Automatic Speaker Recognition System Using Fuzzy C-Means Artificial Neural Networks

MOHD ZAMRI IBRAHIM, MARZUKI KHALID & RUBIYAH YUSOF

ABSTRACT

Speaker recognition is a process of recognizing someone by their voice. The goal of speaker recognition is to extract, characterize and recognize the information about speaker identity. In this paper, we discussed both Fuzzy C-Means (FCM) and Artificial Neural Network (ANN) approach to speaker recognition system. The proposed system comprises of three main modules, a feature extraction module to extract necessary features from speech waves, speaker modeling module to generate the speaker model and FCM and ANN module to classify the speakers whether to accept or reject. The proposed intelligent learning system has been applied to a case study of text-dependent speaker recognition system and the performance is evaluated by applying two types of feature extraction techniques: Mel Frequency Cepstral Coefficients (MFCC) and Linear Predictive Cepstral Coefficients (LPCC). Experiment showed that the new proposed systems provide significantly higher performance compare to conventional method.

Keywords: Speaker recognition, artificial intelligence, fuzzy c-means, artificial neural networks, dynamic time warping and vector quantization.

INTRODUCTION

Crime and illegal access in finance and security applications has reached unpredicted levels. Although a lot of security systems have been developed to detect the fraudulence, but it is still unable to resolve the problem. Conventional keys, access codes, access cards are proving ineffective as they are easily lost, stolen, copied, observed or left at home. Thus the emergence of biometrics has addressed the problems that plague the traditional authentication method. Biometrics refers to the automatic identification of a living person based on physiological or behavioral characteristics. The person to be identified is required to be physically present at the point-of identification. There are many types of biometric technologies available today such as face recognition, fingerprint recognition, hand geometry, iris recognition, signature recognition and speaker recognition. The speaker recognition system has the benefit of being a versatile, easy-to-use and non-intrusive system for verifying personal identity.

The proposed speaker recognition system consists of speaker identification and speaker verification tasks. In speaker identification, the a priori identity is not known: the system has to decide which of the voice stored in a database resembles the most to the voice to recognize In verification task, the system knows a priori the identity of the user, and has to verify this identity, that is, the system has to decide whether the a priori user is an impostor or not.. The primary goal of this paper is to present the performance evaluation carried out using fuzzy c-means and artificial neural network for speaker verification system. The remainder of this paper is organized as follows. Section 2 describes the system process flow and the modules of the proposed speaker verification system. Section 3 elaborates the methodology used for the preprocessing, feature extraction, speaker modeling and classification of the proposed system. Section 4 presents and discusses the experimental results and the conclusions are drawn in section 5.

SYSTEM OVERVIEW

The proposed face recognition system consists of two (2) phases which are the enrollment and identification/verification phases as depicted in Fig. 1. It consists of several modules which are Preprocessing, Feature Extraction, Training, Identification and Verification.

Enrollment phase

The voice is acquired using a microphone and stored in a database. Next, the voice is detected and trained. During training, the voice is preprocessed using endpoint detection and first order FIR filter. The features of the voice are extracted using several feature extraction techniques. The features data is then used to develop speaker model and stored together with the user identity in a database.

Identification/verification phase

A user's voice biometric data is once again acquired and the system uses this to either identify who the user is, or verify the claimed identity of the user. While identification involves comparing the acquired biometric information against templates corresponding to all users in the database, verification involves comparison with only those templates corresponding to claimed identity. Thus, identification and verification are two distinct problems having their own inherent complexities.
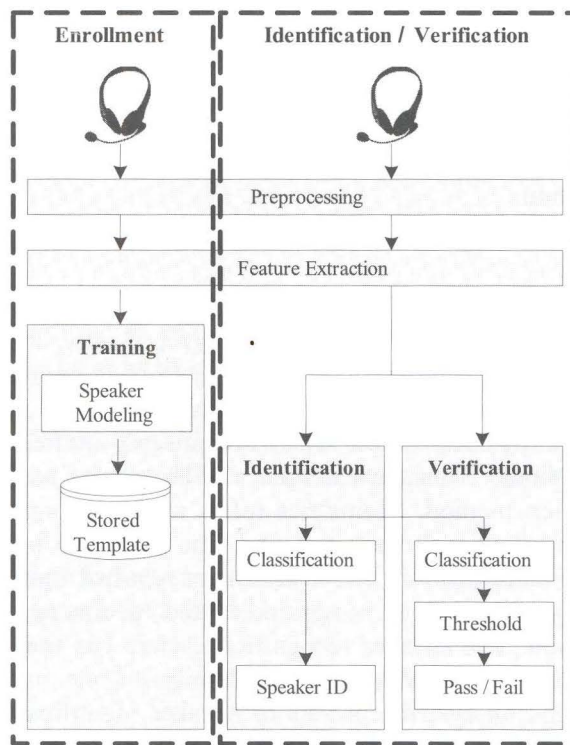

Figure 1: Speaker Recognition System

METHODOLOGY

Preprocessing

The purpose of the pre-processing module is to make the speech signal as clean as possible, noise is eliminated, long-time spectral trends are removed, and the signal is spectrally flattened to give the best immunity to measurement imperfections [1]. The preprocessing is crucial as the robustness of a speaker recognition system greatly depends on it. The type of preprocessing used is endpoint detection and preemphasis (first-order FIR filter).