IJMLNCE JOURNAL

International Journal of Machine Learning and Networked Collaborative Engineering

Journal Homepage: http://www.mlnce.net/home/index.html

DOI : https://doi.org/10.30991/IJMLNCE.2018v02i04.004

A Machine Learning Approach for Speech Detection in Modern Wireless Communication Environment

[a]Shibanee Dash[*], [b]Mihir Narayan Mohanty

[a]R.V.R & J.C College of Engineering(Autonomous), Andhra Pradesh

shibaneedash.3@gmail.com , https://orcid.org/0000-0002-3828-1325

[b] ITER, Siksha 'O' Anusandhan(Deemed to be University),Bhubaneswar

mihir.n.mohanty@gmail.com , https://orcid.org/0000-0003-1252-949X

Abstract

Modern wireless communication has gained a improved position as compared to previous time. Similarly, speech communication is the major focus area of research in respective applications. Many developments are done in this field. In this work, we have chosen the OFDM modulation based communication system, as it has importance in both licensed and unlicensed wireless communication platform. The voice signal is passed though the proposed model to obtain at the receiver end. Due to different circumstances, the signal may be corrupted partially at the user end. Authors try to achieve a better signal for reception using a neural network model of RBFN. The parameters are chosen for the RBFN model, as energy, ZCR, ACF, and fundamental frequency of the speech signal. In one part these parameters have eligibility to eliminate noise partially, where as in other part the RBFN model with these parameters proves its efficacy for both noisy speech signals with noisy channel as Gaussian channel. The efficiency of OFDM model is verified in terms of symbol error rate and the transmitted speech signal is evaluated in term of SNR that shows the reduction of noise. For visual inspection, a sample of signal, noisy signal and received signal is also shown. The experiment is performed with 5dB, 10dB, 15dB noise levels. The result proves the performance of RBFN model as the filter. The performance is measured as the listener's voice in each condition. The results show that, at the time of the voice in noise environment, proposed technique improves the intelligibility on speech quality.

Keywords

Wireless Communication,

Orthogonal Frequency Division Multiplexing,

Speech detection,

Radial Basis Function Networks,

Bit Error Rate.

1. Introduction

In 70’s the Wireless and cellular concept was developed and become more popular than expected at the time. Since then the focus in this area increased till now. The clarity at the receiver end is highly essential. Day-by-day the communication techniques are developed along with the type of modulation. Researchers focus on the capacity as well as the reception capability [1]. The technology including OFDM,

[*] Corresponding author

Shibanee Dash

Email: shibaneedash.3@gmail.com

MIMO and coding types are to be observed for higher bandwidth utilization with better response in less time. Congestion in licensed spectrum increases in gradual manner for which unlicensed spectrum utilization has been occurred that is developed by researchers [2-3].

Many different factors are included in wireless communication, such as environmental factors, channel condition, types of speakers, and the way we transmit in the channel. Research of speech technology requires changing the way of transmission, and reception in specific type of communication [4].

The Speech is a common mode of Communication between of human being. Speech signal is an important data for communication network. The detection and recognition accurately is most important criteria. To maintain wireless communication successfully speech of human being should listen and speak clearly. As well the matching among network should be maintained perfectly. In such cases the synthetic speech is also verified in many problems [5-7] .

Speech has potential of being important mode of interaction with computer for this evaluation. For errorless communication and noiseless signal reception, initially speech signal is processed for noise suppression and enhancement. It helps in choosing the technique in digital age point of view [8-9].

It is a challenge for the next technological development to make the natural speech reception through HCI at the user end. Speech processing is exciting areas of research in signal processing and one type of pattern recognition problem. The choice and use of features should relevant for the purpose of detection. Again it must well manage at the time of training and testing on use of machine learning techniques.

Different Technologies are used for faster communication in both licensed and unlicensed spectrum utilization. These are

(i) OFDM is a FDM scheme where digital multicarrier modulation method is utilized. The data is subdivided into parallel data streams sharing every sub-carrier. Further every sub-carrier is modulated with a modulation method like QAM, and variant of PSK.

(ii) MIMO technique is used with multiple antennas along with transmitters and receiver to improve performance. It offers increase in data and receivers throughput and link range without additional bandwidth.

