IMPLEMENTATION LEARNING VECTOR QUANTIZATION AND LINEAR PREDICTIVE CODING FOR PRONUNCIATION RECOGNITION OF NAME INDONESIAN HERO

Hernanda[1], Susmini Indriani L., MT.[2]

[1,2]Informatics Engineering – Indonesian Computer of University
Jl. Dipatiukur 112-114 Bandung
E-mail : hernanda17@email.unikom.ac.id[1], lestariningati@gmail.com[2]

ABSTRAK

"Ejaan yang disempurnakan" (EYD) is spelling language for Indonesia valid until today. that previously used the spelling "Ejaan Republik" or "Ejaan Soewandi". "Ejaan Soewandi" is difference like "tj" as "c","oe" as "u", and many more. For pronunciation which has a similar pronunciation in pronounced difference will produce a meaning such as "guna" and "tuna". Name of hero Indonesia with the old spellinf can produce different pronunciation.

Based on similarity. A lot of research have been done related similarities of voice recoginition. The results of recognition obtained varied training data, are affected by the extraction methods and classification. In this study the characteristics extraction methods used are the Linear Predictice Coding and classification methods are used is Learning Vector Quantization for recognition pronunciation. Stage of processing of sound conducted namely extraction characteristics Linear Predictice Coding with stages Preemphasis, Frame Blocking, Windowing, Auto Correlation, LPC analysis, and cepstral coefficients. The next stage of classification done is training and testin Learning Vector Quantization. The name of the hero in test as much 8 names. Voice data used for training data 50 sample whose meaning 40 x 8 as many 400 voice used. voice data used for testing data is 10 sample whose meaning 10 x 8 as many 80 voiced used.

Based on the results of test performed on test using K-fold Cross Validation, then obtained best test accuracy amounting to 54.17%. Testing the accuracy of the results is not satisfactory because of its many noise from training data is used. The combination that used is learning rate (α) 0.0001, reduced ratio = 0.001, error minimum = 0.0002 and maximum epoch = 100

Keywords: Artificial Intelligence, Neural Network, Linear Predictive Coding, Learning Vector Quantization, Voice Recognition

1. INTRODUCTION

"Ejaan yang disempurnakan" (EYD) is the spelling of the language of Indonesia valid until today. This replaces "Ejaan Republik" or "Ejaan Soewandi". Before the EYD the Institute of language and literature in 1967 issued a new spelling, new spelling is basically a continuation of the changes seen in previous spellings is "tj" became "c", "dj" to "j", "nj" into "ny", "sj" becomes "sy", "ch" to "kh", "oe" became "u". Pronunciation is the way a person or group of people to pronounce the sounds of language. In writing, pronunciation is not visible clearly. This pronunciation was more reflected in the spoken language [1] differences of similar pronunciation when saying the following, for example, 'Tuna' with 'Guna', exactly. When pronounced with not clear, then the meaning of pronunciation can be interpreted incorrectly. The names of the heroes of Indonesia also have pronounced that if mispronounced in the wrong with it will have the meaning wrong or different, as Tengku Umar with Teuku Umar, Supriyadi with Suprijadi, Thaha Saifuddin with Thaha Syaifuddin, Fakhruddin by Fahruddin, and others. If the pronunciation and pronunciation of the name of the hero of Indonesia is wrong it will be wrong on illegal or improper writing of the name of the hero.

Voice recognition is one of the efforts to recognize or identify the sound so that it can be utilized for a variety of applications. To solve the problem of the pronunciation can be interpreted wrongly or purposely to be wrong, then the voice recognition can be one method that can be used in order for the name of the hero can be defined and pronounced correctly, there are a variety of voice recognition algorithms that have ever been tested. Case study of the introduction of the tone on the strings of the violin [2] get accuracy reach 88.33%, the study tried to match the right data, in the form of 60 tones of G, D, A, E on violin and generate the coefficient cepstral i.e. 73.33%, and in other research that is voice recognition instrument [3] with LVQ method and extraction of characteristics produce the best accuracy of LPC 88,333% with instrument ratings percentage of 50% , a musical instrument at 40% and 66.67% wind instrument, and for case studies using human language, namely the introduction of pengucap could not rely on text that

becomes the reference of some paper to use extraction characteristics of LPC but in such research using the method of Vector Quantization [4], the accuracy obtained in research by 60%. The same study using LVQ with extraction of characteristics to identify gender LPC [5] the research getting accuracy of 75.76%.

