Applying the Artificial Neural Networks with Multiwavelet Transform on Phoneme recognition

Baydaa Jaffer AlKhafaji, May A. Salih, Shaymaa AbdulHussein Shnain, Omar Adel Rashid, Abdulla Adil Rashid, Moheeb Tariq Hussein

Computer Science Department, College of Education for Pure Science/Ibn Al-Haitham, University of Baghdad Iraq.
Math and Computer Science Department, Basic Education College, University of Babylon, Hilla, Iraq.

baydaa.j.s@ihcoedu.uobaghdad.edu.iq
bjkh68@yahoo.com
baydaa.khafaji@gmail.com
may.abd@uobabylon.edu
omar.adel.rashed.b@gmail.com
abdullah.adil@ihcoedu.uobaghdad.edu.iq
username.mt77@gmail.com

Abstract. There are several advantages of Phoneme recognition identification. It is easier to use speech for data entrance spoken communication for data ingress than other tools. It allows writing user-friendly data entrance exploiter friendly data ingress programs. There are several difficulties in speech voice communication recognition. One of these difficulties is noise. Variability in speech is another problem. Even the speech of same speaker varies. The ability of artificial neural networks to generalize and optimize more quickly than some conventional algorithms algorithmic rule has been observed in different areas of research inquiry such as speech and pattern convention recognition, financial forecasting prognostication, image data compression and noise reduction simplification in signal processing. Neural networks take advantage of the redundancy incorporated in their distributed processing structures the proposed system depends on Artificial Neural Networks Network as decision making qualification algorithm to find the best match peer for the tested phonemes. Phoneme. The data used in this project are Arabic phonemes language phoneme stored as 8-bit mono infectious mononucleosis 8000Hz PCM WAVE Sound Auditory sensation file. The results showed that the accuracy of the proposed system is 98% recognizes the phonemes efficiently.

Keywords. MWT, ANN’s, LVQ, PCM WAVE

1. Introduction
One of the most recently developed approaches to pattern recognition has been the use of artificial neural networks (ANN's), which are now finding application in a wide variety of scientific disciplines [2]. Neural networks are fundamentally instructive frameworks comprised of various basic and profoundly
interconnected handling components, which measure data through the dynamic state reaction to outside information.[3]

Neural networks are massively parallel-distributed processing systems, which can improve their performance through dynamical learning. The weighted inputs (or the states) of the network are processed nonlinearly by the processing elements (or the neurons) to produce a single output. The weighted inputs vary through a given learning rule [1].

The dynamical learning property of the neural network can be applied in recognition systems. The states of the neurons can represent the elements of the adjustable model, which is compared with the system to be recognized [5].

The learning rule is designed in such a way as to minimize the error between the system and the model. There exist many different types of neural networks, and a variety of learning or training algorithm associated with them. They can be applied both to unsupervised or supervised pattern recognition [14].

2. A proposed system for speech recognition using MWT and ANN's

The proposed system serves the same objective of the system proposed which is mainly, recognize the Arabic phonemes. In the Figure 1. Shown the user interface of the proposed program

Where the data passes through many processing illustrated in the block diagram shown in Figure 2.
1- Input Data

The data structure of the input of the system is discussed in detail in section 2 which properties is 8-bit mono 8000 Hz.

2- Data framing

The data framing is putting the data stream in fixed-size block which is 128 each for further processing.

3- Feature extraction

Feature extraction is done by implementing the Multiwavelet Transform using the Over sampling algorithm.

This procedure will cause the data to be compressed as well so to use this property the used output of each frame would be 128 samples which have all the main features of the speech signal.

4- decision making
This block of the system processes the extracted properties of the data speech signal using the Learning Vector Quantizer Neural Network.

The Figure 3. Shows the architecture of the LVQ neural network as each output unit has a class to represent it.

![Learning vector quantization neural net](image)

**Figure 3. Learning vector quantization neural net**

The motive behind using the LVQ network algorithm is to find the closest output unit to the input vector in order to achieve this end. If x, w they belong to different classes, the weights will move away from the input vector. If x, w they belong to identical classes, then weights move towards New input vector

- \( x \) Represents the training vector \((x_1..x_j...x_n)\)
- \( T \) Denotes the correct class of the training vector.
- \( w_j \) Represent the weight vector \( j \) represents the output unit \((w_{ij}...w_{ij}...w_{nj})\).
- \( C_j \) Refer to a class represented by the \( j \)th output unit.
- \( ||x - w_j|| \) Represents the Euclidean distance between the weight vector of the output unit \( j \) and the input vector

3. **The Algorithm**
4. Initialization strategies
The simplest method of instating the loads reference is to take the main \( m \) vector and use them as weights vectors, the rest of the vectors are then utilized for preparing. Where there is another way to determine the initial weights and classification randomly.

Where the reference file is a text file. When the this system ask for name, then the process on the data in the file will take first the number of frames stored in 2 bytes because it is defined as Integer. second, the first 128 bit frame placing it in a matrix to be collected with the others ) then skip the rest blocks to the 2 bytes that specify the number of the next Phoneme.

