← Back to **the profile of Julian McAuley** (/profile?id=~Julian_McAuley1)

StylePitcher: Generating Style-Following and Expressive Pitch Curves for Versatile Singing Tasks

(/pdf?id=fIlSXkkuNs)

Jingyue Huang (/profile?id=~Jingyue_Huang1),
Qihui Yang (/profile?id=~Qihui_Yang1),
Fei-Yueh Chen (/profile?id=~Fei-Yueh_Chen1),
Julian McAuley (/profile?id=~Julian_McAuley1),
Randal Leistikow (/profile?id=~Randal_Leistikow1),
Yongyi Zang (/profile?id=~Yongyi_Zang1)

Published: 22 Sept 2025, Last Modified: 08 Nov 2025 AI4Music Everyone
Revisions (/revisions?id=fIlSXkkuNs) BibTeX

Keywords: F0 Estimation, Singing Voice Synthesis and Conversion, Pitch Correction, Rectified Flow Matching
TL;DR: We propose StylePitcher, a general-purpose pitch curve generator that learns singer style from reference audio while preserving alignment with the intended melody.
Abstract:
Existing pitch curve generators face two main challenges: they often neglect singer-specific expressiveness, reducing their ability to capture individual singing styles. And they are typically developed as auxiliary modules for specific tasks such as pitch correction, singing voice synthesis, or voice conversion, which restricts their generalization capability. We propose StylePitcher, a general-purpose pitch curve generator that learns singer style from reference audio while preserving alignment with the intended melody. Built upon a rectified flow matching architecture, StylePitcher flexibly incorporates symbolic music scores and pitch context as conditions for generation, and can seamlessly adapt to diverse singing tasks without retraining. Objective and subjective evaluations across various singing tasks demonstrate that StylePitcher improves style similarity and audio quality while maintaining pitch accuracy comparable to task-specific baselines.

Track: Paper Track
Confirmation: Paper Track: I confirm that I have followed the formatting guideline and anonymized my submission.
Submission Number: 68

Filter by reply type... Filter by author... Search keywords... Sort: Newest First
Everyone Program Chairs Submission68 Reviewers Submission68 Authors
Submission68... Submission68... Submission68... Submission68...
4 / 4 replies shown
Add: **Withdrawal**

Paper Decision

Decision by Program Chairs 22 Sept 2025, 18:07 (modified: 22 Sept 2025, 19:58)
Program Chairs, Reviewers, Authors Revisions (/revisions?id=PjoDYUxEyM)

Decision: Accept

StylePitcher is a diffusion-based model that generates pitch curves with singer-specific style and musical accuracy.

Official Review by Reviewer gacq 20 Sept 2025, 08:47 (modified: 22 Sept 2025, 21:36)
Program Chairs, Reviewers Submitted, Reviewer gacq, Authors Revisions (/revisions?id=3sXY3kzyub)

Strengths:
The method is clearly illustrated in Figure 1.

Experiments demonstrate promising performance with the proposed pitch curve generation strategy.

The approach achieves strong results in pitch correction, singing synthesis, and voice conversion with a plug-and-play strategy.

Weaknesses:
The paper lacks implementation details for the APC, SVC, and SVS models, and does not provide a plot showing how the proposed method integrates into each subtask, which reduces clarity for readers unfamiliar with these pipelines.

Rating: 4: Strong accept
Confidence: 5: The reviewer is absolutely certain that the evaluation is correct and very familiar with the relevant literature
Justification:
The method is clearly presented in Figure 1, and experiments show promising performance with the proposed pitch curve generation strategy. It delivers strong results in pitch correction, singing synthesis, and voice conversion with a plug-and-play design.

To improve clarity, the authors could include implementation details of the APC, SVC, and SVS models in the appendix, along with a plot illustrating how the method integrates into each subtask. While the approach is understandable given my background in singing and speech synthesis, readers without sufficient experience in APC, SVC, or SVS pipelines may find the integration details hard to follow.

I would recomment strong accept.

Suggestions:
Please add a plot or paragragh illustrating how the proposed method integrates into each subtask to improve clarity.

Review of StylePitcher

Official Review by Reviewer 89Xi 19 Sept 2025, 06:55 (modified: 22 Sept 2025, 21:36)
Program Chairs, Reviewers Submitted, Reviewer 89Xi, Authors Revisions (/revisions?id=g9N2CLPDsB)

Strengths:
This work uses a generative model to produce stylistic pitch curves. The work is motivated by the limitation of some f0 extraction methods that capture pitch but lose information about the stylistic details of the singer. Their approach is interesting as it leverages a conditional generative model for this task. The proposed approach is both well motivated, explained, and evaluated.

Weaknesses:
While the proposed method is well motivated and clearly has many potential applications, this work could be strengthened by somewhat simplifying the evaluation and scope of the applications. In its current state, as a four page manuscript, it is somewhat difficult to give sufficient detail on all applications that are presented. The work could be stronger if just one evaluation is the focus. However, expanding on the current format into a full conference paper could also be fruitful.

Some elements of the presented method are unclear. How does this approach enable variable length inference? It seems like the DiT is trained on 1024 frames or about 20s. If you have a 1 min song how do you perform inference on this? It also appears that the more context you use for the infill will also decrease the size of the prediction. Was there any investigation into how changing the amount of context changed the performance?

Rating: 4: Strong accept

Confidence: 4: The reviewer is confident but not absolutely certain that the evaluation is correct

Justification:

This work is well motivated, presents an interesting model formulation, and has sufficient evaluation to be of interest to the workshop community.

Suggestions:

See the Weaknesses section. Right now the paper is of a somewhat large scope for a four page manuscript. The work could be strengthened by focusing down the scope, or expanding this work into a full conference paper.

Review of StylePitcher

Official Review by Reviewer EnRk 16 Sept 2025, 04:45 (modified: 22 Sept 2025, 21:36)

Program Chairs, Reviewers Submitted, Reviewer EnRk, Authors Revisions (/revisions?id=aWptJXENUa)

Strengths:

The method is straightforward and introduces meaningful applications for singing voice-related tasks. The authors have well done the evaluations with a couple of baselines and its variant both objectively and subjectively. They also provided further analysis on each downstream tasks and discussed them well.

Weaknesses:

The work is decently solid for a workshop paper. I'd love to see the detailed version of this work once it's published at a proceedings.

Rating: 5: Strong accept (Nominate for best paper/demo)

Confidence: 4: The reviewer is confident but not absolutely certain that the evaluation is correct

Justification:

The current work is worth sharing with the community with the reasons I provided in the Strengths section.

Suggestions:

n/a