

Speaker Verification Using Adapted Gaussian Mixture Models

Dissertation proposal

Eduardo Martins Barros de Albuquerque Tenório (*student*)

Tsang Ing-Ren (*advisor*)

George Darmiton da Cunha Cavalcanti (*reviewer*)

{embat,tir,gdcc}@cin.ufpe.br

Centro de Informática,
Universidade Federal de Pernambuco

October 16, 2014

Abstract

This dissertation is a reproduction of MIT Lincoln Laboratory's Gaussian Mixture Model (GMM)-based speaker verification system used successfully in several NIST Speaker Recognition Evaluations (SREs) implemented by Douglas A. Reynolds. The system is built around the likelihood ratio test for verification, using simple but effective GMMs for likelihood functions, an Universal Background Model (UBM) for alternative speaker representation, and a form of Bayesian adaptation to derive speaker models from the UBM. Additionally, a new image processing technique named Fractional Principal Component Analysis (FPCA) will be used in order to improve experimental outcomes, with no guarantee of success due to the lack of experiments in SREs.

keywords: *speaker recognition; Gaussian Mixture Models; likelihood ratio test; Universal Background Model; Fractional Principal Component Analysis*

1 Introduction

Speaker recognition is a field of *Digital Speech Processing*, a set of techniques aimed to identify who is speaking through the analysis of their voices' characteristics, also called *voice recognition*. It is different from *speech recognition*, in which the aim is to recognize what is being said. Additionally, there are two different subfields of speaker recognition: *speaker verification* and *speaker identification*.

This project is focused on the speaker verification field, when a speaker claims to be of a certain identity and the voice is used to verify this claim (the other is

concerned about what is said). The dissertation is a reproduction of the main idea presented in [1], in which a likelihood test is performed using an UBM with adapted GMMs for speakers. The project's scope is reduce, using one speaker per track in the base [2].

Additionally to Reynolds' work, there will be an attempt to improve outcomes using FPCA, described in [3]. This technique is recent and was originally developed to *Digital Image Processing*, with no guarantees of good experimental results (or anyone) in speech. As image and speech processing are just specializations of signal processing, we have confidence to receive some interesting outcome.

2 Objectives

This project main objective is to provide a decent speaker verification system, with experimental results similar to (or better than) those presented in [1], as well as a possible improvement given by the addition of FPCA. The specific objectives include:

- Develop a speaker recognition system with outcomes similar to those presented in [1].
- Study the differences in performance between the adapted speaker GMM and the simple speaker GMM.
- Verify the possibility of implement FPCA, and if yes, try to improve the outcomes with this modification.

3 Methodology

The project's idea is easily findable [1], as well as the base used [2] to train, test and validate the system. The

same to the additional proposal [3]. All tools used are open source (Python, NumPy, SciPy, etc.) and requires no training, just basic programming knowledge.

The work will be divided in the following stages, with schedule detailed in the next section:

- Study of the [1] to properly understand the idea, as well as a survey of the state of the art in speaker verification algorithms (particularly in works related to [1]), in order to improve the dissertation if possible.
- Development of Voice Activity Detector (VAD) and feature extractor, using the tools specified before.
- Development and training of the GMM-UBM.
- Development of the adapted speaker GMM. Test the speaker's models.
- Attempt to improve the project using the idea of FPCA [3]. If not possible, detail the process and try to find a new solution.
- Write the report while executing each stage.

4 Schedule

The schedule is divided in months, going from October (when this proposal is written) until February (when the dissertation is delivered).

October 2014

- Literature research.
- Report writing.

November 2014

- Literature research.
- Voice Activity Detector (VAD) and feature extractor development.
- Report writing.

December 2014

- Speaker GMM adaptation.
- System testing.
- Report writing.

January 2015

- FPCA research.
- FPCA development.
- FPCA testing.
- Report writing.

February 2015

- Project review.
- Presentation.

References

- [1] D. A. Reynolds et al., "Speaker verification using adapted gaussian mixture models," *Digital Signal Processing*, vol. 10, (1-3) pp. 19-41, 2000.
- [2] R. H. Woo et al., "The MIT Mobile Device Speaker Verification Corpus: Data Collection and Preliminary Experiments," in *The Speaker and Language Recognition Workshop (IEEE Odyssey 2006)*, San Juan, Puerto Rico, 2006.
- [3] C. Gao et al., "Theory of fractional covariance matrix and its applications in PCA and 2D-PCA," *Expert Systems with Applications*, vol. 40, (1-3) pp. 5395-5401, 2013.