

Adapter-based Debiasing of Pre-trained Speech Recognition Models for Low Resource Languages

Introduction

Attribute-specific bias remains a challenge in automatic speech recognition (ASR) systems. If not mitigated, bias has the potential to limit certain groups of people from accessing speech technology systems based on demographic information such as gender, age and speaker type.

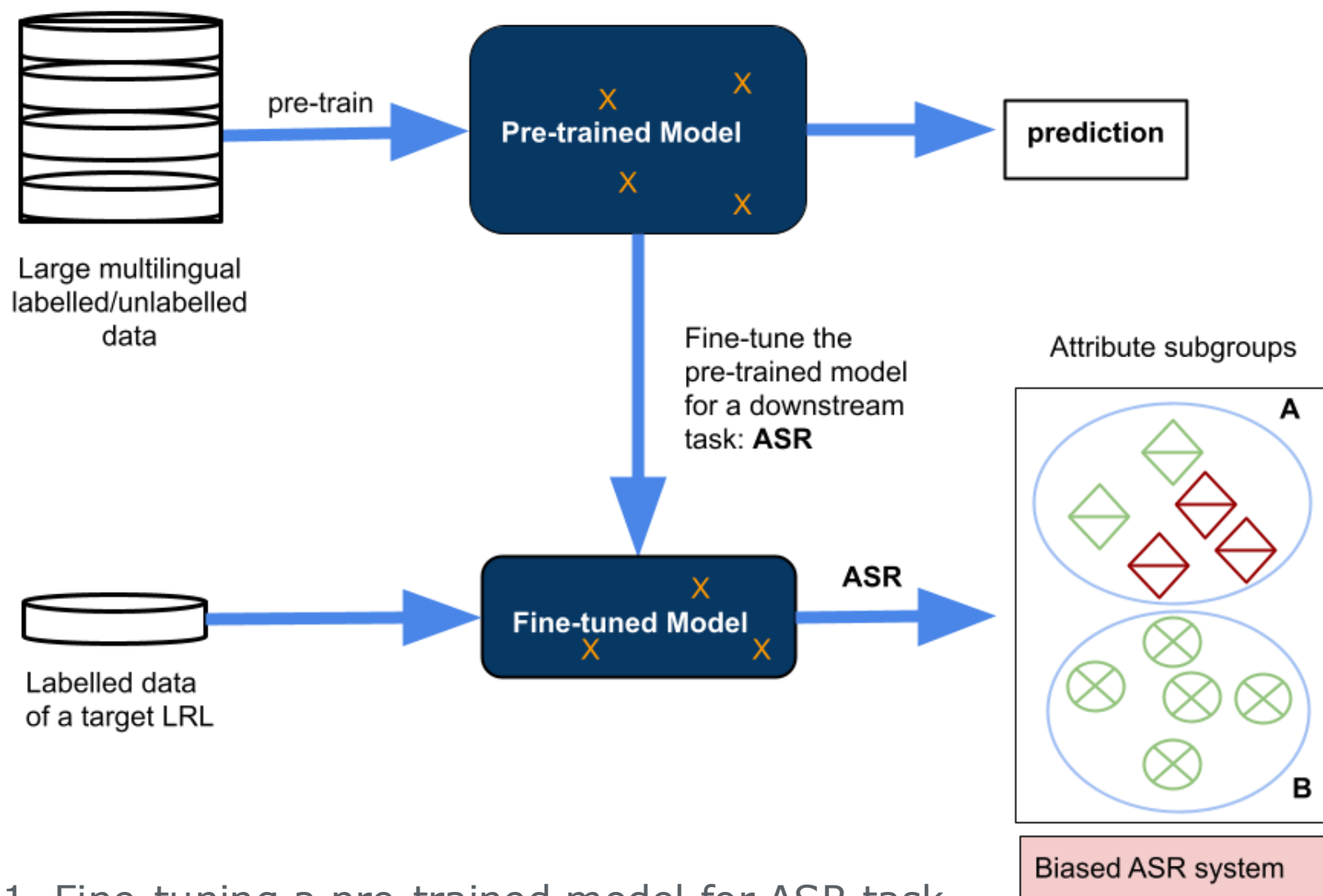


Fig. 1. Fine-tuning a pre-trained model for ASR task.

Aim

To develop a fine-tuning algorithm that minimizes speaker attribute-specific bias in ASR systems for low resource languages, while preserving speech recognition accuracy.

