Regional Dialect Synthesis Pipeline

Hindi-to-Bangru (Haryanvi) Translation & Speech Synthesis Architecture

Data Science & AI Lab — Group 12

Abhishek • Satyam Srivastava • Sanket Agrawal • Md Fazlur Rahman

The Linguistic Divide

- While major scheduled Indian languages enjoy robust support from digital translation tools, widely spoken dialects like Haryanvi (Bangru) remain severely low-resource.
- Current State-of-the-Art (SOTA) systems normalize speech and text to Standard Hindi, effectively erasing unique vocabulary, structural grammar, and distinct tonal prosody.
- This normalization creates a severe linguistic accessibility barrier, particularly for rural learners whose native comprehension and cognitive framing align tightly with Haryanvi (Bangru).

Core Project Objectives

Dialect-Accurate Translation

Develop a specialized LLM pipeline utilizing QLoRA to convert Standard Hindi into Bangru. The system models systematic lexical substitutions and morpho-syntactic shifts rather than basic semantic equivalence.

Authentic Speech Synthesis

Engineer a TTS architecture based on VITS and transfer learning to generate natural Bangru speech. Explicitly captures distinct syllabic compression, burst energy, and forceful pitch contours.

Identifying Linguistic Gaps

- **Lexical Divergence:** Bangru exhibits systematic, rule-governed shifts (e.g., 'कहाँ' to 'कड़े', 'नहीं' to 'कोनी') rather than sporadic colloquialisms.
- **Morpho-Syntactic Variation:** Paradigm shifts in auxiliaries (e.g., 'है' to 'सै') and progressive markers alter fundamental grammatical structures.
- **Prosodic Deviation:** Bangru features higher plosive burst energy, heavy syllabic compression, and a perceptually forceful, flatter F0 contour that standard TTS smooths out.

Lit Review: Translation Architecture

Legacy LSTM Systems

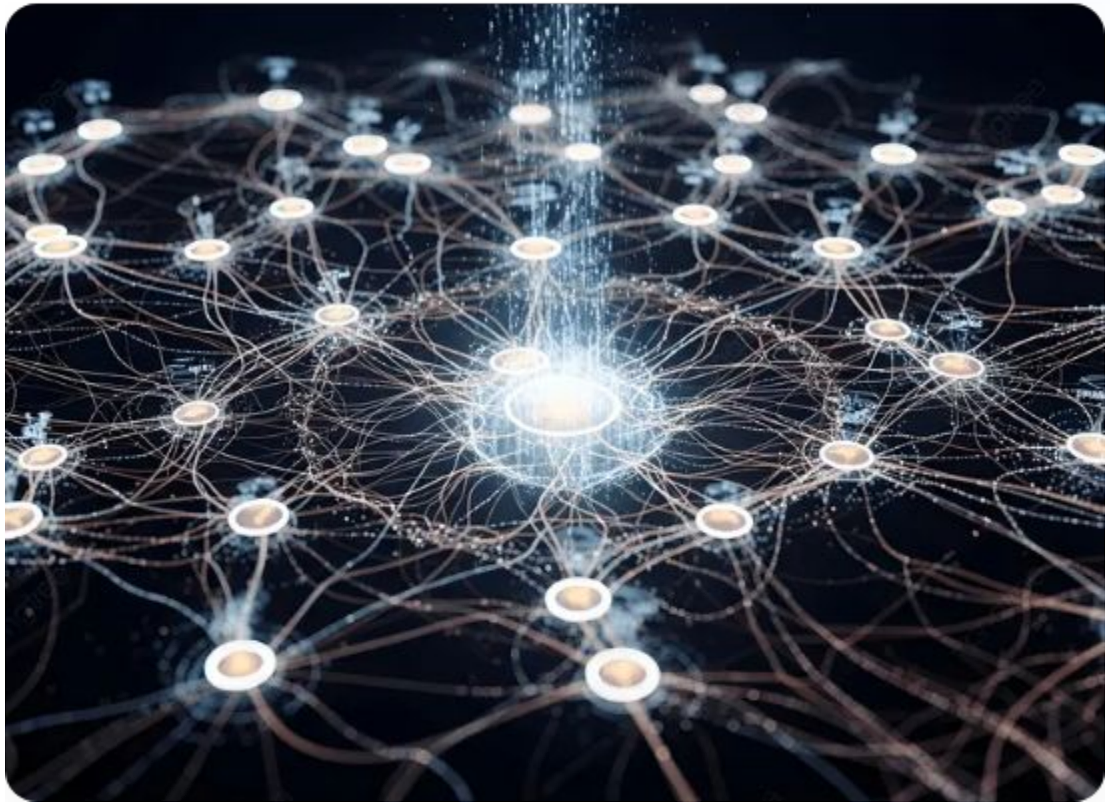
Historically relied upon for translation, sequence-to-sequence LSTMs struggle heavily with long-range dependencies and morphologically rich Indian languages. They require massive, highly-curated parallel datasets to establish baseline weights, making them unviable for zero-shot or few-shot dialect engineering where data is scarce.