(iii) Turbo Code-It is a category of high performance error correction codes which was developed in 1993.The coding technique is one of the significant technologies in communication. It is helpful for maximum information transmission without error/ relatively small error.

The respective generations of wireless communication have many advantages like higher bandwidth, Better response time. It works at 2.6 GHz frequency implies that better coverage even though with same tower. The gradual developed generations provide higher flexibility as compared to already existing technologies [10].New technology with less cost and better usage needs to simplify hardware with effective design so that the versatility can be maintained with the same handset with better reception capability for different generations.

The paper is organized as follows. Section 1 introduces the work. Section 2 provides the methodology proposed in this work. It explains the principle of hearing the speech communication model. Through the model the speech is communicated. For detection purpose Radial Basis Function Network (RBFN) is used and explained with its parameters as the problem formulation. Section 3 explains the result and section 4 concludes the work.

2. Methodology

One important mechanism for received signal is source separation that has the capability to remove time-frequency regions where the speech signal is less distorted [11]. To increase the success of communication, adaptation is required at different levels, such as subject, place and vocabulary. As a result the Lombard effect [5] can be analyzed. The parameters for Lombard speech like intensity, vowel duration, speaking rate, energy distribution, spectral tilt, formant frequency are observed by researchers.

OFDM System

OFDM technique is used due to better transmission capacity and high bandwidth efficiency in wireless communication for both licensed and non-licensed spectrum. Such system is based on spreading technique with low rate carriers. The spacing between the orthogonal components is generated using the Fast

Fourier Transform technique [12].The data is converted to parallel stream and grouped. Further, it is modulated using either Quadrature Amplitude Modulation (QAM), or Quadrature Phase Shift Keying (QPSK), or Binary Phase Shift Keying (BPSK). Finally, required spectrum is then converted back to its time domain signal using an Inverse Fast Fourier Transform (IFFT). At the receiver end it is converted from parallel to serial for transmission of data. With this technology the system is designed by considering the Gaussian noise channel [13].

The basic model of OFDM system is presented in Fig. 1. Input signal as considered for transmitted symbols through the wireless Gaussian channel.

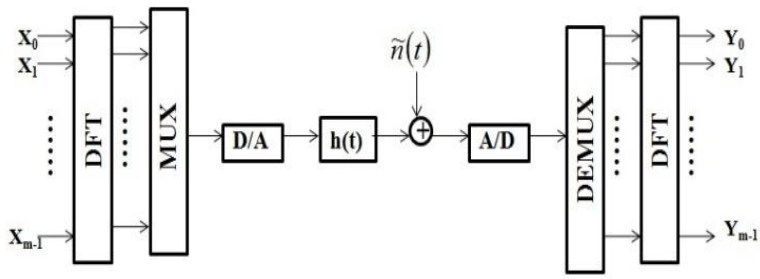

Fig. 1. Basic OFDM system

The impulse response of the channel can be expressed as,

$$\mathrm{h}(\mathrm{t}) = \sum_{m} a_m \delta(\mathrm{t} - \tau_{\mathrm{m}} \mathrm{T_s}) \tag{1}$$

Where α_m represents the amplitudes. It is formed by a N-point $\mathrm{DFT_N}$ and is expressed as,

$$\mathrm{y} = \mathrm{DFT_N}(\mathrm{IDFT_N}(\mathrm{x}) \otimes \frac{h}{\sqrt{n}} + n) \tag{2}$$

For N-independent channel the expression will be, $y_i = h_i X_i + n_i, i = 0.....N-1$

Where $h=(h_0,h_{1.....}.h_{n-1})^T$ that can be considered as attenuation of the channel and $n=(n_0,n_{1.....}.n_{n-1})^T$ is a noise vector. The system can be formulated as,

$$\mathrm{y} = \mathrm{XFg} + \mathrm{n} \tag{3}$$

Where X is the input data and can be expressed in terms of twiddle factor as,

$$F = \begin{pmatrix} W_N^{00} & \cdots & W_N^{0(N-1)} \\ \vdots & \ddots & \vdots \\ W_N^{(N-1)0} & \cdots & W_N^{(N-1)(N-1)} \end{pmatrix} \tag{4}$$

