

FlashSpeech Training Efficiency in Low-Resource Cross-Lingual Voice Cloning

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: How does the training data efficiency of FlashSpeech scale relative to large-scale diffusion TTS models when evaluated on cross-lingual voice cloning performance for languages with under 10 hours of. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 6.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Cross-Lingual F5-TTS: Towards Language-Agnostic Voice Cloning and Speech Synthesis. Research question: How does the training data efficiency of FlashSpeech scale relative to large-scale diffusion TTS models when evaluated on cross-lingual voice cloning performance for languages with under 10 hours of labeled data?.

2 Methodology

Systematic literature search across multiple databases yielded 14 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 6.2/10.

3 Results

14 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 6.2/10.

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/2602.04160v2>
- <http://arxiv.org/abs/2406.04630v1>
- <http://arxiv.org/abs/2509.14579v4>