FlashSpeech Scalability in Memory Usage and Generation Quality for Long-Form Speech Synthesis

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: How does FlashSpeech scale in terms of memory usage and generation quality when increasing the context window for long-form speech synthesis. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: FlashSpeech: Efficient Zero-Shot Speech Synthesis. Research question: How does FlashSpeech scale in terms of memory usage and generation quality when increasing the context window for long-form speech synthesis?.

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 3.7/10.

3 Results

12 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.7/10.

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/1909.03965v1
- http://arxiv.org/abs/2508.02112v1
- http://arxiv.org/abs/2404.14700v4