Conversion of Whispered to Normal Voice

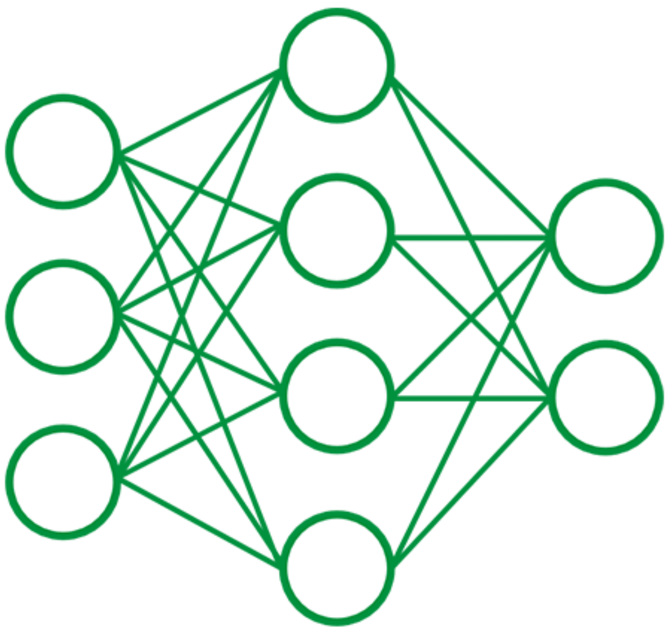
Richard Gajda, supervisor: Ing. Jan Brukner

The thesis is based on an article published by G. Nisha Meenakshi from the Indian Institiute of Science. The article proposes a method which converts whispered speech into neutral using neural networks. The thesis reconstructs the conversion method and deploys it in Android application environment.

First, a dataset for training and testing the neural network was obtained. It was recorded by 3 male and 3 female professional actors from the National Theatre Brno. They were asked to record the 1990 New Year's speech by Václav Havel. The speech was recorded in both whispered and normal voice.

The obtained dataset then underwent extensive editing and silence removal. Those treated utterances were analysed using WORLD vocoder-based speech analysis and synthesis system. The training parameters - spectral envelope (sp), aperiodicity levels (ap) and fundamental frequency (F0) were obtained.

WORLD



The obtained parameters were further treated and transformed and were passed into BLSTM models. Each neutral speech parameter prediction was trained on an separate BLSTM. The resulting trained models are used in the whispered speech conversion application for Android.

Android application was developed to employ the trained models for on-demand voice conversion. Although the computational capabilities of smartphones are increasing every year, the conversion is done in the cloud, on a Flask python implemented server.