Jurnal Ilmiah Komputer dan Informatika (KOMPUTA)
Edisi. .. Volume. .., Bulan 20.. ISSN : 2089-9033

Voice identification High And Low In Men And Women Seen From Pitch and Formant Method Using Support Vector Machine

Indra Tri Prabowo

Teknik Informatika - Universitas Komputer Indonesia
Jl. Dipati Ukur No. 112-114 Bandung
Email : indra3p@gmail.com

ABSTRACT

Voice recognition is the process of identifying the sounds by the words spoken by someone who captured the sound input device to be recognized and then translated into a data that is understood by the computer. When humans make a sound, that's when the voice convey some information in spoken words through sound waves. The voice information can be known through the voice feature itself, including the pitch and formant, where the pitch is the fundamental frequency of the sound signal produced by the vibration of the vocal cords, and formant is the acoustic resonance frequency of the human voice field. Both of these features is a voice feature that is very important to identify the spoken voice of a person.

Voice identification analysis stage is the stage of pre-processing, feature extraction and classification. Stages of pre-processing with pre-emphasis, frame blocking, and hamming window of the sound signal. Stages feature extraction with autocorrelation to determine pitch and linear predictive coding to determine formant, as well as the stages of classification by support vector machine to classify voice features that will be used to identify high and low voices.

Based on the results of testing of the simulator with the conclusion that the average value of a man's voice pitch high and low is lower than women, the average value of formant formant 1st and 2nd women's high and low sounds higher than men, the mean average formant all three men to the sound of high and low is higher than women, the voice feature is better suited to represent the voice high for men and women is the pitch, the range of pitch and formant for a male voice is 191.1435 Hz (F0), 503.9955 Hz (F1) , 956.9225 Hz (F2), and 1561.43 Hz (F3). For a female voice is 268.51 Hz (F0), 534 871 Hz (F1), 1080.4475 Hz (F2), and 1690,835 Hz (F3). Simulator able to predict high and low noise for both men and women with an accuracy percentage of 70% for the voices of men and 100% for the female voice.

Keywords*: Pitch, Formant, Linear Predictive Coding, Support Vector Machine.*

1. INTRODUCTION

Issues discussed in this thesis regarding the low and high sound human. The analysis was done on the basis to be able to identify high and low sounds by seeing the values of sound features, namely the pitch and formant. Pitch and formant value is obtained through the process of feature extraction. Values pitch and formant sound high and low then stored as training data and compared with the value of pitch and formant contained in the prediction data to be classified by using support vector machine (SVM) in order to know whether the results of the classification value pitch and formant are entered into the high category or lower.

The study of some of the literature explains that the method of support vector machine is a method of classification of types of assisted (supervised), which works using a mapping linear and non-linear and very good if used to classify the features that have two classes or group, where the class or group here is sound high and low. The resulting solution of support vector machine method for determining the high voice is the voice of identifying high accuracy of voice features we tested, which tested sound features are pitch and formant.

1.1 Autocorrelation

Autocorrelation is the cross-correlation of the signal with itself. Autocorrelation value of a speech signal will show how the sound waves that form a correlation to himself. The forms are the same at any given time delay shows the repetition of the pattern of the sound signal. Thus it would be able to estimate the value of the fundamental frequency.

1.2 Linear Predictive Coding

Linear predictive coding is one sound modeling methods that are based on the theory that that a human voice signals at time n, the sound signal can be approximated as a linear combination of previous human voice signal p. The goal of the method is to separate the effects lpc formant pitch or frequency of human nature.

1.3 Support Vector Machine

Support vector machine (SVM) is a classification of the types of assisted method (supervised) because

when training, required specific learning targets. SVM can be used for classification that can be applied to handwriting detection, object recognition, voice identification and others.
SVM is a machine learning method that works on the principle of Structural Risk Minimization (SRM) with the goal of finding the best hyperplane that separates two classes in the input space.
In contrast to the neural network strategy which seeks class separating hyperplane, SVM trying to find the best hyperplane in the input space. SVM concepts can be explained simply as an attempt to find the best hyperplane serves as a separator are two classes in the input space.

