Personal Identification using Voice Recognition with Neuro Fuzzy Method

Ariyawan Sunardi, Rezky Mahardika, Sunarko, Heri Suherkiman
Pusat Reaktor Serba Guna (PRSG) BATAN, Kompleks Puspiptek Gd. 31 Setu Tangsel - Indonesia

Email: ariyawan@batan.go.id

Abstract. The Experimental Power Plant (RDE) is one of the vital object that will be built by the Indonesian nation. During the operation time, RDE needs to pay attention to the nuclear security aspect. Personnel whose accessing the RDE is need to be controlled especially in the restricted area. It needs hardware systems and supporting software to ensure the nuclear security. The purpose of this research is to design personnel access system using voice recognition to support the development activity of detection and response system of RDE test facility. This is related to the development of nuclear security. Several parameters used in this research are voice samples with sampling frequency 8000 Hz and 8 bit per sample with High Pass Filter wavelet filter (HPF). We use wavelet coiflet level 2 for decomposition level and Shannon entropy is used to calculate wavelet optimization, by used that we get a characteristic vector of each speaker by value Feature vector (Shannon Entropy). We used Neuro-Fuzzy N-Input and 1 Output, with 19 class classification (19 personnel) to identify personnel. In this research, we get time about 4.7379 s for identification personnel.

Keywords: personnel identification, voice recognition, neuro fuzzy, RDE, Shannon Entropy

1. Introduction

RDE Detailed Design Development is a challenge for BATAN to continue to develop engineering design capabilities, especially nuclear reactor engineering design, and pursue the ideals of becoming a Technology Provider in the field of nuclear energy [1]. We have to support the RDE program by a system that can ensure nuclear security. Through of them, access control personnel are recorded. Access personnel can be done through a biometric system. The biometric system is a pattern recognition system, which creates personal identification by determining the authenticity of certain physiological or behavioral characteristics of users [2]. It consists of a unique method of recognizing human beings based on one or more intrinsic physical or behavioral traits. Biometrics are used as a form of identity access management and access control. It is also used to identify individuals in groups under supervision.

Voice is a popular and easy biometric in hardware applications [2]. The purpose of this study was to identify people using voice recognition by neuro fuzzy method. Voice recognition application can be identification of personnel by voice database is recorded. The author sees the possibility of developing a personnel access system to ensure the safety and security of the installation. In the discussion of Intan in reference [3] get 75% accuracy in personnel identification.
This research was limited to personnel identification systems to ensure safety and security. The idea to design a voice recognition system that uses wavelet transforms as feature extraction and neuro fuzzy as a classification method. Neuro fuzzy method is expected to increase the level of personnel recognition.

2. Theory
Voice recognition becomes one of the potential things that concern the interaction between human and computers. The design of voice recognition system requires special attention and handling as there are several constraints: the realization of phoneme acoustics, the diversity of the acoustics, the diversity of language pronunciation/dialect, the speaker's emotional state, the speed of speech and the environment [2].

Voice Recognition is a process of converting voice signals into words or commands. In this case the required algorithm is implemented into a computer program to execute the command. The purpose of voice recognition is to create a technique and system to enter voice commands into the machine, so that the machine understands what human say and obeys what it says.

2.1. Signal Pre-processing
In the pre-processing stage, all the first sound signals are converted to a sampling rate of 8 kHz with 16-bit resolution. In practice, the sound of the speaker is contaminated with noise/disturbance components. The sound recording conditions, devices and environment act as additional annoying unknown sources. Therefore, it needs to be filtered by a digital filtering signal as a pre-processing for the analysis and sound classification of the speaker.

\[ X = FS . dt (det ik) \left( \frac{bit}{8} \right) . j \]

2.2. Wavelet Transformation

Wavelet is an analytical tool that commonly used to present data or function or operators into different frequency components, and then examine each component with a resolution that matched the scale. [3].

The wavelet transform makes it possible to solve resolution problems that cannot be solved with Short Time Fourier Transform (STFT). The basic function of the wavelet allows to exchange resolutions between time and frequency.

![Wavelet Transform](image)

**Figure 1.** Resolution of Wavelet Transform.

\[ \int_{-\infty}^{\infty} \psi(t) \, dt = 0 \]

This function is called wavelet or mother wavelet function.
2.3. **Wavelet Decomposition**

The decomposition process can be oriented, with the successive approach being a decomposed signal, so that one signal is broken down into some of the lower resolution components. This is called the wavelet decomposition tree.