Based on the background of the problems that have been described in studies to test the name of the hero of Indonesia with a different pronunciation of the word, whether a different pronunciation can be recognized by the method of extracting characteristic Liniear Predictive coding and Vector Quantization method of classification of Learning. The resulting output from the system that is created is the result of introduction and accuracy, the test case with the hope of extracting characteristic Liniear Predictive coding and Vector Quantization method of Learning classification can be used for the development of voice recognition method, and for the introduction of the name of the hero of Indonesia so that at the time of writing no one written due to a wrong pronunciation.

1.1 Sound

The sound is a physical phenomenon that is produced by the vibration of objects. The vibration form of analog signal and amplitude that change continuously over time. Sound or the sound is usually propagated by air and cannot propagate through a vacuum [6]..

1.1.1 Voice Signal Representation

Analog voice signal wave can not be directly represented on the computer. The computer measures the amplitude at a given time unit to produce a number of numbers. Each unit of measurement is called sample. Sampling is the process of mencuplik analog signal into several chunks of time, then the exact value was given the sample and the value is given in the form of digital [7].

1.1.2 *Analog Signal to Digital Signal*

Natural signal in General as the voice signal is the signal continue where an unlimited value. While on the computer, all signal that can be processed by the computer is just another signal discrete or often known as the term digital signal. In order for a natural signal can be processed by a computer, then it must be changed in advance of the data signal continue into discrete [7]. It can be done with 3 processes, including the following:

1. The process of sampling is a process to extract the data signal continue for each period. In the process of sampling data, apply rules, the name that is Nyquist frequency sampling (sampling rate) must be at least 2 times higher than the maximum frequency to be sampling. If signal sampling less than twice the maximum frequency of the signal that will be sampling, aliasing effects will arise then. Aliasing is an effect where the generated signal has a frequency different from the original signal.
2. The process of quantization is the process of value made up my data into certain numbers have been determined in advance. The more level used the more accurate data also signal stored but will generate large data sizes and a long process.
3. The process of coding is the process of granting code for each data signal is quantized based on levels that have been assigned.

1.2 Linear Predictive Coding

Linear Predictive Coding (LPC) is a digital method to perform the process of encoding analog signals, where a certain value in the prediction by a linear function of the previous values of the signal [2]. There are two main reasons the use of a method of LPC, i.e.

a. LPC can be used to predict the route of a signal. In the frequency domain, the number equivalent to the prediction modeling spectrum signals [2]
b. LPC were able to eliminate sections that can be predictable in an effort to avoid redundancies at the time of re-transmitting that can be predicted by the receiver and simultaneously can save storage, bandwidth, and time [2]

The steps of signal processing with the LPC to obtain the coefficient cepstral i.e.

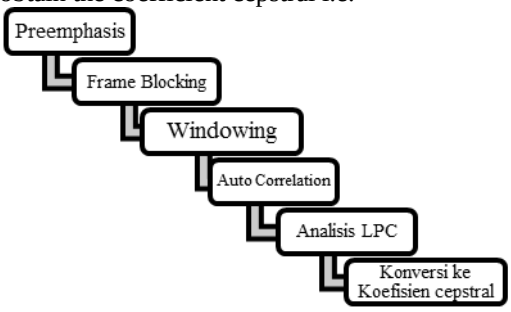


Figure 1 Block Diagram

1.2.1 *Pre-Emphasis*

At this stage used to flatten spectral signal and increase the authenticity of the signal in the signal processing. The equation for the calculation of the pre-emphasis can be seen in equation 1 here:

$$s'(n) = s(n) - 0.9375\ x\ s(n-1) \qquad 1$$

Description
s^' (n) : Fungtion of pre-emphasis
s(n) : Signal to n
s(n-1) : signal to n-1
:

1.2.2 Frame Blocking

The stages of this grouping process results signal pre-emphasis into frames-frames with size of N data entered. This sequential frames with the

separation between these two frames of M data, so that M = 1/3 N, where N was symbolic of the amount of block signals. As for the example of the frame blocking can be seen in Figure 2 here

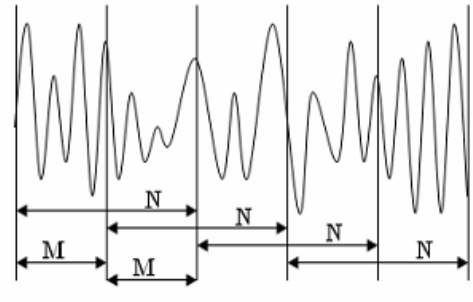


Figure 2 Example Frame Blocking

1.2.3 *Windowing*

After the frame blocking is done, then the next process, namely the process of windowing. The process uses Hamming equation, where each of the frame data is entered into the blocking process Hamming [8]

$$W(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N} - 1\right) \quad 2$$

Decription:
W(n) : Hamming Fungtion
N : Number of sample block signals

1.2.4 Auto Correlation

Correlation is a method to find out the level of density of the relationship of two or more variables represented by the size of the correlation coefficient [4].

The correlation between the two waveforms is a measure of similarity of both. Waveform comparison at different time interval, and "similarity" is calculated on the lapse of time respectively. The results of a correlation is a measure of similarity as a function of time lapse between the beginning of the second wave form. Autocorrelation function is the correlation of a waveform with himself. The similarities just in an interval of zero, while inequality continues to increase with time lapse. Mathematical definition to describe the function of autocorrelation is shown in Equation 1 for a function of discrete infinity x (n) whereas the definition of the autocorrelation function of a discrete mathematics is limited to x (n) to size N [9]

Each set of windowing in the autocorrelation coefficient obtained a subset so that p-1, with p is the order of the LPC reply expected [4]. Here is the equation of the function of auto correlation

$$rl(m) = \sum_{n=0}^{N-1-m} xi(n) * xi(n+m) \quad 3$$

1.2.5 Analysis LPC

At this stage each frame with autocorrelation (p + 1) will be converted to a shape parameter LPC LPC coefficients which can be. Here the equation of the LPC analysis

$$E(o)=r(0) \quad 4$$

$$km=\{r(m)-\Sigma m-1j=1\alpha j(m-1)r(|m-j|)/E(m-1), \quad 5$$
$$1 \le m \le p$$

$$\alpha m(m)= km \quad 6$$

$$\alpha j(m) = \alpha j(m-1)- km\ \alpha m-j(m-1),\ 1 \le j \le m-1 \quad 7$$

$$E(m)=(1- km2)\ E(m-1) \quad 8$$

Description:
E = Error Value
r = auto correlation Value
m = position autocorrelation
p = orde LPC
k = Analysis LPC Value

1.3 Artificial Neural Networks (ANN)

ANN (Artificial Neural Networks) is a huge parallel spread processor that has a tendency to store experiential knowledge and making it ready for use [16]. ANN resembles a human brain in two ways, namely:

1. Knowledge gained through learning
2. The strength of the relationship between nerve cells (neurons) known as synaptic weights are used to store knowledge.

ANN is one the human effort to model the behavior or function of the human nervous system to perform certain tasks. This modeling is based on the human brain's ability to organize its constituent cells called neurons, so as to execute certain tasks, in particular the introduction of a pattern with very high effectiveness.

1.3.1 Learning Process ANN

Learning in the context of ANN is a process where the parameters are non ANN adapted through a process of continuous stimulation by the environment in which the network is located. there are two learning process in ANN, namely:

1. Supervised Learning

   Supervised or active learning is a learning process requires teachers. Here the teacher means something that has knowledge of the environment. Teachers can be represented as a set of sample input-output. The construction of knowledge is done by teachers by giving the desired response to the ANN. Network parameters fluctuate based training vectors and the error signal (error signal is the difference between the output ANN and the desired response). This change process is done repeatedly, step by step, with the goal of keeping ANN can have similar capabilities with the teacher. In other words, ANN is trained to be able to map a set of sample input-output with high accuracy. In this learning process, the expected output has been previously known.