2 RESEARCH CONTENTS

2.1 Method of Data Collection

Stage of data collection used in this fieldwork is as follows:

a. *Literature*
   The study of literature is a method of collecting data obtained by studying the theory associated with the introduction of sound, including the sound signal processing applications and support vector machine.
b. *Variable Data and Research*
   The data used in this research is data about the value of pitch and formant frequency of the voices of men and women were obtained by recording. In this study, used only two categories (group), namely the category of high-and low-noise sound.
c. *Steps Analysis*
   Steps voice identification analysis is as follows:
   1. *To collect sound data of men and women with high and low categories.*
   2. *Perform predictive classification category of high noise and low-noise analysis method Support Vector Machine (SVM) with the completion of the steps shown in Figure 1 below:*

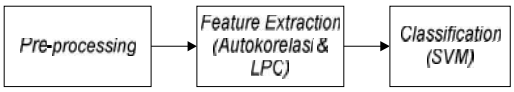


Figure 1 Stages Analysis Identification Vote

2.2 Software Development Method

The method used in the making of this software using the waterfall model as in figure 2 below.

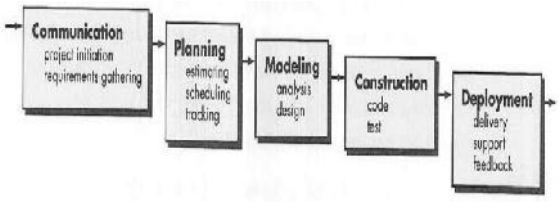


Figure 2 Model Waterfall

This model is a classic model that conduct systematic approach, sequentially thanks to a decline in building software from one phase to another phase. Stage of this model are as follows:

1. *Communication*
   Stage communication is an analysis of the needs and the simulator to conduct the data collection phase of the human voice by recording voices of men and women aged 20-35 years.
2. *Planning*
   Planning stage is a continuation of the process of communication (requirement analysis). This phase will produce a document or user requirement can be regarded as data relating to the wishes of the user in the plan of the simulator that will be done. Software requirements specification will be divided into two parts of SKPL-F (Functional Requirement Specification Software) and SKPL-NF (Software Requirement Specification Non-Functional).
3. *Modeling*
   Modeling stage will translate into a design requirement needs simulator that can be expected before made coding. This process focuses on the design detail (algorithm) procedural. This stage will produce a document called software requirements. The requirement specification is based simulator simulator needs a voice identifier for the application of algorithm analysis and the results of observation. Software requirements specification will be divided into two parts, namely SKPL-F (Functional Requirement Specification Software) and SKPL-NF (Software Requirement Specification Non-Functional).
4. *Construction*
   Stage construction is the code generation process. Coding design or coding is a translation in a language that can be recognized by the computer. This stage which is a stage significantly in doing a simulator, meaning that in this stage of computer use will be maximized. After the coding is completed it will be testing on simulators that have been made. The purpose of testing is to find errors on the simulator to later be repaired.

5. *Deployment*
   Can be said to be the final stage of deployment of manufacture simulator. After doing the analysis, design and coding, the simulator already finished will be used by the user. Then the software that has been made to do maintenance on a regular basis. In the simulator this voice identification, deployment stage is not necessary.

2.3 Method Analysis

2.3.1 Pre-processing

A preprocessing stage to obtain the data required for the identification of high and low sounds. At this stage, the acquisition of samples of the voice signal by recording the respondents using the software audacity, eliminating the effects of noise with a filter pre-emphasis, putting the sound signal into a frame by frame blocking, and minimize the effects of discontinuities at discounted frame voice signal with a hamming window.

a. Pre-emphasis

Pre-emphasis is done to eliminate the irrelevant information and noise by using a low pass filter calculation on the voice signal. Pre-emphasis refers to the process memaksimakan analog signal quality by minimizing the effects of noise such as distortion during the recording and transmission of data.
Pre-emphasis is required in order to obtain a frequency spectral shape of the voice signal is more subtle.
The results of the voice signal pre-emphasis can be seen in Figure 3.

