Speech Understanding

CODE-SWITCHING SPEECH RECOGNITION

MENTOR: Dr. Richa Singh

IIT JODHPUR

Team Members

- SOUVIK MAJI: B22CS089
- ARJUN BHATTAD: B22AI051

INTRODUCTION

What is code switching?

The act of alternating between two or more languages within a conversation.

Why is Code-Switching ASR important?

Common in multilingual societies (e.g., India, Singapore, Canada).
- Traditional ASR systems struggle with multilingual speech.

IMPORTANCE IN REAL LIFE

- Enhances user experience by enabling seamless multilingual speech recognition.
- Helps in language preservation for endangered languages.
- Supports media and entertainment by improving automatic subtitling.
- Expands accessibility for individuals using assistive technologies.
- Improves customer service by preventing misunderstandings in call centers.

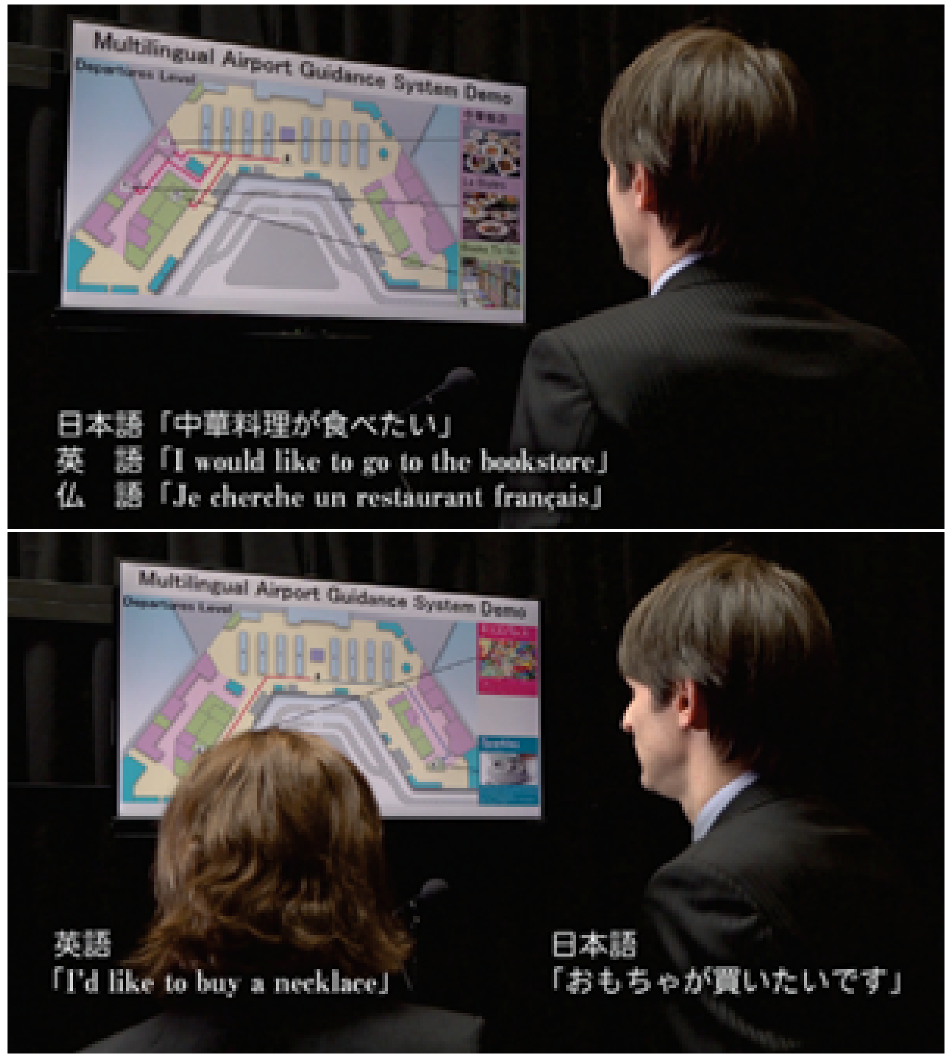
CHALLENGES IN CODE-SWITCHING ASR

- Lack of annotated datasets
- Language boundary identification difficulties
- High Word Error Rate (WER) in mixed-language settings
- Poor generalization to unseen language pairs

