

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/357126549>

# Design Engineering SNR Improvement in Voice Activity Detection

Article · December 2021

CITATIONS

0

READS

125

5 authors, including:



[Anurag Sharma](#)

KIIT University

65 PUBLICATIONS 341 CITATIONS

[SEE PROFILE](#)



[Pravin Kshirsagar](#)

Raisoni Group of Institutions

12 PUBLICATIONS 15 CITATIONS

[SEE PROFILE](#)



[Rajendra R Patil](#)

GSSS Institute of Engineering and Technology for Women

20 PUBLICATIONS 16 CITATIONS

[SEE PROFILE](#)

## SNR Improvement in Voice Activity Detection

Shilpa Sharma<sup>1</sup>, Rahul Malhotra<sup>2</sup>, Anurag Sharma<sup>3</sup>

<sup>1</sup>Department of Computer Science & Engineering, CT University, Ludhiana, India  
and Lovely Professional University  
E-mail: shilpa13891@gmail.com

<sup>2</sup>Department of Electronics Communication Engineering, CT Group of Institutions, Jalandhar, India.  
E-mail: [blessurahul@gmail.com](mailto:blessurahul@gmail.com)

<sup>3</sup>Faculty of Engineering, Design and Automation, GNA University, Phagwara, India.  
E-mail: [er.anurags@gmail.com](mailto:er.anurags@gmail.com)

### Abstract

Over the years, the aim of error-free continuous speech recognition has remained controversial. Environmental robustness has acquired widespread acceptance as one of the key topics of research in the voice recognition field in recent years. Several methods have been investigated. Window framing, speech feature extractions, feature filtering techniques, and other algorithms have all been examined and shown to improve recognition accuracy. One of the neuron connectivity strategies for artificial intelligence applications is to distinguish between sound signals in form of voice and unvoiced. The device first applies fixed weights to these audios before providing output for each format and high speed. Communication through Voice over Internet Protocol (VoIP) in modern-age speech is attracting many researchers. Voice Activity Detector (VAD) is a technique of distinguishing the articulated portion of speech and the silence component in the original voice signal. In many applications such as speech recognition, voice compression systems, mobile communication, this approach plays an important role for obtaining better speech quality and reducing bandwidth complexity.

### Keywords

Voice Activity Detection, Signal Processing, Artificial Neural Network

### 1 Introduction

VAD avoids needless coding or transmission of silence packets in VoIP applications and use huge amount of bandwidth due fewer computations. For designing a perfect VAD, we require self-sufficient background for noise which is a most difficult part as speech often gets corrupted due to background noise and other ecological reasons. Therefore, VAD is used to improve the speech detection robustness in raucous atmosphere. For stationary backdrop noise, averaging the spectrum of raucous signal and computationally is approximated. However, for an active/non stationary situation noise spectrum becomes a quite complicated as noise diverges quickly over time and continuous updating is necessary.

### 2VAD Trends and Applications

The use of voice recognition has spread well beyond customer service phone lines. It can be found in our cell phones and on our computers, and it is employed in a wide range of different businesses as well. The applications for voice recognition appear to be virtually limitless. When it comes to voice activity detection, the following are some of the most popular trends and applications.

## **(i)Voice-activated mobile payments are becoming increasingly popular**

Do you ever become worried about how you're going to pay for things in the future? What if you could make payments using your voice instead of cash or your credit card? Would you believe it if I told you that it would be possible one day to pay with your voice instead of cash or a credit card? It's possible that day will arrive very soon! There are companies working on voice recognition for money transactions right now, and they are based in the United States. When you shop from your smartphone or computer, this method is convenient because you won't have to seek for your wallet as you would otherwise.

Using voice recognition, you can make payments in the following ways: Prior to completing an online purchase, this technology allows you to speak a one-time password rather of typing in your password or personal identification number (PIN). Consider captchas and other one-time passwords that help to improve internet security, but this is stated aloud instead of on a screen. It is considerably superior to using the same password every time since the random nature of the password assures that someone will not be able to overhear you reciting your password and then make purchases in your name on your behalf. Mobile payments that use voice recognition are on the verge of becoming widely accepted and popular.

