

School of Computing, Creative Technology and Engineering

**Module: Fundamentals of Computer Science**

**Academic Year: 2021/22**

**Level 4: Term 1**

**Assignment Title: Academic Report**

**Date Due: 9th January 2022**

**Tutor: Saroj Shakya**

**Student Name: Karan Chaudhary**

**Student ID: 77297911**

**Enhanced Voice recognition system using Artificial Neural Networks**

Contents

[Introduction 3](#_Toc272490025)

# Voice Recognition System 3

## Issues of Voice Recognition 4

## Artificial Neural Networks 5

## Pattern Matching using Neural Networks 5

## Back-Propagation Neural Networks 6

## Self Organizing Maps 6

## Enhancing voice recognition using back-propagation algorithm 7

## Conclusion 9

# References 10

Introduction

In today's environment, speech recognition and voice recognition are becoming increasingly popular. It is significantly more convenient to utilize the voice recognition technology to operate all of the devices. However, there are a few flaws in this approach, which are detailed below. We employ Artificial Neural Networks to overcome these problems. There are numerous algorithms in ANNs. The following are some of the algorithms. One of the algorithms, Back-propagation, is significantly better at enhancing the voice and has a 100% success rate. It filters the speech and uses a matching method to match it to samples kept in the database in order to identify the speaker.

Voice Recognition System

A voice recognition system is computer software that allows users to manage their computers with their voices rather than using the keyboard and mouse. This technology is the first step toward the next generation of human-computer interfaces, with the potential to improve the efficiency and usability of computer programs. (Zmalt .J, 2005)

When voice recognition software receives analog audio signals and converts them to digital signals is known as Analog-to-digital conversion. For a computer to decode signals or parameters provided by users, it must have a digital database that stores and checks speech patterns on a memory.

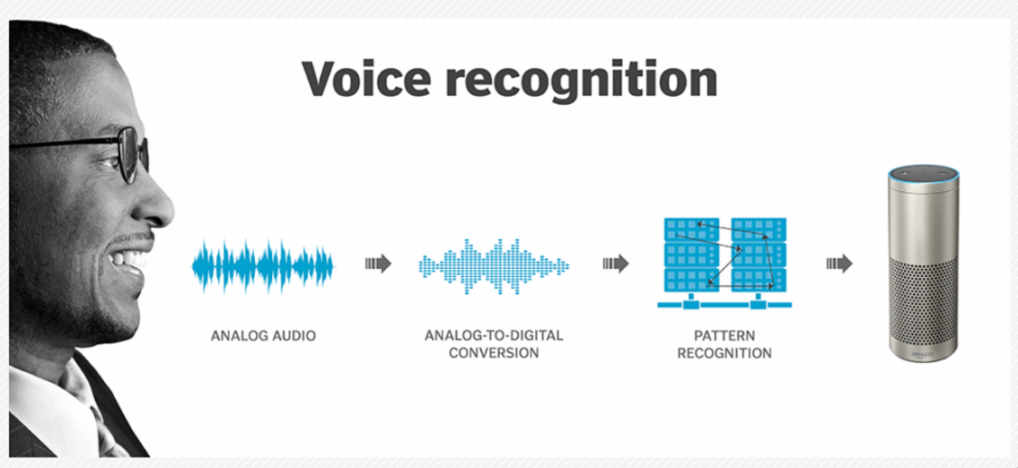


Fig.1. Work flow of voice recognition

This system detects a user by comparing the input data to the memory's relevant templates. The user is identified when the match happens. This approach should be contrasted with speech recognition, which uses vocal patterns to identify the user's voice. To identify the users, they all used the same test or steps. Using sound recording software, a group of people's voices are captured in voice files on the computer. Digital signal processing techniques are used to convert the data from the time domain to the frequency domain. The audio signal's frequency spectrum is used to train a neural network to create input and determine the users' identities. After then, it's employed to separate the user from the group (G.K, K, & V, 2002).

Issues of Voice Recognition

Every system has an issues, like that this system also have an issue. Some of the issues are listed below.

