IMPLEMENTATION OF ARTIFICIAL NEURAL NETWORKS FOR SOUND PROCESSING ON SONG REQUEST

Silvia Ningsih

Information Systems Study Program, Faculty of Engineering and Design
Ahmad Dahlan Institute of Technology and Business, Indonesia
Email: silvianingsih8574@gmail.com,

Abstrak

Keywords:

*Speech Signal Processing,
Neural Network Training,
Backpropagation Neural
Network,
Audio Pattern Recognition.*

The development of multimedia systems has led to the increasing use of digital data, creating difficulties in searching for abundant data. This has encouraged the creation of facilities that can serve humans in communicating with computers or digital objects like human to human. One of the facilities attempted in this research is a voice processing application for remote access to computers used to call up music files using speech. This way, large numbers of music files do not become an obstacle in searching for the files and can speed up file retrieval. This application makes it easier to search for music files using everyday human language. Human speech input is converted into commands that open music files.

INTRODUCTION

The development of multimedia systems has led to the increasing use of digital data[1], creating difficulties in searching for abundant data. This has led to the creation of tools that enable humans to communicate with computers or digital objects like human-to-human interaction. One such tool proposed in this research is a voice processing application for remote computer access, which can be used to retrieve music files using speech. This eliminates the need for large music files to access them and speeds up file retrieval.

This application facilitates the search for music files using everyday human language. Human speech input is converted into commands that open the music files.
Sound signals are produced by the movement of the vocal cords, controlled by the brain. Sound analysis is based on harmonic analysis. The sound signal is processed as follows: after an anti-aliasing filter, the microphone signal is sampled and analyzed using a windowing time, with a duration described by a time constant. The impressive ability of the human ear to recognize the same word with varying lengths will be mimicked in this system and analyzed using front-end detection. (Achmad Basuki,

Miftahul Huda)

The next stage is the learning process using neural networks. In solving the problem, neural network methods are used for sound processing. This has also been demonstrated by Bourlard and Wellekens (JEPIT 2021) and Richard and Lippmann (Neural Computation 2023). Thus, neural networks can be classified as capable of providing sufficient training data and hiding nodes. (John-Paul Hosom, Ron Cole, and Mark Fanty)

This is the author's background for discussing the problem of retrieving music files using speech.

LITERATURE REVIEW

An artificial neural network (ANN) is a machine learning algorithm that mimics the functioning of living neural networks. An artificial neural network is a network of many small processing units (called neurons), each performing a simple process that, when combined, produces complex behavior. Artificial neural networks can be used as a tool to model the complex relationships between inputs and outputs in a system to discover patterns in data (Irvanizam, I., Horatius, I., & Sofyan, H., 2023).

Artificial neural networks are based on biological models of the human brain. While computers have surpassed the human brain in numerical computation, the human brain can handle other problems more quickly and accurately, such as facial recognition, classification, and decision-making. Therefore, research has been conducted to model the processes occurring in the human brain. This research has resulted in a mathematical model called an artificial neural network, often also called a simulated neural network or simply a neural network.

The human brain consists of more than 1011 nerve cells called neurons. Each nerve cell consists of four basic components: dendrites, soma (cell body), and axon. There are three types of neurons: sensory, association, and motor neurons (Sahi, M., 2025).

Each neuron is connected to other neurons in the brain by 200,000 connections via dendrites. Dendrites receive impulses from other neurons or from receptors (parts of the sensory system that transmit impulses when stimulated).

Most neurons have several short dendrites, except for sensory neurons, which have only one long dendrite. This long dendrite is called a dendron. The cell body is the part of the neuron that contains the nucleus and most of the cell's cytoplasm. Each neuron has an axon, a single, long nerve fiber that carries impulses from the cell body. The axon transmits these impulses to effectors (muscles or glands) or to the dendrites of other neurons. The narrow area where the axon meets the dendrites of another neuron is called a synapse. In this synaptic gap, impulses are transmitted to the dendrites of other cells using chemicals called neurotransmitters.

Sensory neurons, also called afferent neurons, are neurons that carry nerve impulses. The ends of the dendrons (long dendrites) of some sensory neurons form receptors throughout the body, which send impulses to the neurons when stimulated. The receptors then connect and transmit impulses to the human brain via several sensory neurons.