## **(ii)Artificial Intelligence Assistants which can detect your specific voice**

The majority of us are now familiar with Siri, Alexa, and other typical artificial intelligence (AI) assistants who aid us with daily chores at home and at the office [23][24]. You may use these artificial intelligence assistants to assist you in getting directions while driving, to Google something or start a music playlist without having to pick up your device, or even to switch the lights on in your home. These assistants normally reply to any voice that triggers them, and they do not respond differently depending on who is speaking to the assistant.

However, there are currently technologies that allow artificial intelligence systems to recognise and understand the identities of their users [25] [26]. Using this technology, it is feasible, for example, to have it activate just when a specific person's voice is heard. It has already been implemented with iPhones in the last few years, according to the company. The Siri feature on your iPhone can be configured so that it will only execute tasks or answer queries when you speak into the device. Unauthorized users will be far less likely to mess with your devices, data, and belongings if your AI assistant is only activated by your specific voice, as seen in the following example [27] [28]. Anyone who is not permitted to use the assistant will simply be unable to activate it. In the near future, this technique will almost definitely be used to a variety of different uses [29][30].

## **(iii)Voice Identification for Security Purposes**

The majority of people have a large number of online accounts that require protection, and some of these online accounts such as online banking applications, pose significant security threats. Now that internet banking has become so widespread, it is critical that strong identification procedures are put in place to ensure that only the account owner has access to important information on the account. The use of voice detection is one of the more recent ways of user identification. A speech authentication factor, similar to artificial intelligence assistants that know your distinctive voice, acts as a one-of-a-kind 'password' that unlocks

secured accounts only when your voice is utilised[31][32][33]. Anyone else will be unable to access the account due to the fact that their voice is different from mine.

Because multi-factor systems, which combine a variety of security elements, may now be used, this provides exceptionally high levels of security. Take, for example, the scenario in which you need to scan your fingerprint and pronounce your password loudly in your own distinct voice in order to gain access to your online banking service. That is significantly more secure than the use of a standard password in most cases. In addition to speech recognition, there are systems that use facial recognition as well as voice recognition.

#### **(iv)Voice detection in Forensics and Criminal Identification**

Using voice recognition detection technology to assist in the identification of criminals is one of the more startling advances in the field of VAD. It is now possible to use audio recordings of a crime suspect as key evidence if they have been made by a third party. There is now a collaboration between AGNITIO and Morpho (Safran) that is bringing Voice ID technology into the forensics business. This device has made it possible for voice biometrics technology to be utilised all over the world (in conjunction with fingerprints and other methods) to assist in the identification of persons and the verification of their background history.

### **3.Artificial Intelligence**

Since the development of computers or machines, their ability to execute a wide range of jobs has increased tremendously in both quantity and quality. As humans progressed in their understanding of computer systems, they increased their power in terms of their diverse working areas, growing speed, and shrinking size as a function of time. Artificial Intelligence is a discipline of computer science that is dedicated to the development of computers or machines that are as intelligent as human beings.

When a machine can mimic the human brain and execute tasks and make decisions in the same way that the human brain can, this is referred to as artificial intelligence. Artificial intelligence is the term used to describe the machine's intelligence. Artificial intelligence is an integrated area that encompasses a wide range of disciplines such as computer science, mathematics, neurology, and other related fields. In general, there are three steps in the development of artificial intelligence.

To begin, there will be learning, wherein necessary data will be taken from a source, and this data will be used to assist in the formulation of rules. Second, reasoning in which formed rules are thrown in reaction to a given input and, after mapping all of the rules, a definite conclusion is generated in response to the supplied inputs. Thirdly, self-correction [16], in which the system corrects itself using heuristic data.

### **4. Neural Network**

The neural network is basically a parallel computing device that, just like human brain, is capable of performing multiple computations and making decisions. The biological nervous system of humans serves as the primary inspiration for the neural network[34][35][36]. It is a

collection of various processing elements that are interconnected with one another, as illustrated in Figure 1. The examples are used to train the neural network.

