Simon Kassian Locher

Maximilian Schwendeler

Graduate Candidates	Simon Kassian Locher, Maximilian Schwendeler
Examiner	Prof. Dr. Guido Schuster
Co-Examiner	Fabian Riesen, Schaffhausen, SH
Subject Area	Sensor, Actuator and Communication Systems

NLP for Flight Simulation

Using Deep Learning for Air Traffic Communication Training

Introduction: Complementing pilot training with the use of simulators yields many benefits, one of which is increased flight safety. However, most of the simulation training systems focus purely on flight skills and overlook the communication between the pilot and the Air Traffic Control (ATC). Inaccurate use of the strictly defined phraseology may lead to dangerous situations. Advances in Automatic Speech Recognition (ASR) over the last decade opened up new possibilities for speech-based training, so it might be used in an ATC environment.

The goal in this thesis is to implement an ASR system which adds another key training feature to the flight simulator developed by VRMotion AG. Even though it is possible to use the simulator casually and fly around the virtual world freely, the main goal is to prepare pilots for various scenarios, ranging from round trips to emergency maneuvers. This means the dialog is predetermined which makes it easier to determine whether or not the pilot followed the correct phraseology. Currently, evaluating the pilots communication performance relies solely on the flight instructor. The system aims to relieve this person by tracking the communication errors made by the trainee and point out frequent mistakes.

Approach: In the first part of the thesis the State-of-the-Art (SOTA) methods are researched and compared. Based on these findings, an ASR system is designed which best meets the desired specifications. The development cost of training a modern Deep Neural Network (DNN) for Natural Language Processing tasks is immense. To evade this limitation, NVIDIA's new toolkit NeMo is used. Not only does NVIDIA offer a selection of pre-trained models, it simplifies the process of fine-tuning DNN's which leads to significant performance improvements, especially in target domains where the amount of data is limited, e. g. ATC. To further improve the Word Error Rate, a statistical n-gram Language Model is built using sentences which match the phraseology and context of a specific training scenario. To complete the system, a modern Text-to-Speech synthesis architecture is used to synthesize the voice of an ATC operator.

Result: By using a simple input device and a combination of SOTA DNNs and domain-specific applied knowledge, a system which transcribes the spoken input with high accuracy could be built. To increase robustness of the DNN which is pre-trained on native English speakers, the network is fine-tuned with German, Swiss-German and -French speakers. Comparing the results to a recently released benchmark shows that the achieved prediction accuracy is comparable to the current SOTA methods. It is worth mentioning that all of the components used, i. e. software and datasets, may be used free of charge for commercial applications. Furthermore, the system allows to effortlessly improve its overall performance by adding more training data. All the results are bundled in an app which offers great demonstration possibilities.