

An Intelligent Text Independent Speaker Identification Using VQ-GMM Model Based Multiple Classifier System

Authors : Ben Soltane Cheima, Ittansa Yonas Kelbesa

Abstract : Speaker Identification (SI) is the task of establishing identity of an individual based on his/her voice characteristics. The SI task is typically achieved by two-stage signal processing: training and testing. The training process calculates speaker specific feature parameters from the speech and generates speaker models accordingly. In the testing phase, speech samples from unknown speakers are compared with the models and classified. Even though performance of speaker identification systems has improved due to recent advances in speech processing techniques, there is still need of improvement. In this paper, a Closed-Set Tex-Independent Speaker Identification System (CISI) based on a Multiple Classifier System (MCS) is proposed, using Mel Frequency Cepstrum Coefficient (MFCC) as feature extraction and suitable combination of vector quantization (VQ) and Gaussian Mixture Model (GMM) together with Expectation Maximization algorithm (EM) for speaker modeling. The use of Voice Activity Detector (VAD) with a hybrid approach based on Short Time Energy (STE) and Statistical Modeling of Background Noise in the pre-processing step of the feature extraction yields a better and more robust automatic speaker identification system. Also investigation of Linde-Buzo-Gray (LBG) clustering algorithm for initialization of GMM, for estimating the underlying parameters, in the EM step improved the convergence rate and systems performance. It also uses relative index as confidence measures in case of contradiction in identification process by GMM and VQ as well. Simulation results carried out on voxforge.org speech database using MATLAB highlight the efficacy of the proposed method compared to earlier work.

Keywords : feature extraction, speaker modeling, feature matching, Mel frequency cepstrum coefficient (MFCC), Gaussian mixture model (GMM), vector quantization (VQ), Linde-Buzo-Gray (LBG), expectation maximization (EM), pre-processing, voice activity detection (VAD), short time energy (STE), background noise statistical modeling, closed-set tex-independent speaker identification system (CISI)

Conference Title : ICMAP 2014 : International Conference on Multimedia Applications and Processing

Conference Location : London, United Kingdom

Conference Dates : October 24-25, 2014