global thresholding framework with a fuzzy entropy tool [4], acoustic decoy that use deep learning [5], and support vector machine (SVM) classifier with a radial basis function kernel [6]. Besides these approaches, some other methods have also been investigated to exploit the variability of noise and voice properties. Additionally, few researchers have used the integration of multiple features with the help of linear combination, principal component analysis, and linear discrimination analysis. The major limitation of supervised VAD method is requirement of abundant trained dataset which may lead to mismatch between the trained and untrained testing data [7]. Therefore, unsupervised approaching is attaining the attention of the researchers in this domain and various approaches have been development. All the techniques are developed to improve the voice detection process.

To detect the presence of noise in a given speech frame, all of the previously proposed algorithms use a window size of audio frame and a feature extraction method. The complexity of proposed model is that whereas if noisy signal estimate is substantial, the speech signal's intelligibility degrades due to the loss of frames containing primary sounds, because the prediction fails to discriminate between low-energy initial low-energy speech and noises.

The proposed method will play a key role in future researches which will base on voice recognition algorithm. Because the proposed research provide a new hybrid approach for speech and non-speech classification that combines a hybridisation model of artificial neural networks (ANNs) and K-means classifiers to increase the noise level in uncontrolled speech signals.

The window size selection processor, on the other hand, has not been investigated in order to obtain the best window size. As a result, in this study, the overlap window size was taken into account in order to find a way to improve the voice frame detection procedure. The technique is assessed in terms of voice detection efficiency, noise presence likelihood, precision, and consumption of time.

The VAD module was described in the first Global System for Mobile communication [8] standards for low sampling voice coding, which was less noise resistant than later versions. The capacity to endure noise is also a significant step forward, as it has the potential to increase automatic speech recognition performance in noisy situations. The main problem in the previous supervised VAD approach is that it necessitates a large number of training datasets, which might result in mismatches between the trained and untrained testing data. The previous works are less effectiveness in term of voice detection, precision, and more consumption of time. A variety of solutions have been described to stay up with the technology in speech activity detection. The most of these proposed methods may be recognised by the properties that they contain [8]. Short term energy, Mel Frequency Cepstral Coefficient (MFCC), pitch, and low frequency variability (LFV) have come to the top of the list due to its easy. They are, however, effective in loud contexts, and these qualities were employed in this study [9]. Between the years 2000 and 2022, all of these changes occurred. All of these tests improved the robustness of the system in a variety of scenarios while lowering complexity and enhancing overall efficiency. Combinations of multiple features-based VAD algorithms, such as Computer Assisted Radar Tomography [10] and ANN [11], have also been tested, however the complexity of these methods has risen. Furthermore, certain attempts at noise characterisation have been made, culminating in the enhancement of voice spectra using Wiener-filtering based noise parameters [12], which are already in use.

Because these approaches were designed with static noise in mind, they are more sensitive to fluctuations in SNR. Noise estimates with modification have been utilised in other studies to improve the resilience of VAD while increasing the computational complexity of the equations. Many new qualities, such as wavelet based transformed images, spectrum information, and wavelet entropy energy ratio, are being introduced as the number of VAD applications grows [13]. Such features are simple to execute and are better suited to randomised signal analysis, but they have the drawback of being unable to accurately locate the end point of a voice transmission in a noisy environment. In terms of improving judgement performance, a Teager energy (TE) based on power spectral density is obtained.

The contribution of VAD in speech identification in simulated acoustic-electric hearing arises from the fact that they serve as trustworthy acoustic landmarks, allowing listeners to efficiently integrate fragments of the message glimpsed through temporal gaps into a single coherent speech stream. The following is a list of the contents of this paper: In Section 2, important research is mentioned; in Section 3, Methodology; in Section 4, result is evaluated; and in Section 5, the conclusion and future scope.

