FlashSpeech Latency-Accuracy Trade-offs on VCTK Under Varying Noise Schedules

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 10 peer-reviewed papers addressing the following research question: What is the trade-off between inference latency and speaker verification accuracy for FlashSpeech when evaluated on the VCTK dataset with varying noise schedules. 7 claims were extracted from source literature; 4 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 6.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: A Framework for Robust Speaker Verification in Highly Noisy Environments Leveraging Both Noisy and Enhanced Audio. Research question: What is the trade-off between inference latency and speaker verification accuracy for FlashSpeech when evaluated on the VCTK dataset with varying noise schedules?.

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 6.8/10.

3 Results

10 papers retrieved. 7 claims extracted; 4 independently verified. Quality review score: 6.8/10.

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.

5 Extracted Claims

Claim	Verified	Confidence
The proposed method combines embeddings from noisy and enhanced audio to improve speaker verification in highly noisy en	✓	0.24
Generative DNNs for speech enhancement can produce superior speech quality but may distort speaker characteristics under	✓	0.24
The proposed framework is lightweight and agnostic to specific speaker verification and speech enhancement techniques.	✓	0.35
The proposed method outperforms both noisy and enhanced audio alone in speaker verification tasks under various noise co	✓	0.17
The triplet loss function used in the proposed method aims to distinguish between similar and dissimilar speaker example	×	0.09
The proposed framework reduces computation complexity compared to methods that employ a learning-based interpolation age	×	0.04
The proposed method delivers reliable speaker verification performance even in severe noisy conditions where previous me	×	0.10

References

- http://arxiv.org/abs/2002.03562v2
- http://arxiv.org/abs/2508.18913v1
- http://arxiv.org/abs/2204.02609v1