Jestr
Journal of Engineering Science and Technology Review 15 (3) (2022) 193 - 196
Research Article
JOURNAL OF
Engineering Science and Technology Review
www.jestr.org

Identification of Noises and Speech Signals by Artificial Neural Networks

Ivelina Balabanova[1], Stela Kostadinova[2], Valentina Markova [2], Panagiotis Kogias[3], Dionisia Daskalaki[1], Stanimir Sadinov[1] and Georgi Georgiev[1,*]

[1]Technical University of Gabrovo, Department of Telecommunications Equipment and Technologies4 Hadji Dimitar Str., 5300 Gabrovo, Bulgaria
[2]Technical University of Varna, Department of Communication Engineering and Technologies, 4 Hadji Dimitar Str., 5300 Varna, Bulgaria
[3]Department of Physics, International Hellenic University,Ag. Loukas, 65404 Kavala, Greece

Received 23 October 2020; Accepted 9 June 2022

Abstract

According to the studies, the most commonly used mathematical apparatuses for signal recognition tasks are Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) and Deep Neural Networks (DNN). This paper presents an innovative approach related to the possibility for identification of accidental noise impacts and human speech with superimposed presence of noise by Backpropagation Neural Networks (BNN) in different transfer functions. BPNs with linear, tangent-sigmoid and log-sigmoid transfer functions in the output layers are tested. A neural architecture for noise recognition in 6 neurons defined in the hidden layer with "tansig" activation output and achieved accuracy of 98.3% is selected. In the speech processing, an identical efficiency of 93.7% in 4, 3 and 4 hidden neurons for all types of output transfer functions was observed.

Keywords: noise identification; noisy speech signals recognition; accuracy; mean squared error.

1. Introduction

Convolutional Neural Networks are proving to be one of the most commonly used models in the field of signal recognition. An example of this is the registration of the presence of Gaussian noise and the establishment of its levels in the processing of standard test images added by the LPG-PCA method. Another area of appliance is the identification of real-time voice activity in mobile phone applications based on the analysis of noisy speech signals. Voice activity registration is also based on the widespread use of multilayer Recurrent Neural Networks by analyzing the acoustic characteristics in each test frame. In speech recognition, studies with combined approaches such as Deep Neural Networks and Hidden Markov Models (HMMs) are often found, related to the analysis of voice data in heterogeneous groups of speakers. Another study examined the methodological combination of CNN, RNN and DNN for speech detection in the processing of large audio data for music event and speech detection [1-5].

Another significant problem in information acommunication systems is related to the reduction of the effect of noise effects. In this regard, according to studies the processes of noise assessment and reduction, there is the applicability of specialized approaches and algorithms, among which can be specified, respectively [6-10]:

- Independent Component Analysis, Recursive Least Squares (RLS), and Recurrent Neural Networks in automated speech recognition systems;
- Support Vector Machine (SVM) method, k-means cluster analysis, k - Nearest Neighbors (k-NN), in optical communications;
- DNNs and Convolutional Neural Networks in Orthogonal Frequency-Division Multiplexing (OFDM) and Two-Dimensional Magnetic Recording (TDMR) systems;
- RNN neural architectures in combination with Long Short Term memory (LSTM) - a variant of RNNs, and Gated Recurrent Unit (GRU-GRU) in Micro-Electro-Mechanical System Inertial Measurement Units (MEMS-IMU);
- Deep-Learning Neural Networks (DLNNs) in the development of biomedical electronic devices;
- Principal Component Analysis, CNN, Feed-Forward Neural Networks in case of parameter reduction - SNR; SNR RMS; Peak SNR and SSIM in image diagnostic systems;
- Linear regression analysis, Discriminant analysis, Naïve Bayes algorithm, Decision tree method, Adaptive neural-fuzzy interface systems in electronics.

The paper discusses the problem of identification of simulated noise effects – Uniform White Noise (UWN), and Periodic Random Noise (PRN), and speech signals with presence of Gaussian White Noise (GWN) and PRN based on the study and selection of artificial backpropagation neural networks, trained by the Levenberg-Marquardt algorithm.

2. Experimental Noise Identification by Artificial Neural Networks

*E-mail address: givanow@abv.bg
ISSN: 1791-2377
doi:10.25103/jestr.153.21

Object of the task is the some of the most common types of noise, as follows:

- Uniform White Noise;
- Periodic Random Noise.

