

Recognition of Voice and Noise Based on Artificial Neural Networks

Shilpa Sharma¹, Rahul Malhotra², Anurag Sharma³

¹Department of Computer Science & Engineering, CT University, Ludhiana, India. E-mail: shilpa13891@gmail.com

²Department of Electronics Communication Engineering, CT Group of Institutions, Jalandhar, India.

E-mail: blessurahul@gmail.com

³Department of Computer Science & Engineering, GNA University, Phagwara, India.

E-mail: er.anurags@gmail.com

ABSTRACT

We present a model-based approach to voice activity detection (VAD) for a variety of applications including medical, education, and online video conferencing. The artificial neural network is optimally trained to provide a consistent model by using a Mel Frequency Cepstrum Coefficient) factor derived from clean or noisy speech samples. One of the strategies of neuron networks for artificial intelligence applications where this method can differentiate between abnormal users' sound signals and neural networks. The device first trainers fixed weights on these audios and then gives the output for each format and high speed. The proposed neural network analysis is focused on speech recognition solutions, signal detection by means of angular modulation and modulated techniques detection. The results show that: (i) the proposed artificial neural system classification system delivers fairly high scores under a variety of noise conditions; (ii) in terms of different classification steps, the invented model exceeds other VAD methods.

KEYWORDS

Neural Networks, Signal Processing, Artificial Networks.

Introduction

Artificial intelligence applications have risen throughout significantly in recent years, especially neural network applications, where they are an ideal method for solving a variety of problems with distinct styles and classification. The year 1943 is known as the beginning of artificial neural network evolution. During the vacuum tube era, it was not technologically feasible or rational to create the first formal model of neurons using a computer model that included all necessary elements, as well as to complete and enforce the electronic version of this model. It's worth noting that this model has been used to characterise vacuum tube-based computer hardware extensively [1]. The proposed tutorial to upgrade nerve cell connections originally referred to the law educational learning rule HYIP, which stated that knowledge could be stored in links and connections. It is widely accepted that educational technology has proven its worth in the sector's future development. Hip education Act's first approach to neural network theory was designed and deployed in the 1950s in the first study of the neurological machine, where the application automatically contacts and during this stage, the term preceptor was used to define the unit described for neural cell to construct the terminology and extracted from the neuron, he pioneered the concept in 1958. Through the use of modulated indicators, this innovation offered a viable method for training machine learning and classification of specific models. As a result, it became a part of engineers' and scientists' imaginations, as well as a context for calculations relating to this type of machinery that is still used today. In the early 1960s, a newly evolved technique referred as an extremely desirable principle [2].

The Science of Recognition of Pattern

Recognition, identification, and grouping of patterns automatically are important criteria in a wide range of engineering and scientific disciplines, including artificial intelligence, biology, remote sensing, psychology, marketing, medicine and computer vision. The prototype may be a picture of a fingerprint, cursive handwriting, the voice signal or a human face. When presented with a pattern, one of the following two tasks may be required [3]. a) Classification, discriminated analysis under the supervision of a classifier. Supervised classification, clustering, in which the input pattern is classified into a predefined class; b) Unsupervised classification, cluster analysis, in which the class template is unknown. The issue is referred to as a classified or classification problem in this instance with classes identified either by the designer of computer system in a regulated classification or by unsupervised classification using similar models. Among these applications is data mining, which is used to define the term "plan."

now for instance, they must efficiently classify text documents, undertake appropriate forecasting, arrange and retrieve multimedia databases, and perform biometrics in billions of multidimensional models. The rapid increase in computing capacity, which allows faster processing of large amounts of data, has also enabled the use of diverse and complex data classification and analysis methods. Simultaneously, demand for automated pattern recognition is growing as large databases become available and stringent requirements for cost, accuracy and speed become more stringent. The creation of a blueprint for a recognition scheme is essentially composed of three main parts: a) Data collection and pre-treatment, as well as analysing of data b) Decision-taking process c) The pre-treatment protocol used is determined by the nature of the project. Analyses and decision-making models based on schemas It is recognised that addressing the issue of clearly defined and adequately restricted recognition would require the creation of a compact model and an uncomplicated decision-making strategy. The ability to learn from examples is a crucial and necessary function of the vast majority of recognition template systems. The following are the most frequently used pattern recognition techniques: a) Pattern matching; b) Classification using statistics; c) Structural or syntax conformity; and d) algorithms for Neural networks.