Research Questions

- To what extent do ASR models fine-tuned from multilingual pre-trained models exhibit attribute-specific bias in the target languages?
- To what extent does debiasing pre-trained speech models using adapters via adapter fusion minimize attribute-specific bias in ASR models of target languages?
- To what extent does the application of meta adapters minimize attribute-specific bias in ASR models of target languages?

Methodology

- **Attribute-specific bias:** gender, speaker type and age
- **Target languages:** Bemba, Nyanja, Lozi and Tonga
- **Proposed methods:** Adapter fusion-based and meta adapter-based models
- **Training objective:** Domain adversarial training

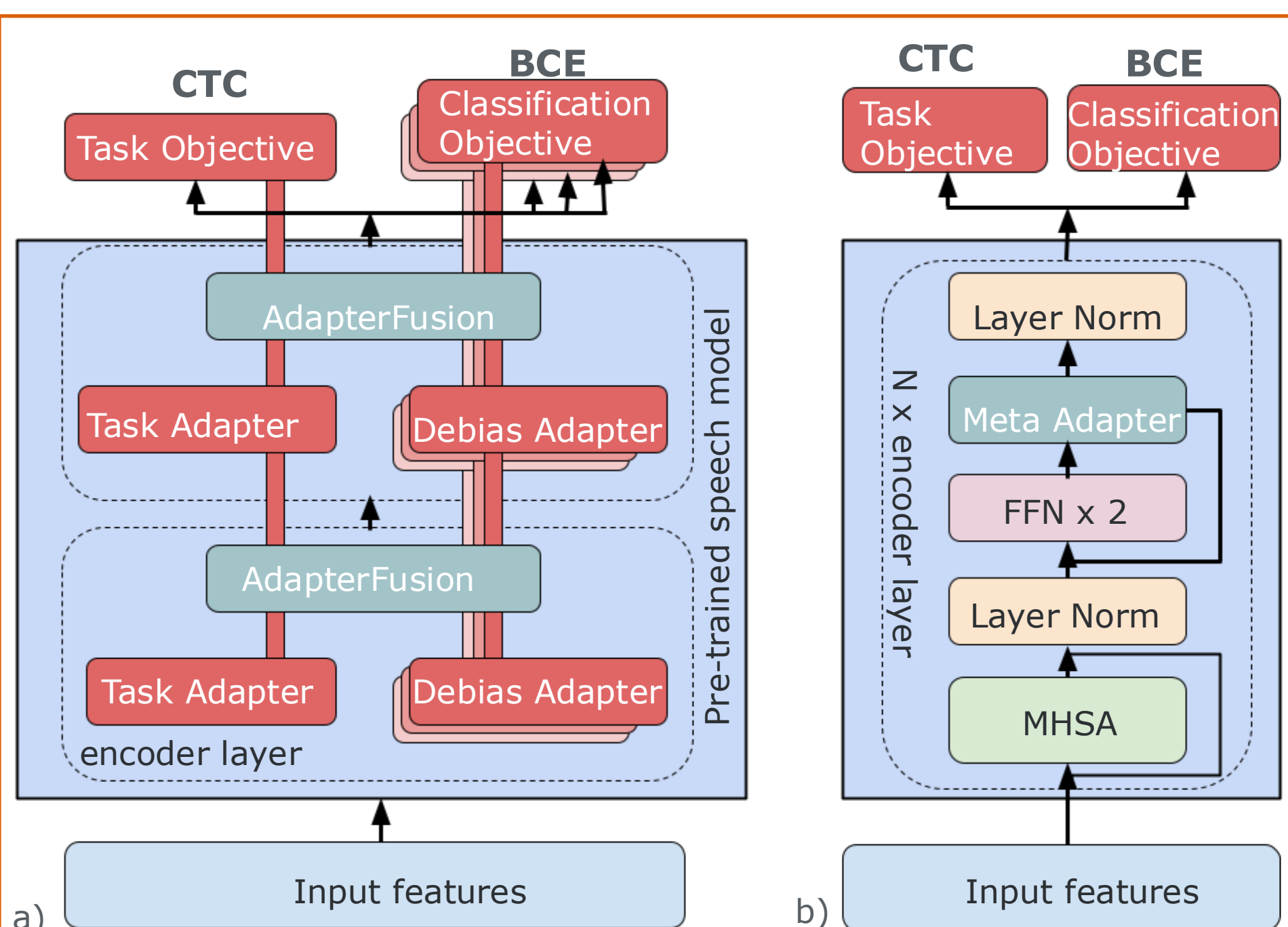


Fig. 2. Adapter fusion-based (left) and meta adapter-based (right) models

Results (Study 1)