The twiddle factor is defined as,

$$W_N^{nk} = 1/\sqrt{N} e^{-j2\pi nk/N} \tag{5}$$

The MMSE estimate of h becomes,

$$\hat{h}_{MMSE} = R_{hy} R_{yy}^{-1} y \tag{6}$$

Where,

$\mathrm{R_{hy}} = \mathrm{E}\{\mathrm{hy^H}\} = \mathrm{R_{hh}F^H X^H}$

$$R_{yy} = E\left\{yy^{H}\right\} = XFR_{hh}FHX^{H} + \sigma_n^2 I_N$$

signifies the cross covariance matrix and the auto covariance. Again R_{hh} is the auto covariance matrix of h and σ_n^2 denotes the noise variance. Assuming these quantities to be known, the MMSE estimates (h_{MMSE}) will be,

$$\hat{h}_{MMSE} = F\hat{h}_{MMSE} = FQ_{MMSE}F^{H}X^{H}y \tag{7}$$

Where Q_{MMSE},

$$Q_{MMSE} = R_{hh}[\left(F^{H}X^{H}XF\right)^{-1}\sigma_n^2 + R_{hh}]^{-1}(F^{H}X^{H}XF)^{-1} \tag{8}$$

The LS estimator for channel impulse response h is analyzed as follows,

Similarly the least square channel estimator can be formulated as,

$$\hat{h}_{LS} = FQ_{LS}F^{H}X^{H}y \tag{9}$$

where,

$$Q_{LS} = \left(F^{H}X^{H}XF\right)^{-1} \tag{10}$$

considering the two equations we have,

$$\hat{h}_{LS} = X^{-1}y \tag{11}$$

From equation (7) and (11) it is shown the LS estimate has a high mean square error as compared to MMSE estimation technique.

Parameters for Speech Detection

Energy

It is defined as the squared signal. In speech signal case it is analyzed frame wise. Hence the short time energy is to be evaluated considering different windowed signal [14]. The energy of the speech signal reflects the amplitude variations. Short-time energy can define as:

$$Energy = \sum_{i=-\infty}^{\infty}[s(i)w(m-w)]^2 \tag{12}$$

where , *s(i)* represents the signal, *w(m)* represents the window and *En* represents the energy.

Zero-Crossing Rate

The rate of change of signal from positive to negative is defined as the Zero crossing Rate (ZCR). It is a measure of number of times in particular time interval/frame. As a result the amplitude of the speech signals passes through a value of zero. The zero crossing rate of a signal can be found by using

$$ZCR = \sum_{-\infty}^{\infty}|\text{sgn}[s(i)]\text{sgn}(i-1)|w(n-w) \tag{13}$$

where, $\text{sgn}[s(i)] = 1$ if $s(i) \geq 0 = -1$ if $s(i) < 0$

and $w(n) = \frac{1}{2N}$, $0 < n < N-1$(N is the length of signal) $= 0$, otherwise.

The model for speech production suggests that the energy of voiced speech is concentrated about 3

kHz as the spectrum fall of glottal wave and for unvoiced speech, the energy is found at higher frequencies. Since high frequencies imply high zero crossing rates, and energy. There is a strong correlation between zero-crossing rate and energy distribution. Therefore, another parameter is considered as autocorrelation co-efficient to keep relevancy at received signal.

Autocorrelation

For clean speech and separation of noise autocorrelation coefficient has a major role alike to energy and ZCR. It works not only for noise elimination, but also for smoothening the signal. It is a type of cross correlation and convolution. The relation is expressed as,

$$R_{xx}(j) = \sum_n x_n \bar{x}_{n-j} \tag{14}$$

It is of finite energy of Signal. Similarly for the measurement of frequency as low or high fundamental frequency is an important parameter of speech and is considered.

Fundamental frequency.

As the human voice varies over a range of frequencies, the fundamental frequencies cannot be considered as a specific value. Though it is an essential component of speech and speaker recognition, it has a similar application in voice communication and is taken as an attribute.