2 Relevant work

Speech restructuring is the oldest technique of expressing our views, and because of its wide range of uses, it is now attracting the interest of many researchers across the world. The main contribution of proposed work is that it provides a new hybrid approach for speech and non-speech classification that combines a hybridisation model of ANNs and K-means classifiers to increase the noise level in uncontrolled speech signals. As a result, a lot of work has gone into improving voice signal transmission and removing extraneous signals. Voice activity detection is a method for extracting speech signals from a voice signal. It consists of two parts: feature extraction and discrimination. As a result, different studies, methodologies, and schemes on these two crucial aspects of VAD have been offered earlier. Pitch, MFCC, autocorrelation-based features [9], line spectral frequencies [14], periodicity-based features [15], linear prediction residual, statistical model, Gaussian Mixture Model, Hidden Markov Model, and Multi-Layer Perceptions are some of them [15]. The majority of these methods have chosen a sequence of steps such as taking a sample and sampling at a pace of 5000 per sample, followed by sample framing [16]. In most cases, the window takes about 30 frames each window. After then, frames are relocated to a fixed window size of 40% of the real window size. Before analysing the signal, whether it is speech or non-speech, the study team chopped the speech data signal into frames, signal discontinuity as shown in Figure 1; the frame size is 10–30 ms, and frames can be overlapped properly; the overlapping region spans from 0% to 50% of the frame size [16]. Because there is an amplitude discontinuity at the window's endpoints, we adopted the hamming windowing method for the research. The signal discontinuity is shown in Figure 1. Windowing the signal data guarantees that the signal ends line up while maintaining everything smooth; this reduces 'spectral leakage' significantly. The Hamming window has slightly better sidelobe suppression than other window techniques. We also used two strategies that were already in use. VAD-ANN and VAD-Kmeans are two VAD models.

Then an extracting features approach is utilised, and a variety of feature extraction methods have been developed by a number of researchers. MFCC, pre-emphasis, frames block windows, DFT spectrum, Mel spectral response, and Dynamic MFCC are some of the most widely examined aspects in prior research. The VAD model with the combination of four feature extraction techniques is shown in Figure 2. However, with the introduction of MFCC, signal energy, pitch, and a new parameter i.e. LFV, as illustrated in Figure 2, we added a novelty to this work.

To locate the voice segments in low frequency, LFV is used. Because human ears are more responsive in the low frequency, we used all signal segments detected in the lower frequencies as a voice in this document. LFV indicates whether or not there is speech in the input signal. If the LFV is big, the frame is considered unvoiced; if the rate is low, the frame is considered voiced.

Figure 1 Signal discontinuity

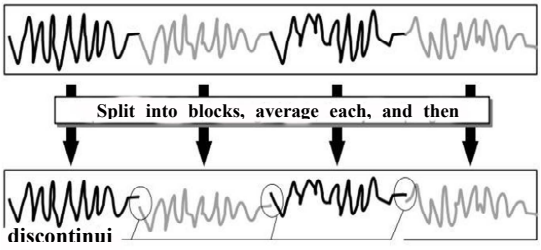
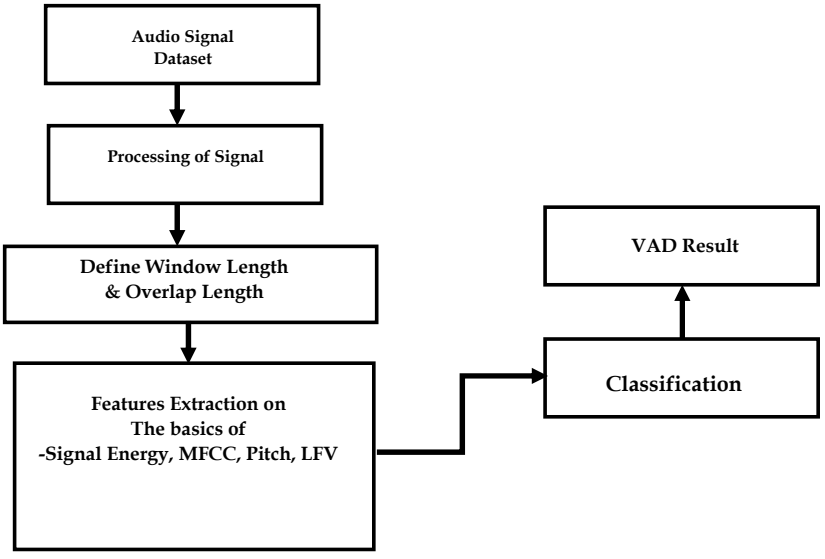


Figure 2 VAD model with the combination of four feature extraction techniques



A SNR of 1000 decibels, for example, indicates that the audio signal is 1000 decibels louder than the component's specified noise level. Thus, a 1000 dB SNR requirement is noticeably preferable to one of 70 dB or less in this regard which contains less noise value.