The training process prepares the neural network for a specific task or application. Weights are adjusted during the training phase in order to reduce error. These changes are made between neurons. Subtracting the observed value from the required value yields the error. An input layer, an output layer and one or more hidden layers are the three main layers of a neural network. [37] [38][39]

- **Input layer:** The input layer of a neural network is made up of numerous input neurons that are present in a given network [23][24][25]. The neural network's initial data is provided by the input layer. In other words, the input layer accepts data in the various formats specified by the programmer or user.
- **Hidden Layer:** The hidden layer is the second layer of the neural network. This layer sits between the network's input and output layers. A lot of hidden neurons constitute the hidden layer. These neurons are in charge of all computations and calculations to determine the features of the user-supplied input to the input layer. A neural network may have one or more hidden layers.

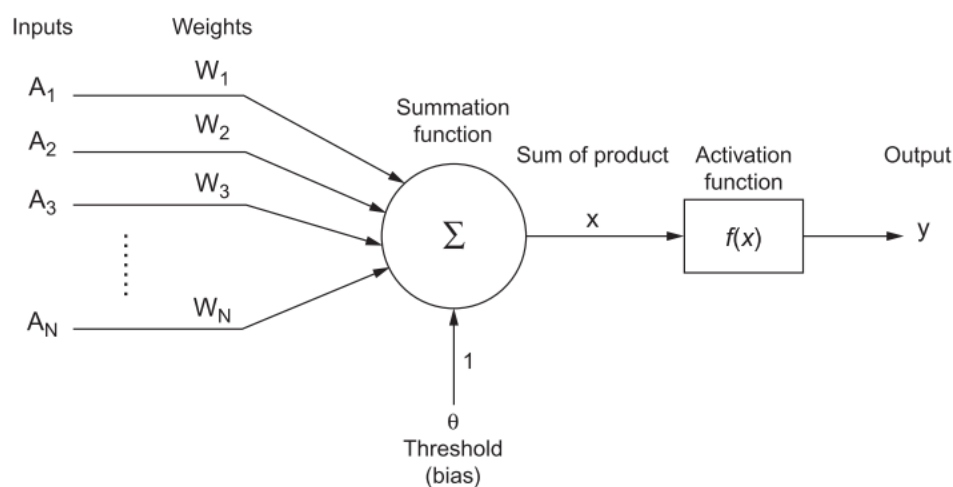


Figure 1. Artificial neural network

- **Output Layer:** A neural network's output layer is its last layer. It consists of one or more output neurons. The user's input is transformed and computed using the neural network's hidden layer. The output layer communicates the network's final outcome or result. The input is acquired by the artificial neural network, which then computes the weighted total of the provided inputs and bias[26][27][28].

In Table 1, it is simple to distinguish between the most famous machine learning applications and their source types, as well as between some related applications, as illustrated by the examples. [5] [6].

Table 1: Different Neural Network types

Common uses	Input method	Input type	Types of neural networks
Associated memory to distinguish ASCII characters	Supervised	Binary	Hopfield-Net
Connect with similar dual channel	Supervised	Binary	Hammin_Net
Assembly (adaptive resonance theory)	Supervised	Binary	Carpenter/grassbery classifier
Discrimination and classification of simple shapes	Supervised	Continuous	Perceptron
Featuring complex shapes and classification	Supervised	Continuous	Multi-layer perceptron
Evaluation of vector and speech, and analogy to biological neural networks	Supervised	Continuous	Kohonenself organizing feature map

## 5. Problem Definition

In this manuscript, proposed an improved model for unsupervised Voice Activity Detection. An improved model begins from analysing the optimal window overlapping size to feature extractions with the help of not only one or two merges VAD features extractions methods rather we have used four features extraction methods so that our new model gives better results. After these steps, we introduced a new technique as the classification of speech and non-speech i.e. hybridization model of two classifiers i.e. Artificial Neural Network and K-means clustering methods as shown in figure 2 which further enhance the signal to noise ratio in the unsupervised speech signal.