EVOLUTION OF CODE-SWITCHING ASR

Early Methods (1997–2015):

- Rule-based and statistical models.
- Required explicit language segmentation.
- High WER due to limited training data.



End-to-End Neural Models (2015–Present):

- Transformer-based ASR models.
- Self-attention mechanisms for better context-switching.

STATE-OF-THE-ART MODELS

Conformer-RNNT Model

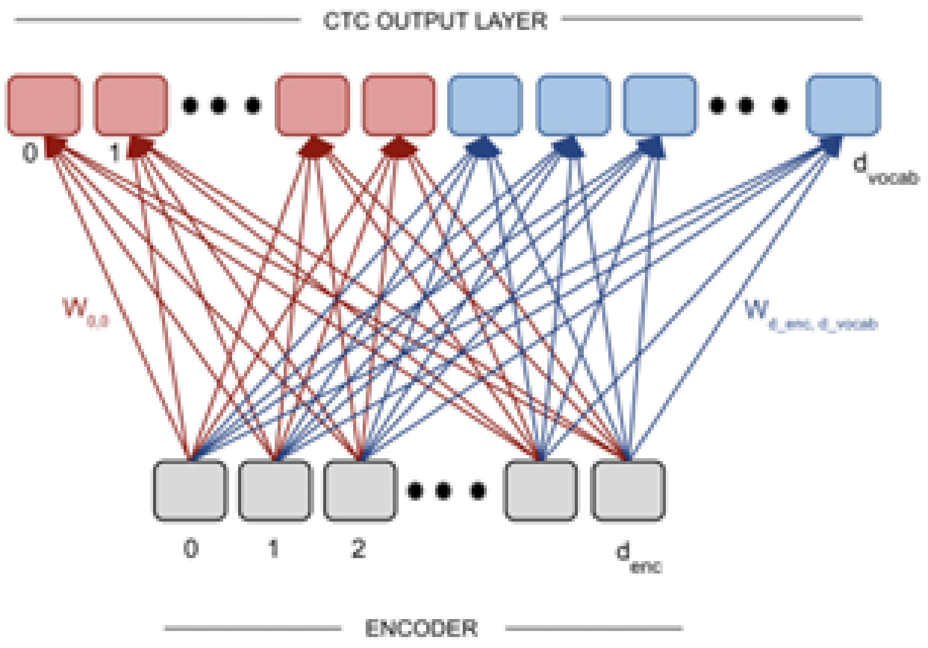
- Uses transformers + CNNs + RNNT for better accuracy.

No External Language Model (LM)

- Learns acoustic and linguistic features directly.

Concatenated Tokenizer

- Distinct token ID spaces for each language.
- Allows automatic language identification at the token level.



STRENGTHS OF SOTA MODELS

- Handles intra-word language shifts
- Preserves language identity at the token level
- Achieves real-time language identification
- Improves multilingual ASR accuracy
- Scalable with new languages

