How does the integration of phonological and semantic embeddings in OpenPangu-MLA affect its zero-shot emotion

Assignee Research

June 10, 2026

Abstract

Speech inherently contains rich acoustic information that extends far beyond the textual language. In real-world spoken language understanding, effective interpretation often requires integrating semantic meaning (e.g., content), paralinguistic features (e.g., emotions, speed, pitch) and phonological characteristics (e.g., prosody, intonation, rhythm), which are embedded in speech. While recent multimodal Speech Large Language Models (SpeechLLMs) have demonstrated remarkable capabilities in processing audio information, their ability to perform fine-grained perception and complex reasoning in

1 Introduction

This paper examines: MMSU: A Massive Multi-task Spoken Language Understanding and Reasoning Benchmark. Research question: How does the integration of phonological and semantic embeddings in OpenPangu-MLA affect its zero-shot emotion recognition accuracy on the MMSU benchmark compared to prosody-only baselines?.

2 Methodology

Systematic literature search across multiple databases yielded 4 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.8/10.

3 Results

4 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.8/10.

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/2111.15271v1
- http://arxiv.org/abs/1709.03820v1
- http://arxiv.org/abs/2506.04779v3