Transformer LLMs (Llama-3.1)

Utilizing Grouped-Query Attention (GQA), instruction-tuned LLMs leverage "in-context learning." Rather than building language from scratch, Llama-3.1 possesses deep Indo-Aryan representations. Parameter-Efficient Fine-Tuning (QLoRA) allows us to efficiently map this existing knowledge to Haryanvi markers without catastrophic forgetting.

Proposed Methodology

- **Text Module:** Llama-3.1-8B-Instruct fine-tuned via QLoRA. We enforce deterministic lexical and morphological shifts without semantic drift.
- **Data Augmentation:** Rule-constrained synthetic data generation creates parallel pairs, allowing the model to generalize dialect rules.
- **Speech Module:** VITS initialized with AI4Bharat Indic-TTS Hindi checkpoints, solving the extreme cold-start problem of low-resource data.
- **Contingency Plan:** Matcha-TTS utilized as a flow-matching backup to mitigate potential VITS GAN instability or phoneme-skipping.



Dataset Preparation

1. AI4Bharat Indic-TTS Corpus

Over 10,000 hours of high-quality Hindi and Rajasthani speech utilized for robust phonetic transfer learning, providing the baseline acoustic foundation.

2. Custom Bangru Audio

A curated, low-resource 5-hour dataset of authentic Bangru speech. This ground-truth audio is used to precisely fine-tune the TTS architecture's tonal identity and syllabic compression.

3. Synthetic Parallel Text

Generated via structured Gemini prompting to enforce deterministic lexical replacements, creating a Hindi-to-Bangru corpus to stably fine-tune Llama-3.1 via QLoRA.

Rigorous Evaluation Metrics

Empirical & Subjective Validation

The generative pipeline is evaluated against standard Hindi baselines (Meta MMS-TTS, AI4Bharat) using a strict multi-metric framework:

- **MCD (< 6.0 dB):** Quantifying spectral distance and distortion.
- **MOS (> 3.8/5.0):** Native Bangru speaker panel grading naturalness and forceful prosody.
- **WER (< 15%):** Ensuring the synthesized speech remains highly intelligible via ASR evaluation.

Deployment Milestones

Milestones 1 & 2

Feb 26 - Mar 5: Literature review, architecture selection, synthetic parallel text generation, and audio data preprocessing/dataloaders.

Milestones 3 & 4

Mar 19 - Mar 26: Set up QLoRA scripts, build end-to-end inference piping, execute Llama-3.1 training, and begin VITS transfer learning.

Milestone 5

Apr 2: Extract objective metrics (MCD, WER), conduct MOS blind listening panels, and compile comprehensive error analysis reports.

Milestone 6

Apr 16: Finalize inference notebooks, write comprehensive documentation, and deploy the live web demo / dubbed video prototype.

Team contribution of Milestone-1

Satyam Srivastava (21f1000629@ds.study.iitm.ac.in)

- Initialized GitHub repository
- Configured branch structure
- Created worklog.md file for Team contribution

Abhishek (22f3000978@ds.study.iitm.ac.in)

- Drafted core problem statement
- Reviewed LLaMA 3.1 paper
- Compared LLM vs LSTM
- Integrated team contributions

Md Fazlur Rahman (23f1001897@ds.study.iitm.ac.in)

- Reviewed VITS paper
- Reviewed Indic-TTS
- Defined MCD & MOS metrics

Sanket Agrawal (23f1001709@ds.study.iitm.ac.in)

- Listed vocabulary differences
- Documented grammar shifts
- Described tonal characteristics

Conclusion & Q&A

By transitioning from generic translation to highly structured, low-resource dialect engineering, we can establish a scalable foundation to overcome educational linguistic barriers in Haryana.

Image Sources

https://www.indiancentury.in/wp-content/uploads/2023/08/Digitalization-in-Rural-India.jpg

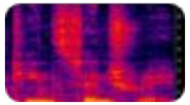
Source: www.indiancentury.in

https://static.vecteezy.com/system/resources/thumbnails/074/816/059/small/abstract-visualization-of-glowing-blue-and-purple-particles-forming-undulating-waves-against-a-dark-background-representing-data-flow-or-sound-waves-photo.jpg

Source: www.vecteezy.com

https://png.pngtree.com/thumb_back/fw800/background/20251124/pngtree-abstract-visualization-of-a-neural-network-with-glowing-nodes-and-connections-image_20578701.webp

Source: pngtree.com



https://upload.wikimedia.org/wikipedia/commons/c/c5/Spectrogram-19thC.png

Source: en.wikipedia.org