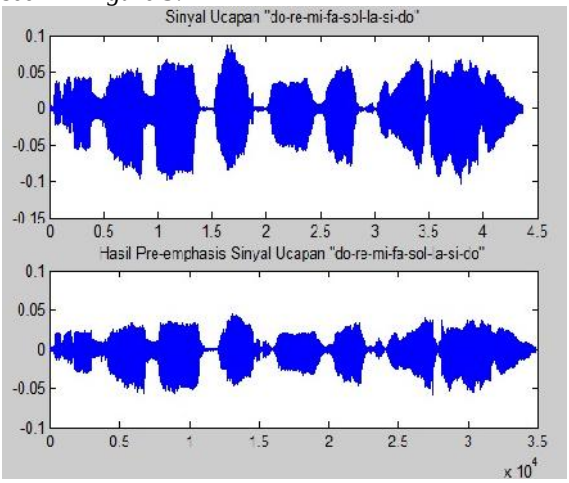


Figure 3 The voice signal pre-emphasis Results

b. Frame Blocking

Sound signal pre-emphasis results are then placed into a frame into sections, where each frame as long as 30 milliseconds and 20 milliseconds separated as far as that would facilitate the calculation and analysis of sound.

c. Hamming Window

Hamming window is needed to reduce the effects of discontinuities of pieces - pieces of sound signals that are on every frame.
Sound signal multiplication results with hamming window is shown as in Figure 4.

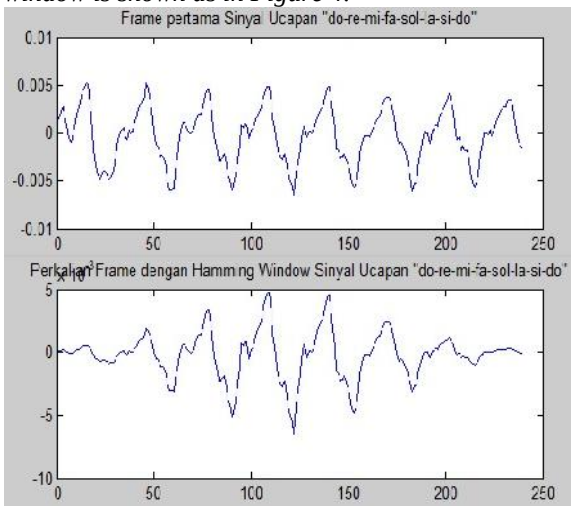


Figure 4 Results of Voice Signal Multiplication with Hamming Window

2.3.2 Feature Extraction

Feature extraction or feature extraction is the process to find the value of voice features, wherein the features of votes cast is pitch and formant. The method used to obtain the value of pitch is autocorrelation, while to get the value of the formant is a linear prediction coding.

2.3.2.1 *Pitch*

Pitch is the fundamental frequency (F0) of the sound signal that is the result of acoustic vocal cord vibration velocity, the greater the vibration of the vocal cords, the higher the pitch value. Pitch period ranges from 10 to 20 milliseconds. Every human being has its own pitch range, depending on the base of the throat owned. The range of a typical pitch (habitual pitch) shared by most men at 50Hz - 250Hz, while women have a pitch (habitual pitch) higher than men, it ranged between 120 - 500Hz. The fundamental frequency changes constantly and give someone linguistic information such as distinguishing between intonation and emotion.
In men when voiced trachea and larynx in the throat opening is wider than in women. The size of the vocal cords in men ranged from 17.5 mm to 25 mm, while in women the size of his vocal cords ranged from 12.5 to 17.5 mm. because of the size of the smaller female vocal cords, the noise generated by the women will be higher.

2.3.2.2 *Formant*

Formant is the natural resonant frequency of the cavity is happening in the field of sound, depending on the shape and size of the sound field and more like an echo. Formant frequencies are not

limited however, to identify a person at least 3 (three) format are analyzed, namely, first formant (F1), formant formant 2 and 3 (F3).

2.3.3 Classification

Classification or classification is the process of data classification features of voice, where the voice feature in this case is the pitch and formant will be classified by support vector machine classification method for generating voice information generated from both the voice features.

a. *Training*
   Training is a stage made to establish a data classification model voice, which data this model will be used as reference data classification sounds from training data. Training data used in the simulator pengideintifikasi this voice is the voice feature data and the group which is the high and low sound features 40 respondents consisting of 20 men and 20 women.
b. *Prediction*
   In this process first enter the data that would have predicted that the data pitch and formant, then matched with the model that has been gained from the training process.