2. Unsupervised Learning
   Unsupervised or self-organized learning does not require the teacher to monitor the learning process. In other words, there is no set of sample input-output or a particular function to be learned by the network. One example of unsupervised learning is a competitive learning. In this method, it is not specified whether the results were as expected during the learning process.

In this research, learning process that used is supervised learning.

1.3.2 Learning Vector Quantization (LVQ)

Learning Vector Quantization (LVQ) included in the supervised learning process. LVQ is a pattern classification method that each unit of output represents a specific category or class. In this method, a competitive layer will automatically learn to classify input data vectors. The classes were obtained as a result of competitive layer is only dependent on the distance between the input vector. If two vectors approximately equal, then the competitive layer would put both the vector into the same class. Here is the architecture of Learning Vector quantization methods.

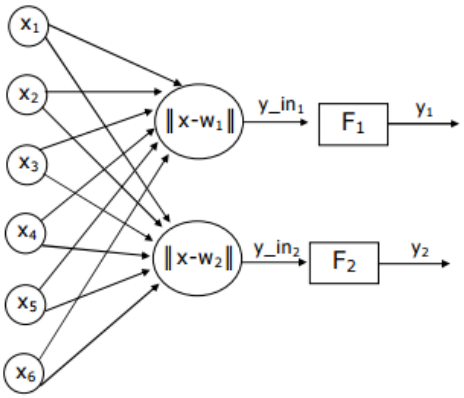


Figure 1 LVQ Networks

While LVQ algorithm is as follows.

1. Initialization: weight (W) Maximum epoch (MaxEpoch), error minimum that expected (Eps), *learning rate (α).*
2. Input:
   Input : x(m,n)
   Target m : T(1,n)
3. Set initial conditions:
   epoch = 0
   err=1
4. do if: (epoch<MaxEpoch) or (α > eps)
   a. epoch=epoch+1
   b. do for i=1 to n
      i. set J such that ||x − wj|| minimum (referred to as C_j)

$$c_j = \sqrt{\sum_{i=1}^{i=n} (x_i - w_{ji})^2} \quad (10)$$

ii. fix w_j with provisions
if T = C_j then:

$$w_j(baru) = w_j(lama) + \alpha[x - w_j(lama)] \quad (11)$$

if T ≠ C_j then:

$$w_j(baru) = w_j(lama) - \alpha[x - w_j(lama)] \quad (12)$$

c. decrease α value.

Here is a description formula of the algorithm in the above.

T : Target class training data
c : Difference distance the smallest weight
x : Training data
W : Data weight class
j : index of data
m : Lots of training data
n : Much of the data in the input layer / the number of pixels the sound

1.4 K-Fold Cross Validation

Accuracy is how close a number of results measurement of the number of actual (true value or reference value). In this technique the dataset is divided into a number of K-partition fruit randomly. Then performed a number of K-times experiments, where each experiment using the partition data as data testing and utilizing the rest of the other partitions as training data.

2. RESEARCH CONTENTS

In this research will be built applications introduction to the pronunciation of the name of the hero of Indonesia with the method of Learning Vector Quantization (LVQ), the name of the hero in this research is.

1.H.O.S Tjokroaminoto
2.Tjiptomangunkusumo
3.Koesoemah Atmadja
4.Teuku Tjik Ditiro
5.Soeprijadi
6.Mohammad Hoesni Thamrin
7.Maskoen Soemadiredja
8.Gatot Mangkoepradja

The sound is recorded as 15 respondents each pronounce the names of heroes. For the data the respondents 15 trainer can mention the right name heroes with different pronunciation, pronunciation differences as differences of pronunciation, still using the old spelling, different length of the commemoration, the difference in intonation and different length of the pronunciation. Test data for as many as 10 respondents, can be selected from 5 persons training data, respondents and 5 new respondents that mention each 8 name of the hero.

