Digital Signal Processing Architecture for Large Vocabulary Speech Recognition

by

Professor Wonyong SUNG

Seoul National University

Seoul, Korea

Abstract:

Large vocabulary continuous speech recognition (LVCSR) demands very high computing and memory resources, and thus optimized architecture is greatly needed for real-time, high-throughput or low-power implementations. A traditional speech recognition algorithm employs the GMM (Gaussian Mixture Model) for estimating the likelihood of phonemes at each speech frame, while utilizing the HMM (hidden Markov model) to explore the long-term information.

In recent years, DNN (Deep Neural Network) has shown very good performance in phoneme recognition, and has replaced the GMM in most state of the art LVCSR applications. The language model has been mostly implemented with a statistical N-gram back-off model, which demands a very large memory space when increasing the N-gram order. These language models are recently being replaced by the RNN (Recurrent Neural Network), which contains short-term memory inside. Still, the state of the art LVCSRs are based on the HMM although the GMM and the language model are replaced by recently developed neural network based algorithms. In order to know the possibility of non-HMM based LVCSR, we have conducted an experiment that uses only one RNN for implementing a whole LVCSR. The experimental results show that the RNN is capable of learning long-term dependencies of linguistic structures without an external language model.

The recent algorithmic advancement, such as from GMM to DNN and also from HMM to RNN, implicates drastic change of the optimum DSP architecture for implementing LVCSR. Although the new LVCSR algorithms are much more demanding in terms of arithmetic operations when compared with the traditional GMM and HMM based ones, they require low arithmetic precision, little conditional or branch operations and allow

fully distributed processing. With the help of non-volatile memory based logic, the new LVCSR algorithms can be operating with very low energy.