Domain Analysis Report

1. Application Context

The chosen application domain is Voice and Language Processing for Local Accessibility.

This domain addresses critical challenges in Lesotho, where many citizens face barriers to digital services due to low English literacy levels, visual impairments, or limited comfort with English-based interfaces. Voice and language technologies provide natural, inclusive interaction with mobile devices, making them accessible to a wider population.

The relevance of this domain is significant: it promotes financial inclusion (e.g., securing mobile money accounts with voice authentication), improves communication (e.g., sending messages in Sesotho by dictation), and broadens access to information (e.g., translating and reading English alerts aloud). These applications empower individuals who are often excluded from the digital economy, reducing the digital divide and enhancing participation in critical services such as banking, healthcare, and government communication.

2. Representative Applications

Application	Short Description	Key Computational Tasks	Notes (Relevance)
Sesotho Voice Control & Dictation.	Control the phone and send messages by speaking in Sesotho.	Voice-to-Text (FFT, MFCCs, acoustic pattern matching). MAC-heavy vector arithmetic.	Removes literacy barriers and supports disabled users, making essential phone features universally accessible.
Voice Biometric Security.	Authenticate payments or unlock the phone using a voiceprint.	Speaker Verification: audio feature extraction + pattern matching against stored templates.	Improves mobile money security and convenience, reducing dependence on vulnerable PIN codes.
Translation & Read-Aloud Assistant.	Translate English text to Sesotho and read it aloud.	Machine Translation (sequence processing) + Text-to-Speech (waveform synthesis).	Bridges the language gap for vital information (e.g., health, banking alerts).

3. Workload Characteristics

Across applications, workloads are dominated by digital signal processing and sequential data analysis.

The primary data types are audio streams (16-bit samples) and text sequences (8-bit characters or tokens). Processing involves extracting features from audio, performing pattern matching, and generating speech, all of which rely heavily on Multiply-Accumulate (MAC) operations. These operations appear in FFTs, MFCC calculations, vector similarity checks, sequence modeling, and waveform synthesis.

Arithmetic intensity is high, as many calculations are performed per data sample, while data movement is frequent: audio buffers must be streamed in real time, and model parameters (voiceprints, dictionaries, neural weights) must be repeatedly accessed from memory. The overall pattern is vector and matrix operations with high temporal locality, which makes efficient memory handling as critical as computation throughput.

4. Initial Insights for Processor Design

- Optimized execution of MAC-heavy vector and matrix operations.
- Support for SIMD/vector processing to accelerate frame-based audio and text workloads.
- An efficient memory subsystem to stream real-time audio without stalls.
- Low-power operation to suit mobile devices.
- Flexible support for both DSP kernels and small AI models.

5. References

- Speech Recognition: Everything You Need to Know
- Artificial intelligence needs assessment survey in Africa - UNESCO Digital Library