Radial basis function Network (RBFN) Model for Detection

Both RBFN is used and tested for detection accuracy. Different possible hybridization of features has been attempted. Radial basis function Network (RBFN) consists of an input layer, a hidden layer and a linear output layer. In this case, the Gaussian kernel as activation function is used and the distance is evaluated [15]. The hidden layer depends on a non-linear RBF activation function [16-17]. The output of the network is found as the distance between the input vector and the vector of the centre of the Gaussian function and can be expressed as [18-20].

$$\begin{bmatrix} y1 \\ y2 \\ \vdots \\ y_j \end{bmatrix} = \begin{bmatrix} R\|x_1 - c_1\| & R\|x_1 - c_2\| & \cdots & R\|x_1 - c_j\| \\ R\|x_2 - c_1\| & R\|x_2 - c_2\| & \cdots & R\|x_2 - c_j\| \\ \vdots & \vdots & \vdots & \vdots \\ R\|x_j - c_1\| & R\|x_j - c_2\| & \cdots & R\|x_j - c_j\| \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_j \end{bmatrix} \tag{15}$$

where, R is the RBF, c_j is the center, $\|x - c_j\|$ is the distance between input and the center. $x_1, x_2, \ldots x_j$ are represented as the inputs, $y_1, y_2 \ldots . y_j$ are the outputs and $w_1, w_2, \ldots w_j$ are the weights of the network. The target output is obtained by updating the corresponding weights. The output to weight and input is given as,

$$y = \sum_{j=1}^{N} R(\|x - c_j\|) w_j \tag{16}$$

where, w_j is the weight of the j^{th} center and N is the length of the signal. The structure of the network is shown in Figure1. The network is operated with the activation function that is Gaussian and is expressed as,

$$R(\|x - c_j\|) = \exp\left[-\frac{(x - c_j)^2}{2\sigma^2} \right] \tag{17}$$

where, σ is the width of the center. The network is trained with adaptive learning method and is described in following subsection as proposed method.

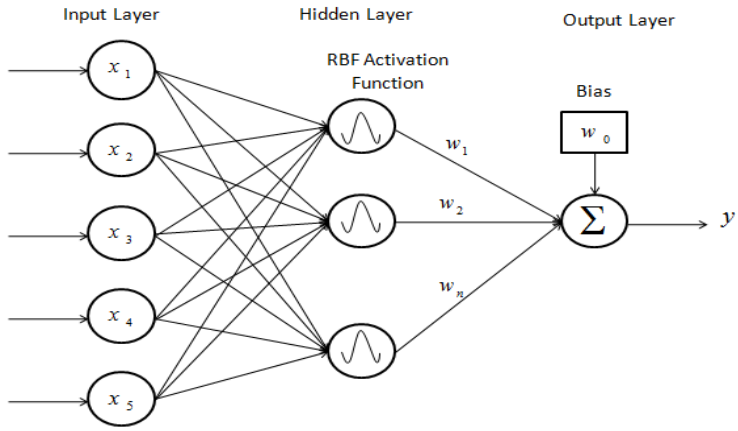

Fig 2. Structure of RBF Network

3. Result

The work consists of two parts as modulation technique and noise elimination through neural network model. The results are obtained from both the techniques and depicted in this section. The bit error rate is obtained to validate the OFDM system and is shown in Fig. 3. To strength it the MSE is found and is shown in Fig. 4.

Once the system found suitable, the chosen parameters are given to the RBFN model. One of the sample of speech is shown in the Fig.5 for visual aid. The corresponding outputs for original signal, noisy signal and obtained result with noiseless signal are shown in Fig. 6 and Fig. 7. From this result it is clear that the voice signal is well communicated and can be suitable for next generation wireless network.