2.1 General System Architecture

The process is performed in the system is divided into sections where each process has its respective role in performing he pronunciation of the

name of the hero Indonesia. The following process is performed in Figure 4.

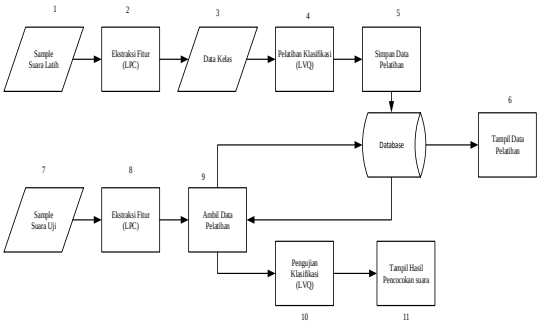


Figure 4 System Architecture

2.1.1 Linear Predictive Coding

In the process of extraction characteristic LPC, some voice in the process so that it can be better. Sound data that's already been doing extraction characteristics after it can be placed into the process of classification method using Learning Vector Quantization (LVQ). Here is a description of Linear predictive coding with the architecture of the system are made.

1. Pre-Emphasis
   At this stage used to flatten spectral signal and increase the authenticity of the signal in the signal processing next, in this process using the equation 1 results from the process of Pre-emphasis is different to the original data but does not change the data contained in the original data
2. Frame Blocking
   Frame blocking is a process on the stages of signal preemphasis has in past on the block into pieces by the number N, and each part on separate with a number of M sample. Divide the result of preemphasis into frames-frames, each of which contains N samples are separated as far as M sample. The better the estimates sprektral LPC from frame to frame, each frame on the first position will be equal to the contents of the previous frame in the position of the last called is overlapping
3. Windowing
   After the frame blocking the next process, namely the process of windowing, this process on every window on the part of the signal that has been previously made. This is done to minimize at the beginning and end of the signal. If defined a windowing w (n) and the signal of each part is x (n) then his signal process of windowing, and in the process uses hamming function in equation 2
4. Auto Correlation
   At this stage each set of windowing in the autocorrelation coefficient obtained a subset so that p-1, with p is the order of the LPC reply expected
5. Analysis LPC
   At this stage each frame with autocorrelation (p + 1) will be dikonvesi to form the LPC parameters can be either LPC coefficient, a coefficient cepstral's reflection, or other transformations as required
6. Conversion Parameters LPC to Cepstral Coefficients
   A very important parameter sets, which can be derived directly from the set of coefficient cepstral coefficients, LPC is an LPC coefficient cepstral generating appropriate size order, order LPC LPC can be selected from 8 to 16.

2.1.3 *Learning Vector Quantization* (LVQ)

Classification methods that used in this research is Learning Vector Quantization (LVQ). Classification is done after the sound already gone through the process of extraction characteristic. The process that performed at this stage namely training and testing process. Here is an explanation of each process.

1. Training LVQ
   In the process of training the training data used LVQ consists of data classes and training data. The number of data training used in this study is., data classes as much as 240 votes (8 x 30 hero's name) and the training data as much as 80 votes (hero'ss name 8 x 10). Later also used the parameters of training (learning rate, reduction in learning rate, Max. Epoch and the minimum error). After this training process is done, the end result will renew data weighting classes from all data classes. Data weighting this is later used in the testing phase. Testing LVQ
2. Testing LVQ
   After the extraction process characteristics, then proceed with the testing process by comparing the values of the weights of each sound test with weights class that is already updated on the previous stages of the training. The result of this test is the name of the hero of the match with a voice say. However, in the process done there is a possibility the results obtained less than perfect. The process of testing this is to be used on the K-fold Cross Validation

2.2 Use Case Diagram

Use case diagrams are modeling that shows the interaction between the system by the user (actor). In the use case that is made using five use cases, where there is a use case that is related to

another use case. Here is use case diagram for the application is built.

