



Sylva Sadovská
Faculty of information technology
Brno University of Technology
Božetěchova 2
612 66 Brno

Report on the thesis
PHONEME RECOGNITION BASED ON LONG TEMPORAL CONTEXT
by
Petr SCHWARZ

This thesis presents an in depth study and develops new techniques for improving the recognition performance of a phoneme recognition system base on Artificial Neural Networks (ANNs).

The topic of the thesis is important in the field of person-machine interaction. This field is receiving more and more interest in the Information Technology community due to the level of performance recently achieved by automatic speech recognizers, and due to the still open challenges. This work contributes with originality to the present level of knowledge, by clarifying several mechanisms that contribute to the quality of a recognizer, and proposing new phonetic models and architectures for achieving the best performances with the number of model parameters appropriate with respect to the dimension of the training data.

I fully confirm the claims of originality given by the candidate at the beginning of the thesis, in particular the definition of the split temporal context ANN architecture (STC) is an original and clever way of exploiting at the best the available data for training accurate models.

The quality of the work done is witnessed by several contributions in the last years to the most important international conferences in this field and also by the very good results obtained in the most important worldwide evaluation of language recognition systems, where the phonetic recognizer plays a major role for obtaining state of the art performance.

The contributions of the candidate to the field are well known and referenced in the speech recognition literature, confirming that he is a very good researcher, that he collaborates with an outstanding research team, and is able to compare his work with the best scientists in this field.

The thesis is well organized and clearly presented: the background is carefully revisited in Chapter 3, and its original contributions are given in the successive chapters.

The candidate has developed a relevant amount of very good quality scientific research to highlight all the factors that affect the performance of a phoneme recognizer, such as the features, the number of parameters, the length of the context, the architecture of the ANNs, etc.



This work has not only allowed the candidate to confirm the effectiveness of long temporal contexts for phonetic recognition, but has put forward the knowledge of the elements that contribute to the success of a system.

Every claim in this thesis is supported by a massive amount of well planned experiments focusing on a specific aspect of the system, but not ignoring the overall performance.

I appreciated very much the tutorial presentation of the work, that demonstrates that the candidate masters all the aspects involved in a complex system, that he was able to plan his experiments, to explain why they were performed, and to give convincing interpretations of their results.

Summarizing my review, I can state that the candidate has developed a very good scientific work, that gives an innovative contribution to the state of the art in speech recognition, with many practical fallbacks.

I therefore recommend to accept without reserve this thesis, which meets without any doubt the requirements for obtaining a PhD degree.

Torino, 9/12/2008

Pietro Laface

Professor of Computer
Science