Embedded Robust Speech Recognition Engine

Project Investigators

Dr. Waleed H. Abdullah and Professor Zoran Salcic
Electrical & Electronic Engineering Department

Synopsis

Natural man-machine interaction is currently one of the most unfulfilled pledges of automatic speech recognition (ASR). The purpose of an automatic speech recognition system is to accurately transcribe or execute what has been said.

The aim of this research is to search for a robust speech recognition engine that can be embedded in different applications. We have achieved encouraging advancements in the algorithmic side of implementation while the embedded part needs further investigation which might even require modifying the algorithms to be more suitable for embedded hardware and software implementation.

State-of-the-art speech recognition systems consist of four basic modules: the signal processing, the acoustic modelling, the language modelling, and the search engine. In the signal processing module we will be looking for extracting prominent features from the speech signal. The feature extraction comprises computing representations of the speech signal that are robust to acoustic variation but sensitive to linguistic content. We will be using our newly developed technique called Gammatone cepstral coefficients (GTCC) [8]. These features will also be used in extracting the speech segments from the non-speech background and for gender discrimination [1,2,5,9].

In the acoustic modelling we will be using hidden Markov modelling (HMM) as it is the state of the art technique [3,7]. HMM is a statistical framework that supports both the acoustic and the temporal modelling. The neural network will be hybridised with the HMM to alleviate some of its suboptimal modelling assumptions. The multiple stream paradigm will be implemented to improve the recognition rate [4,6]. The language model and the search engine will also be based on the HMM paradigm.

Embedded systems have become one of the major research topics as they are involved in almost all implementations of computing systems. Our results in embedded systems have been oriented towards some general concepts that support hardware and software partitioning [10,11,12,13] and studies of specific applications and their requirements for embedded
applications [14,15,16,17]. Those applications have some common elements with the ASR, but further research is required to determine the unique requirements of the ASR in order to enable its embedding (implementation of individual modules, their interconnections, power requirements, to mention just a few).

The whole process should be implemented using simplified algorithms to reduce the computational and overall cost for the sake of simplifying the synthesis process to produce the targeted speech recognition engine. We expect that some unique intellectual property (IP) will be developed within the research, which can be re-used in a number of applications that require ASR.