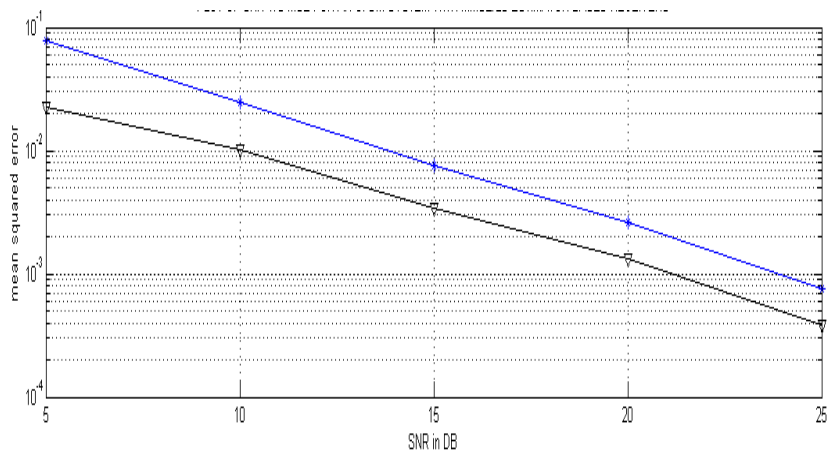

Fig 3. Plot of SNR vs. MSE for OFDM system

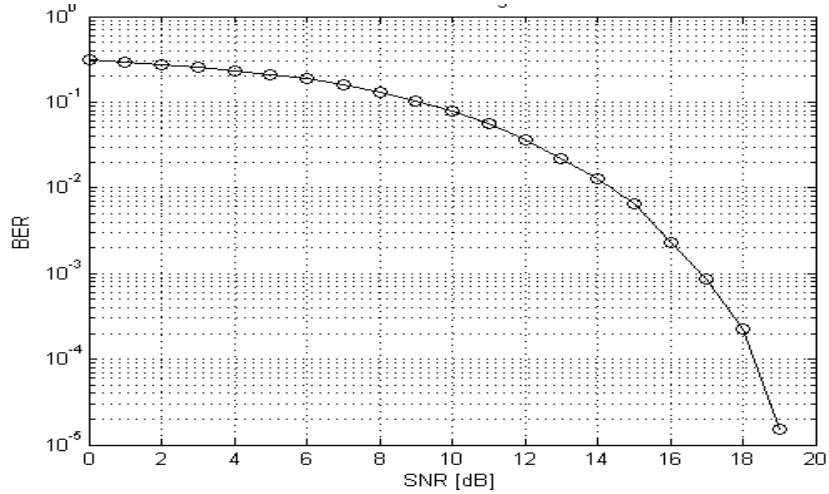

Fig 4. BER vs. SNR of OFDM system

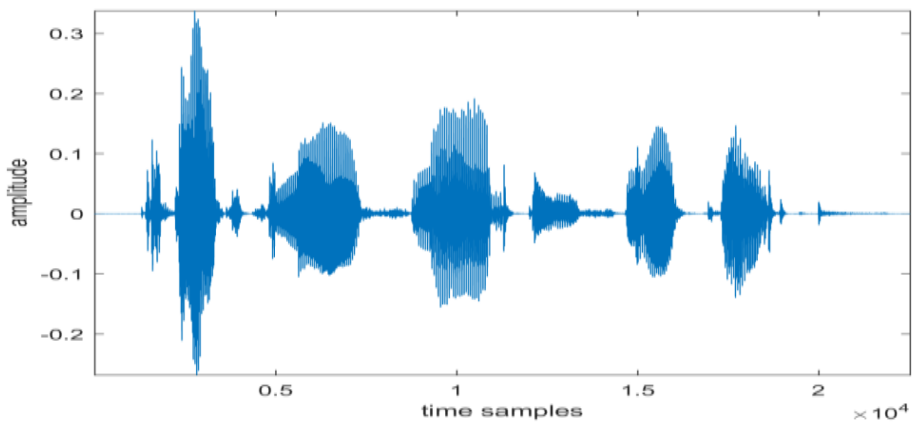

Fig. 5.Original Speech signal

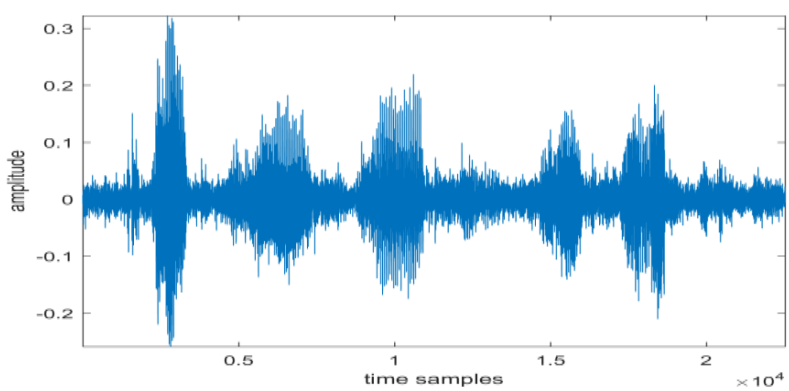

Fig. 6.Noisy Speech signal with SNR of 5dB

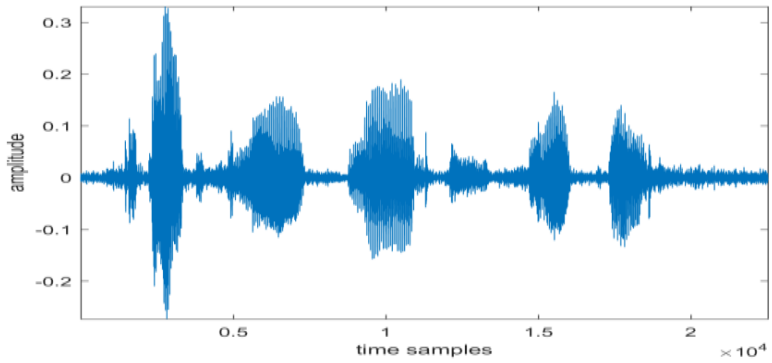

Fig.7: The Enhanced Signal at the Receiver End

Table 1: Accuracy performance of RBFN with different combinational features

Features	Average Accuracy	MSE
	80.44%	0.706
Z + ACF	80.53%	0.683
E_{sT} +$F0$	75.03%	0.813
$F0 + Z$	75.40%	0.808
E_{sT} + ACF	80.55%	0.680
$F0 + ACF$	75.33%	0.810
$E_{sT} + Z + ACF$	85.82%	0.436
$E_{sT} + F0 + ACF$	84.16%	0.487
$ACF + Z + F0$	83.14%	0.535
$E_{sT} + Z + F0$	83.18%	0.514
$E_{sT} + F0 + Z + ACF$	**90.38%**	**0.352**

E_{sT} = STE, $F0$= Fundamental Frequency,

ACF=Autocorrelation Coefficient,

Z= Zero Crossing Rate

4. Conclusions

From the work, it is concluded that the detection accuracy depends largely on the types and size of the features fed as input. Graphical analysis shows that Intensity or energy appears to be the best feature for detection. Recognition of emotional speech in communication can provide a new future direction.

References