This experiment uses a TIMIT dataset with 156 samples that has been thoroughly vetted. The investigations are also reported using the unsupervised VAD approach. Although classification accuracy for noisy and spoken signals decreases as the energy level of the quiet segment increases with a lower signal to noise ratio, the present approach suffers from loss of identifying capability. The VAD assigns a supervised or unsupervised classification to speech detection. The supervised VAD technique is less complex but even more accurate than unsupervised alternatives since it requires difference can be observed that has not been labelled. The remaining steps are carried out in the same manner as those chosen by other researchers and stated in this study [17]. We also focused on improving the SNR ratio with the aid of K-means clustering and classifications. Table 1 shows the confusion matrices:

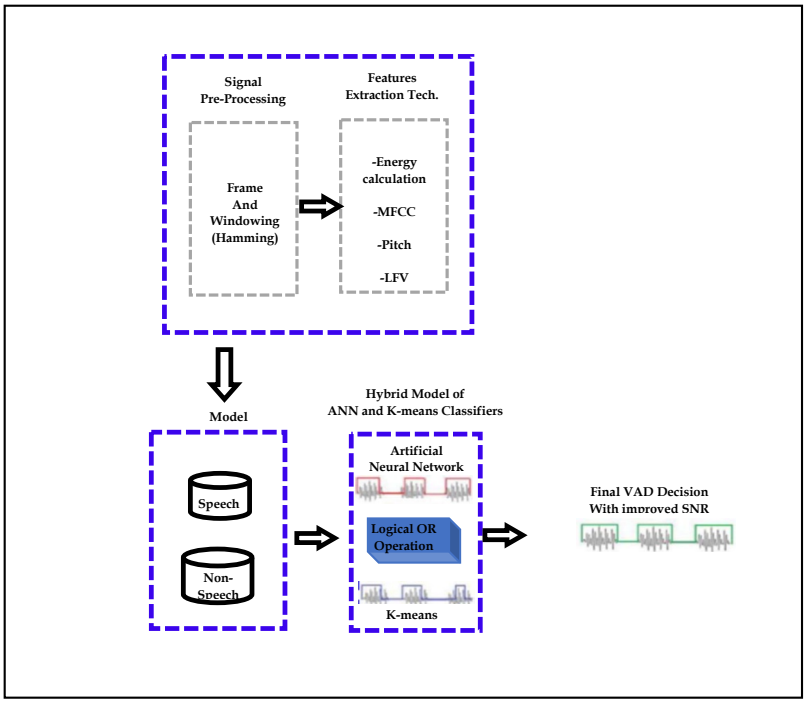
Table 1 Performance matrix based upon K-means classifier

Performance parameters	*K-means*
Accuracy	99.7263
Recall	99.5072
Fscore	99.5606
Error rate	0.27367
False alarm rate	0.13684
SNR	25.6158

3 Methodology

This work is focused on the accurate and faster unsupervised VAD method. As demonstrated in Figure 3, this research effort provided an improved model for unsupervised VAD. The proposed technique builds a detection model in a high SNR environment by extracting features for noisy and voiced segment classification. Moreover, the model is less affected by noise for speech features, and identification efficiency is better by combining new and old feature extraction approaches. The appropriate window length and window overlapping size are determined by this enhanced model. Four feature extraction methods were employed to analyse the audio segment in order to improve the suggested model's outcomes.

Figure 3 An improved model for unsupervised voice activity detection (see online version for colours)



A new hybrid technique for classification of speech and non-speech is employed in the research effort, which uses a hybridisation model of ANNs and K-means classifiers to improve the signal to noise ratio in the unsupervised speech signal. Clustering is an unsupervised machine learning technique that can be used to construct clusters as features to improve classification models. They are sufficient for classification on their own, as the outcome demonstrates. When incorporated as features, however, they aid in the model's accuracy using a hybrid model of ANN with K-means classifications also improves the SNR ratio. An improved model for unsupervised VAD is shown in Figure 3.

