

Title: Neural networks for automatic speaker, language, and sex identification

Author: Bich-Ngoc Do

Department: Institute of Formal and Applied Linguistics, Faculty of Mathematics Physics, Charles University in Prague; Department of Linguistics, Faculty of Arts, University of Groningen

Supervisor: Ing. Mgr. Filip Jurek, Ph.D., Institute of Formal and Applied Linguistics, Charles University in Prague and Dr. Marco Wiering, Institute of Artificial Intelligence and Cognitive Engineering, Faculty of Mathematics and Natural Sciences, University of Groningen

Abstract: Speaker recognition is a challenging task and has applications in many areas, such as access control or forensic science. Moreover, in recent years, the deep learning paradigm and its branch, deep neural networks have emerged as powerful machine learning techniques and achieved state-of-the-art performance in many fields of natural language processing and speech technology. Therefore, the aim of this work is to explore the capability of a deep neural network model, recurrent neural networks, in speaker recognition. Our proposed systems are evaluated on the TIMIT corpus using speaker identification tasks. In comparison with other systems in the same test conditions, our systems could not surpass reference ones due to the sparsity of validation data. In general, our experiments show that the best system configuration is a combination of MFCCs with their dynamic features and a recurrent neural network model. We also experiment recurrent neural networks and convolutional neural networks in a simpler task, sex identification, on the same TIMIT data.

Keywords: speaker identification, sex identification, deep neural network, recurrent neural network, convolution neural network, MFCC, TIMIT