

# COMPARISON OF CLUSTERING ALGORITHMS FOR SPEAKER IDENTIFICATION

*Nikolay Lubimov, Evgeny Mikheev, Alexey Lukin*

Moscow Lomonosov State University, Moscow, Russia

In this paper we consider the problem of text-independent speaker identification that refers to acoustic recognition research. Many different techniques have been presented over past several decades. A state-of-the-art technique uses Gaussian Mixtures (GMM) for modeling speaker data distribution presented by MFCC [1] or LPCC [2] features. The classification is obtained by choosing the speaker class with maximum likelihood on the observed data. More complex approach considers the discriminative capability of methods like Support Vector Machine (SVM) in order to separate different acoustic classes. A hybrid system for speaker identification presented in [4] successfully combines advantages of GMM's generative capability and SVM's discriminative power by introducing Fisher kernel.

We examine the simplest scheme for construction of a speaker identification system having 3 major stages: 1) pre-processing, 2) initial clustering in feature space, 3) Gaussian mixture model parameter re-estimation. There are many different successful techniques proposed for the pre-processing step. The Expectation-Maximization (EM) algorithm used for Gaussian mixture parameter re-estimation is also well documented. On the other hand, it is not obvious how to initialize the recurrent formula of the EM algorithm in this task, because it is known that convergence properties of the EM algorithm strongly depend on initial approximation. An interesting problem is: which type of initial clustering in feature space should be used to obtain better results? In this paper we describe some existing methods for making an initial approximation of in the EM procedure, and show how these methods affect to final speaker recognition rate. Using different algorithms for feature space clustering we construct several classifiers for speaker identification. We perform comparison between them using identification rate error on a speaker database with telephone-quality signals. Our main goals are to compare the performance of fuzzy and hard clustering methods, and also examine the influence of deterministic and random initializations of the EM algorithm.

We have performed the comparison of different clustering methods for speaker identification: the standard K-means clustering, K-means++, Linde-Buzo-Gray, Fuzzy C-means, and Gustafson-Kessel algorithms. It has been found that Gaussian mixture model performance depends on deterministic properties of EM initialization method. Linde-Buzo-Gray (LBG) method outperforms other non-fuzzy clustering approaches probably because of natural arrangement of cluster centers along principal components of the data, rather than random choice used in K-means or K-Means++. Fuzzy clustering algorithms show better results because they are more deterministic and use complete dataset during clustering iterations. All of the tested clustering algorithms except Gustafson-Kessel divide the dataset into spherical clusters. Gustafson-Kessel finds ellipsoids, so it shows best result.