

ABSTRACT

BONILLA C, D A. *Automatic Speech Recognition of Phonemes in Portuguese Using Mixtures of Neural Networks Experts*. 2016. 130f. Dissertação (Mestrado em Engenharia Eletrônica) – Faculdade de Engenharia, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2016.

The automatic speech recognition based on detection of phonemes provides advantages for online recognition of a speech represented by a sound signal. Studies in computational intelligence have allowed associating the advantages of information parallel processing to the computational workload distribution, aiming at simplifying the use of complex models. The development of a system for automatic speech recognition involves various processes, covering several areas of research, such as linguistics, signal processing and computational intelligence. In this dissertation, the process starts with a speech signal pre-processing to extract and represent more succinctly, the main characteristics of the speech signal at a given instant of time. Inspired by the principle that recommends to “divide and conquer”, a model of mixture of neural network experts is explored, allowing to divide the decision space of the complex problem of speech recognition so that each expert takes care only of a delimited area of this decision space. It is noteworthy to point out that each expert included in the composed model takes handles in the decision regarding each one of the preprocessed input sample. This set of decisions thus obtained is weighted. So, the expert system with the highest weight for the output will determine the final classification of the considered sample. After that, a dynamic post-processing step, implemented as a recurrent network, is executed. It aims at mitigating the oscillatory effect that occurs during the recognition of classes with similar characteristics. In this dissertation, two models of mixture of experts are investigated. The first is based on the grouping of similar phonetics classes while the second considers the imbalanced distribution of samples in the training set. The comparison of the model proposed in this dissertation with the work related to automatic speech recognition indicates a gain of 7.62 % in terms of accuracy.

Keywords: Automatic speech recognition. Phonetic recognition. Artificial neural networks. Mixture of experts. Recurrent neural networks.