Speech Recognizing for Presentation Tool Navigation Using Back Propagation Artificial Neural Network

Nur Hasanah[a], Dessy Irmawati, Fatchul Arifin and Eko Marpanaji

Electronics Engineering Education Department, Yogyakarta State University, Karangmalang, Yogyakarta 55281, Indonesia

Abstract. Backpropagation Artificial Neural Network (ANN) is a well known branch of Artificial Intelligence and has been proven to solve various problems of complex speech recognizing in health [1], [2], education [4] and engineering [3]. Today, many kinds of presentation tools are used by society. One popular example is MsPowerpoint. The transition process between slides in presentation tools will be more easily done through speech, the sound emitted directly by the user during the presentation. This study uses research and development to create a simulation using Backpropagation ANN for speech recognition from number one to five to navigate slides of the presentation tool. The Backpropagation ANN consists of one input layer, one hidden layer with 100 neurons and one output layer. The simulation is built by using a Neural Network Toolbox Matlab R2014a. Speech samples were taken from five different people with wav format. This research shows that the Backpropagation ANN can be used as navigation through speech with 96% accuracy rate based on the network training result. Thesimulation can produce 63% accuracy based on 100 new speech samples from various sources.

1 Background

Artificial Neural Network (ANN) is a branch of Artificial Intelligence which has been widely applied to various applications and is proven to solve various complex problems. Backpropagation is one of the popular ANN architecture and is widely applied in research of speech recognizing, for example, speech recognition of a recently born baby [1], speech recognition of cardiac abnormalities [2], speech recognition to control robots [3] and speech recognition to learn foreign language [4].

Presentation tools usually consists of slides that can be displayed on the screen and the user can navigate using a keyboard or other controller devices such as mouse, wireless mouse and wireless pointer. The control device intended to facilitate users in giving presentations. But the use of a keyboard and mouse in the transition between slides still limit the movement of user because there is a maximum distance between the devices with a laptop or computer that is used. While wireless pointer provides only know next and previous orders and can not refer directly to a specific slide the user wants. The transition process between slides will be more easily done through speech, the sound emitted directly by the user during the presentation. The easier use of the device, the better support for the presentation process. Therefore, this research aims to create software for presentation tool navigation through speech, using Backpropagation ANN.

2 Theoretical basis

2.1 Fast fourier transform algorithm

FFT is a method for transforming the speech signal into a frequency signal, meaning that the speech recording is recorded in digital form in the form of wave-based speech frequency spectrum. The FFT algorithm is designed to perform complex multiplications and additions, even though the input data may be real valued [5].

2.2 Artificial neural network

An artificial neural network is an information-processing system that has certain performance characteristics in common with biological neural networks. Artificial neural networks have been developed as generalizations of mathematical models of human cognition or neural biology [6].

A neural net consists of a large number of simple processing elements called neurons, units, cells or nodes. Each neuron is connected to other neurons by means of directed communication links, each with an associated weight. The weights represent information being used by the net to solve a problem [6].

[a] Corresponding author: nurhasanah@uny.ac.id

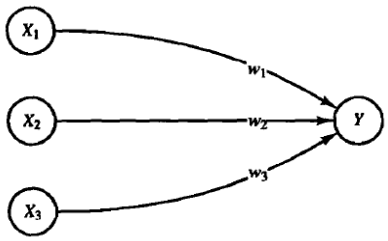

Figure 1. A simple (artificial) neuron.

Each neuron has an internal state, called its activation or activity level, which is a function of the inputs it has received. Typically, a neuron sends its activation as a signal to several other neurons. For example, consider a neuron Y, illustrated in Figure 1, that receives inputs from neurons X_1, X_2, and X_3. The activations (output signals) of these neurons are x_1, x_2, and x_3, respectively. The weights on the connections from X_1, X_2, and X_3 to neuron Y are X_1, X_2, and w_1, w_2 and w_3 respectively. The net input, y_in, to neuron Y is the sum of the weighted signals from neurons X_1, X_2, and X_3, i.e.,

$$y\_in = w_1 x_1 + w_2 x_2 + w_3 x_3 \quad (1)$$

The activation y of neuron Y is given by some function of its net input, y = f(y_in) [6].

