PopMelodyGen：版权意识型可控流行旋律生成平台

Di Lu[1,2,3*]，**Xinran Zhang**[1,2,3*]，**Qingwen Zhou**[1,2,3*]，**Yan Gao**[1,2,3]，**Yuan Zhang**[1,2,3]，**XiaoqingWang**[1,2,3]，**Yuming Sun**[1,2,3]，**Xiaobing Li**[1,2,3]，**Maosong Sun**[1,2,3,4†]

[1]中央音乐学院人工智能音乐与音乐信息技术系 [2]中央音乐学院音乐与脑科学重点实验室，教育部重点实验室，中国[3]中央音乐学院人工智能音乐实验室，教育部哲学与社会科学实验室，中国[4]清华大学计算机科学与技术系 {ld, zxinran, zqingwen}@ccom.edu.cn, sms@tsinghua.edu.cn

arXiv:2611.28686v1 [cs.SD] 21 Nov 2026

摘要

由于版权限制，大规模可再分发的符号流行旋律数据仍然稀缺，这成为音乐生成在实际部署和可复现研究方面的瓶颈。我们介绍 PopMelodyGen，这是一个面向可控流行旋律生成的版权感知平台，将基于规则的合成与神经生成相结合。我们的方法首先利用专家合成器，根据显式音乐知识和模板生成大规模带标签的合成旋律，提供可扩展且来源可审计的预训练数据。随后，神经模型在该语料库上进行预训练，并使用小规模的授权真实流行音乐语料库进行适配，从而在保留可控性的同时提升旋律自然度。该平台支持创作者交互式创作旋律、播放伴奏并导出作品，也支持研究人员批量构建带注释的符号数据集。PopMelodyGen 展示了一种实用的创意人工智能训练与部署范式，适用于大规模可复用真实世界数据有限的情形。

代码 — https://github.com/ld04080106/PopMelodyGen**视频**— https://youtu.be/U_R1MEZFOjI

引言

生成式人工智能的最新进展高度依赖大规模训练语料库，但这一假设并不适用于符号流行旋律，因为商业流行歌曲受版权限制，难以大规模再分发。

现有系统探索了不同架构下的可控符号生成（Choi 等，2021；Huang 和 Yang，2020；Tian 等，2025；Wang 等，2025），但可复现的开发仍依赖合适的符号训练语料库。现有面向流行音乐的数据集，例如 POP909 以及 RWC（Wang 等人，2020；Goto 等人，2002；Goto、Balke，以及 Mueller2026）在可控流行旋律预训练方面的规模和条件覆盖范围仍然有限，而更大规模、可复用的符号语料库（如 PDMX（Long 等人，2025））并非专为这一应用场景设计。

为解决这一瓶颈，我们推出 PopMelodyGen，这是一个版权意识型平台，以合成预训练和获许可的真实流行音乐适配为核心。专家合成器首先依据一般音乐理论规则以及由音乐人策划的结构、和声与节奏模板构建带标签的训练数据，而非使用现有流行旋律。神经模型先在这一合成语料库上进行预训练，随后适配真实流行音乐旋律模式。我们将这一工作流程部署为网络平台，支持交互式旋律创作和批量构建带注释的符号数据集，从而降低可复现音乐人工智能研究所面临的版权相关障碍。

系统概述

图1展示了 PopMelodyGen 的四阶段流程。

合成旋律生成。 我们实现的专家合成器可在明确的音乐控制条件下生成流行风格旋律，并采用标准符号音乐格式（MIDI）：速度类别、调性/调式、人声音域、歌曲结构和和弦进行。每个输出都配有结构化元数据，其中记录了生成设置和注释。该合成器基于通用音乐理论规则和音乐人精选的元素级模板（结构、和声、节奏），不会摄取或转换现有流行旋律，因此每个生成样例的来源信息都完全可查。

合成预训练。 生成的 MIDI–元数据对会转换为基于事件的词元表示，并遵循既有的符号音乐编码惯例（Huang 和 Yang，2020）。随后，使用下一词元预测对仅解码器 Transformer 进行预训练，使其学会可控的自回归旋律生成，从而得到 Synth-Base 模型。

