****

Engineering Center Building 1st Floor

Faculty of Engineering University of Indonesia

New Campus UI Depok 16424

ISAS

(Information Search and Analysis Skills)

**Feature Modelling using Support Vector Machine and Gaussian mixture Model Applied on Voice Based Biometric Authentication**

Name : 1. Dina Nurhayati

2. Julian Maulana

Faculty : Mr. M. Octaviano

Semester : 4(four)

# PREFACE

AssalamualaikumWr.Wb

First at all, give thanks for God’s love and grace for us. Thanks to God for help us and give us a chance to finish this ISAS timely. And we would like to say thank you to Mr. M. Octaviano P. as the faculty that always teach us and give more knowledge about how to make the ISAS well.

The purpose of this ISAS is to finish our tasks and to explain about Support Vector Machines, Mel-Frequency Cepstral Coefficients and the Discrete Cosine Transform Applied on Voice Based Biometric Authentication . In order to people know more about Voice Recognition. We realize that without the guidance from all of the parts we couldn’t finish this ISAS timely.

We feel sorry whenever this manuscript lacks something. However, we do appreciate for your critics and constructive advice to make this paper better soon. Additionally, may this paper be useful for all of us.

Depok, March 2017

Writers

# TABLE OF CONTENTS

[PREFACE i](#_Toc476639442)

[TABLE OF CONTENTS ii](#_Toc476639443)

[TABLE OF TABLES ii](#_Toc476639444)

[TABLE OF FIGURE iii](#_Toc476639445)

[CHAPTER I 1](#_Toc476639446)

[INTRODUCTION 1](#_Toc476639447)

[I.1 Background 1](#_Toc476639448)

[I.2 Problem Domain 3](#_Toc476639449)

[I.3 Objective 3](#_Toc476639450)

[I.4 Methodology 3](#_Toc476639451)

[I.5 Writing Structure 3](#_Toc476639452)

[CHAPTER II 5](#_Toc476639453)

[BASIC THEORY 5](#_Toc476639454)

[CHAPTER III 6](#_Toc476639455)

[ANALYSIS 6](#_Toc476639456)

[III.1 The Methodology of Gaussian Mixture Model 6](#_Toc476639457)

[III.2 the methodology of Support Vector Machine 8](#_Toc476639458)

[III.2.1 Generating Support Vector Machine 11](#_Toc476639459)

[CHAPTER IV 13](#_Toc476639460)

[SUMMARY 13](#_Toc476639461)

[IV.1 Conclusion 13](#_Toc476639462)

[IV.2 Suggestion 13](#_Toc476639463)

[BIBLIOGRAPHY 14](#_Toc476639464)

# 

# TABLE OF TABLES

[Table III.1 Summary of tasks in speaker verification](#_Toc382344579) 9

# TABLE OF FIGURE

Figure II.1Speaker Enrollment......................................................................................................

[Figure III.1overview of GMM features from signal.](#_Toc382344579) 8

[Figure III.2 overview of SVM features from signal. 8](#_Toc382344580)

[Figure III.3Words and its different coefficients combinations 1](#_Toc382344583)6

# CHAPTER I

INTRODUCTION

I.1 Background

In the last few years, Technologies have developed large scale adoption of biometric technologies,Biometrics technology is a security technology that uses parts of the body as identity.Several kinds of biometric technology, such as fingerprint scanners on laptops and smartphone, cameras with built-in face recognition capabilities at airport terminals and stadiums,Smartcard with iris recognition[1], and voice based authentication technologies for account access on smartphones [2].

Voice verification comes in two modalities: text dependent and text independent[3].Text independent voice verification, speaker verification, is not concerned with the text that is spoken. In contrast, in text dependentsystems, the verificationrequires a match on the spoken text as well as a match on the user.

Voice authentication divided in two phases, there are voice enrollment and voice verification. and the feature method divided into some methods. on feature extraction, such as MFCC and LFCC. On feature modelling, such as Gaussian Mixture Model, Support vector machine, etc[4].

