GTCRN-Enhanced Speech Integration in Multimodal Large Language Models on CHiME-6

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: How does integrating GTCRN-based speech enhancement into multimodal large language models impact word error rate on the CHiME-6 multi-speaker benchmark compared to end-to-end baselines. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Survey of End-to-End Multi-Speaker Automatic Speech Recognition for Monaural Audio. Research question: How does integrating GTCRN-based speech enhancement into multimodal large language models impact word error rate on the CHiME-6 multi-speaker benchmark compared to end-to-end baselines?.

2 Methodology

Systematic literature search across multiple databases yielded 12 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.7/10.

3 Results

12 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 4.7/10.

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/2006.07898v1
- http://arxiv.org/abs/2505.10975v3
- http://arxiv.org/abs/2508.19583v2