Neural networks is the collection of nodes that collectively represent a special form and all nodes represents the unit of account and ability to operate in parallel is dependent on their interactions with one another and many researcher described in the following way:

- Analytical models that simulate the properties of concurrently processing biological systems and are composed of relatively simple named components.
- Is a fundamental entity class of methods represented as clusters (cluster group a huge proportion of methods together, and all of these methods solve a set of different issues) [4].

To emphasise neural networks' activity, grouping and encoding are used; to emphasise neural networks' properties, the follows are mentioned: a) Noise resistance; b) Efficiency in interacting with blurred images; c) Full resistance to dismembered or partially decomposed images; d) Variations in distinct techniques involving a large number of functional areas that enhance through interdecision. They are an important source of rankings and identification (classification predication) due to its non operations, i.e. their ability to establish non-linear relationships involving noise charts; e) Their maximum bandwidth for internal adjustment of the matrices and forces of system efficiency allows the use of structural improvement in the region of transition period.

Neural Network Types

As illustrated in Table 1, it is easy to distinguish the most famous applications of machine learning and their source types, as well as some related applications. [5] [6].

Table 1. Neural Networks and different 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

Problem Definition

The approach involves iteratively choosing the score that is the most dissimilar to the mean. If this score exceeds a predefined threshold, it is deleted and the standard deviation and mean are recalculated. When a small number of

utterances are used to estimate the standard deviation, this approach significantly improves efficiency. Using a telephonic multisession database, We performed experiments that were content and content independent. The article explores the mapping between system development and algorithmic, drawing on the speaker's experience with mini-problems encountered during the system development process, as well as a paradigm of voice recognition depend upon artificial neural networks [8]. Figure 1 illustrates the detection of speech signals.