2.3 Common activation functions

1. Identity function (Purelin)

$$f(x) = x \text{ , for all } x \quad (2)$$

Single-layer nets often use a step function to convert the net input, which is a continuously valued variable, to an output unit that is a binary (1 or 0) or bipolar (1 or -1) signal [6].

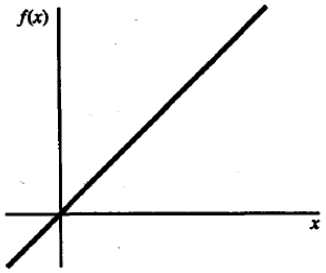

Figure 2. Identity function.

2. Binary Sigmoid function (Logsig)

$$f(x) = \frac{1}{1 + \exp(-\sigma x)}$$

$$f'(x) = \sigma f(x)[1 - f(x)] \quad (3)$$

During feedforward, each input unit (X_i) receives an input signal and broadcasts this signal to the each of the hidden units $Z_1, \ldots, Z_p$. Each hidden unit then computes its activation and sends its signal (z_j) to each output unit. Each output unit (Y_k) computes its activation (y_k) to form the response of the net for the given input pattern [6].

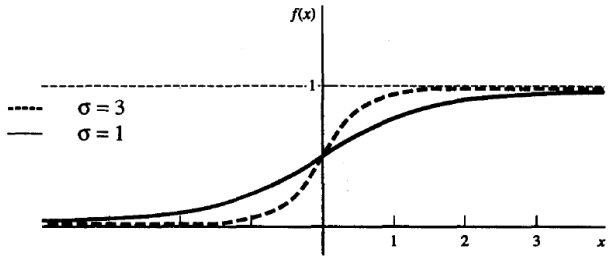

Figure 3. Binary Sigmoid. Stepness parameters a = 1 and a = 3

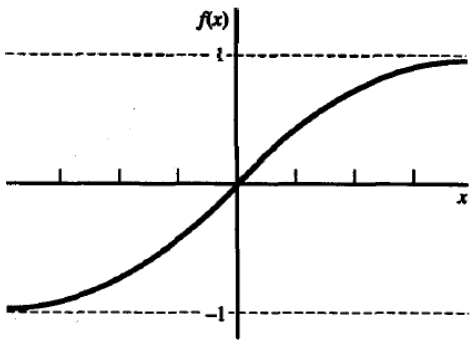

Figure 4. Bipolar Sigmoid.

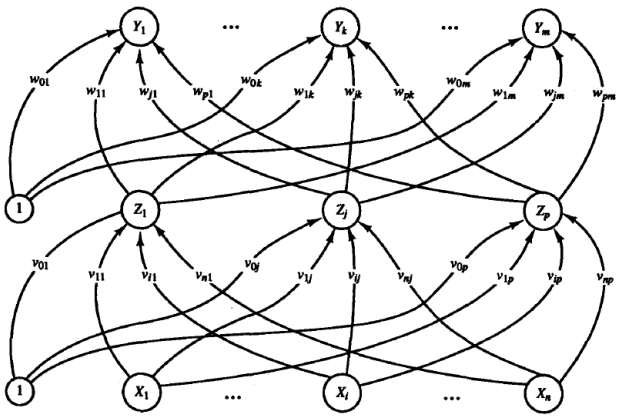

Figure 5. Backpropagation ANN with one hidden layer.

During training, each output unit compares its computed activation y_k with its target value t_k to determine the associated error for that pattern with that unit. Based on this error, the factor δ_k (k = 1, … m) is computed. δ_k is used to distribute the error at output unit Y_k back to all units in the previous layer (the hidden units that are connected to Y_k). It also used (later) to update the weights between the output layer and the hidden layer.

After all of the δ factors have been determined, the weights for all layers are adjusted simultaneously. The adjustment to the weight w_{jk} (from hidden unit Z_j to output unit Y_k) is based on the factor δ_k and the

activation z_j of the hidden unit Z_j. The adjustment to the weight v_{ij} (from input unit X_i to hidden unit Z_j) is based on the factor δ_j and the activation x_i of the input unit [6].