The speech reorganisation method is very useful for researchers. Because this method plays a crucial role in extracting speech signals from a voice signals. This method is sufficient for classification of speech. Its ability to withstand noise is also a major advance, since it has the potential to improve automated speech recognition performance in loud environment.

4 Results and discussion

Further, the investigations are carried out for the complexity matrices of SVM, ANN, Updated ANN and proposed approach with K-means. The study employs a new hybrid approach for speech and non-speech classification that combines a hybridised model of ANNs and K-means classifiers to increase the signal to noise ratio in unsupervised speech signals. Table 2 shows a comparison of VAD approaches [18].

Table 2 VAD techniques comparison

Technique	*Pros*	*Cons*	*Solutions*
Multimodal VAD	In a noisy and dynamic context, this is extremely successful	Incorporation of the video signal	Multimodal compact bilinear pooling (MCBP)
Gaussian Mixture HMM	Its functional design is simple	For nonlinear functions, this is an inefficient strategy	Neural network-based VAD
Spectral Entropy Wavelet Transformation Wavelet Energy Entropy Ratio	It's easier to implement, and better for randomised signal analysis	Limitations in detecting the end state of a spoken transmission in a loud background	Teager Energy is based on power frequency response
GMM-based Voice activity model	Integration of a log probability value and a voice activity recognition feature based on short time energy	Unsupervised VADs only	Nonlinear quantised strategy based self-adaptive VAD
Zero Crossing Rate Signal Energy	Less complexity	Ineffective in noisy environment	Mel-frequency, delta line spectrum frequencies, and features-based higher order statistics are all based on auto-correlation
CART, ANN	In a loud environment, effective	Complexity was increased.	Wiener filtering

Source: Shilpa et al. [18]

As a consequence, the overlap window size was considered in this research in order to optimise the voice frame identification technique. The approach is evaluated in terms of speech detection effectiveness, likelihood of noise presence, accuracy, and time consumed. Following conclusions shown in Figure 4 could be drawn from the simulated results of LFV-KANN proposed model:

Performance metric of hybrid model of ANN and K means i.e., proposed model is compared with the existing techniques and it is shown in the Figure 5.

Figure 4 Simulated results of proposed model (see online version for colours)

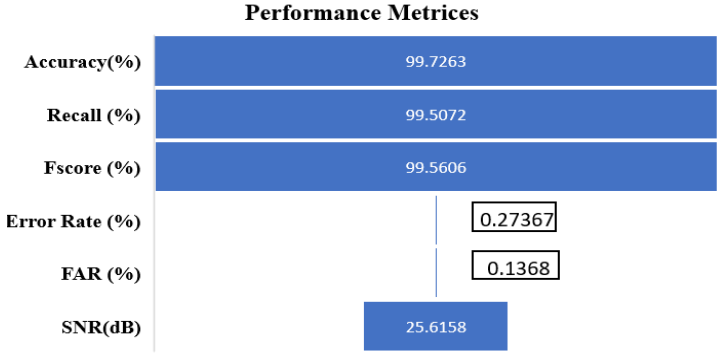
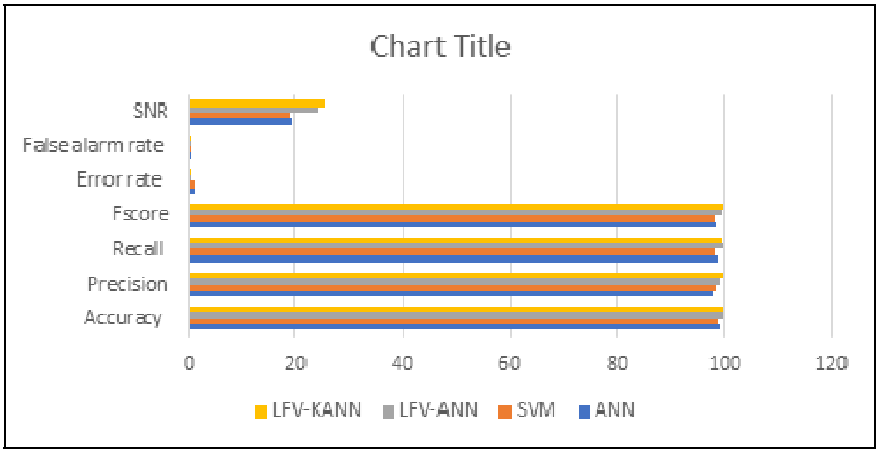
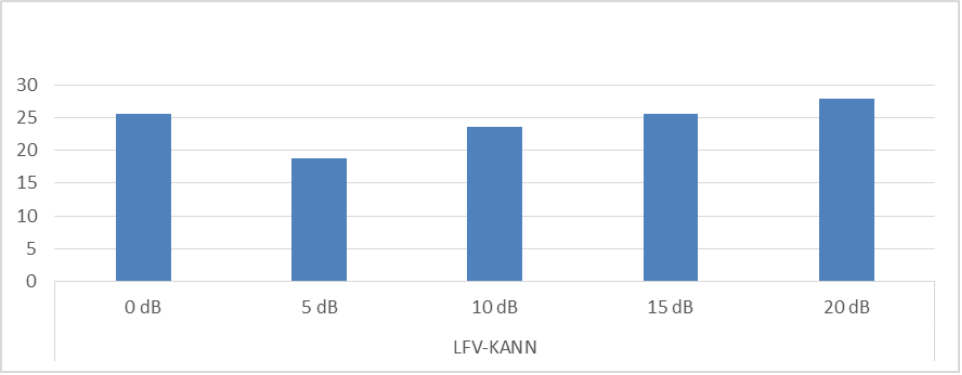