The implementation of a Support Vector Machine (SVM) is proposed based on automatic systemfor voice biometric authentication, recognizing the speaker usingMel-Frequency Cepstral Coefficients (MFCC) and the Discrete Cosine  
Transform (Felipe Gomes Barbosa et al). The voice recognition problem can be modelled asa classification problem, where the objective is to obtain the bestdegree of reparability between the classes which represent thevoice. Building an automated speech recognition system capableof identifying the speaker, has many techniques using artificialintelligence and general classification at disposal, Support VectorMachines (SVM) being the one used in this work. The voice samples usedare in the Brazilian Portuguese language and had its featuresextracted through the Discrete Cosine Transform (DCT). Extractedfeatures are applied on the Mel-frequency Cepstral Coefficients (MFCC) to create a two-dimensional matrix used as input to the SVM (Support Vector Machine) algorithm. This algorithmgenerates the pattern to be recognized,leading to a reliable speaker identification using few parameters and a small dataset.

The Gaussian Mixture Model (GMM) (D.A. Reynolds et al) is one of the most widely used voice models in text-independent speaker verification. Based on the GMM, many other methods were derived. One of the most popular methods is Joint Factor Analysis (JFA).

With all this deployment of voice authentication technologies, it becomes crucial to evaluate voice authentication technologies from a principles point of view. Specifically, we are interested in Gaussian Mixture Model and Support Vector Machine.

on this ISAS we want to know the best fature modelling for voice authentication, whether Gaussian Mixture Model or Support Vector Machine.

I.2 Problem Domain

What this ISAS will discuss in this article is about the security concept and the techniques of voice Recognition.

As for the restrictions on the issue of the writing of this article:

1. Basic Principles of Voice Recognition
2. History of Voice Recognition
3. Concept Security of Voice Recognition
4. Algorithm of Voice Recognition

# I.3 Objective

As for the purpose of writing this article is to provide information to the reader about Voice Detection, because the author would also like to learn more about this Technology.

# I.4 Methodology

The methodology used is research methods in library (Library Research) by way of collecting Journal both in the form of digital books, articles, and the opinions of experts on this subject.

# I.5 Writing Structure

To know the description of this paper, the ISAS divides it into four chapters. Each chapter in this study are interconnected between chapters with chapter one another by systematic writing as follows:

**ChapterI Introduction**

This chapter was published about the background of the problem, writing objective, problem domain, writing methodology, writing framework.

**Chapter II Basic Theory**

This chapter describes the basic theory used to solve problems that would be presented, the detailed description of Mel Frequency Cepstral Coefficients, Discrete Cosine Transform, Gaussian Mixture Model, and Support Vector Machine

**Chapter III Problem Analysis**

This chapter describes the Methodology of Gaussian Mixture Model and Support vector machine.

**Chapter IV Conclusion and Suggestion**

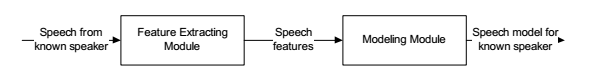
This chapter contains a description of the conclusions and suggestions

# CHAPTER II

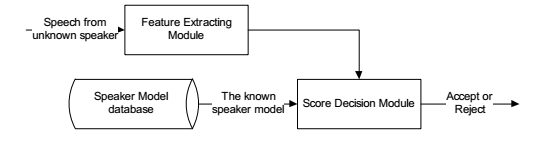
BASIC THEORY

II.1 Voice Authentication phases

Speaker verification systems work in two phases: enrollment and verification.  
During enrollment, a speaker is asked to contribute speech samples whose featuresare then extracted as shown in Figure III.1. The speech features are then used todevelop the users’ speech models. The speech model is stored for future comparison.At a later time, when verification is required, see Figure III.1.2, fresh samples arecollected from the user. After similar extraction phases, the resulting extractedfeatures are compared against the model stored during enrollment.