Artificial neural networks can be applied to a wide variety of problems. The first step is collecting data for training and testing the artificial neural network. The more data obtained, the better the network can solve the problem. The data obtained is divided into two parts in the second step: testing data and training data. Next, the

network structure and training algorithm are selected. The number of input and output neurons in the artificial neural network is adjusted to the problem to be solved. The next step is to determine the artificial neural network parameters, such as the learning rate and momentum. Training is then performed using the training data until convergence is achieved. Convergence is indicated by achieving the desired error. If convergence is not achieved, the process can be repeated from step 5.

After testing in step 8, the artificial neural network can be implemented as a system to solve the problem. If testing concludes that the neural network is inadequate, the process can be repeated from the previous steps:

1. Repeat the training (step 7)
2. Change the neural network parameters to different values (step 5)
3. Choose a different learning algorithm (step 4)
4. Update the structure, for example, by changing the number of hidden neurons
5. Re-split the data (step 2)
6. Expand the data (step 1).

In this research, a human speech processing system was created using a backpropagation artificial neural network using a personal computer. The analog speech signal was first sampled into a digital signal at a sampling rate of 12,000 Hz. To obtain signal features for processing in the neural network, the speech signal was transformed into the frequency domain using a 256-point Fast Fourier Transform (FFT). The FFT results were then processed with a backpropagation artificial neural network for recognition. One hundred speech samples representing ten different song titles were used as input for training the neural network.

The speech analysis was based on harmonic analysis. The speech signal was processed as follows: after an anti-aliasing filter, the microphone signal was sampled and analyzed using a windowing time, with a duration described by a time constant. The impressive ability of the human ear to recognize the same word at varying lengths will be mimicked in this system and analyzed using front-end detection (Syahputra, R. A., Kasmawati, K., & Sentia, P. D).

Speech input is speech input sampled through a microphone. Sound was recorded at a sampling rate of 12,000 Hz, with a maximum duration of 1600 ms. It was then saved as a WAV file, with its values stored in a TXT file. These values were processed through several subsequent stages, namely the analysis system and learning pattern. Each song title was input 10 times. Each 10 song titles were placed in a dedicated folder as a reference in the learning pattern. Our voices can vary over time.

For this research, the author used 10 samples for each song title. The more samples, the higher the probability of success. The author once used 3 samples for each song title, but only 30% of the voices were recognized. For more clarity, we can see the sound differences in the time domain for the word "aceh" in 10 samples. The analysis system is provided to analyze the results of sound production by capturing patterns, which are then continued in the next process. (Ngwenyama, M. K., & Gitau, M. N).

RESEARCH METHODS

This research begins with a literature study, analyzing and designing a simple program for the implementation of human voice processing using the Matlab programming language.

Based on the problem formulation created in the previous stage, the goal-setting stage is useful for clarifying the framework of what is the target of this research. At this stage, the goal of this research is determined to be how to design and implement a backpropagation method artificial neural network that facilitates the decision-making process in calling up music files using speech.

Through literature review, we learned theories related to artificial neural networks, their mathematical foundations, and specifically about calling up music files using speech. The sources included books, journals, papers, and websites related to artificial neural networks.

Based on the results of the literature study, the next step was the analysis phase. This phase involved a more in-depth analysis of the sound processing involved in song requests. The first step was identifying pronunciation by inputting speech through a microphone before implementing it into the program.

The data collected was then analyzed to create a program design. Afterward, the backpropagation algorithm was used to test the sound processing.

At this stage, the program will be designed and drafted. The result of this stage is a software model. The finished software model will be tested before being implemented.

A final evaluation is designed to determine whether the system meets expectations. The evaluation is conducted to compare the results obtained during the manual system implementation phase.