Figure 5 Performance metric comparisons (see online version for colours)



Above Figure 5 concludes that proposed model LFV-KANN outperforms existing techniques in terms of accuracy, precision, recall, F score, error rate, false alarm rate and SNR. Figure 6 illustrate concludes that proposed model LFV-KANN efficiently handles increase in noise power by hybridisation of two classifiers: ANN and K-means clustering.

Figure 6 Performance metric during different noise added (see online version for colours)



5 Conclusion and future scope

This paper relates recognition performance in noisy environments to strong voice detection algorithms using feature extraction. Voice recognition performance degrades because the feature extraction parameters are susceptible to noise signals in a realistic noisy environment with a variety of environmental disturbances and for a speech signal with a low SNR. The proposed method will play a key role in future researches which will be based on voice recognition algorithm. Because the proposed research provides a new hybrid approach for speech and non-speech classification that combines a hybridisation model of ANNs and K-means classifiers to increase the noise level in uncontrolled speech signals. As a result, we've developed a feature-based method for recognising voice features in noisy environments: the proposed method is based on extracting features for categorising voiced and unvoiced signals with a greater signal to noise ratio (SNR). Furthermore, for voice features, the model is less impacted by noise, and recognition performance is increased by merging new and old feature extraction methodologies. Virtual assistants will rule our daily lives in the future, as voice will allow us to speak with household gadgets such as alarm systems, lights, sound systems, and even culinary appliances. There will be a significant increase in the number of voice-controlled gadgets in our workplaces. In hospitals, laboratories, and manufacturing facilities, hands-free mobility will be critical. In future, there will be intelligent voice-driven automobiles, entertainment, and location-based searches, and passengers will be able to travel hands-free.

References

1 Gelly, G. and Gauvain, J.L. (2018) 'Optimization of RNN-based speech activity detection', *IEEE/ACM Transactions on Audio Speech and Language Processing*, Vol. 26, No. 3, pp.646–656, https://doi.org/10.1109/TASLP.2017.2769220

2 Elton, R.J., Vasuki, P. and Mohanalin, J. (2016) 'Voice activity detection using fuzzy entropy and support vector machine', *Entropy*, Vol. 18, No. 8, https://doi.org/10.3390/e18080298

3 Ghahabi, O., Zhou, W. and Fischer, V. (2018) 'A robust voice activity detection for real-time automatic speech recognition', *Proc. of ESSV. 2018*, p.644283, http://essv2018.de/wp-content/uploads/2018/03/30_OmidGhahabi_ESSV2018.pdf

4 Elton, R.J., Mohanalin, J. and Vasuki, P. (2021) 'A novel voice activity detection algorithm using modified global thresholding', *Int. J. Speech Technol.*, Vol. 24, No. 1, pp.127–142.

