From Raw Audio to Fair Recognition: Pre-processing Strategies for Reducing Gender Bias in Automatic Speech Recognition

Andrew Murphy

M.S. Artificial Intelligence Student

School of Information

acm7552@rit.edu

Steven Szachara

M.S. Data Science Student

Department of Software Engineering

ss9270@rit.edu

Automatic Speech Recognition (ASR) systems have advanced rapidly in recent years, but persistent issues of bias across demographic subgroups remain a critical concern. In particular, gender bias has been shown to affect Word Error Rate (WER) performance, with female and male speakers often experiencing unequal levels of recognition accuracy. Such discrepancies raise both ethical and technical challenges for the deployment of ASR technologies in real-world contexts. Building on recent work by Raes et al. (2024)[1], which introduced the WER parity metric to quantify predictive gender bias in ASR systems, this project investigates whether targeted pre-processing of raw audio can reduce gender-related disparities.

Our goal is to create a novel normalization algorithm designed to operate directly on raw speech waveforms prior to model ingestion. The algorithm seeks to minimize systematic differences between male and female voices by normalizing measurable and editable acoustic features such as pitch, formant distributions, frequency, silence ratios, and spectral characteristics.

The process to construct, test, and verify our algorithm is highlighted in Figure 1 below. The reduction in bias will be measured using the WER parity framework proposed by Raes et al. (2024)[1]. Through this design, our objective is to demonstrate that pre-processing can play a critical role in mitigating gender bias in ASR, offering a lightweight and model-agnostic intervention strategy.
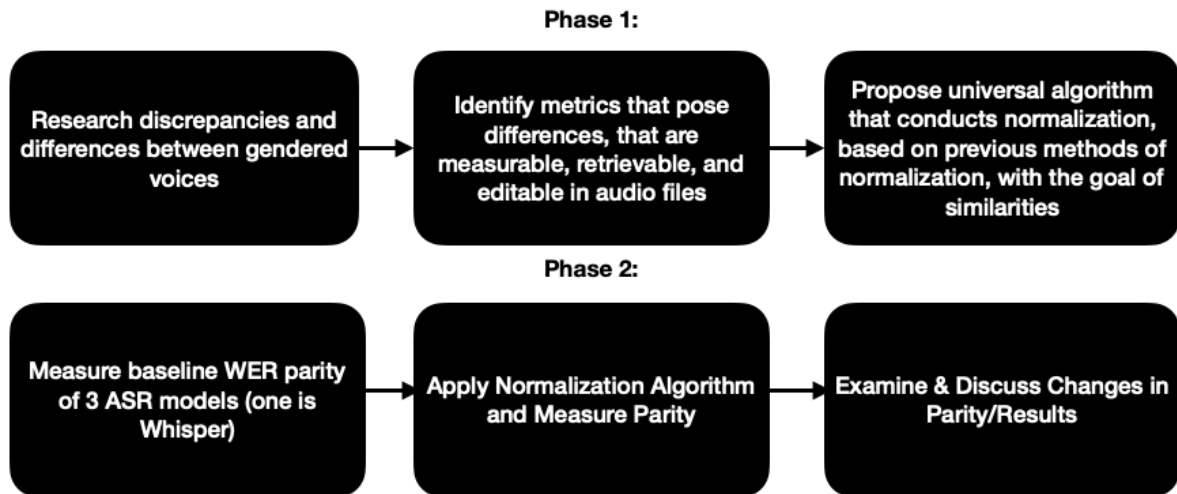

Figure 1: Proposed research process for mitigating gender bias in ASR.

[1] R. Raes, S. Lensink, and M. Pechenizkiy, *Everyone deserves their voice to be heard: Analyzing Predictive Gender Bias in ASR Models Applied to Dutch Speech Data.* arXiv:2411.09431, 2024. Available at: `https://arxiv.org/abs/2411.09431`