**FigureII.1Speaker Enrollment**



**FigureII.2Speaker Verification**

Feature extraction, also known as speech parameterization, is dominated by  
the cepstral family. Mel-frequency cepstral coefficients (MFCC) and Linear predictive coding coefficients (LPCC) are the representative technologies. Feature  
modeling methods can be classified as generative approaches and discriminative  
approaches based on the training mechanism. The generative approaches capture  
within class features, including Gaussian Mixture Models (GMM), Hidden MarkovModels (HMM), Vector Quantization (VQ), and the well known Joint Factor Analysis (JFA). Note that the HMM take into the consideration the temporal sequence ofthe feature therefore widely used in text-dependent speaker verification tasks. Othertechniques do not model temporal information, mainly used for text-independentspeaker verification tasks. The discriminative approaches capture the boundary between two classes. The representation ofdiscriminative approaches are artificialneural networks (ANNs) and Support Vector Machines (SVM). Table III.1 summarizes the popular feature extraction and feature modeling methods

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature extraction methods** | **Feature Modelling Methods** | | |
| **MFCC (Mel-frequency cepstral coefficients)**  **LPCC (Linear predictive coding coefficients)** | **Generative approaches** | | **Discriminative approaches** |
| **Text-Independent** | **Text-Dependent** | **SVM (Support Vector Machines) ,**  **ANN (artificial neural networks)** |
| **GMM (Gaussian Mixture Models), VQ (Vector Quantization), JFA (Joint Factor Analysis)** | **HMM (Hidden Markov Models)** |

**Tables II.1Summary of tasks in speaker verification**

II.2 Voice Authentication system works

As a recognition default we proposed the classification andidentification of the voice of a speaker by a keyword, in a textdependent system. The speech signal is sampled and encodedin mel-cepstral coefficients and coefficients of DiscreteCosineTransform (DCT)in order to parametrize the signal witha reducednumber of parameters. Then, it generates two dimensional matrices referring to the Discrete Cosine Transformcoefficients. The elements of these matricesrepresenting twodimensional temporal patterns will be classified bySupportVector Machines (SVMs). The innovation of this work isin the reduced numberof parameters which lies in the SVMclassifier and in the reduction of computational load causedby this reduction of parameters.

II.3Feature Extraction

The most popular feature extraction technique used in voice verification systemsis based on short term cepstral analysis including MFCC (Mel-Frequency Cepstral Coefficients and DCT(Discrete Cosine Transform).

II.4 Feature Modelling

In voice verification, the extracted feature is not directly used as the voice tem  
plate. Instead, a more compressed probabilistic representation will be generated  
based on the voice feature. This process is called feature modeling. The fature modelling have many types, such as Gaussian Mixture Modelling, SVM( Support Vector Machine) , etc.

II.4.1 Gaussian Mixtture Model

The Gaussian Mixture Model (GMM) is one of the most widely used voice models  
in text-independent speaker verification. Based on the GMM, many other methods  
were derived. Gaussian Mixture model is a generative classifier. Chapter 3 describes the Gaussian Mixture Model(GMM).

II.4.2Support Vector Machine

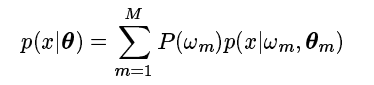
The support vector machine (SVM) is a discriminative classifier that is simple in concept but has some extensions that make it very powerful. Support vector machines solve nonlinear problems by transforming the input feature vectors into a dimensionally higher hyperplane, where the linear separation becomes possible. Chapter 3 describes the SVM in details.

# CHAPTER III

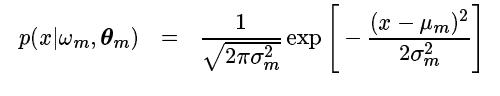
# ANALYSIS

III.1 The Methodology of Gaussian Mixture Model

The Gaussian mixture model for speech representation assumes that a M component mixturemodel with component weights *P(wm)*and parameters *m*can represent the spectral shape.The general form of a univariate mixture model is



where*P(wm)* is the prior probability of component *m* and *m*the componentparameters. Themixture components in this case are Gaussian, and hence can be described by



where m is the mean and the standard deviation for component .