These noises are one of the most common effects in communication channels, which is why they were chosen as the object of study. At the initial stage, the interference effects can be modeled according to specific parameters such as amplitude, sampling rate, etc., using different software environments. The signals were obtained after a simulation via a virtual application using the LabVIEW engineering software.

In the second phase, after generating the test noises, the synthesis of models for their identification by artificial intelligence was started. The main goal of this stage of the study is related to the analysis of:

- the applicability of BNNs with Levemberg-Marquardt training algorithm as an analytical tool for noise detection in identical Noise amplitude = 0.02;
- the efficiency of the neural models for identification when submitting only single input effects without the need for more input variables compared to accepted quality indicators – Classification accuracy and Mean-Squared Error (MSE);
- the indicated quality indicators according to different types of activation functions in the output layers of BNNs, as follows – linear “purelin”, hyperbolic tangent sigmoid “tansig” and log-sigmoid “logsig”.

Three-layer BNN architectures with different types of activation functions in output layers were used for the test signals recognition, and a number of the neurons are ranging from 5 to 20 in the hidden layers. Regarding the experiment partial results are presented.

Table 1 contains the data from the process of synthesis of architecture with an initial tangent-sigmoidal function, in which the best indicators for UWN and PRN recognition were observed. From the given results it can be seen that the first indicator changes in the range of 94.0% in 12 to the highest 98.3% in 6 intermediate neurons. A maximum of 0.0488 and a minimum 0.0162 threshold for the Mean-Squared Error with 13 and 6 hidden neurons were registered for MSE, the latter being selected as the most suitable for defined noise categories recognition.

Table 1. Results in ANNs with “tansig” transfer function in output layer for noise identification

Hidden neurons	Accuracy, %	MSE
5	96.70	0.0353
6	98.30	0.0162
7	95.70	0.0439
8	97.70	0.0220
9	96.0	0.0380
10	98.0	0.0250
11	97.6	0.0296
12	94.0	0.0484
13	95.0	0.0488
14	97.3	0.0281
15	97.7	0.0204
16	95.3	0.0434
17	95.7	0.0375
18	96.0	0.0314
19	97.0	0.0261
20	95.3	0.0378

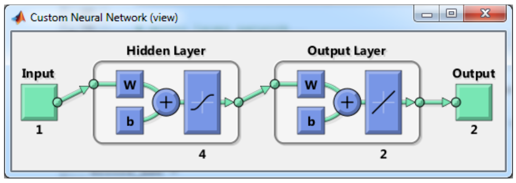
3. Analysis of Artificial Neural Networks of Noisy Speech Signals

In view of the found positive indications when using the neural pair with backpropagation of the error, we proceeded to the analysis of different groups of test speech signals with the presence of GWN and PRN in different RMS noise values. The activities here refer to the establishment of close or higher levels of the observed quality indicators, giving grounds for confirming or rejecting the adequacy of artificial intelligence in identifying different types of signals.

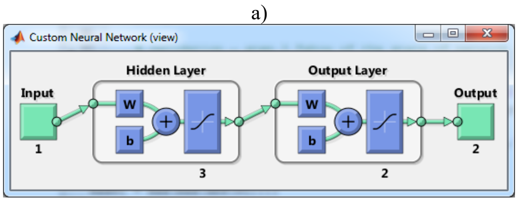
By analogy with the one suggested in the previous section, “Speech with GWN” and “Speech with PRN” were simulated in LabVIEW in identical RMS values, respectively 0.01. A study was done, at an interval of neurons in the hidden network layers, identical to the indicated output types of functions, respectively from 3 to 15. According to the analysis of the quality criteria in a sequential order, at the output activation type:

- “purelin” – accuracy ranges from 88.3% to 93.7% at 8 and 4, while MSE ranged from 0.0621 for 4 to 0.0949 for 8 intermediate neurons;
- “tansig” – maximum of 93.7% and minimum of only 51.7% accuracy and respectively lowest rms errors of 0.0598 and highest of 0.2833 were recorded at 3 and 8 hidden neural units;
- “logsig” – it is observe a general trend of significant increase of MSE for all experimental architectures. Despite the highest success rate of 93.7%, the best error found is 0.1560 with 3 hidden neurons, giving reason for the corresponding ANN and output activation type to be evaluated with the worst adequacy.