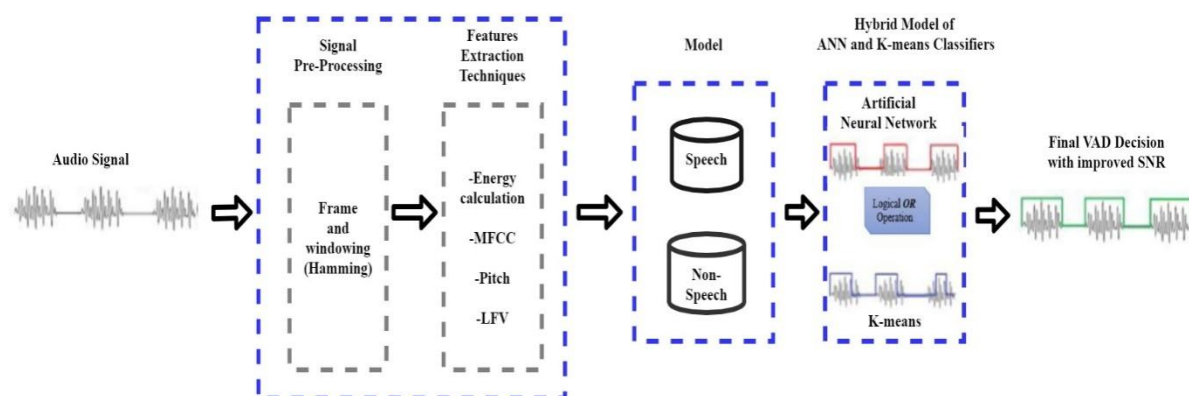


Figure 2: Proposed Model of VAD

And SNR ratio is also improved by using hybrid model of ANN and K-means classifications. As voice segment recognised by both classifiers will further considered for voiced labelled. So chances of missing voice segments will be less in this case.

## Conclusion

The development of a speech recognition model made use of artificial neural networks. The goal was to employ a genetic algorithm to develop a learning neural network that could learn new things. This technology has been applied to the device identifying numbers, resulting in the establishment of a

voice command recognition system that recognises commands spoken by the user. The development of these features is based on the fact that recorded conversation analysis has already shown features that can be used to lay the groundwork for automatic detection of speech parameters. Predicting accuracy was usually higher when forecasting was done using current data set knowledge.