![Diagram of the decomposition tree](image)

Each time the wavelet decomposition causes the length of the sound signal to be reduced by half of the original signal's length. The result of the wavelet decomposition is the feature vector of the sound signal in the form of a low frequency signal. This feature vector will be the input for neuro fuzzy.

2.4. **Shannon Entropy**

Information is a measure of uncertainty, or entropy, in a situation. The greater the uncertainty, the greater the information available in the communication process. When a situation is complete can be ascertainable or predictable (posable predictable), then no information at all that can be recorded. This condition is called negentropy.

By looking at information as entropy, Shannon sparked the idea that an information source is probabilistic.

The calculation of entropy can be done in various ways, in this decomposition used is Shannon entropy.

\[
E_t(s_i) = - s_i^2 \log (s_i^2)
\]

2.5. **Neuro Fuzzy**

The neuro fuzzy model is a merger of two systems, the artificial neural network (ANN) or "neural network" and fuzzy logic or "fuzzy logic". In neuro fuzzy, a step in the fuzzy system is formed using artificial neural networks.

The fuzzy inference system used is the fuzzy model of Takagi, Sugeno and Kang (TSK) first order with the consideration of simplicity and ease of computing. This consideration is important because the system will go through a learning process that has a large computational load.

In a first order TSK fuzzy inference system with two entries, the rule used is

**Rule 1:** If \( x \) is \( A_1 \) and \( y \) is \( B_1 \) then \( f_1 = p_1 x + q_1 y + r_1 \)

**Rule 2:** If \( x \) is \( A_2 \) and \( y \) is \( B_2 \) then \( f_2 = p_2 x + q_2 y + r_2 \).

The neuro-fuzzy system that is equivalent to the above fuzzy inference system has a network structure with five layers. Each layer has a different function and consists of several vertices. The ANFIS layer is described as follows.
3. Research Methods

The personnel identification system designed in this study can be seen in Figure 1.

We use 19 person voice samples to record for database. The word used in this research is "ijin masuk Pak". For the required input signal is a sound signal with a sampling frequency of 8000 Hz, and samples per bit 8 bits. The next step tests the sound files of the personnel to be identified based on the database. For processing, we used computers with specifications: Intel Pentium CPU P6300 @ 2.27GHz & 2.26GHz, 2 GB of RAM, Windows Operating System 8.1. Enterprise and Matlab R2010a.
4. Results And Discussion

4.1. ANFIS Testing

The design of ANFIS testing for personnel identification conducted with signal input and output.

![Input signal from voice](image1)

**Figure 5.** Input signal from voice

The input signal is required by a voice signal with a sampling frequency of 8000 Hz, and samples per bit 8 bits. The consideration for using the 8000 Hz sampling frequency is based on the information on the sound signal being at a frequency of 0 - 4000 Hz. So the process is done by using the frequency of 8000 Hz. Pre-processing is done to remove the dc frequency component (zero frequency) contained in the voice signal. Therefore, pre-processing is done by pre-emphasis filter to reduce dc frequency by using high pass filter (HPF) as shown in Figure 5.

![Input Signals before and after pre-processing](image2)

**Figure 6.** Input Signals before and after pre-processing

Figure 6 (center) indicates that the frequency response of the filter used is a high frequency pass filter and damp frequency (zero).

Wavelet decomposition process is done to extract the information contained in the sound signal, by dividing or breaking the signal into the frequency bands.

The decomposed sound signal is illustrated by Figure 7 which shows at level = 2 decomposition, the media signal is divided into 2 frequency bands, i.e. low frequency band (approximation component) and high frequency band (detail component).
The next stage of the process is calculation of Entropy Shannon which will be used as the characteristic vector of each band of the decomposition of the sound signal. The mechanism performed is the sound signal of all speakers carried out the process up to this stage to obtain the characteristic vector of each speaker. This feature vector will be used as training data for the identification process performed by Adaptive Neuro-Fuzzy Inference System (ANFIS).

Figure 8 is a comparison of the feature vector (Entropy Shannon) of each speaker that indicates the characteristic vector between one speaker with another speakers is different for further identification and recognition.

