Leveraging Community Speech Corpora to Build a Multilingual Translation Pipeline in Nigeria

# Abstract

This research proposes a modular speech-to-speech translation pipeline for real-time communication across Nigeria’s three major languages: Hausa, Yoruba, and Igbo. By leveraging community-collected corpora like NaijaVoices and recent advances in transformer-based architectures (e.g., Whisper, mBART, Coqui TTS), the study aims to bridge linguistic divides that hinder collaboration, service delivery, and education in multilingual contexts. The pipeline will consist of three main components: automatic speech recognition (ASR), machine translation (MT), and text-to-speech (TTS), each optimized for low-resource settings. A key novelty is the adaptation of pre-trained models through fine-tuning and synthetic data augmentation using code-switching and dialectal variants. The work will also engage local communities in corpus validation to improve model fairness and trustworthiness. If successful, this system could contribute significantly to inclusive digital access and language preservation in Africa’s most linguistically diverse country.

# Introduction

Nigeria is home to over 500 languages, with Hausa, Yoruba, and Igbo serving as its major regional lingua francas. Despite their widespread use, digital tools and machine learning systems remain predominantly English-centric. This linguistic exclusion not only marginalizes large portions of the population but also undermines access to health, education, and legal services. In contexts where real-time communication is vital—such as multilingual classrooms, public health campaigns, and civic engagement—automated, low-latency translation tools could dramatically improve outcomes.

Current speech translation technologies are largely designed for high-resource languages. Tools like Google Translate or Meta’s SeamlessM4T offer limited or no support for African languages, particularly for speech-based tasks. Recent community-driven efforts such as NaijaVoices, a publicly available Nigerian speech dataset, offer a unique opportunity to fill this gap by training and fine-tuning multilingual pipelines grounded in local data.

This proposal addresses three key challenges: (1) the scarcity of labeled speech data in Nigerian languages, (2) the lack of modular pipelines that combine ASR, MT, and TTS components in an extensible way, and (3) the need for trustworthiness and cultural grounding in model deployment. By leveraging pretrained models like Whisper for ASR, mBART for translation, and Coqui TTS for synthesis, and adapting them through community-specific fine-tuning, we aim to develop a practical and scalable solution.

Our research is motivated by the increasing relevance of voice-first interfaces in developing regions, as well as the potential for speech technology to enhance inclusion in governance, education, and public health. The proposed work contributes to both applied machine learning and sociolinguistics, offering a case study in how open AI tools can support linguistic justice.  
  
Research Questions:  
1. How can we effectively adapt pretrained speech and translation models to Nigerian languages using limited community-collected data?  
2. What are the impacts of dialect variation and code-switching on translation quality and user trust?  
3. How does modularity in pipeline design affect scalability and maintainability across additional languages or domains?

# Objectives

• To build a multilingual speech-to-speech pipeline using open-source models adapted for Hausa, Yoruba, and Igbo.  
• To fine-tune ASR, MT, and TTS models on community-collected datasets like NaijaVoices.  
• To evaluate the pipeline’s performance across different dialects and code-switching scenarios.  
• To explore the social impact and user perception of speech translation tools in multilingual Nigerian communities.

# Methods and Activities

Data Collection & Preparation:  
We will use the NaijaVoices dataset for speech samples and explore open-access textual corpora (e.g., JW300, Bible translations). Preprocessing will include forced alignment, language tagging, and augmentation via speed perturbation and pitch shifting. To address dialect variation, we will label subsets by region (e.g., Northern vs. Southern Yoruba).  
  
Model Development:  
- ASR: Fine-tune OpenAI’s Whisper on Hausa, Yoruba, and Igbo speech using transfer learning.  
- MT: Use mBART or NLLB (No Language Left Behind) for fine-tuning translation between language pairs.  
- TTS: Train Coqui TTS models with speaker embeddings to preserve prosody and local intonation.  
All components will be modular, allowing flexible updates and reuse.  
  
Evaluation Metrics:  
- ASR: Word Error Rate (WER) and Character Error Rate (CER).  
- MT: BLEU and chrF scores.  
- TTS: Mean Opinion Scores (MOS) from native speakers and automatic metrics like MCD.  
Trust and fairness will be assessed through user surveys and dialectal robustness tests.

# Anticipated Results

We expect to deliver a working prototype capable of speech-to-speech translation between Hausa, Yoruba, and Igbo. Fine-tuned models should outperform zero-shot baselines, especially on dialect-specific input. Preliminary tests will likely show higher WER in tone-sensitive languages like Yoruba, prompting additional tonal alignment strategies. BLEU scores may be lower than in European languages but should improve with domain-specific fine-tuning. Challenges include data sparsity, dialect drift, and limited benchmarks for subjective evaluations (e.g., intonation quality). Nonetheless, we anticipate strong community engagement and growing interest from NGOs and developers. The long-term impact includes better access to services and broader AI participation for underrepresented groups.