

# Universal Speech Representations vs. Language-Specific Phoneme Layers in Low-Resource ASR

Assignee Research

June 9, 2026

## Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: How does the UML approach compare to language-specific phoneme-based output layers in terms of WER performance on the Common Voice dataset for low-resource languages when trained with varying amounts. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: CUPE: Contextless Universal Phoneme Encoder for Language-Agnostic Speech Processing. Research question: How does the UML approach compare to language-specific phoneme-based output layers in terms of WER performance on the Common Voice dataset for low-resource languages when trained with varying amounts of labeled data?.

## 2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.7/10.

## 3 Results

13 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.7/10.

## 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.

## References

- <http://arxiv.org/abs/2412.10008v1>
- <http://arxiv.org/abs/2005.10441v1>
- <http://arxiv.org/abs/2508.15316v1>