Domain Adaptation with Incremental Learning for Code-Switching ASR Performance

Assignee Research

July 7, 2026

Abstract

Automatic Speech Recognition (ASR) has become a key technology for human–AI interaction. However, code-switching ASR (CS-ASR) remains particularly challenging due to the severe scarcity of multilingual CS speech resources across diverse language pairs. Existing approaches primarily improve CS-ASR performance through synthetic CS speech generation or pair-specific fine-tuning on limited bilingual datasets. Nevertheless, these approaches face an inherent scalability limitation, as support for CS must be developed separately for language pairs whose number grows combinatorially with the number o

1 Introduction

This paper examines: Towards Truly Multilingual ASR: Generalizing Code-Switching ASR to Unseen Language Pairs. Research question: To what extent does domain adaptation with incremental learning improve code-switching ASR performance on unseen language pairs while preserving monolingual WER, as evaluated on LibriMix and CoSwDA datasets?.

2 Methodology

Systematic literature search across multiple databases yielded 12 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 8.5/10.

3 Results

12 papers retrieved. 14 claims extracted; 13 independently verified. Quality review score: 8.5/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
WHISPER-MEDIUM is used as the backbone model for the study.	×	0.07
The study investigates whether code-switching capabilities learned from seen English-centric pairs can improve recogniti	✓	0.23
Three bilingual code-switching datasets are used for fine-tuning: AI-Hub, S. Korea for KO-EN, Shin-nosuke et al. for JA-	✓	0.32
The evaluation is conducted on the human-recorded READ split from Yan et al. (2025).	✓	0.21
No publicly available datasets exist for the unseen KO-JA and KO-DE code-switching pairs.	✓	0.23
For KO-JA, 450 code-switching utterances are collected, written, recorded, and verified by authors proficient in both Ko	✓	0.24
For KO-DE, 387 speech samples are created by translating the English segments of the KO-EN code-switching dataset from P	✓	0.33
Mixed Error Rate (MER) is used to evaluate the performance, accounting for language-specific transcription characteristi	✓	0.16
Fine-tuning on one CS dataset improves recognition not only on the corresponding language pair but also, to some extent,	✓	0.25
The pretrained WHISPER-MEDIUM model exhibits a substantially higher MER for JA-EN, and all fine-tuning configurations yi	✓	0.25
The KO-JA evaluation dataset is available at https://huggingface.co/datasets/thetaone-ai/Korean-Japanese-Code-Switching-	✓	0.29
Publicly available datasets for non-English language pairs, such as Korean-Japanese or Korean-German, remain virtually n	✓	0.25
Recent studies have attempted to synthesize code-switching speech using TTS systems or by concatenating monolingual spee	✓	0.23
Such synthesized approaches often generate acoustically unnatural code-switching speech due to limited CS-aware synthesi	✓	0.22

References

- http://arxiv.org/abs/2107.01573v1
- http://arxiv.org/abs/2105.14779v2
- http://arxiv.org/abs/2606.05846v2