## References

1. Shilpa Sharma, Punam Rattan, Anurag Sharma. "Chapter 39 Recent Developments, Challenges, and Future Scope of Voice Activity Detection Schemes—A Review", Springer Science and Business Media LLC, 2021
2. Sharma, S., Rattan, P., Sharma, A. and Shabaz, M. (2021), "Voice activity detection using optimal window overlapping especially over health-care infrastructure", *World Journal of Engineering*, Vol. ahead-of-print No. ahead-of-print. <https://doi.org/10.1108/WJE-02-2021-0112>
3. S. Sharma, A. Sharma, R. Malhotra and P. Rattan, "Voice Activity Detection using windowing and updated K-Means Clustering Algorithm," *2021 2nd International Conference on Intelligent Engineering and Management (ICIEM)*, 2021, pp. 114-118, doi: 10.1109/ICIEM51511.2021.9445371.
4. Choudhary, A. and Kshirsagar, R. (2012) Process Speech Recognition System Using Artificial Intelligence Technique. *International Journal of Soft Computing and Engineering (IJSCE)*, 2.
5. Ovchinnikov, P.E. (2005) Multilayer Perceptron Training without Word Segmentation for Phoneme Recognition. *Optical Memory & Neural Networks (Information Optics)*, 14, 245-248.
6. Guo, X.Y., Liang, X. and Li, X. (2007) A Stock Pattern Recognition Algorithm Based on Neural Networks. *Third International Conference on Natural Computation*, 2.
7. Dai, W.J. and Wang, P. (2007) Application of Pattern Recognition and Artificial Neural Network to Load Forecasting in Electric Power System. *Third International Conference on Natural Computation*, 1.
8. Shahrin, A.N., Omar, N., Jumari, K.F. and Khalid, M. (2007) Face Detecting Using Artificial Neural Networks Approach. *First Asia International Conference on Modelling & Simulation*.
9. Lin, H., Hou, W.S., Zhen, X.L. and Peng, C.L. (2006) Recognition of ECG Patterns Using Artificial Neural Network. *Sixth International Conference on Intelligent Systems Design and Applications*, 2.
10. Al Smadi, T.A. (2013) Design and Implementation of Double Base Integer Encoder of Term Metrical to Direct Binary. *Journal of Signal and Information Processing*, 4, 370.
11. Takialddin Al Smadi Int. An Improved Real-Time Speech Signal in Case of Isolated Word Recognition. *Journal of Engineering Research and Applications*, 3, 1748-1754.
12. Mc Cowan, D. Dean, M. McLaren, R. Vogt, S. Sridharan: The Delta Phase Spectrum With Application to Voice Activity Detection and Speaker Recognition, *IEEE. trans. Audio Speech Lang. Proc.*, vol. 19, pp. 2026- 2038, 2011.
13. D. Valj, B. Kotnik, B. Horvat, Z. Kacic: A Computationally Efficient Mel Filter Bank VAD Algorithm for Distributed Speech Recognition Systems, *Eurasip J. Appl. Signal Processing*, no. 4, pp. 487-497, 2005.
14. B. Kotnik, Z. Kacic, B. Horvat: A multiconditional robust front-end feature extraction with a noise reduction procedure based on improved spectral subtraction algorithm, in *Proc. 7th Europespeech*, pp. 197-200, 2001
15. T. Kristjansson, S. Deligne, P. Olsen: Voicing features for robust speech detection, *Proc. Interspeech*, pp. 369-372, 2005.



16. J. Haigh, J. Mason: A voice activity detector based on cepstral analysis, Proc. Eurospeech, pp. 1103-1106, 2003
17. S.O. Sadjadi, J. Hansen: Unsupervised Speech Activity Detection Using Voicing Measures and Perceptual Spectral Flux, IEEE Sig. Pro. Letters, vol. 20, pp. 197-200, 2013.
18. M. Marzinik, B. Kollmeier: Speech pause detection for noise spectrum estimation by tracking power envelope dynamics, IEEE Trans. Speech Audio Process., vol. 10, pp. 109-118, 2002
19. E. Nemer, R. Goubran, S. Mahmoud: Robust voice activity detection using higher-order statistics in the LPC residual domain, IEEE Trans. Speech Audio Process., vol. 9, pp. 217-231, 2001.
20. K. Ishizuka, T. Nakatani: Study of Noise Robust Voice Activity Detection Based on Periodic Component to Aperiodic Component Ratio, in Proc. ISCA Tutorial and Research Workshop on Statistical and Perceptual Audition, pp. 6570, 2006.
21. J. Ramirez, J. Segura, M. Benitez, L. Garcia, A. Rubio: Statistical voice activity detection using a multiple observation likelihood ratio test, IEEE Signal Proc. Letters, vol. 12, pp. 689-692, 2005.
22. P. Ghosh, A. Tsiartas, S. Narayanan: Robust voice activity detection using long-term signal variability, IEEE Trans. Audio Speech Lang. Process., vol 19, pp. 600-613, 2011.
23. Sudhir Akojwar, Pravin Kshirsagar, "Performance Evolution of Optimization Techniques for Mathematical Benchmark Functions", WSEAS International conference on Neural Network-2016, Rome, Italy.
24. Pravin Kshirsagar and Dr.SudhirAkojwar, Prediction of Neurological Disorders using Optimized Neural Network, In the proceeding of International Conference on signal processing, Communication, Power and Embedded System ,October (2016).
25. Velvizhi; Satish R Billewar; Gaurav Londhe; Pravin Kshirsagar; Neeraj Kumar, "Big Data for Time Series and Trend Analysis of Poly Waste Management in India", Materials Today: Proceedings, Elsevier, 2020.
26. Pravin R Kshirsagar, Anil N Rakhonde, Pranav Chippalkatti, "MRI IMAGE BASED BRAIN TUMOR DETECTION USING MACHINE LEARNING", Test Engineering and Management, January-February 2020 ISSN: 0193-4120, Vol. 81, Page No. 3672 –3680.
27. Pravin Kshirsagar et.al (2016), "Brain Tumor classification and Detection using Neural Network", DOI: 10.13140/RG.2.2.26169.72805.
28. Pravin R. Kshirsagar, Arpit D. Yadav, Kirti A. Joshi, PranavChippalkatti, Rinali Y. Nerkar (2020) "Classification and Detection of Brain Tumor by using GLCM Texture Feature and ANFIS" Journal of Research in Image and Signal Processing, 5(1), 15-31 <http://doi.org/10.5281/zenodo.3732939>.
29. H. Manoharan, et al., "Examining the effect of aquaculture using sensor-based technology with machine learning algorithm", Aquaculture Research, 51 (11) (2020), pp. 4748-4758.
30. S. Sundaramurthy, S. C and P. Kshirsagar, "Prediction and Classification of Rheumatoid Arthritis using Ensemble Machine Learning Approaches," 2020 International Conference on Decision Aid Sciences and Application (DASA), 2020, pp. 17-21, doi: 10.1109/DASA51403.2020.9317253.
31. P. R. Kshirsagar, H. Manoharan, F. Al-Turjman and K. Kumar, "DESIGN AND TESTING OF AUTOMATED SMOKE MONITORING SENSORS IN VEHICLES," in IEEE Sensors Journal, doi: 10.1109/JSEN.2020.3044604.