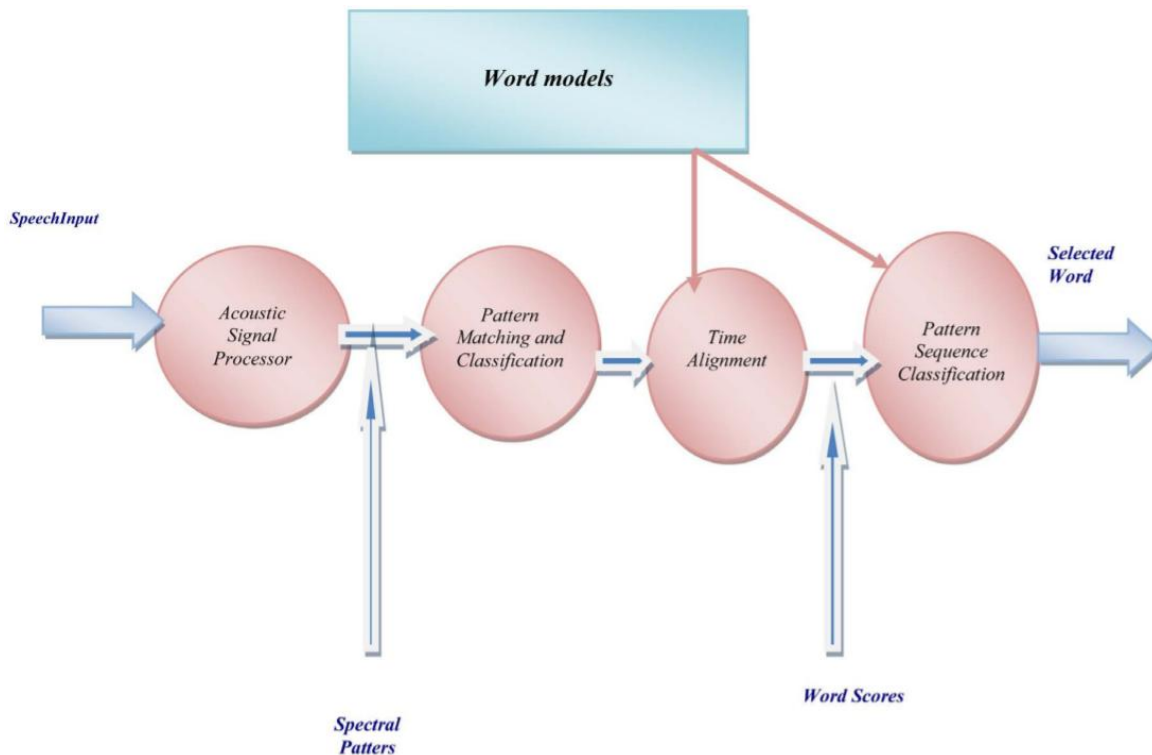


Figure 1. Detection of speech signals planning

- The current research examines the use of artificial networks for voice detection. The scale of the neural network has an impact on the efficacy of detecting phonemes in words. The investigation of techniques for parameterizing speech signals include learning of how and when to use generalised linear analysis to train a neural network for phonetic recognition on a temporary basis. The proposed method of teaching involves the interpretation of items from the trained set and does not require any segmentation of components;
- Methods for diagnosing and detecting modulated signals are being developed and researched;
- Implementation and pilot research of neural network models for real-time signal processing in software.

Recognized Procedure Algorithm for Recognition

- Input signal and word boundary selection;
- Allocation of signal spectrum parameters;
- Evaluation of a intensity in the proximity of acoustic parameters using an artificial neural network;
- Comparison with dictionary standards [8].

A neural network was given an array of signal segments as input after processing the audio data. Each section refers to a series of data that describe the amplitude spectrum of a signal; all of the numbers are written for the preparation for the neural network's signal thresholds calculation. Each row represents a set of frame numbers, as shown in Table 2. The number of unique values in a set of numbers is denoted by I , and the number of unique sets is denoted by N (frame segment post slicing). The number between input and output neurons is recognised; that input neuron corresponds to a unique set of numbers, while the hidden layers only has one neuron that correlates to the received

signals recognition value. Table 3 outlines the parameter meanings used in this analysis, as seen in Figure 2. It is necessary to construct ten neural networks in order to recognise the one number.

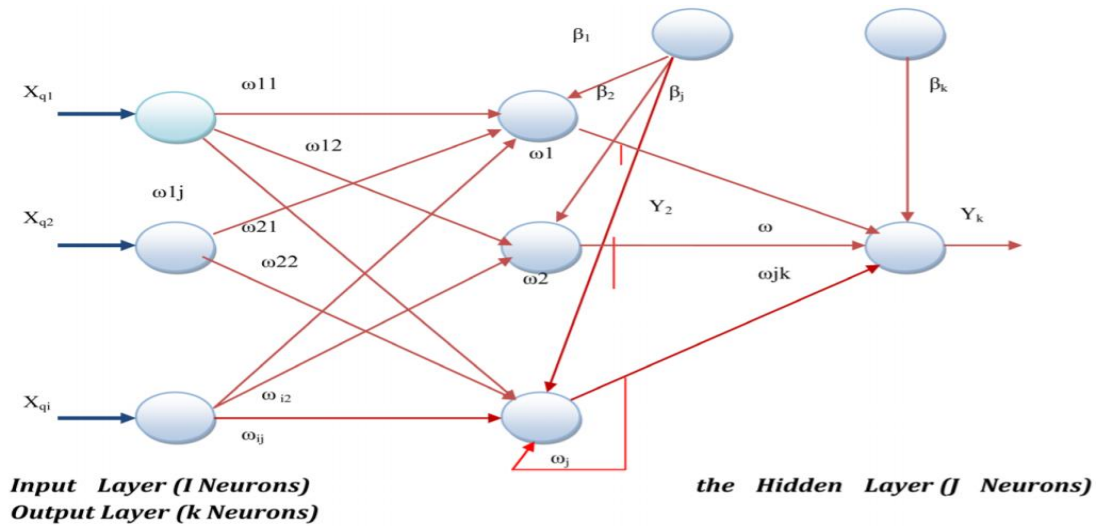


Figure 2. A neural network with feedback.