Objective: To investigate the impact of fine-tuning multilingual pre-trained speech models on gender bias in ASR systems for three low resource African languages; **Bemba, Nyanja and Swahili**.

a) All fine-tuned models (Fig. 3, Whisper) on gender-specific and-balanced datasets of target languages exhibited different degrees of gender bias:

- Female speech is preferred for Bemba-BS, Nyanja and Swahili
- Male speech is preferred for conversational speech in Bemba-BIGC

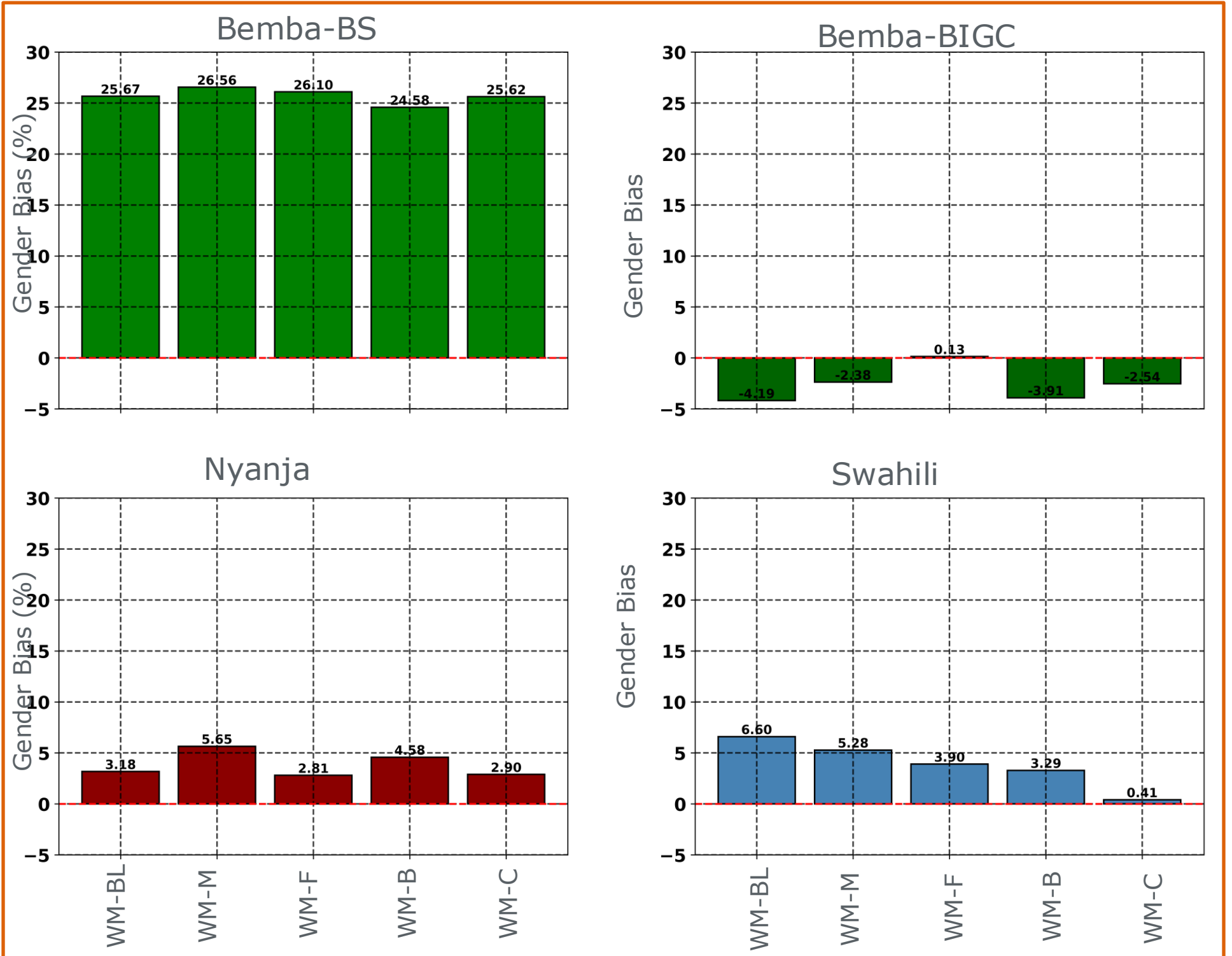


Fig. 3. Gender bias (%) estimates in fine-tuned Whisper models (WM).

b) Ablation study (Fig. 4) shows no correlation between training data size and gender bias.