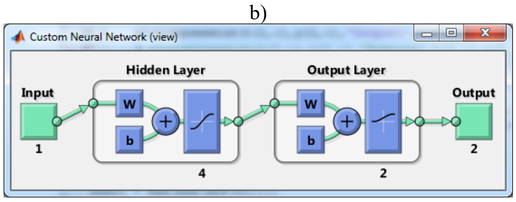
The final synthesized ANNs are presented in Fig. 1.



a)



b)



c)

Fig. 1. Selected ANNs with a) linear, b) tansig and c) logsig output transfer functions for identification of noisy speech signals

Fig. 2 gives a notion of the nature of the change in the mean-squareд errors in the implementation of training, validation and test procedures on the selected networks within 49 for purelin ANN, 21 for tansig ANN and 27 iterations for ANN with logsig activations.

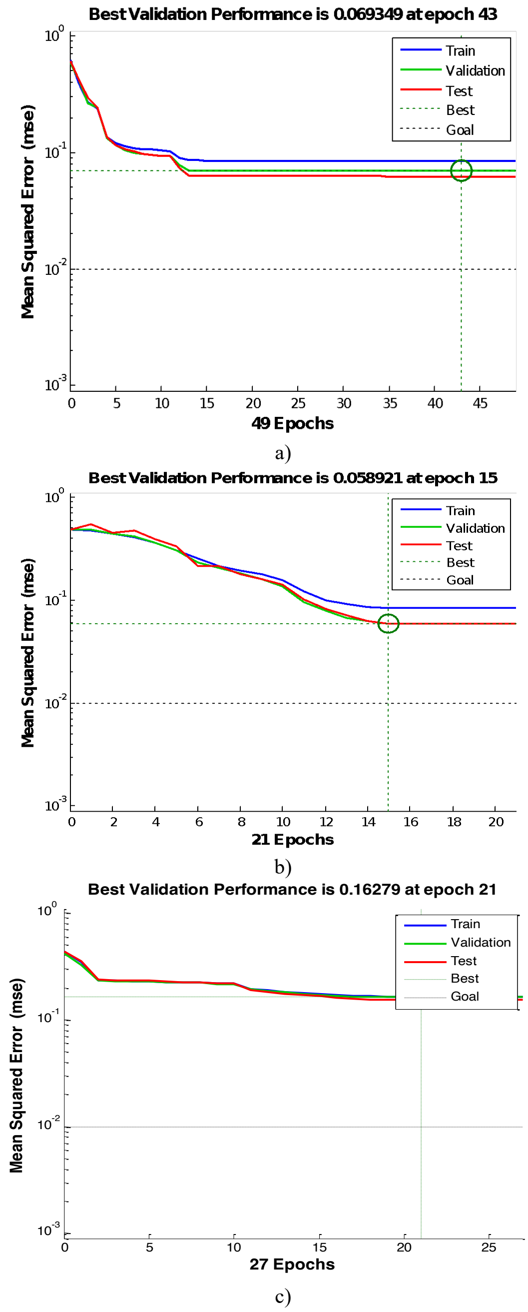


Fig. 2. MSE for ANNs with a) purelin, b) tansig and c) logsig output transfer functions for speech recognition

First of all, it can be said that there are no indications of emerging "neural retraining", i.e. No sharp increases in the studied types of MSE were observed. An identical feature of achieving the best validating network performance is observed, as follows 0.069349, 0.058921 and 0.16279 at 43 for ANN with purelin, 15 for BNN with tansig and 21 epochs within the training for ANN with logsig transfer functions.

The linear regression diagrams of Fig. 3 confirm the existence of good similarity between set targets and the results calculated by ANN. In sequential order at:

- validation process – the correlation coefficients R = 0.93234, R = 0.94327 and R = 0.92481 were observed at ANN with purelin, tansig and logsig activation functions;
- test procedure R = 0.9365 for ANN with purelin, R = 0.95041 for ANN with tansig and R = 0.92373 iterations for ANN with logsig activations are ontained.