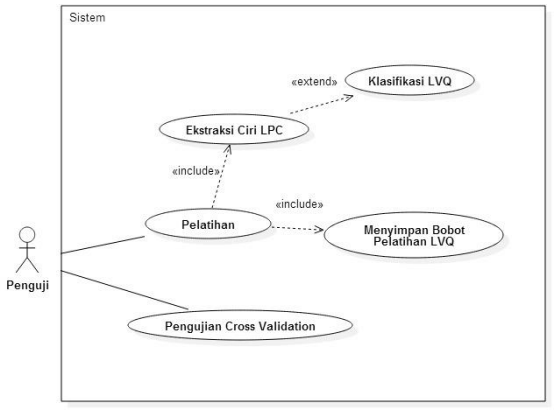


Figure 5 Use Case Diagram

2.3 Database Design

Database in this research is used to store training data that used for the LVQ classification process. Tables were used that table data_kelas and data_latih. Here is a database design in form of schematic relation table.

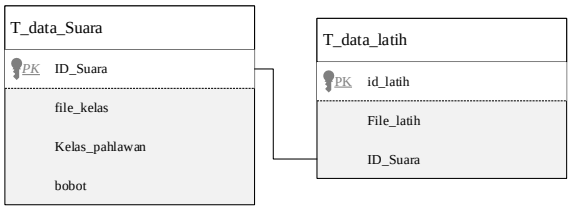


Figure 6 Relation scheme

2.4 Testing Accuracy K-Fold Cross Validation

On testing the accuracy will be calculated the value of accuracy or suitability of the application results of Indonesia hero name recognition. In this test previously searched parameters in LVQ best, and get Learning rate = 0.0001, reduction ratio = 0.001, Error = 0.0002, minimum and maximum epoch = 100. Next is testing the K-Fold Cross validation, the amount of data used as many as 240 data. Each dataset (D) consists of 24 data consisting of 3 votes the same hero name. Experimental testing conducted as many as ten times, where each dataset will alternately be training data and test data. Scenario testing to be performed can be seen in Tabel 1.

Table 1 Testing Scenario

Fold	Data Uji	Data Latih
1	D1	D2,D3,D4,D5,D6,D7,D8,D9,D10
2	D2	D1,D3,D4,D5,D6,D7,D8,D9,D10
3	D3	D1,D2,D4,D5,D6,D7,D8,D9,D10
4	D4	D1,D2,D3,D5,D6,D7,D8,D9,D10
5	D5	D1,D2,D3,D4,D6,D7,D8,D9,D10
6	D6	D1,D2,D3,D4,D5,D7,D8,D9,D10
7	D7	D1,D2,D3,D4,D5,D6,D8,D9,D10
8	D8	D1,D2,D3,D4,D5,D6,D7,D9,D10
9	D9	D1,D2,D3,D4,D5,D6,D7,D8,D10
10	D10	D1,D2,D3,D4,D5,D6,D7,D8,D9

The next process that is testing one by one fold, so the results are obtained, namely an average accuracy of the introduction of the pronunciation of the name of the hero of Indonesia. The following is a table of the test results obtained.

Table 2 Test Result

Pengujian	True	Akurasi (%)
1	12	50.00
2	17	70.83
3	12	50.00
4	10	41.67
5	10	41.67
6	12	50.00
7	14	58.33
8	16	66.67
9	15	62.50
10	12	50.00
Rata-rata akurasi (%)		54.17

From the table above the test results, testing the K-fold Cross validation by 10 fold obtained is 54,17%. The combination of training parameters are used i.e. Learning rate = 0.0001, reduction ratio = 0.001, minimum Error = 0.0002,

3. CLOSING

3.1 Conclusion

Based on the results of the testing that has been done on this research, then the conclusion that Indonesia hero name classification using Vector Quantization method of Learning classification and extraction of characteristics of Linear Predictive Coding, with LVQ Learning rate parameter = 0.0001, reduction ratio = 0.001, Error = 0.0002, minimum and maximum epoch = 100. The parameters used have been tested with testing paramter LVQ and obtained with the best accuracy. But when using different parameters will increase or decrease the value of the accuracy. By using the testing accuracy of K-Fold Cross Validation of 54.17% accuracy is obtained, then it can be concluded that the method of Learning Vector Quantization of Linear Predictive Coding and good enough to use for difference pelafalaan.