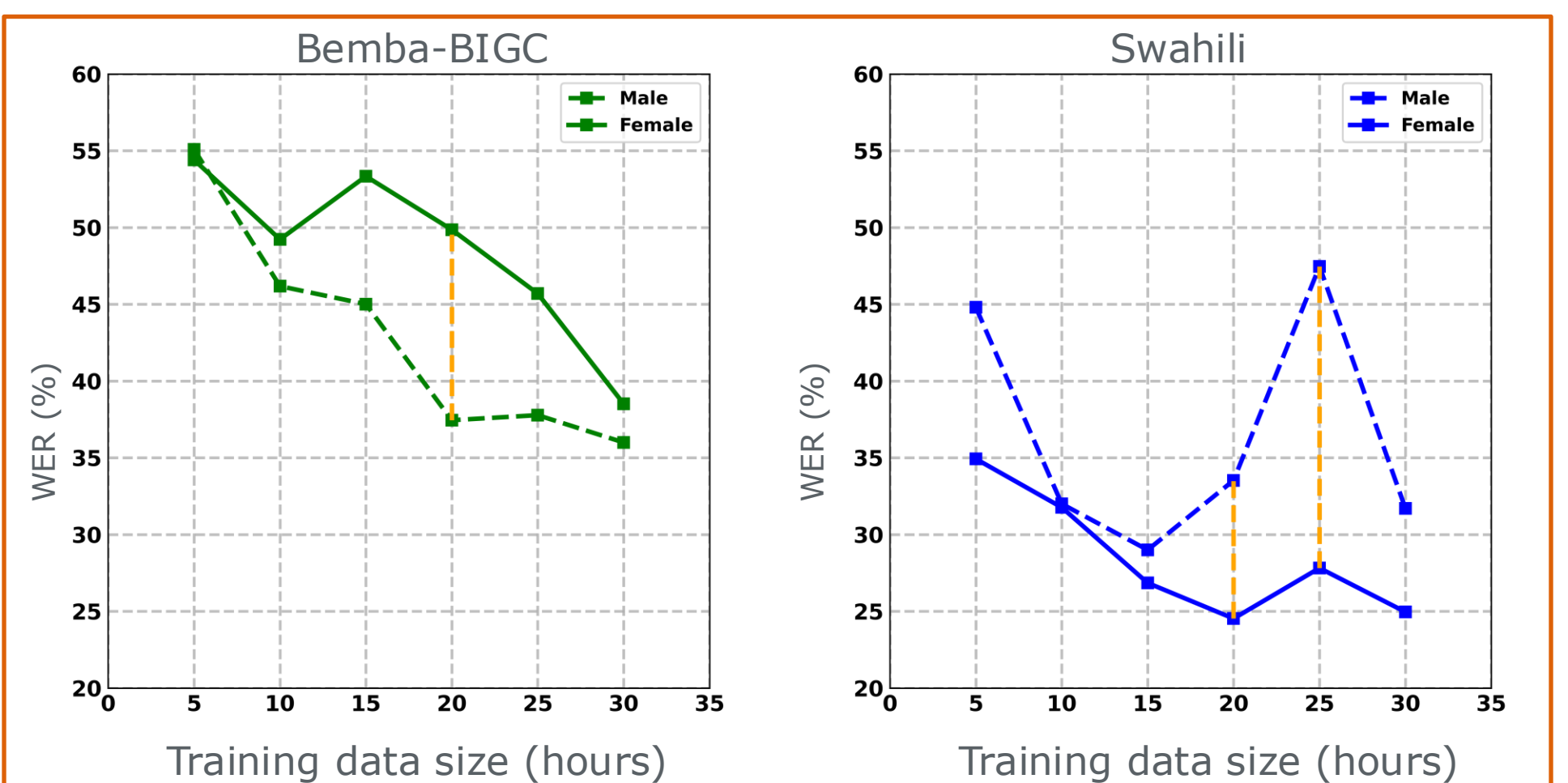


Fig. 4. Ablation study: training data size vs. model performance (WER %)

Findings

- There is bias transfer from pre-trained models during the fine-tuning
- The extent of bias in the fine-tuned model is language & dataset dependent
- Speech type has the potential to determine the gender favoured
- There is no relationship between training data size and gender bias

Next Steps

- Implement the proposed adapter fusion-based model
- Create bias evaluation datasets for our target languages

Claytone Sikasote

- Email: SKSCLA001@myuct.ac.za
- Website: <https://csikasote.github.io>
- Supervisors: Dr. Jan Buys and Prof. Hussein Suleman