Backpropagation learning has emerged as the standard algorithm for the training of multilayer perceptrons, against which other learning algorithms are often bench-marked [7].

3 Research Method

This study uses Research and Development to develop speech recognition simulation of number one, two, three, four and five. Speech samples were taken from five different people with wav format. The total of speech samples for network training is 75. Each number has 15 speech samples.

3.1. Feature extraction

In this stage, software for preprocessing is built using FFT, as shown in Figure 6.

3.2 Network training

The next stage is to design software and implement ANN Backpropagation algorithms to learn speech patterns. Backpropagation ANN is an autoassociative network type, input range that is processed into the same network with the range of output results. Backpropagation ANN development in general can be seen in Figure 7.

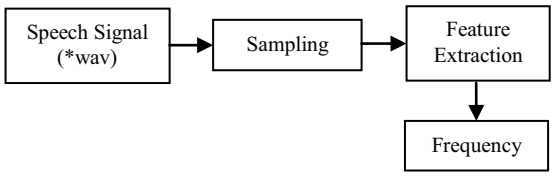

Figure 6. Preprocessing stage.

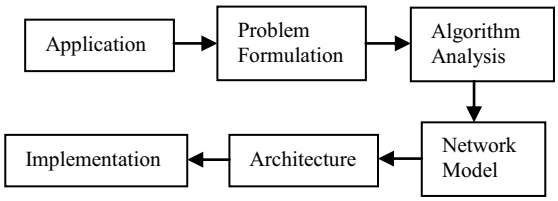

Figure 7. Development process of artificial neural networks.

Thus, the general system of this application follows a path as shown in Figure 8.

The Backpropagation ANN consists of one input layer, one hidden layer and one output layer. Activation function is used to connect the input layer to the hidden layer and also to connect the hidden layer to the output layer. Activation function used in the research is Tansig. A multilayer network learns much faster when the sigmoidal activation function is represented by a hyperbolic tangent [8].

4 Result and discussion

The Backpropagation ANN built has 1 hidden layer with 100 neurons. The activation function used is Tansig. There are two ways to determine the training process stops, by limiting the number of iterations or Mean Square Error (MSE).

The simulation is built by using a Neural Network Toolbox Matlab R2014a. Matlab is an interactive, matrix-based system for scientific and engineering numeric computation and visualization. Its strength lies in the fact that complex numerical problems can be solved easily and in a fraction of the time required with a programming language. The basic Matlab program is further enhanced by the availability of numerous toolboxes [9].

In the simulation, the number of iterations is limited to 1000 while MSE is limited to 1.00e-06. Network training results can be seen in Figure 9. The training process occurs until 23 iteration and training time during 0 second.

While Figure 10 shows the results of Confusion. The conclusion obtained is the level of accuracy in determining the output network is 96 %.

To measure the accuracy of Backpropagation ANN, it is necessary to test some new speech samples using a simulation program that can be seen in Figure 11.

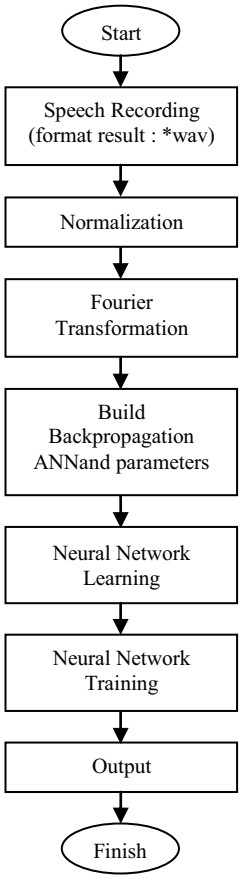

Figure 8. Flowchart of the simulation system.

