PhD Thesis Abstract

"Contributions to the development of the speaker verifications systems on handheld devices"

Author: Eng. Florin-Dragoș CURELARU, *e-mail*: patronixys@yahoo.com
Thesis supervisor: Brig. Gen.(r.) Prof. Eng. Mihai RADU, PhD

The main purpose of this thesis was to find the most robust text-independent speaker verification methods and their associated parameters, which lead to the best performances of the systems trained on limited enrolling data and evaluated on short utterances collected under real conditions. In order to evaluate the speaker verification methods under real conditions, the experiments were conducted with utterances collected on handheld devices, which involve recordings in various acoustic environments with different types of microphones. The constraints related to the limited amount of enrolling data and to the short utterances come from two usability issues of a verification system, namely, to quickly enroll new clients and to avoid asking the clients to spend considerable time at the authentication/evaluation phase.

The experiments were conducted on utterances comprised in two corpora which were made publicly available for free: the *MIT Mobile Device Speaker Verification Corpus* (MIT-MDSVC) and the *CSTR Voice Cloning Toolkit Overview* corpus (CSTR VCTK). The MIT-MDSVC corpus contains limited data composed of short phrases collected on handheld devices in three different acoustic environments with two microphones. The CSTR VCTK corpus is around ten times larger than MIT-MDSVC and it was used especially for investigating the evolution of the performance of the system when the training data are extended with utterances collected after a different schema regarding the acoustic conditions of the recordings.

Considering the modeling used to represent the speakers and the scheme used to classify the speakers, modern approaches of speaker verification generally belong either to the generative, or to the discriminative category. Each category has specific properties which are not present in the other one. In order to take advantage of the properties from both categories, several systems are based on discriminatively trained generative models. Having started from the classification of the methods in generative, discriminative and methods based on discriminatively trained generative models, five representative methods were chosen to be comparatively evaluated, at least one method in each category. The five representative methods are: *Gaussian Mixture Model – Universal Background Model* (GMM-UBM), the i-vector representation in conjunction with Probabilistic Linear Discriminant Analysis (that fall in the generative category), Support Vector Machine (SVM) with polynomial kernel (that fall in the discriminative category), and SVM with Fisher kernel and with GMM Supervector Linear kernel (that are based on discriminatively trained generative models).

For comparing the speaker verification methods in an objective manner and for analyzing the influence of the Voice Activity Detectors (VAD) in the performance of the speaker verification systems, an extensive set of experiments was conducted with four popular VADs: an energy-based VAD included in Alize/LIA_RAL toolkit, a self-adaptive VAD proposed by Kinnunen et al., a statistical-based VAD proposed by Sohn et al. and the standardized G.729B VAD.