[1]. Haykin, S. S. (2011). Modern wireless communications. Pearson Education India.

[2]. Mohanty, M. N., Mishra, L. P., & Mohanty, S. K. (2011). Design of MIMO space-time code for high data rate wireless communication. International Journal on Computer Science and Engineering, 3(2), 693-696.

[3]. Dash, S., & Mohanty, M. N (2018). Voice Detection for Cognitive Radio Receiver in Cooperative Spectrum Sensing Environment, AESPC (Accepted).

[4]. Tse, D., & Viswanath, P. (2005). Fundamentals of wireless communication. Cambridge university press.

[5]. Junqua, J. C. (1993). The Lombard reflex and its role on human listeners and automatic speech recognizers. The Journal of the Acoustical Society of America, 93(1), 510-524, doi.org/10.1121/1.405631.

[6]. Loizou, P. C. (2007). Speech enhancement: theory and practice. CRC press

[7]. Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. The Journal of the Acoustical Society of America, 84(3), 917-928, doi.org/10.1121/1.396660.

[8]. Mowlaee, P., Stahl, J., & Kulmer, J. (2017). Iterative joint MAP single-channel speech enhancement given non-uniform phase prior. Speech Communication, 86, 85-96, doi.org/10.1016/j.specom.2016.11.008.

[9]. Rabiner, L. R., & Schafer, R. W. (1978). Digital processing of speech signals (Vol. 100, p. 17). Englewood Cliffs, NJ: Prentice-hall.