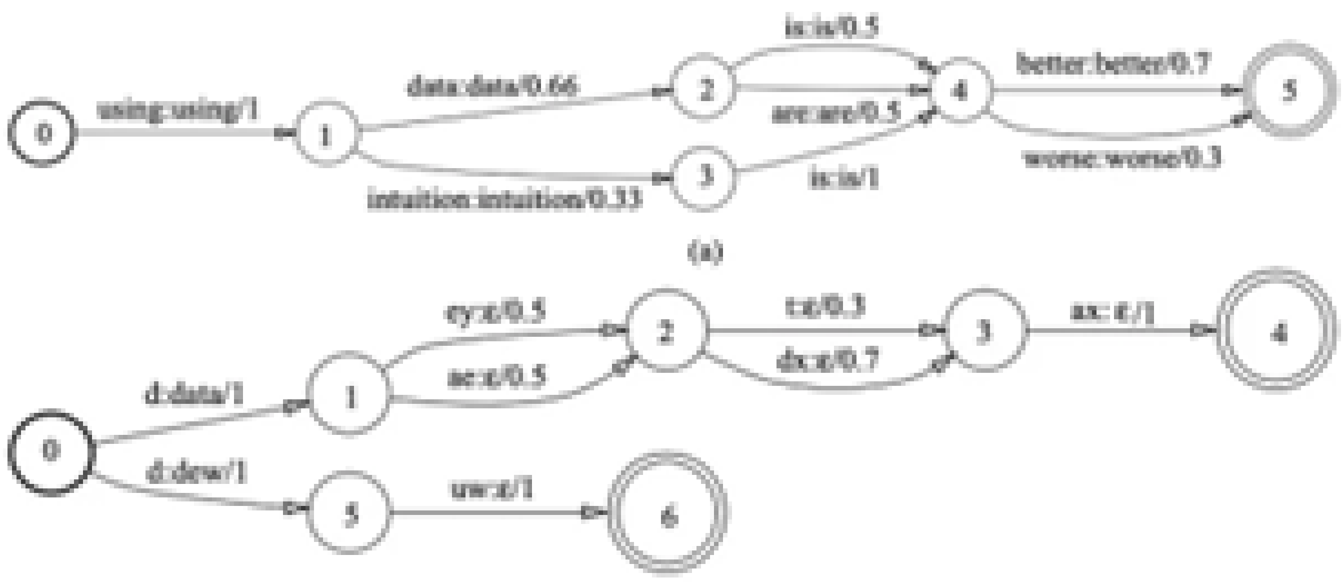


Figure 3. An illustration of WFST.

DATA AUGMENTATION IN ASR

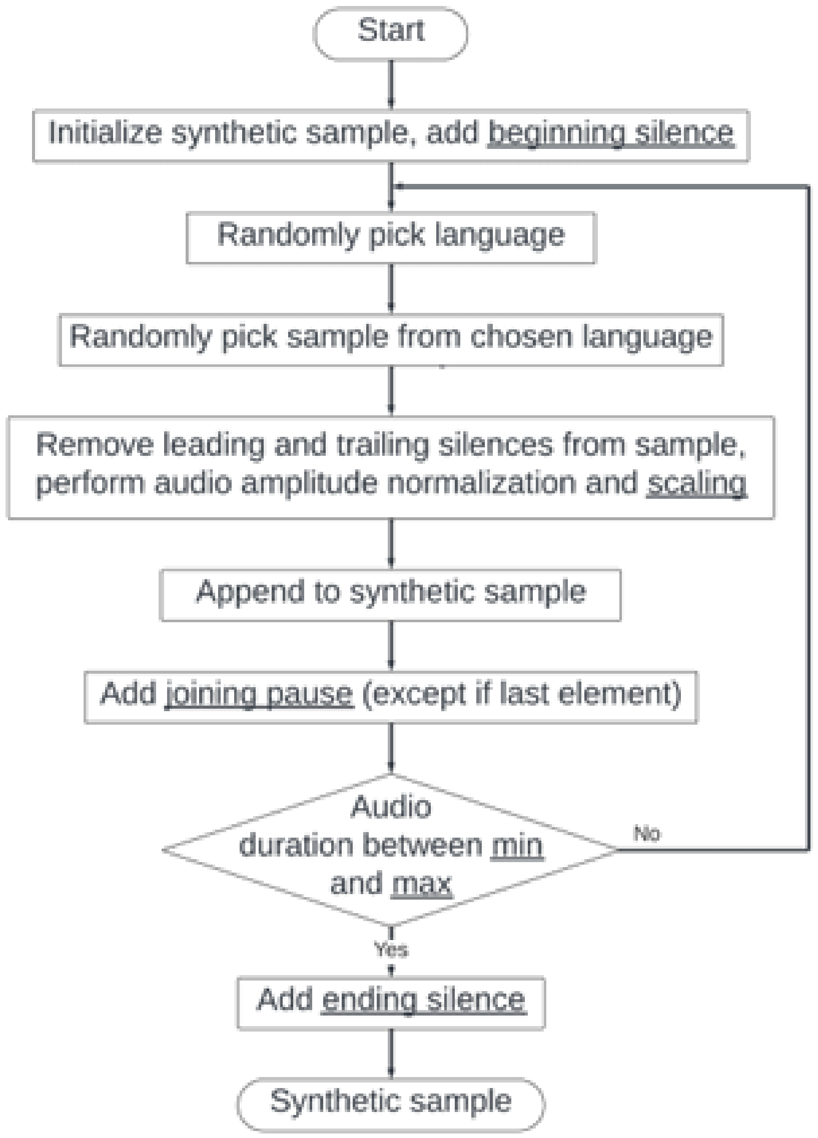
- Synthetic data is essential due to the lack of real-world code-switching datasets.
- Techniques used:

