Discrete Audio Tokens Enhance Data Efficiency in Cross-Lingual Self-Supervised Speech Models

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: Do discrete audio token representations improve the data efficiency of cross-lingual transfer learning in self-supervised speech models when evaluated on standard low-resource speech recognition. 13 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 2.4/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Exploiting Adapters for Cross-lingual Low-resource Speech Recognition. Research question: Do discrete audio token representations improve the data efficiency of cross-lingual transfer learning in self-supervised speech models when evaluated on standard low-resource speech recognition datasets?.

2 Methodology

Systematic literature search across multiple databases yielded 14 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 2.4/10.

3 Results

14 papers retrieved. 13 claims extracted; 0 independently verified. Quality review score: 2.4/10.

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
The MetaAdapter optimization uses a meta step size denoted by μ.	×	0.04
SimAdapter leverages the knowledge of source languages from adapter modules to improve cross-lingual adaptation and mode	×	0.15
SimAdapter uses an attention mechanism where language-agnostic representations serve as the query and language-specific	×	0.04
In the SimAdapter attention operation, the temperature coefficient is denoted by τ.	×	0.02
In SimAdapter, attention matrices WQ and WK are initialized randomly, while WV has a different initialization.	×	0.02
For the Romanian (ro) language, the Trans.(B) method achieved a score of 70.14.	×	0.02
For the Romanian (ro) language, the Trans.(S) method achieved a score of 97.25.	×	0.02
For the Czech (cs) language, the Full-FT method achieved a score of 75.12.	×	0.02
For the Breton (br) language, the Trans.(S) method achieved a score of 97.88.	×	0.01
For the Arabic (ar) language, the SimAdapter method achieved a score of 81.70.	×	0.09
For the Ukrainian (uk) language, the MetaAdapter method achieved a score of 82.71.	×	0.03
The average (AVG) score for the Trans.(S) method across all listed languages is 76.68.	×	0.02
The Weighted Average score for the SimAdapter+ method is 59.43.	×	0.02

References

- http://arxiv.org/abs/1908.02590v3
- http://arxiv.org/abs/2105.11905v2
- http://arxiv.org/abs/2304.11976v1