[10]. Mohanty, M. N., & Mishra, S. (2013, March). Design of MCM based wireless system using wavelet packet network & its PAPR analysis. In Circuits, Power and Computing Technologies (ICCPCT), 2013 International Conference on(pp. 821-824). IEEE, doi/10.1109/ICCPCT.2013.6528867.

[11]. Cooke, M. (2003). Glimpsing speech. Journal of Phonetics, 31:579 – 584

[12]. Mishra, D., Mishra, S., & Mohanty, M. N. (2011). Estimation of MIMO-OFDM Based Channel for High Data Rate Wireless Communication. IJCSIT) International Journal of Computer Science and Information Technologies, 2(3), 1263-1266.

[13]. Mishra, B., Mishra, S., & Mohanty, M. N. (2012). Design of Wavelet Packet Based Model for Multi Carrier Modulation. International Journal of Engineering Science and Technology, 4(04), 1572-1575.

[14]. Mowlaee, P., Stahl, J., & Kulmer, J. (2017). Iterative joint MAP single-channel speech enhancement given non-uniform phase prior. Speech Communication, 86, 85-96, doi.org/10.1016/j.specom.2016.11.008.

[15]. Haykin, S. S. (2009). Neural networks and learning machines (Vol. 3). Upper Saddle River: Pearson.

[16]. Phooi, S. K., & Ang, L. M. (2006, November). Adaptive RBF neural network training algorithm for nonlinear and nonstationary signal. In Computational Intelligence and Security, 2006 International Conference on (Vol. 1, pp. 433-436). IEEE, doi.org/10.1016/j.specom.2016.11.008.

[17]. Palo, H. K., Mohanty, M. N., & Chandra, M. (2015). Design of neural network model for emotional speech recognition. In Artificial intelligence and evolutionary algorithms in engineering systems (pp. 291-300). Springer, New Delhi, doi.org/10.1007/978-81-322-2135-7_32

[18]. Cheng, J. C., Su, T. J., Li, T. Y., & Wu, C. H. (2015, September). The Noise Reduction of Speech Signals Based on RBFN. In Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2015 International Conference on (pp. 449-452). IEEE, doi/10.1109/IIH-MSP.2015.111.

[19]. Mohapatra, S. K., Palo, H. K., & Mohanty, M. N. (2017). Detection of Arrhythmia using Neural Network. Annals of Computer Science and Information Systems, 14, 97-100, doi/10.15439/2018KM42.

[20]. Singer, E., & Lippman, R. P. (1992, March). A speech recognizer using radial basis function neural networks in an HMM framework. In Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on (Vol. 1, pp. 629-632). IEEE, doi/10.1109/ICASSP.1992.225830.

Author's Biography

Shibanee Dash is presently working as a Assistant Professor in the Department of Electronics and Communication Engineering, at R.V.R & J.C College of Engineering (Autonomous), Guntur, Andhra Pradesh, India. She has Master of Technology in Electronics and Telecommunication at Kalinga Institute of Industrial Technology (Deemed to be University), India. She has 3 years of experience in teaching and research.

Mihir Narayan Mohanty is presently working as a Professor in the Department of Electronics and Communication Engineering, Institute of Technical Education and Research. Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India. He has published over 300 papers in International/National Journals and Conferences along with approximately 25 years of teaching experience. He is the active member of many professional societies like IEEE, IET, EMC & EMI Engineers India, ISCA, ACEEE, IAEng, CSI and also Fellow of IETE and IE (I). He has received his M.Tech. Degree in Communication System Engineering from the Sambalpur University, Sambalpur, Odisha and done his Ph.D. work in Applied Signal Processing. His area of research interests includes Applied Signal and image Processing, Digital Signal/Image Processing, Biomedical Signal Processing, Microwave Communication Engineering and Speech Processing.

How to Cite

Dash S., & Mohanty M. N. (2018). A Machine Learning Approach for Speech Detection in Modern Wireless Communication Environment. *International Journal of Machine Learning and Networked Collaborative Engineering*, *2*(04) pp 170-179.

https://doi.org/10.30991/IJMLNCE.2018v02i04.004