3.2 Suggestion

In order to further research on the introduction of the pronunciation of the name of the hero of Indonesia using the method of extraction of Linear Predictive Coding and characteristics extraction methods characteristic of Learning Vector Quantization in the case of the introduction of the pronunciation of the name of the hero of Indonesia with higher accuracy, then here is a saying which can be made into consideration, that is.

1. In the method of extracting characteristic can use a new feature of the extraction as the Mel-Frequency Cepstrum (MFCC), Pitch-syncronous zero-crossing amplitude peak-

(PS-ZCPA), or by other characteristics extraction methods

2. To obtain higher accuracy can use methods noise reduction

REFERENCES

[1] Y. A. Iskak, "Bahasa Indonesia Tataran Semenjana untuk SMK dan MAK kelas X", Erlangga, 2006.

[2] S. A. Suci, S. Esti and Wiharto, "Pengenalan Nada Pada Senar Biola Dengan Metode Jaringan Syaraf Tiruan Learning Vector Quantization", Yogyakarta: Seminar Nasional Teknologi Informasi dan Komunikasi 2013 (SENTIKA 2013), 2013.

[3] S. Indra, "Pengenalan Suara Alat Musik Dengan Metode Jaringan Syaraf Tiruan (JST) Learning Vector Quantization Melalui Ekstraksi Koefisien Cepstral," Tugas Akhir, Semarang, 2013.

[4] A. Fachrudin, "Pengenalan Pengucap Tak Bergantung Teks dengan Metode Vector Quantization (VQ) Melalui Ekstraksi Linier Predictive Coding," 2004.

[5] B. B. Susetyo, "Aplikasi Pengenalan Gender Menggunakan Suara," Seminar Nasional Aplikasi Teknologi Informasi (SNATI 2012), Yogyakarta, 2012.

[6] K. G. Ibnu Gunawan, Pembuatan Perangkat Lunak Wave Manipulator Untuk Memanipulasi File WAV, Jakarta: Universitas Kristen Petra, 2012.

[7] A. R. C, "SUARA DAN AUDIO", Yogyakarta : Fakultas Teknik Informatika Universitas Kristen Duta Wacana , 2005.

[8] B. Munawar, "Pengidentifikasian kata Dengan Menggunakan Metode Hidden Makrov Model (HMM) Melalui Ekstraksi Ciri Linear Predictive Coding (LPC)," 2010.

[9] W. A. H. Aryo Baskoro Utomo, "Analisis Karakteristik Suara Manusia Berdasarakan Frekuensi Fundamental dan Tingkat Usia Pada Pelajar SLTP dan SMA," 2016.

[10] I. K. Meganjaya, "Pengenalan Tulisan Tangan menggunakan metode Learning Vector Quantization dalam Kasus Evaluasi Jawaban Ujian," 2015.

[11] S. M. Rosa. A.S, Modul Pembelajaran Rekayasa perangkat Lunak (Terstruktur dan Berorientasi Objek), Bandung: Modula, 2013.

[12] B. Hartanto, "Memahami Visual C#.Net Secara Mudah", Yohyakarta: Andi, 2008.

[13] D. K. Sinaga, "Implementasi Learning Vector Quantization (LVQ) Dalam Mengenal Pertumbuhan Jamur," 2014.

[14] A. Kristanto, "Jaringan Syaraf Tiruan", Yogyakarta: GAYA MEDIA, 2000.

[15] T. Hidayat, "Pengenalan Pola Tanda Tangan Menggunakan Learning Vector Quatization Dalam IDNentifikasi Kepribadian," 2015.

[16] B. Hariyanto, Rekayasan Sistem Berorientasi Objek, Bandung: Informatika, 2004.