The identification process is done by inserting all feature vectors of all speakers as training data into the neuro-fuzzy network or ANFIS (Adaptive Neuro Fuzzy Inference System) that used as the material for identification and recognition of the characteristic vector of the test input signal. Figure 8 is a membership function (Membership Function) of training data using 19 different voting voices, thus the number of membership functions is NumMF = 19. For each person (member) has different function and scope of entropy value.

Table 1 shows the feature vector (Entropy Shannon) of each speaker for wavelet decomposition of level 1. For level = 1 we get a feature vector of length 4 values, level = 2 length vector = 8, level = 3 with length = 16.
Table 1. Entropy Shannon Wavelet level 2 decomposition results

| No | File Name       | Entropy | Shannon Level Decomposition 2 | Output | Name     |
|----|-----------------|---------|-------------------------------|--------|----------|
| 1  | personel1-1.wav | 13,95   | 10,85                         | 1,18   | 2,18     | 1        |
| 2  | personel1-2.wav | 35,26   | 27,59                         | 2,01   | 4,22     | 1        |
| 3  | personel1-3.wav | 20,65   | 16,15                         | 1,81   | 2,88     | 1        |
| 4  | personel1-4.wav | 27,84   | 21,59                         | 1,77   | 4,10     | 1        |
| 5  | personel1-5.wav | 20,32   | 15,96                         | 1,15   | 2,86     | 1        |
| 6  | personel2-1.wav | 10,37   | 8,09                          | 0,95   | 1,73     | 2        |
| 7  | personel2-2.wav | 8,16    | 6,35                          | 0,65   | 1,26     | 2        |
| 8  | personel2-3.wav | 9,84    | 7,58                          | 0,92   | 1,76     | 2        |
| 9  | personel2-4.wav | 12,86   | 10,42                         | 0,76   | 1,57     | 2        |
| 10 | personel2-5.wav | 10,34   | 8,53                          | 0,42   | 0,84     | 2        |
| 11 | personel3-1.wav | 43,43   | 31,10                         | 5,54   | 8,50     | 3        |
| 12 | personel3-2.wav | 52,99   | 36,91                         | 8,25   | 13,24    | 3        |
| 13 | personel3-3.wav | 62,90   | 43,89                         | 7,78   | 14,95    | 3        |
| 14 | personel3-4.wav | 61,06   | 42,04                         | 8,33   | 15,33    | 3        |
| 15 | personel3-5.wav | 215,02  | 178,80                        | 2,21   | 6,58     | 3        |
| 16 | personel4-1.wav | 125,23  | 91,71                         | 11,35  | 23,65    | 4        |
| 17 | personel4-2.wav | 97,42   | 72,02                         | 7,24   | 16,33    | 4        |
| 18 | personel4-3.wav | 46,01   | 33,25                         | 5,60   | 10,66    | 4        |
| 19 | personel4-4.wav | 60,21   | 43,39                         | 6,41   | 12,91    | 4        |
| 20 | personel4-5.wav | 58,28   | 42,28                         | 6,29   | 13,13    | 4        |
| 21 | personel5-1.wav | 61,25   | 51,58                         | 1,34   | 2,77     | 5        |
| 22 | personel5-2.wav | 20,28   | 16,97                         | 0,35   | 0,82     | 5        |
| 23 | personel5-3.wav | 20,75   | 17,69                         | 0,59   | 1,13     | 5        |
| 24 | personel5-4.wav | 165,33  | 127,36                        | 5,77   | 11,14    | 5        |
| 25 | personel5-5.wav | 64,85   | 54,20                         | 0,60   | 1,69     | 5        |
| 26 | personel6-1.wav | 103,91  | 67,82                         | 11,87  | 22,03    | 6        |
| 27 | personel6-2.wav | 99,15   | 69,96                         | 11,02  | 24,06    | 6        |
| 28 | personel6-3.wav | 101,70  | 63,94                         | 12,00  | 26,91    | 6        |
| 29 | personel6-4.wav | 113,28  | 77,29                         | 12,79  | 26,71    | 6        |
| 30 | personel6-5.wav | 102,15  | 70,78                         | 12,02  | 22,14    | 6        |

