# Optimizing Screen Readers for Visually Impaired Users: A Multi-Stage Approach Utilizing Neural Networks and TTS

Arhaan Khaku, Apoorva Devarakonda

## 1 Introduction

Screen readers are essential tools that enable visually impaired individuals to access digital content. However, existing solutions often struggle with processing multilingual and transliterated text, leading to inaccurate pronunciations and comprehension issues. Many languages have Romanized versions that do not adhere to standard spelling conventions, making it difficult for traditional screen readers to interpret and vocalize the text correctly. This research addresses these limitations by proposing an advanced screen reader system that integrates neural networks for language detection and transliteration, followed by an optimized Text-to-Speech (TTS) module. The goal is to enhance accessibility by ensuring accurate pronunciation and natural speech synthesis for diverse languages and scripts.

## 2 Research Aim and Problem Statement

The primary aim of this research is to develop a robust screen reader system capable of accurately detecting and converting transliterated text into its native script before synthesizing speech. Current screen reader technologies are limited in their ability to process mixed-language content and transliterations, often leading to misinterpretations. This issue affects millions of visually impaired users who rely on digital communication in multiple languages, especially in regions where code-switching between languages is common.

The key challenges include:

- Reliable identification of transliterated text and its intended language.
- Conversion of Romanized text into its native script with high accuracy.
- Generating natural-sounding speech that preserves linguistic nuances.

To address these challenges, this research introduces a multi-stage pipeline that integrates machine learning models for language detection and transliteration, ensuring seamless and accurate speech output.

## 3 Proposed Solution and Methodology

The proposed system follows a structured pipeline consisting of three major components: language detection, transliteration, and Text-to-Speech synthesis.

### 3.1 System Pipeline

1. Extract text from chat interfaces and other digital sources.
2. Call backend API endpoints for processing.
3. Detect the language of the extracted text using a neural network-based classifier.
4. Convert Romanized text into the corresponding native script using an encoder-decoder model.
5. Feed the native script and detected language into a TTS model to generate speech.
6. Deliver the synthesized speech as a .wav file to the frontend for playback.

### 3.2 Language Detection

A neural network-based classifier is trained on labeled datasets containing Romanized text and their respective native languages. The training process involves:

- Preprocessing text data by cleaning and tokenizing.
- Encoding labels using a LabelEncoder.
- Training a sequential model with LSTM layers for text classification.
- Optimizing performance via class balancing and hyperparameter tuning.

### 3.3 Transliteration

Once the language is detected, the Romanized text is converted into its native script using an encoder-decoder neural network with an attention mechanism. The model is trained on paired datasets of Romanized and native script text, enabling accurate conversion that preserves linguistic integrity.

### 3.4 Text-to-Speech (TTS) Synthesis

The final stage of the pipeline involves generating high-fidelity speech using a TTS model trained on native language datasets. The system dynamically selects the appropriate TTS model based on detected language, ensuring accurate and natural pronunciation.

## 4 Implementation Plan

To successfully implement this system, the following tasks will be undertaken:

- Development of backend APIs for text extraction, language detection, transliteration, and TTS processing.
- Backend testing and detailed metric evaluation to assess system performance.
- Optimization of neural network models for improved accuracy and efficiency.
- Integration of the pipeline into a user-friendly interface.
- Evaluation using real-world datasets and feedback from visually impaired users.

## 5 Expected Outcomes

The proposed system aims to significantly improve the accessibility of multilingual screen readers by:

- Enhancing language detection accuracy for transliterated text.
- Improving transliteration fidelity to ensure proper pronunciation.
- Producing high-quality, natural-sounding speech that accommodates diverse linguistic patterns.

By addressing these aspects, the system will provide a more inclusive digital experience for visually impaired users.

## 6 Conclusion

This research seeks to bridge the gap in screen reader technologies by developing an intelligent, multilingual, and script-aware TTS pipeline. Future work includes expanding support to additional languages and improving real-time processing efficiency. By leveraging deep learning techniques, this study aims to enhance accessibility tools and empower visually impaired individuals with seamless, accurate, and natural-sounding screen reader technology.