1. Noise from background

Sound waves created by speech are detected by voice recording machines. Multiple voices from the background can be used to pick up with the user's voice. It makes it difficult for the system to interpret and differentiate between the different sound waves produced in background and user's voice. This makes the system difficult to understand, confused, and also limits the system's processing ability. (RAF100STEAM, 2017)

1. Speedy Talking, Dialect and More

A clear and distinct voice is required when using voice recognition software so that it can understand the user. Things like rapid speech or a jumbled word on the software. If the user spoke too quickly, the system might miss the words and steps to execute.  Fast speech causes mispronunciation and makes it difficult for the program to understand. (SAmin, 2019)

1. A Speaker’s Distance from the Microphone

All voice recognition systems and devices need the microphone to receive sound as a wave. The words spoken through the microphone are processed by the program. The program picks up confused speech if the speaker speaks too close to the microphone. On the other hand, if users speak too far away from the microphone, then risk missing words. (Kikel. C, 2021)

1. Similar Sound and Machine error

Some of the terms sound the same but are written differently. When the system is unable to determine the word the user is attempting to say, the user must repair the errors by hand. Voice detection accuracy levels have substantial error rates. Machines still make around 8% to 12% of the time, which is more than twice as much as people do in their daily speech. (RAF100STEAM, 2017)

Input: sound wave saying “there”

Speech processing and recognition (detail later)

Output: “their”

Fig.2. Voice Recognition Issues in Block Diagram

Artificial Neural Networks

A neural network is a computational framework based on biological neural processing research. Neural networks come in a variety of configurations, ranging from simple to deep. The ability to learn complex nonlinear input-output relationships, implement sequential training techniques, and adapt themselves to the data are the main characteristics of neural networks.

1. Pattern Matching Using Neural Networks

Artificial neural networks (ANNs) are intelligent systems that are connected in some way to a simplified biological model of the human brain. They are made up of several little components called neurons that work in parallel and are coupled in the forward direction by connection weights, which are multipliers. To train neural networks, the values of these connection weights between network sections are modified (G.K, K, & V, 2002). Neural networks have applications in system identification, pattern recognition, classification, speech recognition, natural language processing, and other disciplines because they are self-learning, fault tolerant, and noise immune. (Malur,K.S, 1998)

**Speech Signal**

**Set of Connections**

**Computing Unit**

**Result**

**Training Set**

Fig.3. Artificial Neural Network Approach Block Diagram

1. Back-Propagation Neural Networks

Back-propagation neural networks have used to identify bird species using recordings of birdsong. In this application, pattern matching is accomplished using a three-layer feed-forward neural network with a sigmoid hidden layer followed by a linear layer. The back-propagation algorithm is used to train the neural network. To obtain a faster global convergence, the back-propagation method includes a momentum factor (Garcia, et. al). An adaptive back-propagation learning method is used in this application, in which the learning gain is changed during training to improve faster and global convergence. The three-layer feed-forward neural network architecture for this application. To allow each neuron to fire 100% of the time, a bias value of 1 is employed. (G.K, K, & V, 2002)

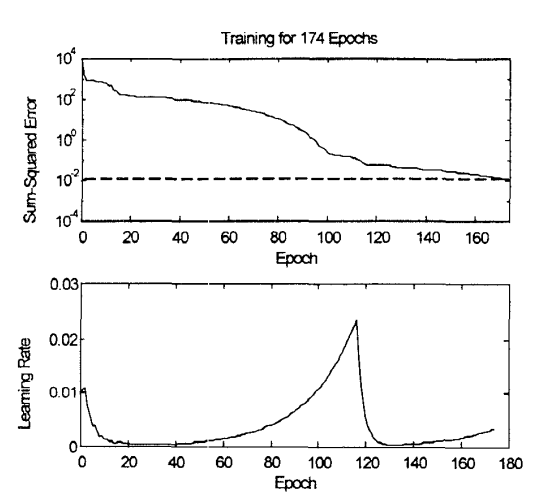


Fig.5. Back-propagation training error curve and learning curve

The first diagram provides a plot of the sum squared error versus the number of epoch during the training phase, as shown in Fig. 5. In only 174 epochs, the sum squared error objective was met. The different learning rates employed during the training are depicted in the second diagram of Fig.5. (G.K, K, & V, 2002)