2.3.4 Implementation

Here is the implementation of application identifiers voice simulator.

a. *Testing Modeling Classification*
   *Modeling classification is made to establish the stage of data classification voice training model, where the model's data will be used as reference data classification sounds from training data. Training data used in the simulator pengideintifikasi this voice is the voice feature data and the group which is the high and low sound features 40 respondents consisting of 20 men and 20 women who were kept on file with format * .dat.*

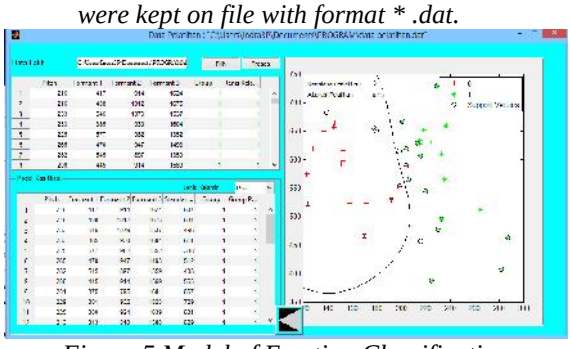


Figure 5 Model of Emotion Classification

b. *Testing Voice Recognition*
   Testing is done by voice recognition voice feature extraction and noise prediction of the sound file recording the sound of respondents amounted to 10 people, consisting of five respondents male and 5 female respondents, where respondents respectively say one sentence.

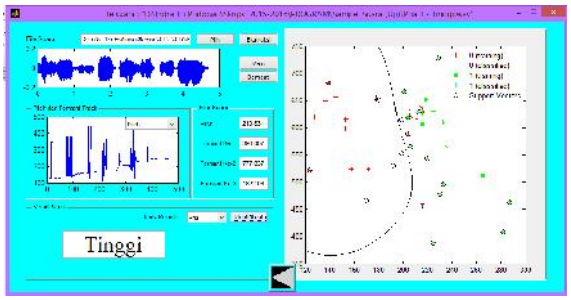


Figure 6 Voice Recognition

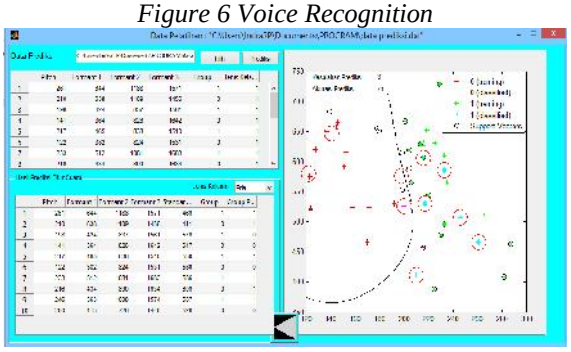


Figure 7 Classification Results Male Voice Features

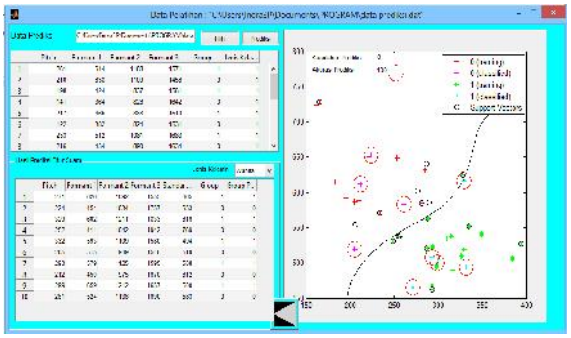


Figure 8 Classification Results Female Voice Features

Based on the results of testing of the simulator software can be concluded that:

1. *The average value of a man's voice pitch high and low is lower than women.*
2. *The average value of formant formant 1st and 2nd women's high and low sounds higher than men.*
3. *Average formant all three men to high and low sounds higher than women.*
4. *Voice features are better suited to represent the voice high for men and women is the pitch.*
5. *The range of pitch and formant for voices of men and women are as follows:*
   men:
   a. *F0*
      Minimal = 121 649 Hz
      Maximum = 260 638 Hz
      Average = 191.1435 Hz
   b. *F1*

Minimal = 363 971 Hz
Maximum = 644.02 Hz
Average = 503.9955 Hz

c. *F2*
Minimal = 725 905 Hz
Maximum = 1187.94 Hz
Average = 956.9225 Hz

d. *F3*
Minimal = 1440.13 Hz
Maximum = 1682.73 Hz
Average = 1561.43 Hz

Women:

a. *F0*
Minimal = 204 869 Hz
Maximum = 332 151 Hz
Average = 268.51 Hz

b. *F1*
Minimal = 410 921 Hz
Maximum = 658 821 Hz
Average = 534 871 Hz

c. *F2*
Minimal = 948 775 Hz
Maximum = 1212.12 Hz
Average = 1080.4475 Hz

d. *F3*
Minimal = 1548.47 Hz
Maximum = 1833.2 Hz
Average = 1690.835 Hz

Simulator is able to predict the sound of high and low for both men and women, as well as the software simulator has been built quite meet the goal of beginning construction simulator software identifiers sound, is evidenced by the training data and predictive data to generate a percentage accuracy of 70% for prediction of sound features men and 100% for the prediction of a female voice.