32. Pravin Kshirsagar et. al., "OPERATIONAL COLLECTION STRATEGY FOR MONITORING SMART WASTE MANAGEMENT SYSTEM USING SHORTEST PATH ALGORITHM", Journal of Environmental Protection and Ecology, Vol. 22, Issue 2, pp. 566-577, 2021
33. Kshirsagar, Pravin R., et al. "Automation Monitoring With Sensors For Detecting Covid Using Backpropagation Algorithm." KSII Transactions on Internet and Information Systems (TIIS) 15.7 (2021): 2414-2433.
34. Jude, A.B., Singh, D., Islam, S. et al. An Artificial Intelligence Based Predictive Approach for Smart Waste Management. Wireless Pers Commun (2021). <https://doi.org/10.1007/s11277-021-08803-7>.
35. Padmaja, M., Shitharth, S., Prasuna, K. et al. Grow of Artificial Intelligence to Challenge Security in IoT Application. Wireless Pers Commun (2021). <https://doi.org/10.1007/s11277-021-08725-4>.
36. S. Shitharth, Pratiksha Meshram, Pravin R. Kshirsagar, Hariprasath Manoharan, Vineet Tirth, Venkatesa Prabhu Sundramurthy, "Impact of Big Data Analysis on Nanosensors for Applied Sciences Using Neural Networks", Journal of Nanomaterials, vol. 2021, ArticleID 4927607, 9 pages, 2021. <https://doi.org/10.1155/2021/4927607>
37. Kshirsagar P., More V., Hendre V., Chippalkatti P., Paliwal K. (2020) IOT Based Baby Incubator for Clinic. In: Kumar A., Mozar S. (eds) ICCCE 2019. Lecture Notes in Electrical Engineering, vol 570. Springer, Singapore.
38. Oza S. et al. (2020) IoT: The Future for Quality of Services. In: Kumar A., Mozar S. (eds) ICCCE 2019. Lecture Notes in Electrical Engineering, vol 570. Springer, Singapore
39. Kshirsagar P., Pote A., Paliwal K.K., Hendre V., Chippalkatti P., Dhabekar N. (2020) A Review on IOT Based Health Care Monitoring System. In: Kumar A., Mozar S. (eds) ICCCE 2019. Lecture Notes in Electrical Engineering, vol 570. Springer, Singapore