1. Self Organizing Maps

When it comes to autonomous data classification, Self-Organizing Maps outperform other neural networks, and they're commonly utilized in condition monitoring and fault detection. Sigmoid feed-forward neural networks have a different structure than SOMs. There is no output layer in this sort of ANN, and the hidden layer neurons are organized in a grid pattern. The similarity of hidden neurons' weights to the input values activates input vectors. Similar inputs stimulate the same or neighboring neurons after SOM training modifies and rearranges these weights. As a result, enormous volumes of data can be categorized into unlabeled categories that match to the hidden layer grid map's activation clusters. SOMs are trained using the organizing mechanism of Kohonen learning. (Araujo .A, 2016)

Enhancing voice recognition using back-propagation algorithm

This paper applies the back propagation algorithm in neural networks to design a voice recognition system. By comparing the speaker's voice signal to previously recorded voice signals in the database and extracting the key aspects of the voice signal using Mel-frequency cepstral coefficients, which is one of the most essential factors in obtaining high identification accuracy.

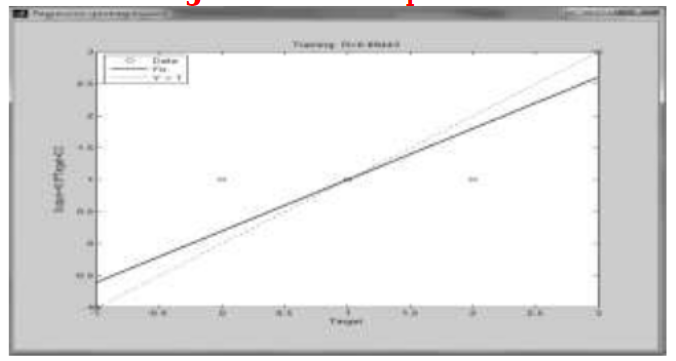


Fig.6. Learning rate of back-propagation network

In the database 20 samples are recorded from the different person. For training 5 samples are taken from every person.

The following figures showing the models of recording the word (“Hozayfa”) 5 times.

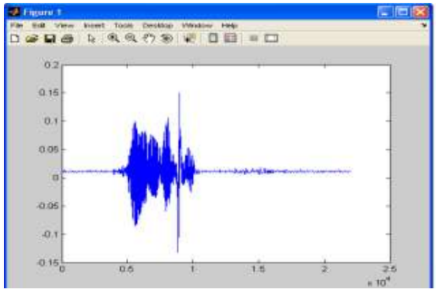
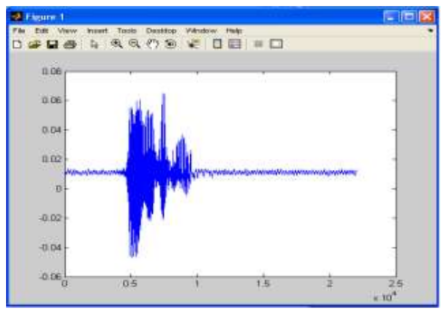
 

Fig.7.1. First sample Fig.7.2 Second sample

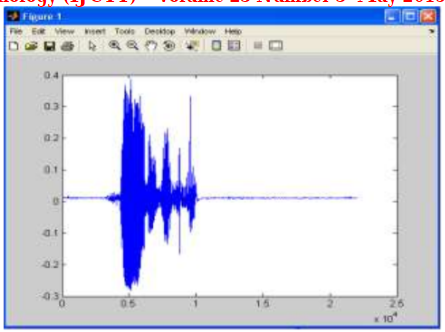
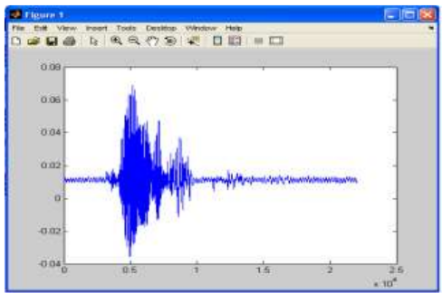
 