Figure 9. shows the process for running an introduction program with ANFIS, where program_utama as the name of the program created to run the recognition process, angga1.wav as the tested voice signal, coif2 is the type of wavelet filter used, 4 = is the decomposition level used, 1 = decimation (down sampling) and epoch_n = 20, as an iteration parameter in ANFIS. The below process gives the output 4 which means recognized as the fourth speaker of the personel4.
>> [out, dikenali] = program_utama('personel4.wav',4,'coif2',1,20)

| Iteration | Objective Function Value |
|-----------|--------------------------|
| 1         | 41690.495858             |
| 2         | 29358.026274             |
| 3         | 23703.831848             |
| 4         | 16723.797194             |
| 5         | 11778.395824             |
| 6         | 9595.180473              |
| 7         | 8798.958454              |
| 8         | 8447.129436              |
| 9         | 8193.405772              |
| 10        | 8045.034590              |
| 11        | 7954.453934              |
| 12        | 7842.587712              |
| 13        | 7720.431902              |
| 14        | 7602.765408              |
| 15        | 7469.161769              |

ANFIS info:
- Number of nodes: 407
- Number of linear parameters: 198
- Number of nonlinear parameters: 352
- Total number of parameters: 550
- Number of training data pairs: 110
- Number of checking data pairs: 0
- Number of fuzzy rules: 22

Start training ANFIS ...

| Step | Objective Function Value |
|------|--------------------------|
| 1    | 0.0456631                |
| 2    | 0.0432162                |
| 3    | 0.040749                 |
| 4    | 0.038247                 |
| 5    | 0.0356972                |
| 6    | 0.033088                 |
| 7    | 0.0301379                |
| 8    | 0.027098                 |
| 9    | 0.0239667                |

Step size increases to 0.011000 after epoch 5.
5. Conclusion
Based on the results of these research, test of wavelet filters shows an recognition level of 90.19%, false acceptance rate results is 1.9% and the average variable running time test for the recognition process is 4.7379 second. It can be concluded as follows: A system of personnel access system design has been developed using voice recognition in order to improve safety and security guarantee.

Acknowledgements
The authors would like thank to the Management of INSINAS-Flagship BATAN for their supports and appreciate their coordinated work. This research is funded by the INSINAS-Flagship 2018 Program of Indonesian Ministry of Research and Technology and Higher Education.

References
[1]. J.M. Beck and L.F. Pincock. “High Temperature Gas-Cooled Reactors Lessons Learned Applicable to The Next Generation Nuclear Plant”. Prepared For The U.S. Department of Energy Office of Nuclear Energy, USA, 2011.
[2]. Ariyawan Sunardi, Djoko Hari Nugroho. “Perancangan Sistem Akses Personel Untuk Safety dan Security Menggunakan Identifikasi Speech Recognition”. Jakarta, ISTN, 2015.
[3]. A. Vijaykumar, Aruna, M. Vijayapal Reddy. 2011. A Fuzzy Neural Network for Speech Recognition. India: ARPN Journal of Systems and Software

[4]. Intan Verona. 2006. Perbandingan Tingkat Akurasi Antara Algoritma Momentum Back Propagation dan Learning vector Quantization Neural Network Untuk Speech Recognition. Jakarta: Universitas Tarumanegara

[5]. Anny Tandyo, Martono, Adi Widyatmoko. “Speaker Identification Menggunakan Transformasi Wavelet Diskrit Dan Jaringan Saraf Tiruan Back-Propagation”. Jakarta, Universitas Bina Nusantara, 2005.

[6]. Ashish Kumar Panda, Amit Kumar Sahoo. 2011. Study Of Speaker Recognition Systems. Departement Of Electronics And Communications National Institute Of Technology, Rourkela

[7]. Binanto, Iwan. 2010. Multimedia Digital Dasar Teori dan Pengembangannya. Yogyakarta: Andi.

[8]. Joseph P. Campbell, Jr. Speaker Recognition. IEEE : Department of Defense Fort Meade.

[9]. Ni Wayan Sumartini Saraswati. Transformasi Wavelet dan Thresholding Pada Citra Menggunakan Matlab. Universitas Udayana: Jurnal GENERIC.

[10]. Samjin Choi, Zhongwei Jiang. 2007. Wavelet Packet-Based Insufficiency Murmurs Analysis Method. Japan: Yamaguchi University, Ube.

[11]. Sutarno. 2010. Analisis Perbandingan Transformasi Wavelet pada Pengenalan Citra Wajah. Universitas Sriwijaya: Jurnal GENERIC.