  -Text-based augmentation:

  Machine translation for synthetic CS text.

  - Speech-based augmentation:

  Combining monolingual speech segments.



EXPERIMENTAL SETUP

Datasets Used:

- English (LibriSpeech, Fisher)
- Spanish (Common Voice)
- Hindi (ULCA Dataset, MUCS 2021)
- Synthetic CS Data (10,000 hours)

Evaluation Metrics:

- Word Error Rate (WER)
- Code-Switching Performance Gap (CS-PG)
- Language Identification Accuracy

RESULTS - ASR PERFORMANCE

Model	Tokenizer	English WER	Spanish WER
Monolingual	-	5.29%	16.14%
Multilingual	Aggregate	5.00%	16.37%
Code-Switching	Concatenate	5.28%	16.42%

CS models trained on synthetic data generalize well to real-world data.

RESULTS - CODE-SWITCHING PERFORMANCE

Model	Tokenizer	WER(Synthetic)	WER(Real)
CS ASR	Aggregate	5.51%	50.0%
CS ASR	Concatenate	5.50%	53.3%

Concatenated Tokenizer allows explicit token-level language identification.

LIMITATIONS OF CURRENT MODELS

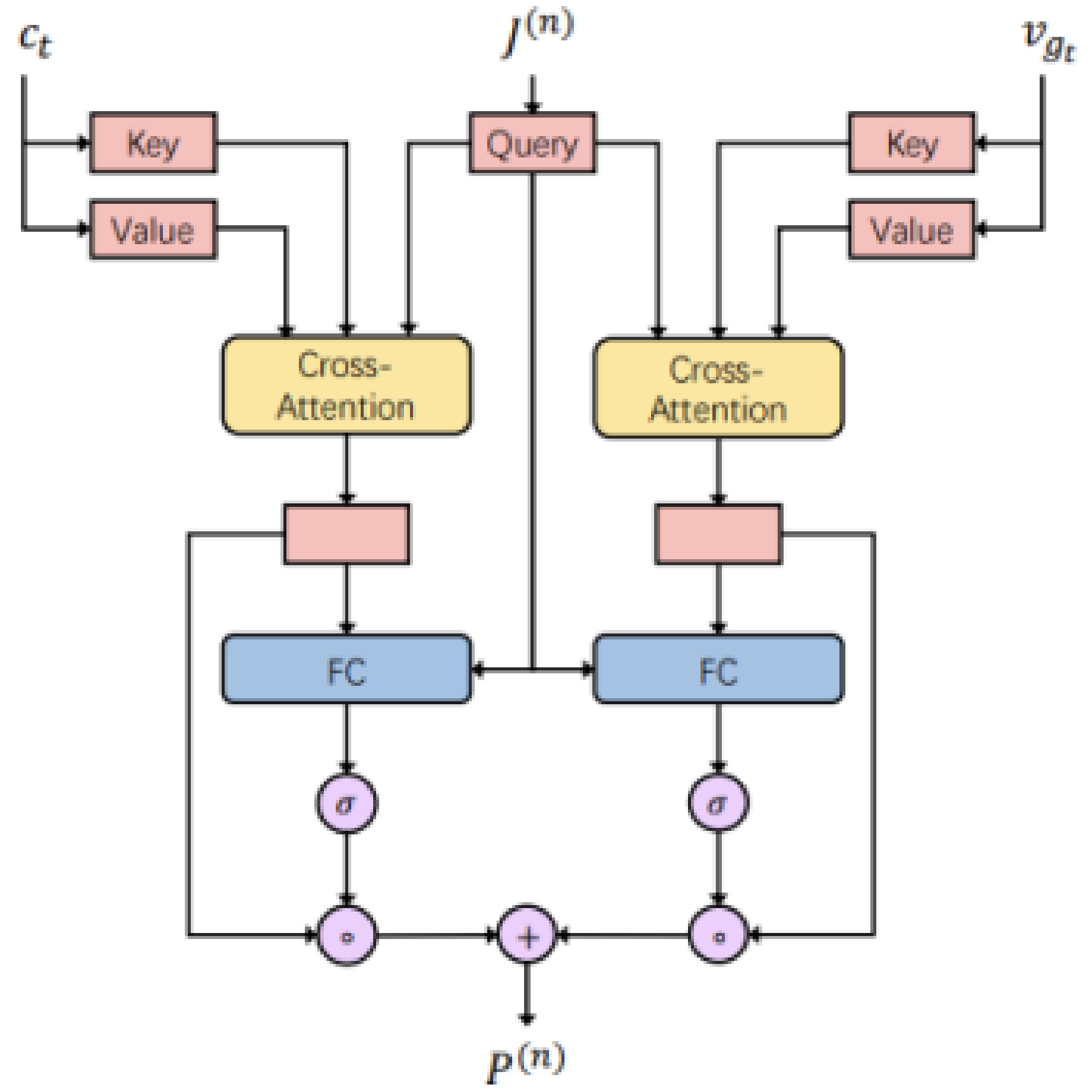
- Scalability Issues – Adding new languages increases memory demands.
- Limited Language Fusion – Struggles to handle linguistic dependencies.
- Data Scarcity – Lack of natural code-switching speech data.
- Performance in Noisy Environments – Struggles with accents and real-world noise

OPEN CHALLENGES & OPPORTUNITIES

- Scalability & Vocabulary Growth
  – Need for compact multilingual tokenization.
- Advanced Language Fusion Techniques
  – Cross-attention-based mixture-of-experts (MoE).
- Better Code-Switching Benchmarks
  – Need for spontaneous multilingual datasets.
- Efficient Cross-Lingual Transfer Learning
  – Reduce retraining costs.

FUTURE RESEARCH DIRECTIONS

- Compact Tokenization Strategies to control vocabulary explosion.
- Gated Cross-Attention for better language transitions in ASR.
- Self-Supervised Learning for low-resource ASR models.
- Developing Realistic Evaluation Datasets for spontaneous code-switching.



CONCLUSION

- Code-Switching ASR is critical for multilingual societies.
- State-of-the-art models improve accuracy but face challenges.
- Innovative tokenization & data augmentation are improving CS ASR.
- Future research must focus on scalability, fusion, and evaluation.

REFERENCES

Based on research from:

- Dhawan et al. (2023) on Concatenated Tokenization
- Miami Bangor, MUCS datasets
- NVIDIA NeMo Toolkit

Thank you