Dynamic Modality Weight Adjustment in OpenPangu-7B-MLA for Noisy Speech Robustness

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 11 peer-reviewed papers addressing the following research question: What is the impact of dynamically adjusting the speech-text modality weight ratio during inference (rather than training) of OpenPangu-7B-MLA on its robustness to noisy speech inputs in MMSU,. 11 claims were extracted from source literature; 2 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.4/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Lightweight speech enhancement guided target speech extraction in noisy multi-speaker scenarios. Research question: What is the impact of dynamically adjusting the speech-text modality weight ratio during inference (rather than training) of OpenPangu-7B-MLA on its robustness to noisy speech inputs in MMSU, evaluated using accuracy and domain-specific degradation metrics?.

2 Methodology

Systematic literature search across multiple databases yielded 11 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.4/10.

3 Results

11 papers retrieved. 11 claims extracted; 2 independently verified. Quality review score: 4.4/10.

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
In training strategy S0, pretraining GTCRN and the backbone separately and then combining them yields an SI-SDR of 7.60,	×	0.04
In training strategy S1, integrating a pretrained GTCRN as the front-end while training the backbone from scratch result	×	0.05
Training strategy E5, which jointly optimizes GTCRN and the backbone via a two-stage scheme, achieves an SI-SDR of 8.32,	×	0.05
Strategy S1 outperforms strategy S0 in SI-SDR, PESQ, and STOI metrics.	×	0.04
Strategy E5 achieves the best performance results among the tested strategies (S0, S1, E5).	×	0.02
Embedding-based and hybrid target speech extraction approaches are limited in practical deployment due to their large si	×	0.07
CIE-mDPTNet, SEF-Net, and SEF-PNet are embedding/encoder-free paradigms that demonstrate state-of-the-art performance.	×	0.10
Recent target speech extraction methods perform poorly in noisy multi-speaker scenarios.	✓	0.21
In noisy multi-speaker scenarios, noise severely corrupts enrollment guidance, often leading to target speech distortion	✓	0.23
Leveraging denoised speech for context interaction effectively suppresses noise components in the guidance spectrogram c	×	0.15
The work was sponsored by the Natural Science Foundation of Shanghai under Grant No. 25ZR1401277.	×	0.01

References

- http://arxiv.org/abs/2506.04779v3
- http://arxiv.org/abs/1809.04553v1
- http://arxiv.org/abs/2508.19583v2