

OpenPangu-7B-MLA and Prosodic-Only Models in Emotion Recognition on MMSU

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 10 peer-reviewed papers addressing the following research question: What is the performance gap in emotion recognition accuracy on the MMSU benchmark between OpenPangu-7B-MLA and models trained exclusively on prosodic features across diverse speaker demographics. 10 claims were extracted from source literature; 1 was independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: MMSU: A Massive Multi-task Spoken Language Understanding and Reasoning Benchmark. Research question: What is the performance gap in emotion recognition accuracy on the MMSU benchmark between OpenPangu-7B-MLA and models trained exclusively on prosodic features across diverse speaker demographics?.

2 Methodology

Systematic literature search across multiple databases yielded 10 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.2/10.

3 Results

10 papers retrieved. 10 claims extracted; 1 independently verified. Quality review score: 4.2/10.

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
MMSU encompasses a wider range of acoustic features spanning 47 distinct tasks.	×	0.09
MMSU increases reasoning complexity by requiring models to integrate paralinguistic, phonetic, and semantic information.	×	0.07
MMSU is the first benchmark to systematically incorporate linguistically grounded phenomena into spoken language underst	✓	0.22
MMSU is evaluated on 22 models, including 12 Speech-LLMs and 10 Omni Large Language Models (OmniLLMs) with audio process	×	0.12
Each instance in MMSU consists of an audio clip and a text prompt, with the model choosing one of four options (A–D).	×	0.04
Answer options in MMSU are randomly ordered and balanced across the dataset to avoid potential positional bias.	×	0.02
All models in MMSU are evaluated with the same optimized instruction-following prompts to ensure fairness and minimize p	×	0.02
The sentence 'It's nice to meet you' is a common greeting that typically ends with a neutral or slightly falling intonat	×	0.01
The first part 'It's nice to meet you,' is spoken in a neutral tone, which is characteristic of a greeting.	×	0.02
The second part, 'you,' is spoken with a rising intonation.	×	0.05

References

- <http://arxiv.org/abs/2506.04779v3>

- <http://arxiv.org/abs/2406.10598v1>
- <http://arxiv.org/abs/2407.09950v2>