

Multimodal Symbolic-to-Audio Model Robustness Across Musical Style Domain Shifts

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 15 peer-reviewed papers addressing the following research question: How robust are multimodal symbolic-to-audio models to domain shifts in musical style when evaluated using objective spectral distortion metrics. 8 claims were extracted from source literature; 3 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.6/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Music SketchNet: Controllable Music Generation via Factorized Representations of Pitch and Rhythm. Research question: How robust are multimodal symbolic-to-audio models to domain shifts in musical style when evaluated using objective spectral distortion metrics?.

2 Methodology

Systematic literature search across multiple databases yielded 15 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 5.6/10.

3 Results

15 papers retrieved. 8 claims extracted; 3 independently verified. Quality review score: 5.6/10.

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
MuseGan allows users to condition generated results on full-length multi-track music.	×	0.04
DeepBach provides a constraint mechanism that allows users to limit the generated results to match composer styles.	×	0.02
Music Transformer supports accompaniment arrangement from an existing melody track in classical music.	×	0.03
Music SketchNet allows users to specify partial musical ideas in terms of incomplete and distinct pitch and rhythm repre	✓	0.26
Music SketchNet completes the missing parts given the known context and user input.	×	0.12
Music InpaintNet completes musical pieces by predicting vector representations for missing measures, then the vector rep	×	0.13
SketchNet outperforms the state-of-the-art both in terms of objective metrics and subjective listening tests.	✓	0.24
SketchNet can successfully incorporate user-specified snippets during the generation process.	✓	0.31

References

- <http://arxiv.org/abs/2008.01291v1>
- <http://arxiv.org/abs/2109.01948v1>
- <http://arxiv.org/abs/1907.02265v1>