Fig 7.3 Third sample Fig.7.4 Forth sample

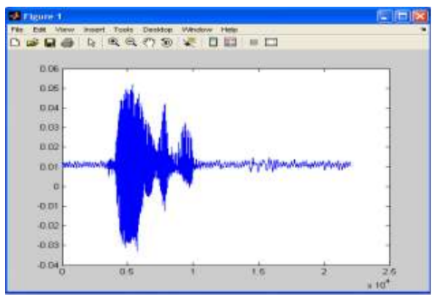


Fig.7.5 Fifth sample

After the 5 samples taken from every person are compared each samples with the stored samples in database to verify the identity of the persons. The number of MFCC coefficients and the recognition rate according to the number of MFCC coefficients are the basic elements that determine system efficiency and accuracy in recognizing and validating a person's identification. The range between (12-15) of MFCC coefficients gives the best results. To improve the voice recognition accuracy we need to increase the number of filters in the filter-bank. System takes 4 attempts out of 5 attempts to identify the speakers. (Mansour H. A et. al, 2015)

Conclusion

It has concluded that this report had a lot of success with neural networks in speech recognition, especially with voice matching using neural networks, back propagation neural networks, and self organizing neural networks. From this, we can conclude a very important fact: the spectrum coefficients have a high specification, indicating their importance in relation to the speaker himself, as well as the method of producing the voice, pronunciation, and style, which can be used in a variety of applications, such as security systems, where people's voices differ like fingerprints. As a result, the back propagation method has been evaluated for voice recognition, with a success rate of around 95%. We demonstrate that this method converges faster in Back-propagation algorithm and has a higher generalization capacity of the technology.

# References

1. Araujo A, (2016, 03 11). Self-organizing Maps for Speech Recognition [Online]. Available from: <https://www.semanticscholar.org/paper/Self-organizing-Maps-for-Speech-Recognition-Ara%C3%BAjo-Araujo/4f3cdd726d86ee919e6e75ac2a3005422fbcd8c8> [Accessed 1 1, 2022]
2. Garcia, et. al. (n.d.). An Online Back-propagation Algorithm with Validation Error-Based Adaptive Learning Rate [Online], from:< https://liris.cnrs.fr/Documents/Liris-6097.pdf>[Accessed 12 31, 2021]
3. G.K, K, & V, et al. (2002, 08 06). voice recognition using neural network [Online], from: <https://ieeexplore.ieee.org/document/736916> [Accessed 1 1, 2022]
4. Kikel. C. (2021, 07 01)., 5 Things That can Interfere with your voice recognition software [Online]. from: <https://www.totalvoicetech.com/five-things-that-can-interfere-with-your-voice-recognition-software/> [Accessed 12 28, 2021]
5. Malur K. S. (1998)., Speech Recognition using Neural Networks [Online]. from: <https://www.informatik.uni-ulm.de/ni/staff/FSchwenker/lit/papers/zegers98speech.pdf> [Accessed 12 31, 202]

# Mansour H. A et. al. (2015, 05). Voice recognition Using back propagation algorithm in neural networks [Online],

1. from: <https://www.researchgate.net/publication/282062349\_Voice\_recognition\_Using\_back\_propagation\_algorithm\_in\_neural\_networks> [Accessed 01 03, 2022]
2. RAF100STEAM. (2017, 08 12). 6 problems AI faces in speech recognition [Online]. from: <https://medium.com/@RAF100STEAM/6-problems-artificial-intelligence-faces-in-speech-recognition-ae705cfa1a72> [Accessed 12 28, 2021]
3. SAmin. (2019, 01 15). Speech recognition is hard [Online], from: <https://towardsdatascience.com/speech-recognition-is-hard-part-1-258e813b6eb7> [Accessed 12 28, 2021]
4. Zumalt .J. (2005, 12 24).Voice recognition Technology [Online], from: <https://www.researchgate.net/publication/289944157\_Voice\_Recognition\_Technology\_Has\_It\_Come\_of\_Age> [Accessed 12 24, 2021]