

Mitigating Speaker Attribute Bias in Automatic Speech Recognition Models for Low Resource African Languages

Claytone Sikasote

SKSCLA001@myuct.ac.za

Supervisors: **Dr. J. Buys** & **Prof. H. Suleman**

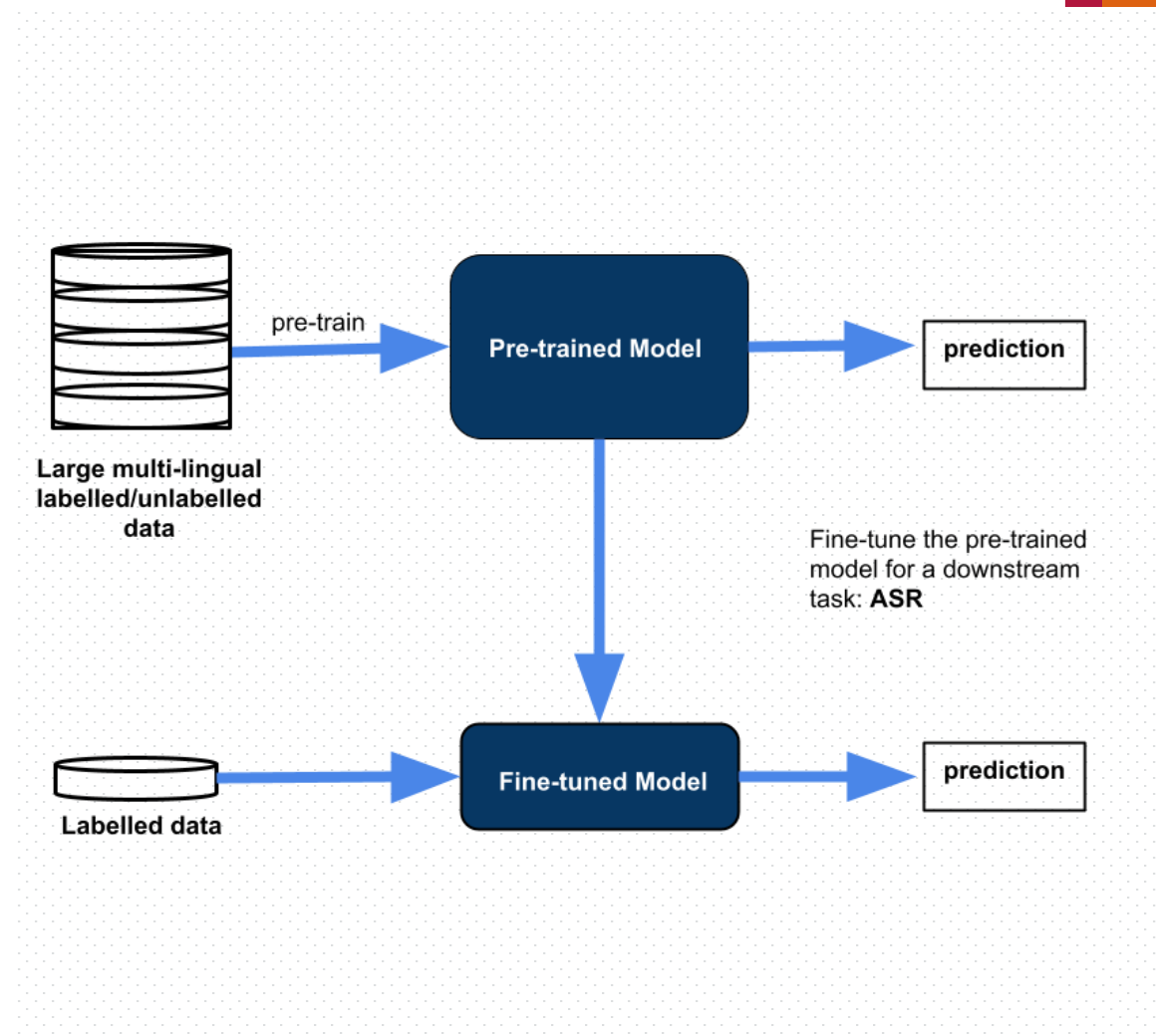
**Design IT.
Create Knowledge.**

www.hpi.de



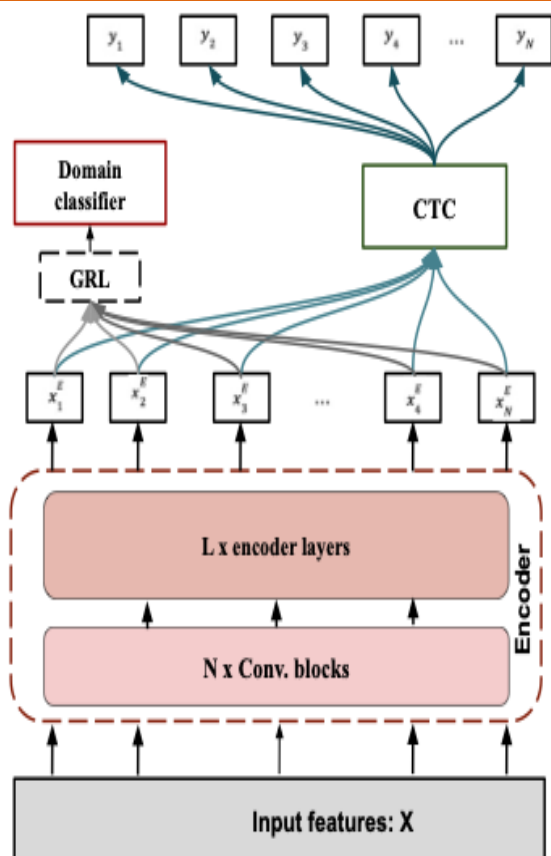
Introduction

- ❖ State-of-the-art approach for **speech recognition** in **low resource languages** is based on fine-tuning pre-trained speech models.
- ❖ **Problem:** This approach increases the risk of introducing **speaker-attribute bias** in resulting speech recognition models such as **gender, age** and **speaker type bias**.