Approximately 300 words (numbers 0–9) with varying pronunciations, randomly divided into two equal sections: tutorial and sample tests. When developing a neural network to recognise a single number, the desired output of the neural network should be unit for the training set containing the number 5, and zero for the remainder. Training of neural networks is done by presenting the training set in a consistent manner, with accordance to the given methods we need a successive tuning scale, until the error reaches an appropriate level after the variety of configuration. Neuron biology is the foundation from which a neuron is constructed. a cell is composed of an organ shell, known as acellular and axendetletal and having two different types known as axon and dendrite. This research postulates that both the nucleus, which stores information on genetic characteristics, and the plasma, which provides molecular tools for neuronal growth, cellular processes, are contained in the body's cells. A node receives signals from other neurons and relays them through its dendrites, which send signals out from the axon terminals, and then bifurcations form in the neuron endings, where the signal is spread through the fibre. [1] [3].

Table 2. A series of speech signals is defined in detail

Frame	1-value	2-value	...	I-value
1-Frame	x_{11}	x_{12}	...	x_{1I}
2-Frame	x_{21}	x_{22}	...	x_{2I}
...
N-Frame	x_{N1}	x_{N2}	...	x_{NI}

Table 3. shows the definitions of the parameters.

Name	Definition
x_{qi}	i -th q is the input value to a set of numbers
y_j	Output j -neuron layer
ω_{ij}	The weight of the link connecting the i -th neuron with the j -th neuron
ω_j	weight feedback
β_j	Weight feedback j -th neuron; the offset of the j -the neuron layer

Figure 3 shows a neuron's block diagram: X_1, X_2, X_3, X_4 represent the set of weights; stimulation functions; output signal represented by y , neuro command works on the basic operations such as labeled summation and treats these outcomes of non-linear threshold values. The structure contains simple homogeneous elements in the neural network approach yield to achieve the challenges of complexities between objects. These quality attributes of network as a whole are described by the structure of relations.

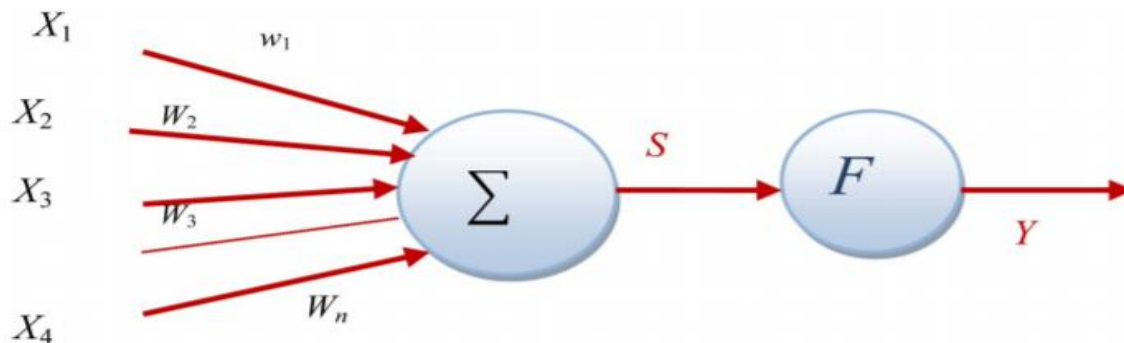


Figure 3. Depicts a technical representation of a neuron

Although the ability to process data in parallel is the most important, the strength of the network is governed by the size of the neural network. This serves to significantly speed up the production of sigma waves, as long as there are several neural connections of foreign character is present. There is an immediate need for real-time signal processing. When it comes to decision-making, the neural networks exhibit features similar to those in-so-called-artificial intelligence (AI) [11].

Conclusion

Artificial neural networks were used to build a speech recognition model. The aim was to use a genetic algorithm to create a learning neural network. This method has been applied to the device identification numbers, resulting in the creation of a voice command recognition system. Since recorded conversation analysis has already revealed features that help form a foundation for automatic recognition of speech parameters, they are developed. On the basis of current data set knowledge, forecasting accuracy was always higher.

