Research Project in the Data Science Profile

Swiss German Text-to-Speech based on phoneme transcription

Situation

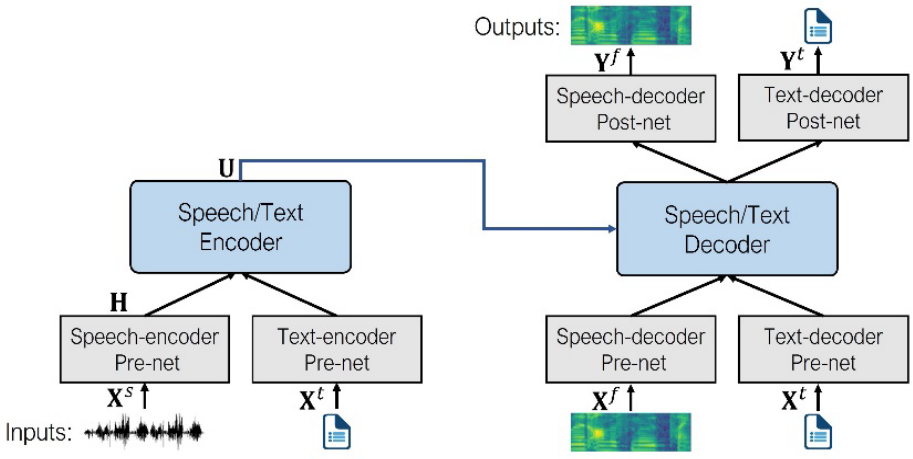
The Institute for Data Science FHNW has been working on Swiss German Speech-to-Text (STT) and Text-to-Speech (TTS) models for several years. The current TTS model is based on two steps: first, High German text is 'translated' into the chosen Swiss German dialect, and then the Swiss German text is synthesized. As there is no standardized orthography for Swiss German, we simply trained a deep learning model to do the translations from High German to Swiss German text.
Demos of the current models are available on https://stt4sg.fhnw.ch/.
This project aims to develop a Text-to-Speech (TTS) system for Swiss German dialects through a completely novel approach based on phoneme transcription. There is strong evidence from previous work that a better quality of the synthesized voices can be achieved with such an approach, and this will be confirmed or disproved in the present project.

Goals / Methodology / Tasks

The goal of this project is to develop a TTS system for Swiss German dialects based on phoneme transcription as an intermediate representation and to compare the quality of this model with a TTS which is based on Swiss German text as an intermediate representation. There are several challenges to be overcome in this project, namely the Low-Resource environment which means that there is limited access to high-quality, labeled audio data for Swiss German dialects. In addition, the lack of a standardized writing system for Swiss German complicates the mapping of phonemes to audio output. The development of effective models that account for dialect variation is also crucial. Finally, comparing TTS models based on phoneme transcription vs. Swiss German text requires a workflow and specialized metrics that take care of code switching ("*ich ha am Friitig en neuä Touchscreen kauft*"), cope with pronounciation of named entities and abbreviations ("*USA*" pronounced as "*UUESAA*") and can ideally be automated rather than relying on human judgment.

The acquisition of the necessary data (audios and transcriptions) is not part of this project, these will be provided by the research team.

You will learn how to work with advanced AI systems for audio and language processing, apply techniques to train and fine-tune cutting-edge models and build efficient AI systems tailored for conversational applications.



Required Skills

Good programming skills, interest in deep learning and especially in Natural Language Processing (NLP), willingness to work closely in a team, knowledge of Swiss German would be an advantage.

Tasks for the Master Student

The project can cover several aspects in 2 to 3 sub-projects (IP7, IP8, IP9).

Full/Parttime: [X] Full time study
[X] Part time study

Location: Windisch

Advisor: Prof. Dr. Daniel Perruchoud