BACHELOR THESIS

ARTIFICIAL INTELLIGENCE



|  |
| --- |
| Adapting pre-trained XLS-R Wav2Vec2 on Air Traffic Control Communication Data |

*First supervisor:*

dr. L.F.M. ten Bosch

Taal en communicatie

[louis.tenbosch@ru.nl](mailto:louis.tenbosch@ru.nl)

*Second supervisor:*

Dr. A. B. LAST NAME

Departement

E-mail

*Author:*

Karan Chand

Student number: s1033357



28 August 2022

**Abstract**

The abstract of your thesis is a brief description of the research hypothesis, scientific context, motivation, and results.

The size of an abstract is usually one paragraph or one page of text.

* Recent work
* domain shift
* data evaluation
* xlsr model – why -> focus on multilingual, so accents good?
* wer and cer reduction
* study impact of size of data on finetune performance
* impact of arpa (maybe on other finetuned models as well)

Contents

[Introduction 4](#_Toc52190693)

[Preliminaries 5](#_Toc52190694)

[Related work 6](#_Toc52190695)

[Research 7](#_Toc52190696)

[Conclusions 8](#_Toc52190697)

[Bibliography 9](#_Toc52190698)

# Introduction

The introduction of your bachelor thesis introduces the research area, the research hypothesis, and the scientific contributions of your work.

Simon Peyton Jones, a researcher at Microsoft Research, gives seven tips on how to write a great research paper which might be worth looking at (Jones, 2016).

The following things are usually in an introduction:

* Describe the problem and state the research question;
* Motivate why this problem must be solved;
* Demonstrate that a (new) solution is needed;
* Explain the intuition behind your solution;
* Motivate why / how your solution solves the problem (this is technical);
* Explain briefly how it compares with related work.

The introduction is closed with a paragraph in which the content in the next chapters is briefly mentioned to give an overview. A guideline is one sentence per chapter.

Mistakes in Air Traffic Control ATC communication. Models for automatic speech recognition ASR exist, but are constant improvement is needed. ASR models should minimize errors, here the word error rate WER, which is still massively in progress. Minimizing errors in communication will reduce the risk of accidents.   
Models of ASR are being designed and updated constantly, which leads to great innovation in this field. An example of a community that specializes in Wav2Vec2 transformer models and data sharing is HuggingFace. Here, individuals can create, modify and publish their own models and datasets for others to use. Big companies such as Meta have shared their models as well, which are trained on massive amounts of data. These models are very recent and have been tested on ATC inadequately. (Has been done before though). In this paper, Wav2Vec2 transformer models are evaluated on their performance on transcribing ATC communication from the atcosim dataset. The models that are considered contain:

* A model that was pretrained on LibriSpeech data (Base)
* A model that was pretrained and fine-tuned on multiple noisy datasets as well as clean data (Robust)
* The finetuned XLS-R model in addition to a language model
* (Finetuned model on ATC data)

**Research questions:**

* How robust are pretrained and fine-tuned automatic speech recognition models on ATC data?
* How does a language model affect the performance of a finetuned XLS-R model on ATC data?
* How does the amount of data used to finetune affect performance of a finetuned XLS-R model on ATC data?

# Preliminaries

This is an optional chapter that contains knowledge that your reader needs to know in order to understand your work. If there is not much you can also embed the preliminaries in one of the other chapters or even discard it completely.

* Transformer models
* WER and CER -> why not other metrics, like Glue or Bleu
* Huggingface

# Related work

Within this chapter you show that you have sufficient knowledge of the problem domain.

* Other models of ATC and their performance
* Wav2Vec2 Model and their performance
* Attention is all you need paper
* XLS-R on accents

# Methodology

* Data Shortage
* Dataset
* Evaluation dataset

**Transcribing**

Pretrained robust model performs a bit better than base model. Reason for this is:

**Training (Fine-tuning)**

**Addition Language model**

# Experiments

## Data

## Settings

In this chapter or in these chapters you write/ show all the details that are required to “prove” your hypothesis. This should be sufficiently detailed and precise such that your fellow students are able to repeat the research and establish the same results and conclusions.

# Results

~~It has been shown from the metrics that the robust model has a slight deviance in performance compared to the base model. The reason for this might be that the complex grammar that is used in ATC communication is too detrimental to performance. Even a model that is trained on more noisy data is not able to significantly increase performance.~~

* Performance of pretrained models
* Performance of finetuned models
* Performance of finetuned model with language model

# Conclusions

In this chapter you present all conclusions that can be drawn from the preceding chapters.

It should not introduce any new material or theories; these should have been written down earlier in the thesis. Because of this, your conclusion can be brief and to the point.

When further research?

* Fine tune on different models, was not the focus of this paper
* Try finetuned models on more data for comparison and generalisation factor
* Finetune on more data or on different training settings, as hyperparameters have not been tuned for our ATC data.
* Add LM to lower finetuned models to see the performance difference

# Bibliography

# Appendix

Appendices are optional chapters in which additional material is covered. This material is required to fully support your research, that would otherwise clutter the presentation of your research.