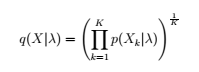
The GMM is a collection of weighted Gaussian distributions λ which reflects the  
real distribution of mass2. A GMM is denoted by λ = {pi, µi, Σi} i = 1, 2, ...Nwherepi gives the weight of ith component. Therefore, P pi = 1. The mean and varianceof the ith component are represented by µi and Σi, respectively. N represents thenumber of Gaussian components. The Gaussian Mixture Density is defined as



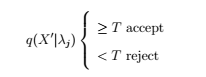
Where X is a random vector, bi(X) is probability density function of ithcomponentexplicitly given as



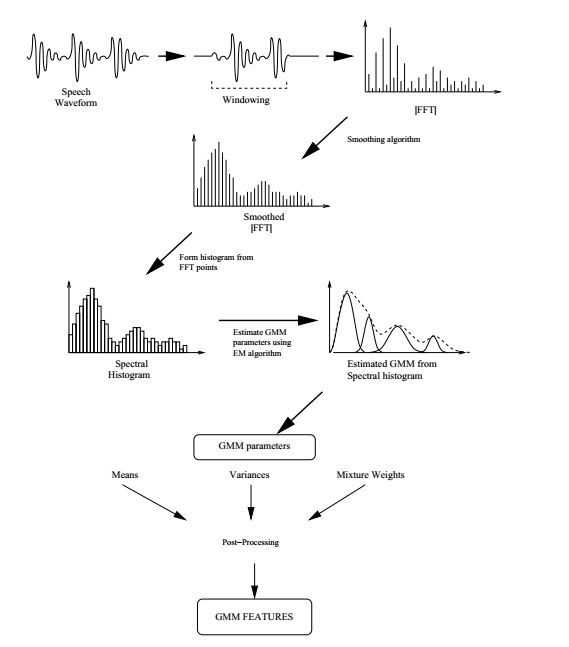
Given K observations of the random vector X, the probability of X following the  
GMM λ can be expressed as



where Xk is the kth observation of X. For a known speaker j, the GMM model λj  
is computed such as to maximize the overall probability q(Xj|λj). Therefore, the  
GMM λj provides a voice template. In GMM based biometric verification system,a two phase scenario is applied. In the enrollment phase, a feature Xjextractedfrom a person j, is used to generate a template GMM λj. In the verification phase, a decision function



is computed. Where T is the pre-defined constant threshold and X′ is a fresh featureextracted from an unknown person who claims to be j. If the likelihood q(X′|λj)is greater than the thresh hold, the unknown person passes the verification as jotherwise the authorization fails. The main advantage of the adapted GMM is that the trainingphase for a speaker is much faster while at the same time it gives a more accurateverification performance. In this dissertation we will base our analysis on the morepopular adapted GMM.



**Figure II1.1: overview of GMM features from signal.**

III.2 The methodology of Support Vector Machine

SVM, as a supervised learning technique, can infer from a set of labelled examples, on which the class is known, a function capable of predicting new labels from unknown examples. The simpler derivation of the SVM algorithm is  
the linear function case, where to illustrate the separation plane generated by it, we can drawn a line that represents the decision boundary that correctly classifies some data set. *Support vector machines* solve nonlinear problems by  
transforming the input feature vectors into a dimensionally higher hyperplane, where the linear separation becomes possible. Maximum discrimination is obtained with an optimal placement of the separation plane between the borders of the two classes. If we assume a set *𝐻* of points *𝑥𝑖 ∈ 𝑅𝑑* with *𝑖* = 1*,* 2*,* 3*, . . . , 𝑛.* Each one of the *𝑥𝑖* belongs to either of two classes labelled *𝑦𝑖 ∈ {−*1*,* 1*}*. Establishing the equation of a hyperplane that divides *𝐻* is the desired goal, and for this purpose we have some preliminary definitions. By taking the set *𝐻*, if linearly separable, there exists *𝑤 ∈ 𝑅𝑑* and *𝑏 ∈ 𝑅* to satisfy