References

- [1] Childer, D.G. (2004). *The Matlab Speech Processing and Synthesis Toolbox*. Photocopy Edition, Tsinghua University Press, Beijing, 45-51.
- [2] Chien, J.T. (2005) Predictive Hidden Markov Model Selection for Speech Recognition. *IEEE Transaction on Speech and Audio Processing*, 13.
- [3] Luger, G. and Stubblefield, W. (2004). *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*. 5th Edition, The Benjamin/Cummings Publishing Company, Inc. <http://www.cs.unm.edu/~luger/ai-final/tocfull.htm>
- [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 Eurospeech*, 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. Marzinzik, 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] Y. Kida, T. Kawahara: Voice Activity Detection based on Optimally Weighted Combination of Multiple Features. *In Proc. Interspeech*, pp. 2621-2624, 2005.
- [24] S. Soleimani, S. Ahadi: Voice Activity Detection based on Combination of Multiple Features using Linear/Kernel Discriminant Analyses, *in Proc. Information and Communication Technologies: From Theory to Applications*, pp. 1-5, 2008.
- [25] H. Singh and A.K. Bathla. A survey on speech recognition. *Int. J. Adv. Res. Comput. Eng. Technol.*, no. 2, no. 6, pp. 2186–2189, 2013.
- [26] M.A. Anusuya and S.K. Katti. Speech recognition by machine: A review. *Int. J. Comput. Sci. Inf. Secur.*, vol. 6, no. 3, pp. 181–205, 2009.
- [27] J. Padmanabhan and M.J.J. Premkumar. Machine learning in automatic speech recognition: A survey. *IETE*

- Tech. Rev.*, vol. 32, no. 4, pp. 240–251, 2015.
- [28] Chung-Ching Shen, William Plishker, and Shuvra S. Bhattacharyya. Design and Optimization of a Distributed, Embedded Speech Recognition System. *In Proceedings of the International Workshop on Parallel and Distributed Real-Time Systems*, Miami, Florida, April 2008.
- [29] G. Zhou, J.H.L. Hansen, and J.F. Kaiser. Nonlinear feature based classification of speech under stress. *IEEE Trans. Speech Audio Process.*, vol. 9, no. 3, pp. 201–216, Mar. 2001.
- [30] C. Fredouille, G. Pouchoulin, J.F. Bonastre, M. Azzarello, A. Giovanni, and A. Ghio. Application of Automatic Speaker Recognition techniques to pathological voice assessment (dysphonia). *In Proc. Eur. Conf. Speech Commun. Technol. (Eurospeech)*, 2005, pp. 149–152.
- [31] V.A. Petrushin. Emotion recognition in speech signal: Experimental study, development, and application. *In Proc. 6th Int. Conf. Spoken Lang. Process. (ICSLP)*, 2000, p.5.
- [32] N. Fragopanagos and J.G. Taylor. ‘Emotion recognition in human– computer interaction. *Neural Netw.*, vol. 18, no. 4, pp. 389–405, 2005.
- [33] E. Douglas-Cowie, N. Campbell, R. Cowie, and P. Roach. Emotional speech: Towards a new generation of databases. *Speech Commun.*, vol. 40, nos. 1–2, pp. 33–60, 2003.
- [34] B. Kingsbury, G. Saon, L. Mangu, M. Padmanabhan and R. Sarikaya. Robust speech recognition in noisy environments: The 2001 IBM SPINE evaluation system. *Proc. ICASSP, 1*, 53-56, 2002.
- [35] T. Kristjansson, S. Deligne and P. Olsen. Voicing features for robust speech detection. *Proc. Interspeech*, pp. 369-372, 2005. ETSI standard document, ETSI ES 202 050 V 1.1.3., 2003.
- [36] M. Marzinzik and B. Kollmeier. Speech pause detection for noise spectrum estimation by tracking power envelope dynamics. *IEEE Trans. Speech Audio Process*, 10, pp. 109- 118, 2002.
- [37] K. Li, N.S. Swamy and M.O. Ahmad. An improved voice activity detection using higher order statistics. *IEEE Trans. Speech Audio Process.*, 13, pp. 965-974, 2005.
- [38] G.D. Wuand and C.T. Lin. Word boundary detection with mel scale frequency bank in noisy environment. *IEEE Trans. Speech and Audio Processing*, 2000.