真实流行音乐适配。 由于规则生成的旋律可能与人类创作的流行音乐有所不同，我们对

*这些作者贡献相同。†通讯作者。版权所有 © 2027，人工智能促进协会（www.aaai.org）。保留所有权利。

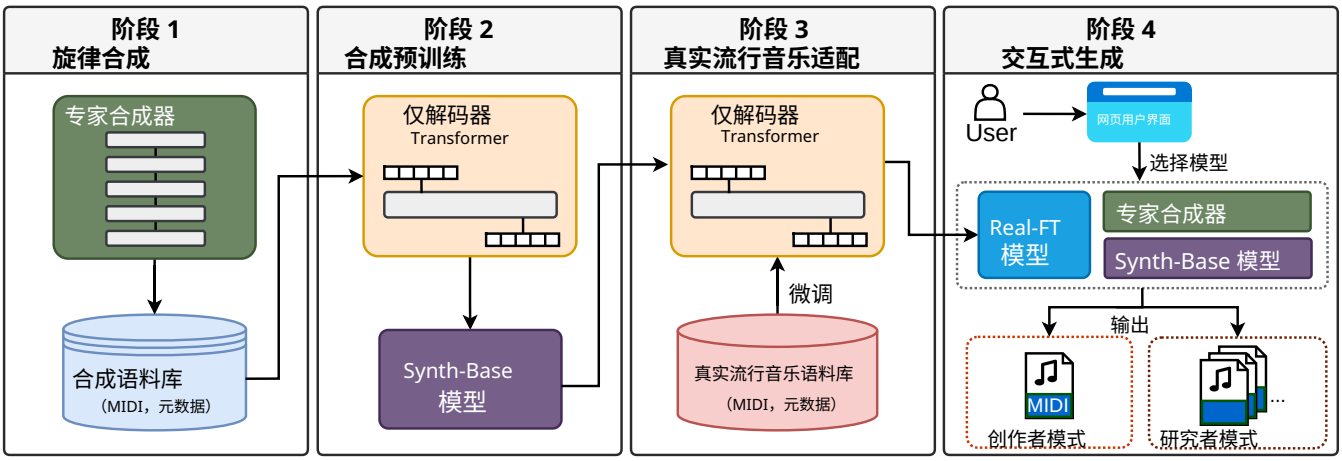


图1：PopMelodyGen 流水线概览，从合成数据生成到交互式部署。

Synth-Base进行适配，使用规模较小且获得授权的真实流行音乐语料库。真实流行音乐序列与采样合成重放示例混合，以保留可控性，从而得到Real-FT模型。

交互式生成。 网络平台将用户选择的音乐条件传递给所选的生成后端。生成的旋律以 MIDI 形式呈现，以便通过网络平台进行可视化、播放和导出。

现场演示

现场演示展示了两种互补的工作流程。在**创作者模式**下，参与者选择生成后端（专家 / Synth-Base / Real-FT），并指定包括调性/调式、速度类别、人声音域、歌曲结构和和弦进行在内的音乐条件。系统生成流行风格旋律，并配以内置鼓组、贝斯、钢琴和吉他伴奏，以便立即试听。参与者可以使用不同后端重新生成相同的音乐条件，比较其输出，并导出结果以便进一步编辑。

在**研究者模式**下，参与者可以配置音乐条件的范围（例如，12个调 × 3 速度类别），并批量生成带标签的MIDI–元数据对。结构化数据集包可以直接下载。这展示了同一系统不仅可以作为创作工具，还可以作为数据生成平台，用于可复现的音乐人工智能实验。

实现与验证

实现。 合成预训练阶段，我们使用 10 万首规则生成的旋律对 Synth-Base 模型进行预训练。真实流行音乐适配阶段，我们使用获得授权的 RWC 流行音乐数据库（Goto 等，2002；Goto、Balke 和 Mueller，2026）。经过预处理和基于移调的数据增强后，98 首经过处理的 RWC旋律生成了 847 个分词后的适配样本。两个阶段均使用 6 层仅解码器 Transformer，隐藏状态维度为 512，注意力头数为 8。Web界面使用 Node.js 实现并部署在专用服务器上，模型推理通过后端API提供服务。

验证。 在客观评估中，Expert、Synth-Base 和 Real-FT 各自在系统性变化的音乐条件下生成 4,320 首旋律。Synth-Base 在和声遵循度和终止式控制方面接近专家合成器；Real-FT 则转向贴近 RWC参考统计数据，同时保留了大部分习得的可控性。在一项有 30 名参与者参加的人体研究中，Real-FT 获得最高的平均意见分（3.55/5）和三选最佳偏好率（48.3%），表明真实流行音乐适配提升了感知旋律质量。

负责任使用

我们将 PopMelodyGen 描述为具有版权意识，而非在法律上版权安全。合成语料库依据通用音乐规则和元素级模板构建，不会摄取或转换现有流行旋律；真实流行音乐数据则仅通过授权的 RWC 数据集引入。

作为实证合理性检查，我们采用基于音程 n 元组的最近邻相似度，将 12,960 条生成旋律与 RWC 参考集进行比较。该方法受到使用局部片段和移调不变的音程表示进行符号旋律相似度分析的方法启发（Urbano 2014；Lattner、Grachten 和 Widmer 2018）。平均最近邻相似度为 0.061，最高为 0.336。对得分最高的配对进行人工检查后，未发现明显的近重复旋律。这项分析有助于识别近似复制风险，但不构成法律保证。