*𝑦𝑖* (*𝑤 ⋅ 𝑥𝑖* + *𝑏*) *≥* 1

where *𝑖* = 1*,* 2*,* 3*, . . . , 𝑛.*

The pair (*𝑤, 𝑏*)defines a hyperplane

(*𝑤 ⋅ 𝑥𝑖* + *𝑏*) = 0

This defines a separating hyperplane, leading to the problem of finding the optimal separating hyperplane, to which we try to minimize *𝑤* as the following



where *𝑦𝑖* (*𝑤 ⋅ 𝑥𝑖* + *𝑏*) *≥* 1.

Then converted to a dual problem by Lagrange multipliers

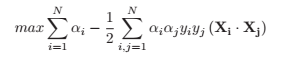


where ∑*𝑁 𝑖*=1 *𝛼𝑖𝑦𝑖* = 0*, 𝛼𝑖 >* 0*.*

When *𝐻* cannot be separated linearly, nonnegative slack  
factor *𝜉* = (*𝜉*1*, 𝜉*2*, . . . , 𝜉𝑁*) is introduced. There is

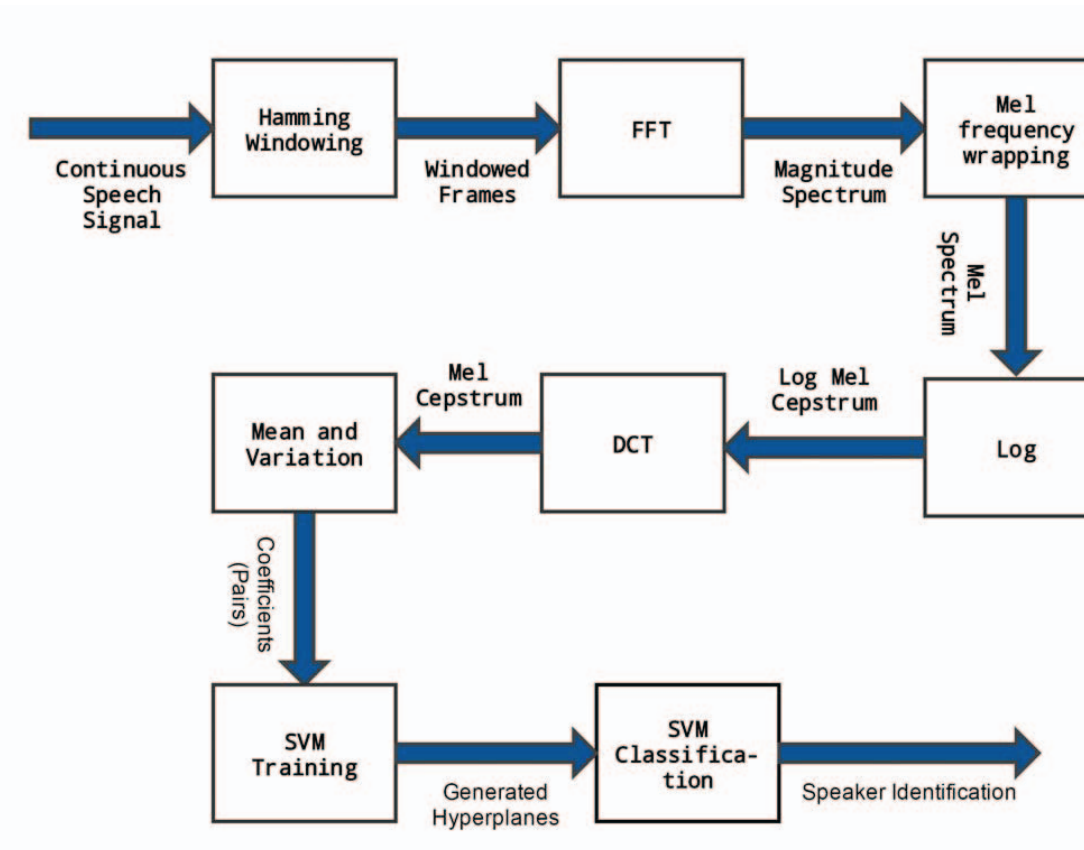
*𝑦𝑖* (*𝑤 ⋅ 𝑥𝑖* + *𝑏*) *≥* 1 *− 𝜉𝑖*

The optimal problem can be described as



where ∑*𝑁 𝑖*=1 *𝛼𝑖𝑦𝑖* = 0*, 𝑖* = 1*,* 2*, . . . , 𝑁,* 0 *≤ 𝛼𝑖 ≤ 𝐶*.

This is the general form of SVM. If *𝐶* tends to infinite, It is a problem that can be solved by quadratic programming using sequential minimal optimization.  
When the data is easily linearly separable, the previous equations are able to classify with minimum error, but when the data is highly nonlinear one needs to use the kernel method, in which the data is put in a higher dimensional plane, where it can be linearly separated. This is possible when we take the dot  
product of *𝑋𝑖⋅𝑋𝑗* and apply another function, validated by the Mercer’s Conditions, that in some cases, like the *Radial Basis* *Function* (RBF).



**Figure II1.2: overview of SVM features from signal.**

III.3 Performance of GMM and SVM

Based on the thesis titled GMM And SVM classifiers for speaker verification by Shanama Afnan, the author compare the False Reject Rate (FRR) and the False Accept Rate (FAR) in GMM and SVM classification method for the same 30 male and30 female speakers. In the case of SVM classifier, we present the results for bothpolynomial and RBF kernels. The following tables contain the 6 sets of results. Eachset includes 10 speakers.

