# Swiss German Speech to Standard German Text

***SwissText.org Shared Task 3***

**Alex Wolf**
University of Zurich
alex.wolf@uzh.ch

**Deborah Noemie Jakobi**
University of Zurich
deborahnoemie.jakobi@uzh.ch

## Abstract

This paper presents a solution to the SwissText Conference (STC) shared task 3, 2021. The shared task challenges participants to implement models for translating Swiss German speech to standard German text. The authors implemented multiple *DeepSpeech* models using the data provided by SwissText.org as well as additional data. The use of the additional data in combination with the official data lead to some promising improvements of the model trained on the official data only. An additional experiment with a sequence to sequence translation model was trained in order to improve our score. We achieved a BLEU score of up to 0.17 on the test set of SwissText. Although the performance of the final model did not score well in comparison with other submissions, there is plenty of room for possible improvements.

## 1 Introduction

Swiss German has a wide variety of different Swiss German dialects, with a huge difference in words, pronunciation, even to the point of sounding like a different language. Swiss German has relatively little speakers (around 5 million) and there is hardly any standardized spelling. This leads to Standard German being one of the official writing language in Switzerland. As there is no official Swiss German spelling, most speakers using written Swiss German just use their own spelling which resembles mostly a phonetical translation. This leads to huge variance within Swiss German writing (**?**). It makes therefore sense to translate spoken Swiss German to Standard German in order to correspond to the official language situation. Tackling a standardized translation of different spoken Swiss German dialects into standardized German text requires a vast amount of data and fine tuning. The STC 2021 proposed a shared task to tackle this problem and provided a dataset (SwissText Conference dataset (STCD) to train and fine tune on. This paper shows what kind of experiments, data, and approaches the authors used to tackle this problem. The proposed task is very complex as it includes not only a Speech-to-Text (STT) conversion but also translation from Swiss German to Standard German which can be refered to as speech translation (**?**). Additionally, it possibly includes domain shift, as the training data stems only from Swiss parliament speeches while the domain of the test set is unclear. This paper presents a quick overview on the previous research done within this area, an introduction to the DeepSpeech model used for tackle the task and our experiments and results.

## 2 Literature review

add more general stuff about STT.

The shared task of a translating spoken Swiss German into standard written German was already presented by the STC in 2020. While the current 2021 task is about reaching the highest BLEU score, last year's submissions where ranked based on the least Word Error Rate (WER). **?** achieved the best WER of 40.29%. The authors used a CNN acoustic model named Jasper. They used additional Standard German data and fine-tuned on the official data set. They used different augmentation techniques and a language model. **?** add text. (**?**) achieved the second best WER. ...
**?** used an end-to-end model called DeepSpeech and achieved an WER of 58.93%. This is the model we decided to use as well. It is publicly available on GitHub and can easily be fine-tuned or used for transfer learning.

## 3 DeepSpeech

explain how deepspeech works...

## 4 Experiments

### 4.1 Datasets

We compared models on both the dataset provided by the STC as well as the ArchiMob corpus. The STCD contains 38 GB of labeled and 65 GB of unlabeled spoken swiss german audio data and an additional validation set containing 1.5 GB of data (**?**). The ArchiMob corpus (Release 2) contains X GB of spoken swiss german data (**?**).

### 4.2 Experiments

The base model of most our experiments was a pre-trained DeepSpeech model (release 9.0, get link!). The first model we trained served as our private baseline. We trained a bare DeepSpeech model on the SwissText labelled dataset.

## 5 Results & Discussion

Interestingly, we noticed that our internal deviations in the BLEU score are not as great as on the official test set.

| Model# | Data | Train BLEU | Test BLEU |
|---|---|---|---|
| 1 | SwissText | 0.23 | 0.0004 |
| 2 | ArchiMob | 0.27 | **0.17** |
| 3 | ArchiMob | 0.24 | 0.07 |

Table 1: Font guide.

## 6 Conclusion