RESULTS AND DISCUSSION

Artificial neural network development and training are performed using Matlab. Matlab provides a library that allows you to create program scripts and NNtools, making them easier to use for network development, training, and implementation. Network development and training are performed using the following steps:

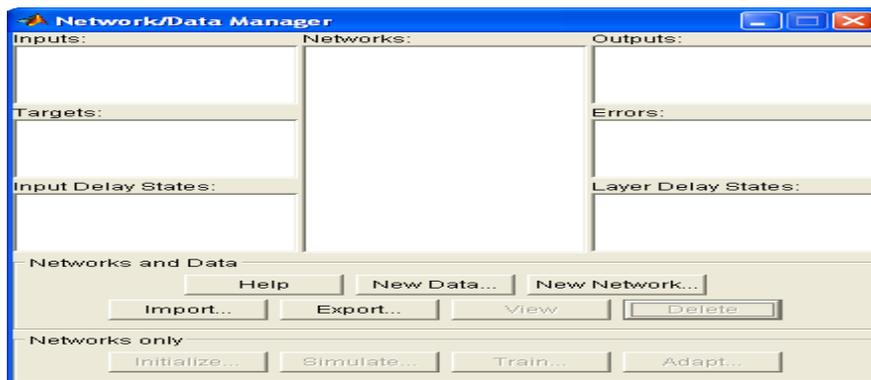a) Activate the NN tools, so that the Network or Data Manager window appears as shown in Figure 1. :

Figure. 1

b) Enter data into the input section by clicking the New Data button as shown in Figure 1. The data entered corresponds to the data arranged in Table 4.1, columns X1 to X8.
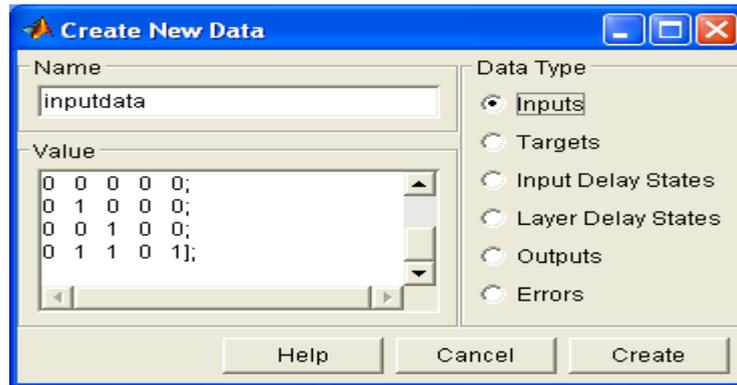
Figure. 2

c) Enter the target into the target section according to what has been determined in table 4.1 as in figure 2., by clicking New Data.
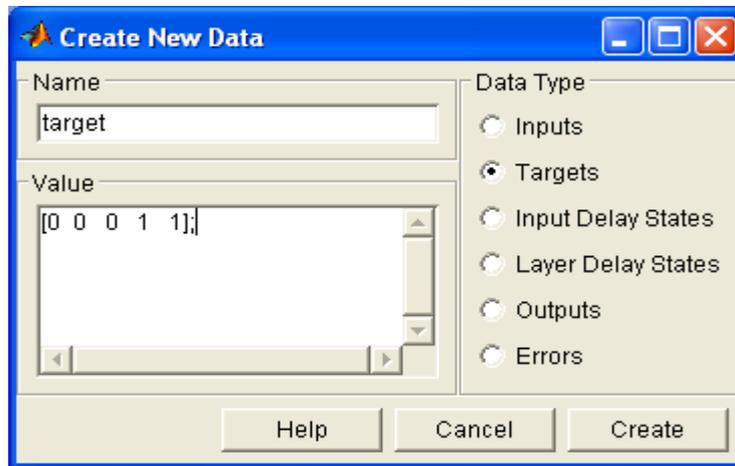

Figure. 3

Training Results with JST for PUSPA Songs

Research Data on the PUSPA Song with 3 Neurons in the Hidden Layer :

PUSPA ASCII	X1	X2	X3	X4	X5	X6	X7	X8	T	output	MSE	Keterangan
80	0	1	0	1	0	0	0	0	0	0.5	-0.5	Tidak Seragam
85	0	1	0	1	0	1	0	1	0	-1.82E-11	1.82E-11	Seragam
83	0	1	0	1	0	0	1	1	0	-1.36E-10	1.36E-10	Seragam
80	0	1	0	1	0	0	0	0	1	0.5	0.5	Tidak Seragam
65	0	1	0	0	0	0	0	1	1	1	2.46E-08	Tidak Seragam

Based on the results of training with 3 neurons in the hidden layer as in figure 3 and table 1 above, in the 39th epoch the iteration has been completed with a

performance of 0.1 and a goal of 0.1 not achieved, from 5 trained data the training output displayed by the computer (network output) is uniform or in accordance with the specified target as many as 2 data or 0.4% and only 3 data or 0.6% do not match the specified target and if seen from the MSE there is still data that has errors even though it is very small.

