I-Vector Speaker Recognition on Android Devices

Christian Vázquez-Machado
Pedro Colón-Hernández

Advisor: Pedro Torres-Carrasquillo
Sponsor: MIT Lincoln Laboratory

Introduction

The current state-of-the-art systems in speaker recognition are often based on a technique known as I-Vector. This technique has resulted in over five-fold performance improvement over previous leading techniques, while also resulting in fast and less complex implementations. This has opened up new ways to use these algorithms on small form factor devices such as handheld and tablet-based mobile devices.

Systems and Performance	
Class	Classification Rate
I-Vector	94%
GMM	81%
SVM	82%

Metrics for Success

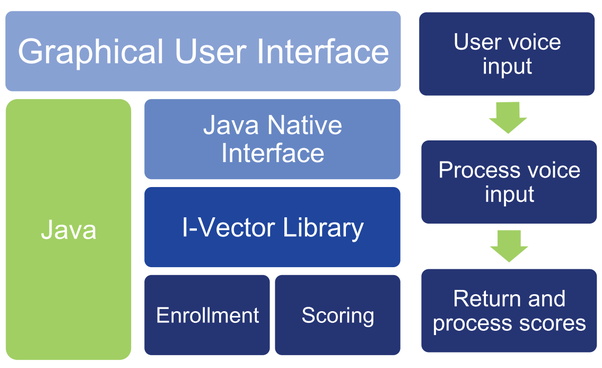
- **Port** the current C++ and initial Java Native Interface (JNI) to perform I-Vector based speaker recognition on Android devices
- **Design** an application for an Android device with an intuitive graphical user interface (GUI)
- **Test** the application to determine performance levels and feasibility of this technology in realistic conditions

Problem

- Existing I-Vector codebase is implemented in **C++**, while Android has limited support for native C++
- The I-Vector system performs data intensive operations that take considerable amounts of **time** and **resources**

Methodology

1. Use Android Native Development Kit to compile Boost C++ Library and I-Vector
2. Fix cross-platform issues
3. Use the Android Development Tools to prototype and test an initial GUI design
4. Test the application with 50 volunteers
5. Address issues that arise from testing
6. Perform final testing and debugging



Porting Results

Scoring Operations		
Size	Scoring	Scoring (Pre-Compute)
64	41.4s	14.3s
512	70.5s	15.2s
2048	Approx. 1h	-

Results were taking *too much time!*

Optimized Scoring Operations			
Size	Subspace Dimensions	Scoring	Scoring (Pre-Compute)
64	400	4.8s	1.7s
512	200	7.6s	1.4s
512	400	28.5s	3.3s
2048	200	26.8s	4.2s

Application Prototype

LLTalk is a prototype that allows easy speaker identification on Android devices

LLTalk's Adaptive Design

3.2" phone | 7" tablet | 10" tablet

Field Testing Results

LLTalk was tested with 50 volunteers:

Results	
Identified	Miss
42	8

Evaluations yielded 84% classification rate under different noise environments and languages

Conclusion

- I-Vector speaker identification is viable for mobile platforms
- Despite I-Vector's data intensive operations, there is room for optimizations that allow real-time operations to be possible
- LLTalk's intuitiveness allows quick, useful results in the field with no need to understand the underlying system
- The application yields a performance of about 84% with one of the lowest performing configurations for the I-Vector (64-400)

Future Work

- Porting I-Vector to other mobile platforms
- Rigorous testing in harsher conditions
- Make identification operations even faster
- Exploit connectivity with external back-end

References

1) N. Dehak, P. Kenny, R. Dehak, P. Ouellet, and P. Dumouchel, "Front-end factor analysis for speaker verification," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 19, pp. 788–798, May 2011.

2) P. Torres-Carrasquillo, J. Acevedo-Aviles, and R. Ford, "Android Application for Language Identification," presented at High Performance Embedded Computing Workshop, Lexington, MA, September 2011.

503698