

UNIVERSITY OF SOUTHAMPTON

Speech Recognition on Embedded Hardware

by

Ricardo da Silva

Technical Report

Faculty of Engineering and Applied Science
Department of Electronics and Computer Science

March 21, 2013

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING AND APPLIED SCIENCE
DEPARTMENT OF ELECTRONICS AND COMPUTER SCIENCE

by Ricardo da Silva

This report presents a proof of concept system that is aimed at investigating methods of performing speech recognition in embedded systems. It makes use of two new electronic boards that are currently under development at the University of Southampton, and implements one part of a speech recognition system using an FPGA and a Linux applications processor.

Contents

Acknowledgements	iii
1 Introduction	1
1.1 Goals	1
1.1.1 Speech Recognition	1
1.1.2 Theoretical understanding	1
1.1.3 The Micro Arcana	1
1.2 Motivation	2
1.3 Contributions	2
2 Background	3
2.1 Speech Recognition Systems	3
2.2 FPGAs	3
2.3 Personal Contribution	3
3 System Design	5
4 Conclusions and Future Work	6
Bibliography	7

Acknowledgements

Thanks to Steve Gunn, Srinandan Dasmahapatra

Chapter 1

Introduction

1.1 Goals

At the highest level, the primary goal of this project is to develop a system that performs one part of the speech recognition process on embedded hardware. In pursuing this goal, the aim is to achieve several other goals that will be beneficial to the author and to the University of Southampton.

1.1.1 Speech Recognition

The

1.1.2 Theoretical understanding

A major goal of the project is to develop a higher level of understanding of the algorithms used in speech recognition, and to get experience designing a large-scale embedded application. This complements the interests of the author and the subjects being studied, in particular, Intelligent Algorithms and Digital Systems Design.

1.1.3 The Micro Arcana

In terms of hardware, one of the project goals is to further development of the Micro Arcana family of boards, and provide a valuable use case example as described in the Motivation section. Part of the project is setting up and configuring these two boards, so that they may be easily picked up by undergraduates. In addition, the aim is to build the entire system on these two boards, making it a self-contained embedded design.

1.2 Motivation

Speech recognition is an interesting computational problem, for which there is no fool-proof solution at this time. Recently the industry for embedded devices and small-scale digital systems has expanded greatly, but in general these devices do not have the power or speed to run speech recognition. Field Programmable Gate Arrays (FPGAs) may present a way of increasing the capability of such systems, as they are able to perform calculations much faster than traditional microprocessors.

The Micro Arcana is a new hardware platform aimed at undergraduate students, being developed by Dr Steve Gunn. As they are still under development, they are very untested, and very little documentation exists. In order to improve their reception by students, it would greatly help to have proven use cases and examples of how these boards may be used individually and together. Using a larger FPGA (such as an Altera DE board) was considered during the planning stages of this project, but it was decided that part of the challenge was to develop and use the Micro Arcana.

1.3 Contributions

The project implements one part of a modern speech recognition system, using two development boards from the Micro Arcana family. It is designed to be a proof of concept exercise, in order to explore the capabilities of the boards, and expand the author's knowledge of the relevant systems. Specifically, the project required substantial research into HMM based speech recognition systems, embedded Linux, and digital design. The resulting system, described in detail in Chapter 3, uses an FPGA to perform the most computationally expensive part of HMM based recognisers – scoring the states of each HMM model for a given input vector. Essentially, the ARM based “L’Imperatrice” is used as the application controller, and is connected to the FPGA based “La Papessa” board which performs the CPU intensive mathematical calculations. Given an observation vector, the FPGA will process it and send back scores for each state in the speech model. The other main processes in a speech recogniser, such as pre-processing and decoding, are tasks that are well suited to software implementation.

Chapter 2

Background

2.1 Speech Recognition Systems

In general, ‘Speech Recognition’ refers to the process of translating spoken words or phrases into a form that can be understood by an electronic system, which usually means using mathematical models and methods to process and then decode the sound signal. Translating a speech waveform into this form typically requires three main steps [1]. The raw waveform must be converted into an ‘observation vector’, which is a set of data that is compatible with the chosen speech model. This data is then sent through a decoder, which attempts to recognise which words or sub-word units were spoken. These are then sent through a language model, which imposes rules on what combinations of words of syntax are allowed. This project focusses on implementing the first stage of the decoder, as this is an interesting task from an electronic engineering point of view.

There are a variety of different methods and models that have been used to perform speech recognition. An overview of the most popular will be described here, along with the chosen approach.

2.2 FPGAs

2.3 Personal Contribution

The initial goal of the project, to build a complete speech recognition system, was very ambitious and had to be narrowed down as more was learnt about the complexities of modern speech recognition systems. Instead of attempting to build a full system, it was decided that development would focus on the decoding stage of recognition. Speech pre-processing is a very established process, and implementing it is usually a case of linking together the appropriate libraries.

The result of this project is a system spread across two development boards from the Micro Arcana: L'Imperatrice and La Papessa. This involved... -

- Benefit from developing my knowledge of SV, speech recognition, intelligent algorithms, etc
- Providing an application for the Micro Arcana family, and showing that it can do stuff
- Providing a solid basis for future work in the area

Chapter 3

System Design

Chapter 4

Conclusions and Future Work

Bibliography

- [1] S.J. Melnikoff. *Speech recognition in programmable logic*. PhD thesis, University of Birmingham, 2003.