结论

PopMelodyGen 通过在规则生成的合成数据上进行预训练，并使用有限的授权真实流行音乐数据进行适配，解决了可控流行旋律生成中的训练数据瓶颈。由此构建的平台既支持交互式流行旋律创作，也支持面向研究的数据集构建，展现了在缺乏大规模可复用训练数据时，创意人工智能的一种实用策略。

致谢

本研究获得国家自然科学基金委员会专项项目（项目编号：T2341003）和北京市属高校高精尖学科建设项目资助。

参考文献

Choi, K.; Park, J.; Heo, W.; Jeon, S.; and Park, J. 2021. 基于变换器解码器的和弦条件旋律生成。*IEEE Access*, 9: 42071–42080. Goto, M.; Balke, S.; and Mueller, M. 2026. RWC音乐数据库。Goto, M.; Hashiguchi, H.; Nishimura, T.; and Oka, R. 2002. RWC音乐数据库：流行、古典与爵士音乐数据库。收录于第三届国际音乐信息检索会议（*ISMIR*）论文集，287–288。Huang, Y.-S.; and Yang, Y.-H. 2020. 流行音乐变换器：富有表现力的流行钢琴作品的基于节拍的建模与生成。收录于第*28*届*ACM*国际多媒体会议论文集，MM ’20，1180–1188。Lattner, S.; Grachten, M.; and Widmer, G. 2018. 从复调音乐序列中学习音程表征。收录于第*19*届国际音乐信息检索会议（*ISMIR*）论文集，661–668。Long, P.; Novack, Z.; Berg-Kirkpatrick, T.; and McAuley, J. 2025. PDMX：用于符号音乐处理的大规模公有领域MusicXML数据集。收录于*ICASSP 2025 - 2025 IEEE*国际声学、语音与信号处理会议（*ICASSP*），1–5。Tian, S.; Zhang, C.; Yuan, W.; Tan, W.; and Zhu, W. 2025. XMusic：迈向通用且可控的符号音乐生成框架。*IEEE*多媒体汇刊，27：6857–6871。Urbano, J. 2014. MIREX 2014符号旋律相似度评测中的MelodyShape。收录于第*10*届音乐信息检索评测交流会（*MIREX 2014*）。Wang, Y.; Wu, S.; Hu, J.; Du, X.; Peng, Y.; Huang, Y.; Fan, S.; Li, X.; Yu, F.; and Sun, M. 2025. NotaGen：采用大语言模型训练范式提升符号音乐生成的音乐性。收录于第三十四届国际人工智能联合会议论文集，IJCAI ’25。ISBN 978-1-956792-06-5。Wang, Z.; Chen, K.; Jiang, J.; Zhang, Y.; Xu, M.; Dai, S.; Bin, G.; and Xia, G. 2020. POP909：用于音乐编曲生成的流行歌曲数据集。收录于第*21*届国际音乐信息检索会议（*ISMIR*）论文集，38–45。

附录
A 专家合成器

专家阶段的目的是生成大规模、多样且可控的合成语料库，用于符号流行旋律预训练。图2展示了这一流程。

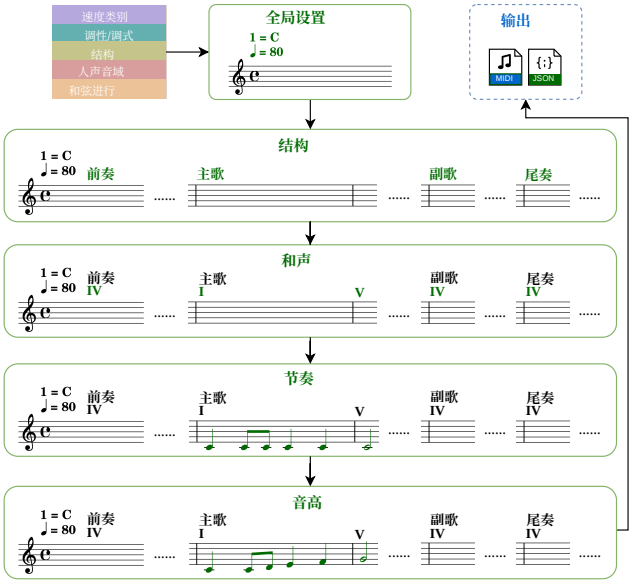


图 2：用于合成旋律生成的专家合成器。该系统实例化全局设置、结构、和声、节奏和音高，并输出带有结构化元数据的旋律 MIDI 文件。

该合成器以可控的音乐条件作为输入，包括速度类别、调性/调式、人声音域、结构和和弦进行。它首先实例化全局设置，然后生成流行歌曲结构，接着将分段和弦进行扩展为逐小节的和声上下文。节奏在乐句级别上以八分音符网格生成，音高则在和声、音域和音程转换约束下采样。强拍位置倾向于采用和弦音，而弱拍位置倾向于采用音阶音。每个生成示例均由一条符号化旋律轨道和结构化元数据组成，用于描述速度、调性、音域、段落边界和和声。