- ❖ **Speaker-attribute bias** is the disparity in performance of an speech recognition systems between speaker attribute subgroups.



Proposed method: Group Distributionally Robust Optimization with Domain Adversarial Training (DRODAT) Method

DRODAT-based Model Architecture



Methodology

(1) Research question: To what extent does a method that couple distributionally robust optimization (DRO) and domain adversarial training (DAT) minimize speaker-attribute bias in fine-tuned speech recognition models for Bemba and Nyanja?

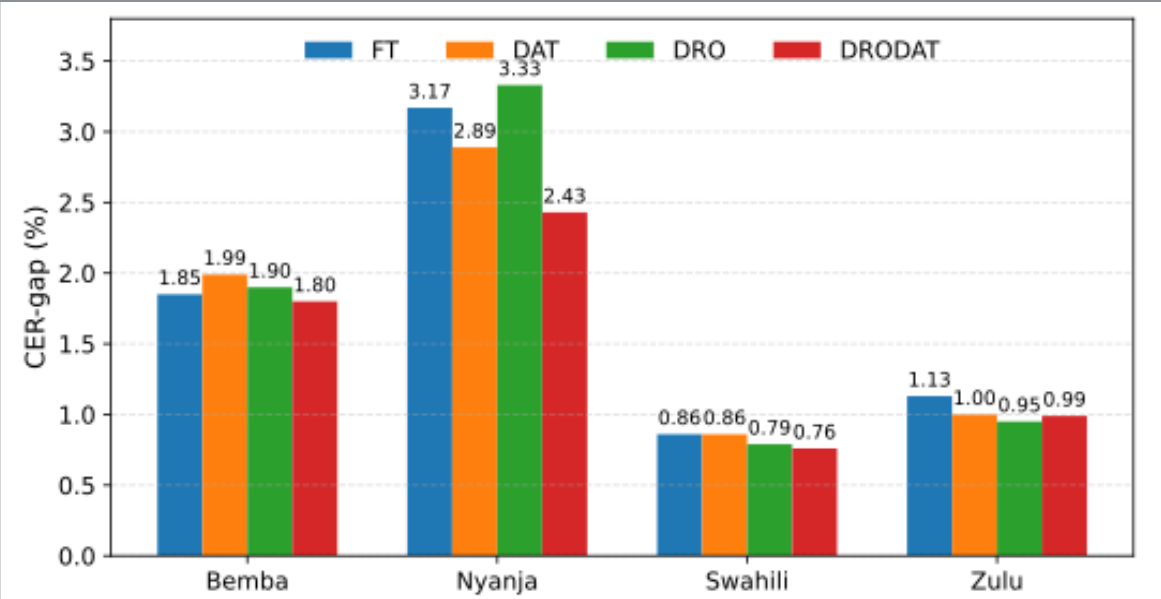
(2) Experimental setup:

- **Pre-trained speech model:** MMS and SSA-HuBERT-Large-60K
- **Datasets:** Zambezi Voice and FLEURS
- **Languages:** Bemba, Nyanja, Swahili and Zulu
- **Speaker attribute:** gender

(3) Evaluation metric: Absolute difference in ASR performance between subgroups (CER-gap)

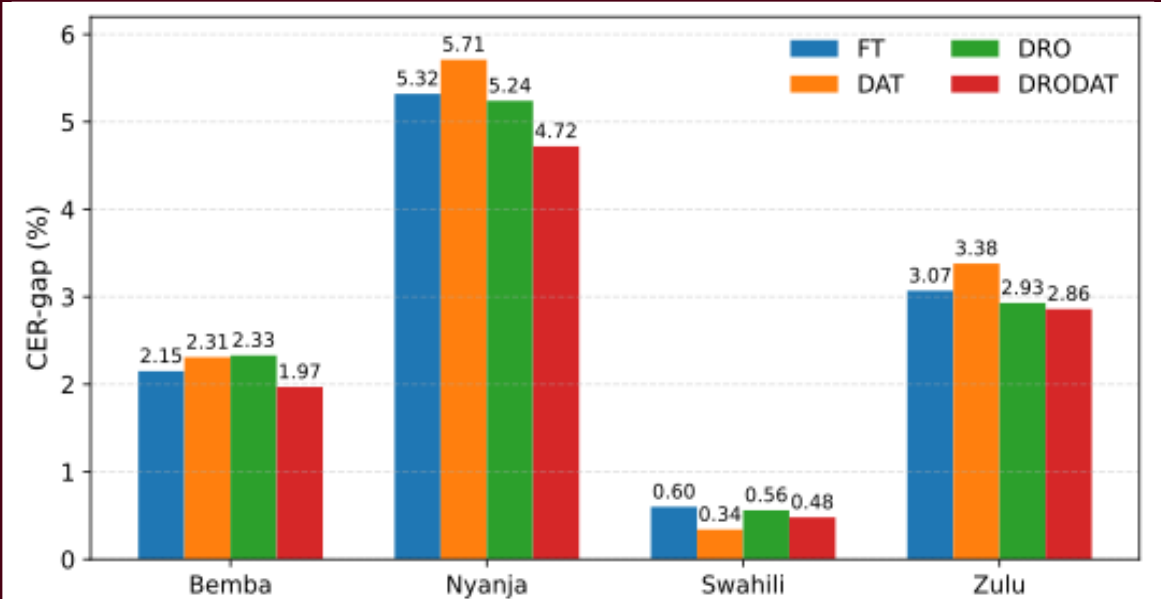
(4) Baselines: Standard fine-tuning (FT), DAT and DRO.

MMS-based Model



Outcome: In MMS-based models, our proposed method outperform the baseline methods in reducing gender bias in **Bemba**, **Nyanja** and **Swahili**.

SSA-HuBERT-Large-60K-based Models



Outcome: In the SSA-HuBERT-Large-60K – based models, our proposed methods outperform baseline methods in reducing gender bias in **Bemba**, **Nyanja** and **Zulu**.

Thank you