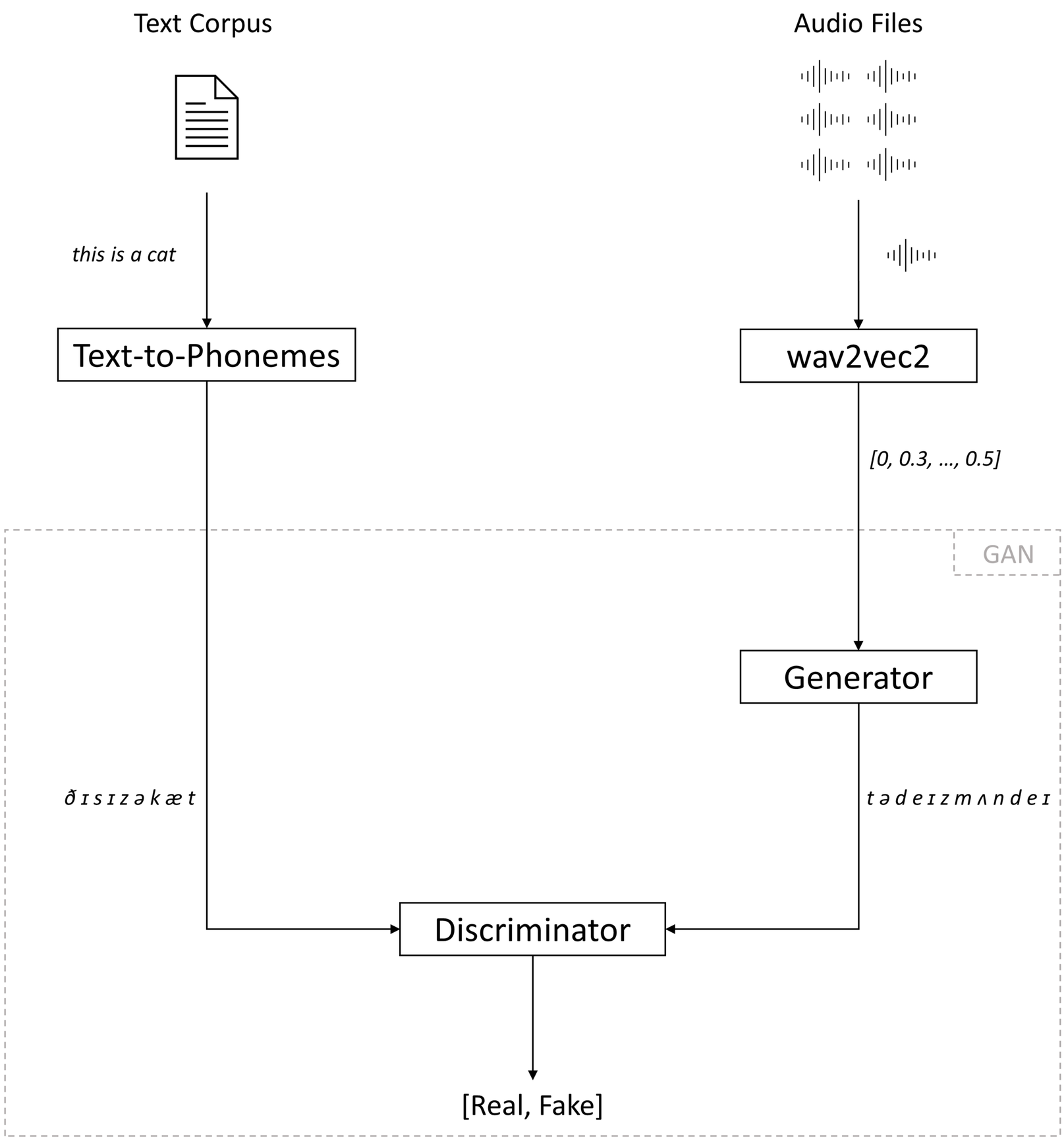
HSLU Hochschule Luzern

Technik & Architektur

Master of Science in Engineering, Profile Data Science

Unsupervised Speech Recognition for Swiss German



Problem Definition

Automatic speech recognition (ASR) is a core building block for voice assistants. ASR has performed strongly in several languages, among them English and German. However, it performs poorly on low-resource languages due to a lack of sizable datasets. Poor performance limits acceptance in countries such as Switzerland because of its unique variants of German. In this work, we explore the possibility of improving Swiss German speech recognition using unsupervised learning.

Methods

The base of our work is wav2vec-Unsupervised. Speech representations are created from audio files using a wav2vec2 model. We trained a Generative Adversarial Network on these representations and unpaired phonemized text. The generator learns transcribing audio, and the discriminator has to distinguish data from our text corpus and data produced by the generator. First, we reproduced the Standard German results published. Then, we performed experiments on Swiss German data.

Results

We report a Phoneme Error Rate (PER) of 17.5% for Standard German and successfully reproduced the published results on a phoneme level. After various attempts to optimize the training stability of the model, our best Swiss German approach achieved a PER of 86.5% using ten hours of audio data. Finally, we proposed suggestions on how our results could be further improved in future works.

Manuel Vogel

Advisor:
Prof. Dr. Andrew Paice

Co-Advisor:
Guido Kniesel

Expert:
Dr. Manuel Dömer