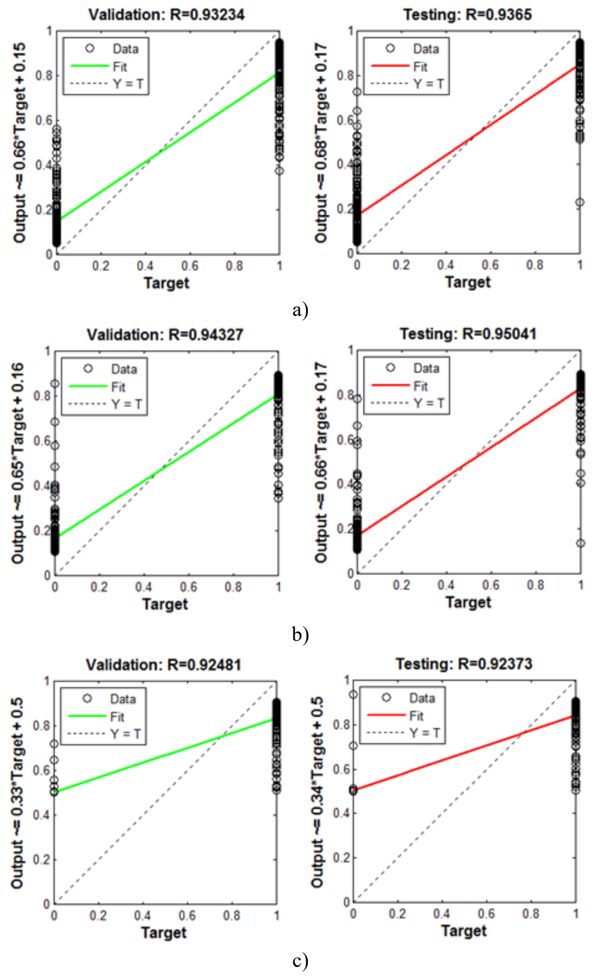
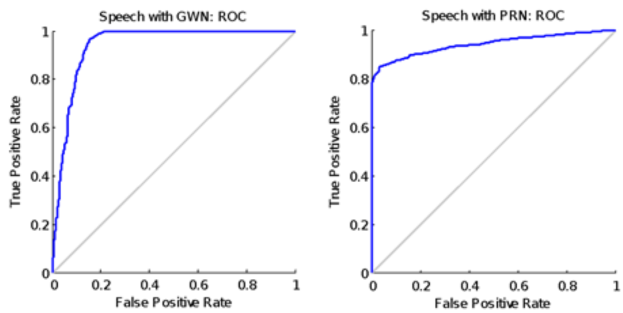


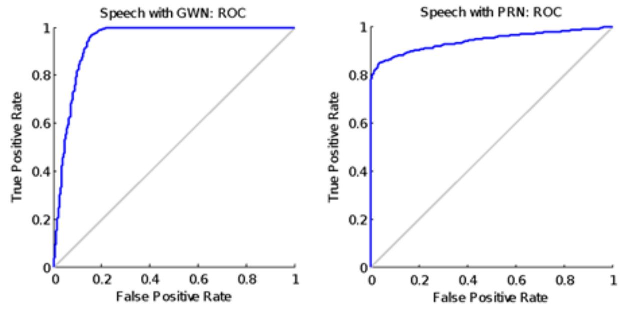
Fig. 3. Linear regression diagrams in ANNs with a) linear, b) tansig and c) logsig output transfer functions for noisy speech signals recognition

The constructed ROC diagrams for visual analysis of the quality of classification of the declared basic signal groups are presented in Fig. 4. Higher recognition efficiency was found in the first classification category.

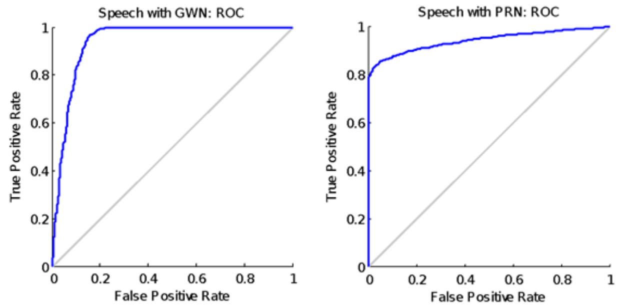
According to the analysis, the calculated differences between the estimated and the desired expected results of the operation of the ANN fall into the following numerical ranges "from -0.9422 to 0.9422", "from -0.8922 to 0.8922" and "from -0.500 to 0.9368" (fig. 5). There are a significantly higher percentage of increased errors in ANN with logsig compared to those with purelin and tansig transfer functions.



a)



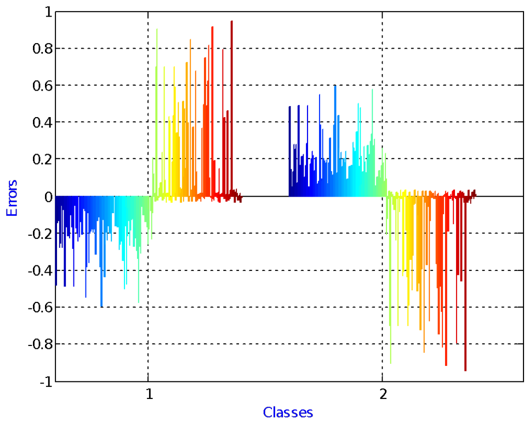
b)



