OpenPangu-7B-MLA vs. Qwen-Audio and Whisper-Large-V3 in Emotion Recognition Under Noise Conditions

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 10 peer-reviewed papers addressing the following research question: How does OpenPangu-7B-MLA compare to Qwen-Audio and Whisper-Large-V3 in emotion recognition accuracy on the MMSU benchmark under varying signal-to-noise ratios. 14 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Qwen-Audio: Advancing Universal Audio Understanding via Unified Large-Scale Audio-Language Models. Research question: How does OpenPangu-7B-MLA compare to Qwen-Audio and Whisper-Large-V3 in emotion recognition accuracy on the MMSU benchmark under varying signal-to-noise ratios?.

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 3.5/10.

3 Results

10 papers retrieved. 14 claims extracted; 0 independently verified. Quality review score: 3.5/10.

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
Qwen-Audio achieves a WER of 1.8 on Librispeech dev-clean and 4.0 on dev-other.	×	0.04
Qwen-Audio achieves a WER of 2.0 on Librispeech test-clean and 4.2 on test-other.	×	0.04
Qwen-Audio achieves a WER of 1.2 on Aishell1 dev and 1.3 on test.	×	0.05
Qwen-Audio achieves a WER of 3.3 on Aishell2 Mic, 3.1 on iOS, and 3.3 on Android.	×	0.04
Qwen-Audio achieves a BLEU score of 25.1 on CoVoST2 en-de, 33.9 on de-en, 41.5 on en-zh, and 15.7 on zh-en.	×	0.02
Qwen-Audio achieves a BLEU score of 39.7 on CoVoST2 es-en, 38.5 on fr-en, and 36.0 on it-en.	×	0.03
Qwen-Audio achieves a CIDEr score of 0.441, SPICE score of 0.136, and SPIDEr score of 0.288 on Clotho.	×	0.03
Qwen-Audio achieves an AAS of 60.3 ms on Industrial Data for SRWT.	×	0.04
Qwen-Audio is initialized with the Whisper-large-v2 model for the audio encoder.	×	0.05
Qwen-Audio is trained in two stages: multi-task pre-training and supervised fine-tuning.	×	0.15
During multi-task pre-training, the weights of the LLM are frozen and only the audio encoder is optimized.	×	0.09
During supervised fine-tuning, the weights of the audio encoder are fixed and only the LLM is optimized.	×	0.04
Qwen-Audio is evaluated on 12 datasets across various tasks including ASR, S2TT, AAC, ASC, SER, AQA, VSC, and MNA.	×	0.05
The evaluation datasets for Qwen-Audio are rigorously excluded from the training data to avoid data leakage.	×	0.05

References

- http://arxiv.org/abs/2506.04779v3
- http://arxiv.org/abs/2311.07919v2

- http://arxiv.org/abs/2602.21464v1