5. The learning phase
This phase include data entering the system and get framed, windowed, transformed then used to learn the system till the last epoch of data is passed in and updating weights is no longer useful.

6. The testing phase
This is the last block of the system that consist of entering a single phoneme to be framed, windowed, transformed then finally finding the best match of this phoneme using the ANN’s.

When the file is chosen it will pass through all the steps of the procedure described above and then the message will appear to give the number of the resulted closest match cluster to the tested phoneme.

7. Conclusions
For so many years researches have been made to create a way that makes computers interact with human users. this kind of interaction is done by making the computer recognize phonemes that form words which in its turn forms sentences that the language consist of.

The phonemes that had been taken as a case study in this project are stored as a PCM WAVE files with properties 8-bit samples, mono channel, 8000 Hz sampling rate.
And for computer processing purposes and extracting phonemes from the carriers carrying them WAVE files are saved as text files because of the changes in the structure of the WAVE file beside the changes made on the type of the data.

after quizzing the data as PCM WAVE files, these files are framed in a 128 bits blocks, windowed with Hamming window, and its features are extracted using Multiwavelet Transform

In this case the decision making is done by Artificial Neural Networks (Learning Vector Quantization Algorithm).

After the reference creation process in the first proposed system, we used the values of the first blocks of each phoneme to generate the weights file. the training phase consist of all the above processes and after the adaptation of the weights are saves in a text file. The weight file saves the properties and the features of all the phonemes. The testing phase begins by getting any test file in the same procedure described above, So the phoneme with the matched features is the matched phoneme.

And the results show that this strategy recognize the Arabic Phonemes with accuracy about 98% for the tested phonemes where the study cases are 35 phonemes for 7 speakers in 8 different experiments

References
[1] Steven E. Golowich "A Support Vector/HMM approuch to phoneme recognition"October, 14, 1998 Bell labs, Lucent Technology.
[2] Al-Khafaji, B.J. Proposed Speech Analyses Method Using the Multiwavelet Transform, Ibn Al-Haitham Journal For Pure And Applied Science.2014, vol.27 No 1.
[3] Mohamed Debyech "Phoneme recognition system basec on HMM with distributed VQ codebook" 2000, IEEE Trans on comm.,vol 28.
[4] BJ AlKhafaji, M Salih, S Shnain, Z Nabat, improved technique for hiding data in a colored and a monochrm images,2020, Periodicals of Engineering and Natural Sciences 8 (2), 1000-1010
[5] Phillip Dermody "The use of Wavelet Transforms in Phoneme Recognition "2000
[6] Sang-Hwa Chung,"A parallel phoneme recognition algorithm based on continuous HMM"2000 Pusan National University.
[7] BJ AlKhafaji, MA Salih, S Shnain, Z Nabat,segmenting video frame images using genetic algorithms2020, Periodicals of Engineering and Natural Sciences 8 (2), 1106-1114
[8] Fabio Arciniegas "Phoneme recognition with staged neural networks" 2000, IEEE. Trans on comm.,vol 14 pp.1033-1037.
[9] Al-Khafaji, B.J. DetectTheInfected Medical ImageUsing Logic Gates. Ibn Al-Haitham Journal for Pure And Applied Science.2014, 27, 2, 260-267.
[10] Jinjin Ye," phoneme classification using naïve bayes classifier in reconstructed phase space "2002, NSF no. IIs-0113508.
[11] Santitham Prom-on, "Consonant Phoneme Recognition using Hidden Markov Model in Thai
Language": International Symposium on Information Theory and its Applications (ISITA 2002), October 7-11, 2002, Xi'an, PRC, pp. 743-746

[12] Erhan Mengusglu "Using multiple codebooks for Turkish Phoneme recognition"2002.M.sc thesis, Hacettepe University.

[13] Mukul Bhatnagar, B.E., "A Modified Spectral Subtraction Method Combined With Perceptual Weighting For Speech Enhancement ", M.Sc.Thesis 2002,The university of Texas At Dallas.

[14] AL-Khafaji, B.J. Image Improvement Using The Combination of Wavelet and Multiwavelet Transform. Ibn Al-Haitham Journal For Pure And Applied Science.2010, 23, 3, 275-282.

[15] Eric Nutt "speech sound production: recognition using recurrent neural networks "December 2003, Phd thesis, University of Massachusessts.

[16] David Auerbach "stop consonant classification using recurrent neural network "2002, Swarthmore College.

[17] Keiichi Tokoda "A very low bit rate Speech coder usins HMM-based speech recognition / synthesis techniques "2000, Nagoya Institute of Technology

[18] John J. Mccarthy "The phonetics and Phonology of semitic Pharyngeals"2002 PhD. thesis, University of Massachusessts.

[19] Vasily Strela, "Multiwavelets: Theory and Applications", 1996, M.Sc. Massachusetts Institute of technology.

[20] Javier Echauz "Elliptic and Radial wavelet Neural networks "Second World Automation Congress (WAC’96). Montpellier, France, May27-30, 1996.TSI Press, Vol.5, pp.173-179.

[21] Ziad J. Mohamed "Video Image Compression based on Multiwavelet Transform "2004 Ph.D. Thesis, University of Baghdad.