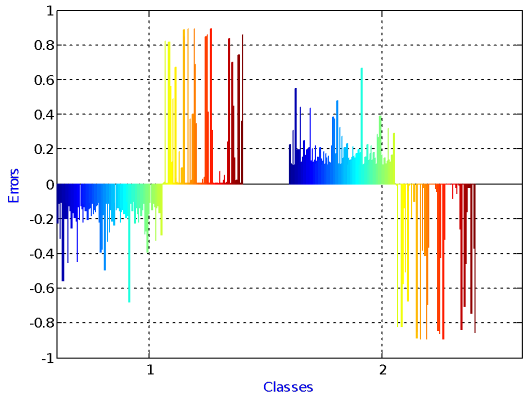
c)

Fig. 4. ROC curves for selected ANNs with a) purelin, b) tansig and c) logsig output transfer functions for identification of noisy speech signals

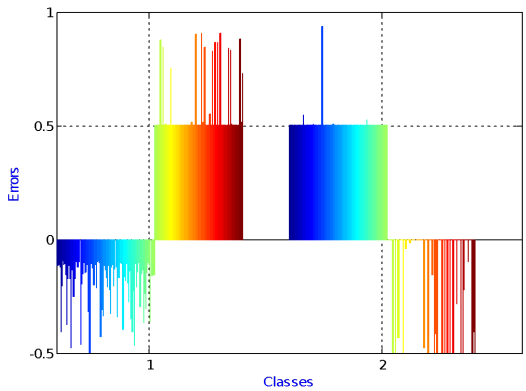
About identification task, ANN with 3 intermediate neurons and tansig transfer function in network output is defined as the most appropriate architecture.



a)



b)



c)

Fig. 5. Errors in ANNs with a) linear, b) tansig and c) logsig output transfer functions

4. Conclusion

The presented study confirms the advantages of artificial intelligence for signal identification. Emphasis is placed here on the potential applicability of ANN backprogation in qualifying interference in communication channels in telecommunication systems and speech recognition in voice analysis systems.

References

1. J. Chuah, H. Khaw, H. Soon, and Ch. Chou, Proc. of the IEEE Region 10 Conference Malaysia, pp. 2447–2450, (2017).
2. A. Sehgal, and N. Kentarnavaz, A Convolutional Neural Network Smartphone ppp for Real-Time Voice Activity Detection, IEEE Access, vol. 6, pp. 9017-9026, (2018).
3. T. Hughes, K. Mierle, Recurrent Neural Networks for Voice Activity Detection, IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 7778-7382, (2013).
4. L. Serizel, and D. Giuliani, Deep Neural Network Approaches for Speech Recognition with Heterogeneous Groups of Speakers Including Children, Natural Language Engineering, vol. 23, pp. 325-350, (2016).
5. D. Benito-Gordon, A. Lozano-Diez, D. Toledano, and J. Gonzalez-Rodroguez, Springer EURASIP Journal on Audio, Speech and Music Processing, pp. 1-18, (2019).
6. F. Avendano, A. Soto, N. Gonzalez, G. Serafino, P. Ghelfi, and A. Bogoni, Machine Learning Techniques for Noise Analysis in Optical Systems, Intelligent System and Computing, pp. 1-15, (2019).
7. A. Hasan, and Th. Shongwe, Impulse Noise Detection in OFDM Communication System Using Machine Learning Ensemble Algorithms, Advances in Intelligent Systems and Computing, pp. 1-9, (2017).
8. C. Fan, X. Yuan, and Y. Zhang, CNN-Based Signal Detection for Banded Linear Systems, IEEE Globecom, pp. 1-27, (2018).
9. Sh. Han, Zh. Meng, X. Zhang and Y. Yan, Hybrid Deep Recurrent Neural Networks for Noise Reduction of MEMS-IMU with Static and Dynamic Conditions, MDPI, Micromachines, pp. 1-23, (2021).
10. Y. Lai, Y. Tsao, X. Lu, F. Chen, and Y. Su, K. Chen, Y. Chen, L. Chen, L. Li, and Ch. Lee, Deep Learning–Based Noise Reduction Approach to Improve Speech Intelligibility for Cochlear Implant Recipients, Wolters Kluwer Health, pp. 1-15, (2018).