

https://dspace.org/

FPGA phonetic speech synthesiser

Mamombe, Allen

2011-05-13
http://hdl.handle.net/10539/9746

Downloaded from a DSpace repository. https://dspace.org/

Abstract

Considerable advancements have been made in the field of speech synthesis and speech analysis. Despite these advancements little progress has been made in the field of embedded speech synthesisers. This can be attributed to the slow pace in the development of Application-Specific Integrated Circuits (ASIC) and the affordability of personal computers in developed countries. The same cannot be said however for Sub-Saharan Africa and developing countries. It is therefore imperative to design low cost, memory and processor efficient devices.

This dissertation discusses the design of such a real time embedded speech synthesiser based on a 400000 system gate FPGA. An extensive literature review is documented on various speech synthesis models used in the FPGA based synthesiser. Significant attention is given to the LPC model, commonly known in the telecommunications circles as the principle behind the Global System for Mobile Communications (GSM) codec.

The challenge posed in designing the embedded speech synthesiser was to optimise the memory requirements of the LPC model to suite the suggested FPGA architecture, whilst maintaining the integrity and the quality of the speech. This challenge was solved by using a speech modelling technique combining LPC source signal modelling with the Harmonic plus Noise Model HNM. The LPC-HNM model was used to synthesise phonemes and words of the English language as required by the objectives of the FPGA based phonetic speech synthesiser. Quality of Service (QOS) and Mean Opinion Score (MOS) based listening tests were conducted on MATLAB^{TM}, VHSIC Hardware Description Language (VHDL) and on an FPGA, by a group of 20 native English speakers. Listening test results showed that the designed model performed better than renowned LPC models obtaining scores of 99% and 4.5 out of 5 on the MOS and QOS scores respectively. All speech used in this dissertation was sampled at 8 kHz.

An FPGA was chosen as the development platform because of its huge multiprocessing structure. Particular attention was given to simplifying LPC algorithms to suite the FPGA structure. This was acheived through the use of popular mathematical models such as the Taylor and the McLaurin's series. The designed system used less than 200000 FPGA system gates.

Results and the work carried out in this dissertation significantly illustrate the contribution made by this work in the field of embedded speech syndissertation.