3 CLOSING

In this chapter contains the conclusions of the research that has been done and suggestions for improvement and development of further research.

3.1 Conclusion

From these results it can be concluded regarding the aspects of the discussion on identifying high and low sounds, namely:

1. *Simulator can identify high and low voices for men and women using support vector machine classification method, with training and prediction accuracy was good.*
2. *The high and low woman's voice more easily identified than the high and low male voice, is proven by testing samples of the sound that produced the testing accuracy of 100% of women and 70% men with voice features the most dominant influence the outcome of the high and low sound prediction in men and women is the pitch.*

3.2 Suggestion

Based on the conclusions that have been described, the expected identifiers sound simulator is developed better, in order to sound more recognizable.

REFERENCES

[1] Endah, Sukmawati Nur dan Dinar Mutiara. 2012. *Analisis Pitch dan Formant Sinyal Ucapan Kata*. Prosiding Seminar Nasional Ilmu komputer. Semarang.

[2] Fadlisyah, Bustami, dan Ikhwanus, M, 2013. *Pengolahan Suara*. Yogyakarta: Graha Ilmu.

[3] Wicaksono, Galieh., Prayudi, Yudi. *Teknik Forensika Audio Untuk Analisa Suara Pada Barang Bukti Digital*. Universitas Islam Indonesia, Yogyakarta: Pusat Studi Forensika Digital.

[4] Pudjo, Widodo, Prabowo., Trias, Handayanto, Rahmadya., Herlawati. 2013, *Penerapan Data Mining dengan Matlab*. Bandung: Rekayasa Sains.

[5] Pan, Yixiong., Shen, Peipei., and Shen, Liping. 2012, *Speech Emotion Recognition Using Support Vector Machine*, International Journal of Smart Home (IJSH).

[6] Roger S. Pressman. 2010. *Software Engineering: A Practitioner's Approach*, 4th ed. New York: McGraw-Hill Companies.

[7] Bagas, Bhaskoro, Susetyo., Riedho, Altedzar. 2012, *Aplikasi Pengenalan Gender Menggunakan Suara*. Seminar Nasional Aplikasi Teknologi Informasi.

[8] Suyanto, S.T., Msc. 2007. *Artificial Intelegence*. Bandung: Informatika.

[9] Rabiner, Lawrence., Juang, Bing-Hwang. 1993, *Fundamental of Speech Recognition*. New Jersey: Prentince-Hall.

[10] Rabiner, Lawrence., W, Schafer, Ronald. 1978, *Digital Processing of Speech Signals*. New Jersey: Prentice-Hall, Inc.

[11] Hadi, Putra, Prabowo. *Penggolongan Suara Berdasarkan Usia dengan Menggunakan Metode K-Means*. Surabaya: Institut Teknologi Sepuluh Nopember.

[12] C, Snell, Roy., Milinazzo, Fausto. 1993, Formant Location From LPC Analysis Data. *IEEE Transaction on Speech and Audio Processing*, vol. I.

[13] Suyanto. 2011, *Artificial Intelligence*. Bandung, Indonesia: Informatika.

[14] Alpaydın, Ethem. 2010, *Introduction to Machine Learning Second Edition*. London: The MIT Press.

[15] Bernhard E. Boser and Isabelle M. Guyon and Vladimir Vapnik. 1992, *A Training Algorithm*

for Optimal Margin Classifiers. Proceedings of the fifth annual workshop on Computational learning theory (COLT).

[16] Ivanciuc, Ovidiu. 2007, *Applications of Support Vector Machines in Chemistry*. Reviews in Computational Chemistry, vol. 23.

[17] Hermawati, Astuti, Fajar. 2013. *Data Mining*. Yogyakarta: Andi.

[18] Kantardzk, Mehmed. 2011. *Data Mining (Concepts, Models, Methods, and Algorithms).* New Jersey: John Wiley & Sons.

[19] McKinney, James (1994). The Diagnosis and Correction of Vocal Faults. Genovex Music Group. ISBN 978-1-56593-940-0.