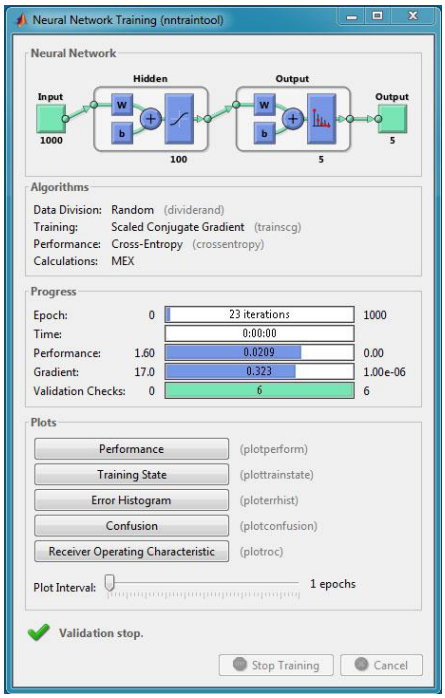

Figure 9. Result of network training.

This program provides menus to recordspeech, play speech, save speech and open a speech that have been saved. After a speech has beenrecorded, the program will display the results in terms of numbers as well as opening the intendedslide presentation tool. The speech will be recorded by the internal microphone on the PC / laptop.

Figure 12 shows part of codes from Matlab program that instructs to go to the slide based on the analysis result from the speech input.

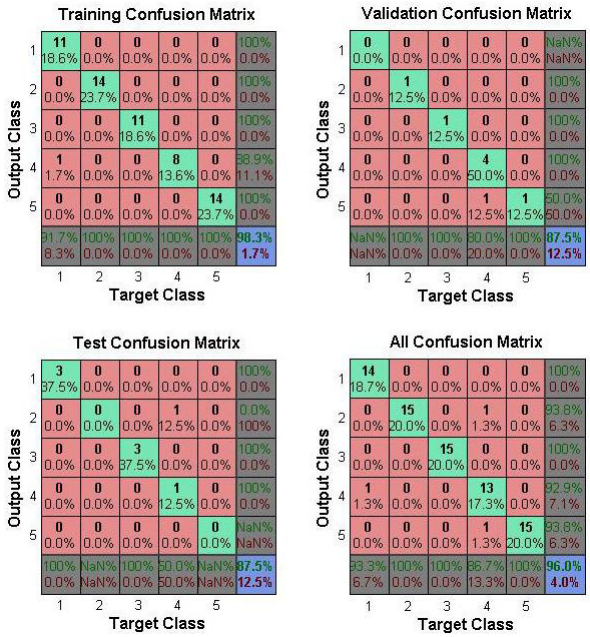

Figure 10. Result of confusion.

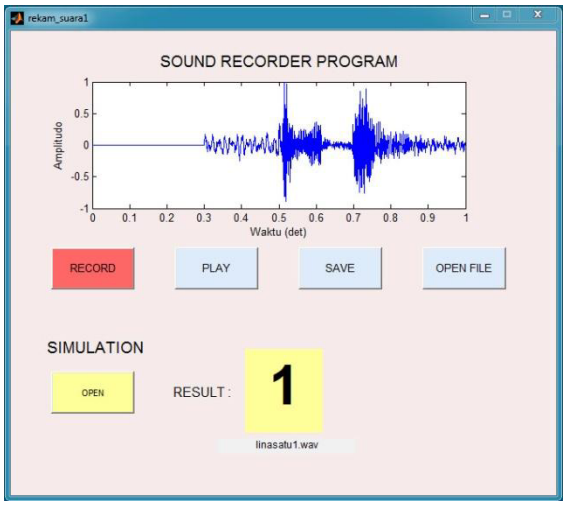

Figure 11. GUI of the simulation system.

```
hShow = s.SlideShowWindows.Item(1);

if y(1)==1
set(findobj('Tag','hasil'),'String','1');
set(findobj('Tag','nama'),'String',filename);
hShow.View.GotoSlide(1);

elseif y(2)==1
set(findobj('Tag','hasil'),'String','2');
set(findobj('Tag','nama'),'String',filename);
hShow.View.GotoSlide(2);
```

Figure 12. Matlab source codes for slide navigation.

In Figure 13, it can be seen that the program record a new speechsample then saved it as Sample_a2.wav. This speech instruction is to open the slide 2. It can be seen that the program succeeded in opening the slide 2.