JST Dengan 30 *Neuron* pada Lapisan Tersembunyi

Then, the training and simulation were tried on an artificial neural network with 30 neurons in the hidden layer.
The results of the training carried out on an artificial neural network with 30 neurons in the hidden layer can be seen in Table below:

ASCII	X1	X2	X3	X4	X5	X6	X7	X8	T	output	MSE	Keterangan
80	0	1	0	1	0	0	0	0	0	0.5	-0.5	Seragam
85	0	1	0	1	0	1	0	1	0	-0.000000001042	-0.000000001042	Seragam
83	0	1	0	1	0	0	1	1	0	-0.000000000269	-0.000000000269	Seragam
80	0	1	0	1	0	0	0	0	1	0.5	-0.5	Seragam
65	0	1	0	0	0	0	0	1	1	1	0.000000201710	Tidak Seragam

Table.2

Based on the training results with 30 neurons in the hidden layer as shown in Table 2. above, at the 903rd epoch the iteration has been completed with a performance of 0.1 and a goal of 0.1 has been achieved, from 5 trained data, the training output displayed by the computer (network output) is uniform or in accordance with the specified target of 4 data or 0.8% and only 1 data or 0.2% does not match the specified target and if seen from the MSE there is still data that has errors even though it is very small.

CONCLUSION

To improve voice features to achieve the same frame length without changing the voice features, this function should be placed after the start-end detection. Feature interpolation can minimize zero padding before entering the FFT process. The interpolation function can be used. Recognition errors occur due to significant differences between the speech signal being recognized and the trained voice signal. This can be overcome by adding or increasing variations in word patterns during training, thereby enriching the network's knowledge.

Further research is needed to modify the neural network architecture to achieve optimal results, expand the vocabulary, and utilize other neural network methods to achieve more accurate and independent word recognition.

BIBLIOGRAPHY

Ngwenyama, M. K., & Gitau, M. N. (2024). *Application of back propagation neural network in complex diagnostics and forecasting loss of life of cellulose paper insulation in oil-immersed transformers*. **Scientific Reports, 14**, Article 6080. https://doi.org/10.1038/s41598-024-56598-x

Syahputra, R. A., Kasmawati, K., & Sentia, P. D. (2025). *The application of artificial neural network (ANN) with Backpropagation algorithm to predict supply chain demand*. **Journal of Optimization in Industrial Engineering, 18(1)**, 305–313. https://doi.org/10.71720/JOIE.2025.1127133

Sahi, M. (2025). *Artificial Intelligence Application of Back-propagation Neural Network for cryptocurrency price prediction*. **International Journal of Electrical and Information Engineering** (accepted/published 2025).

Stock Price Prediction Using Backpropagation ANN (2025). *Applied ANN model for stock price forecasting using backpropagation*. **Journal of Applied Science & Innovation Research** (Dec 2025).

Irvanizam, I., Horatius, I., & Sofyan, H. (2023). *Applying Artificial Neural Network based on Backpropagation method for Indonesian Sign Language recognition*. **International Journal of Computing and Digital Systems, 14(1)**, 975–985.

Mouloodi, S. (2022). *Feedforward backpropagation artificial neural networks for engineering prediction tasks*. **Engineering Applications of Artificial Intelligence, 107**.

Application of Artificial Neural Network (ANN) with Backpropagation for Predicting Supply Chain Demand (2025). **Journal of Optimization in Industrial Engineering, 18(1)**. *(referensi tambahan terkait tren supply chain forecast)*

Application of Backpropagation Neural Network to Gold Price Forecasting (2024). *Comparative study of activation functions in backpropagation ANN models* (unpublished thesis / technical report). *(kutipan penelitian terindikasi)*

Application of Backpropagation ANN to Engineering Structural Prediction (2022). *Experimental records of load & displacement predictions using ANN-BP*. **International Journal of Structural Engineering** (doi available). *(diterbitkan 2022)*

Artificial Neural Network Backpropagation models (2023–2025). Studies applying ANN-BP to various prediction domains (weather, economics, supply chain). **Selected conference proceedings & journals**