5 Kwon, H., Yoon, H. and Park, K.W. (2020) 'Acoustic-decoy: detection of adversarial examples through audio modification on speech recognition system', *Neurocomputing*, Vol. 417, pp.357–370.

6 Meduri, S.S. and Ananth, R. (2011) *A Survey and Evaluation of Voice Activity Detection Algorithms*, Medieteknik.Bth.Se, http://medieteknik.bth.se/fou/cuppsats.nsf/all/a1e356336cee2e3ac125799800566259/$file/BTH2011_Meduri.pdf

7 Wang, Y. and Lee, L. (2014) 'Supervised detection and unsupervised discovery of pronunciation error patterns for computer-assisted language learning', *IEEE/ACM Trans. Audio Speech Lang.*, Vol. 9290, No. c, pp.1–16, https://doi.org/10.1109/TASLP.2014.2387413

8 Mustafa, M.K., Allen, T. and Appiah, K. (2015) 'Research and development in intelligent systems XXXII', *Research and Development in Intelligent Systems XXXII*, https://doi.org/10.1007/978-3-319-25032-8

9 Sunil Kumar, S.B. and Sreenivasa Rao, K. (2016) 'Voice/non-voice detection using phase of zero frequency filtered speech signal', *Speech Commun.*, Vol. 81, pp.90–103, https://doi.org/10.1016/j.specom.2016.01.008

10 Amardeep, A. (2013) *Methods for Improving Voice Activity Detection in Communication Services*, http://www.diva-portal.org/smash/record.jsf?pid=diva 2: 588802%0Ahttp://uu.diva-portal.org/smash/get/diva 2: 588802/FULLTEXT01.pdf

11 Pannala, V. and Yegnanarayana, B. (2021) 'A neural network approach for speech activity detection for apollo corpus', *Comput. Speech Lang.*, Vol. 65, p.101137, https://doi.org/10.1016/j.csl.2020.101137

12 Devi, T.M., Kasthuri, N. and Natarajan, A.M. (2013) 'Environmental noise reduction system using fuzzy neural network and adaptive fuzzy algorithms', *Int. J. Electron.*, Vol. 100, No. 2, pp.205–226, https://doi.org/10.1080/00207217.2012.687192

13 Joseph, S.M. and Babu, A.P. (2016) 'Wavelet energy based voice activity detection and adaptive thresholding for efficient speech coding', *Int. J. Speech Technol.*, Vol. 19, No. 3, pp.537–550, https://doi.org/10.1007/s10772-014-9240-x

14 Sharma, S., Sharma, A., Malhotra, R. and Rattan, P. (2021) 'Voice activity detection using windowing and updated K-means clustering algorithm', *Proceedings of 2021 2nd International Conference on Intelligent Engineering and Management, ICIEM. 2021*, pp.114–118, https://doi.org/10.1109/ICIEM51511.2021.9445371

15 Benzvi, D. and Shafir, A. (2019) 'An ICA algorithm for separation of convolutive mixture of periodic signals', *2018 IEEE International Conference on the Science of Electrical Engineering in Israel, ICSEE. 2018*, Vol. 2, No. 4, pp.273–283, https://doi.org/10.1109/Icsee.2018.8646002

16 Sharma, S., Rattan, P., Sharma, A. and Shabaz, M. (2021) 'Voice activity detection using optimal window overlapping especially over health-care infrastructure', *World J. Eng.,* February, https://doi.org/10.1108/WJE-02-2021-0112

17 Elton, R.J., Vasuki, P. and Mohanalin, J. (2016) 'Voice activity detection using fuzzy entropy and support vector machine', *Entropy*, Vol. 18, No. 8, https://doi.org/10.3390/e18080298

18 Sharma, S., Rattan, P. and Sharma, A. (2021) 'Recent developments, challenges, and future scope of voice activity detection schemes – a review', in Kaiser, M.S., Xie, J. and Rathore, V.S. (Eds.): *Information and Communication Technology for Competitive Strategies (ICTCS 2020)*, Lecture Notes in Networks and Systems, Vol. 190, Springer, Singapore. https://doi.org/10.1007/978-981-16-0882-7_39