In the training phase of the GMM verification system, the used Equal Error  
Rate (EER) for selecting individual threshold. However, in the testing phase author calculated False Reject Rate and False Accept rate as a measure of verification.the accuracy of Support Vector Machine (SVM) is (98.58%). and Gaussian Mixture Model is 98.46%.[5]

# CHAPTER IV

# SUMMARY

IV.1 Conclusion

in this ISAS, we reviewed an authentication model. the Gaussian Mixture Model and Support Vector Machine. There is much literature about extending or improving well known algorithms. For example, DCT(Discrete Cosine Transform), Mel-Frequency Cepstral Coefficients, and Support Vector Machine. Many algorithms are being built around this idea.

as a feature modelling Gaussian Mixture Model, The main advantage of the adapted GMM is that the training phase for a speaker is much faster while at the same time it gives a more accurate verification performance.as a feature modelling Support Vector machine as the classification, revealing the real time application possible, fast, precise and very reliable.

IV.2 Suggestion

for future works, we suggest the next author to review attackenon voice authentication or how an attacker can impersonate anyone in the database after authorization attempts.

**BIBLIOGRAPHY**

[1] M. Abdullah , 2015, "Smart Card with Iris".

[2] Jakub Galka, 2014, “Voice Authentification Embedded Solution”.

[3] Chenghuang Yang, 2014, “Security in Voice Authentication”, Worcester Polytechnic Institute.

[4] R G Maduranga M Jayamaha, 2008, “Human Voice Authentication System Using Hidden Markov Model”, Sri Lanka Institute of Information Technology”, New Kandy Road, Malabe, Sri Lanka

[5] Ajay M. Patrikar; John. P. Baker, " Improving Accuracy of Gaussian Mixture Model Classifiers with Additional Discriminative Training".IEEE. Johns Hopkins University Applied Physics Laboratory. Laurel, MD, USA,