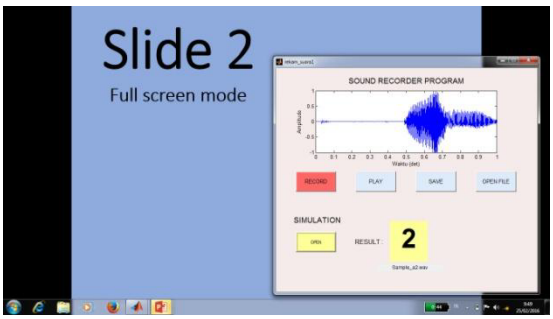

Figure 13. Testing new speech sample in opening slide 2.

While in Figure 14, it can be seen that the program record another new speechsample then saved it as Sample_a3.wav. This speech instruction to open the slide 3. This speech instruction is to open the slide 3. It can be seen that the program succeeded in opening the slide 3.

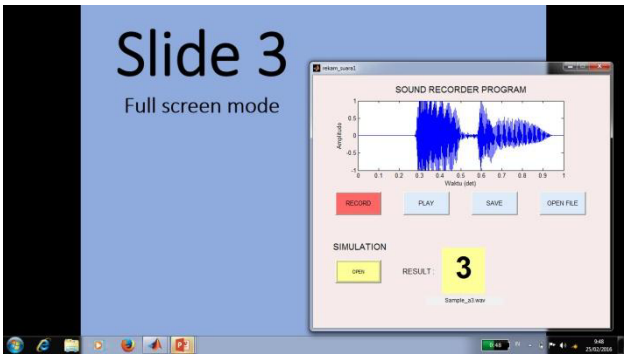

Figure 14. Testing new speech sample in opening slide 3.

The new speech samples tested were 100 speeches from different sources. Some results of the testing can be seen in Table 1.

The result showsthat the simulation can recognize63speeches correctly and can demonstrate the proper slide. While 37speeches can not be recognized properly by the system and showed wrong slides.

Table 1. Result of new speech samples

No	File Name	Purposed Slide	Slide Result	Correct
1	Sample_a1	One	One	Yes
2	Sample_a2	Two	Two	Yes
3	Sample_a3	Three	Five	Yes
4	Sample_a4	Four	Four	Yes
5	Sample_a5	Five	Five	Yes
6	Sample_b1	One	Five	No
7	Sample_b2	Two	One	No
8	Sample_b3	Three	Three	Yes
9	Sample_b4	Four	Five	No
10	Sample_b5	Five	Five	Yes

5 Conclusion

Based on test results and discussion, it can be concluded that the Backpropagation ANN can be used as a presentation tool navigation through speech with 96% accuracy rate based on the network training result. The simulation can produce 63% accuracy based on 100 new speech samples from various sources.

References

1. R. Orion F. and R. Carlos Alberto, SPECOM 9th Conference Speech and Computer, A System for the Processing of Infant Cry to Recognize Pathologies in Recently Born Babies with Neural Network (2004)
2. Asian Network for Scientific Information, Classification of Heart Abnormalities Using Artificial Neural Network, 820–821 (2007)
3. W. Suryo. and Thiang, IPCSIT, Speech Recognition Using Linear Predictive Coding and Artificial Neural Network for Controlling Movement of Mobile Robot, **6** (2011)
4. V. Radha, Vimala C. and M. Krishnaveni, CSEIJ,Isolated Word Recognition System for Tamil Spoken Languange Using Backpropagation Neural Network Based on LPCC Features, **1** (2011)
5. P. John G. And M. Dimitris G, Digital Signal Processing Priciples, Algorithms, and Applications Third Edition. Prentice Hall International Edition (1996)
6. F. Lauren, Fundamental of Neural Network. Prentice Hall International Edition (1994)
7. H. Simon, Neural Network A Comprehensive Foundation. Prentice Hall International Edition (1999).
8. N. Michael, Artificial Intelligence: A Guide to Intelligent Systems, Pearson Education (2002)
9. K. I. Vinay and G.P. John, Digital Signal Processing Using Matlab V.4. PWS Publishing Company (1997)