A.1 全局条件

全局条件包括速度类别（慢速、中速或快速）、调性、人声音域（男声或女声）、歌曲结构和和弦进行。合成预训练所采用的实现枚举了 12 个大调，固定使用四四拍，并依据专家规则采样其他音乐条件。

A.2 歌曲结构

结构家族为前奏–主歌–[预副歌]–副歌– 尾奏。前奏和尾奏固定为八小节，可选的预副歌也固定为八小节。主歌和副歌的候选模式见表1。三种种主歌模式、两种预副歌选项和三种副歌模式的笛卡尔积可生成18种结构模板。实现中还会根据可配置的歌曲时长范围筛选候选结构；报告中的实现将该范围设为60–180秒。

前奏	主歌	预副歌	副歌	尾奏
I8	A8A8	-	B8B8	O8
	A16	C8	B16	
	A16A16		B16B16	

表1：结构模板族。I、A、C、B 和 O 分别表示前奏、主歌、预副歌、副歌和尾奏。数字表示小节数。

A.3 和声模板

和弦进行从音乐人精选的元素级模板中选取，这些模板按段落类型、速度类别和段落长度进行组织。对于长度为 16 小节或以上的段落，合成器会选择适当的较长和弦进行，或重复使用较短的和弦进行。副歌和弦进行也会在前奏和尾奏中重复使用。如表2所示，专家合成器包含 201 个和弦进行模板。实现将这些模板存储在 pop_chordprog.py 中。

段落	速度类别	8小节	16小节
主歌	Slow	20	9
主歌	Mid	23	4
主歌	Fast	8	9
预副歌	Slow	13	–
预副歌	Mid	23	–
预副歌	Fast	14	–
副歌	Slow	9	11
副歌	Mid	15	16
副歌	Fast	4	23
总计		**201**	

表2：按段落类型、速度类别和段落长度划分的和弦进行模板数量。破折号（–）表示副歌前段落不提供16小节模板，因为在我们的结构设计中，预副歌固定为8小节。

A.4 节奏生成

节奏以八分音符网格生成。每个段落包含四个乐句：8小节段落采用2小节乐句，16小节段落采用4小节乐句。乐句的音节数从 $[5, 13]$（2小节乐句）和 $[8, 18]$（4小节乐句）中采样，然后分配给诸如弱起、完整、不完整和结尾等节奏组成部分。每种组件类型都有独特的节奏特征：

- 弱起：开头可能有较长的休止符（弱起）
- 完整：密集，不含休止符（完整小节）
- 不完整：开头可能有短休止符（轻微间隙）
- 结尾：末尾可能有休止符（终止空间）

对于每种组件类型和音节数，我们预先枚举所有八分音符节奏型，在组件时值内放置

指定数量的起音，同时遵循组件的节奏特征。这些节奏型存储在查找表中。在节奏生成过程中，确定组件类型序列和音节分配后，系统会从相应集合中随机选择一个匹配的节奏型。

这些节奏模式和分配规则的实现位于 pop_rhythm.py 中。

A.5 音高生成

我们分两个阶段为每个音符生成音高：首先识别强拍位置和弱拍位置，然后为强音符采样音高，接着为弱音符采样音高。

音高生成首先标记强拍位置。每个强拍是每个小节的下拍，此外还包括每个乐句的最后一个音符，以确保乐句终止闭合。然而，落在强拍上的音符并不总是被视为强音符：倚音被视为弱音符，而其后的音符则获得强音符的地位。

随后，合成器会在音域、和声和音程转换约束下采样强拍音高。其余的弱音符组则依据调性/调式、和弦、周围强拍音高和组长度，使用手工制定规则进行填充。这一两阶段流程生成的旋律具有和声连贯性，并且适于歌唱。完整实现包括所有转换表和规则逻辑，详见 pop_melody.py。

B 数据集构建与 RWC 准备

B.1 合成语料库

对于合成语料库，专家合成器枚举十二个大调、两种人声音域（男声和女声）以及三种速度类别（慢速、中速和快速），并采用固定的四四拍。对于每种配置，流水线会从相应的精选池中选取结构模板和段落级和弦进行，随后生成配对的旋律 MIDI 文件和元数据。同一合成流程生成了用于模型选择实验的 10 万、30 万、100 万和 300 万数据集。

B.2 真实流行音乐适配示例

真实流行音乐适配使用 RWC 流行音乐数据库。我们排除两首非四四拍曲目，并手动准备了 98 个仅包含旋律、覆盖前奏至间奏部分的半首歌曲 MIDI 示例。准备好的数据附有调性/调式、速度类别、段落和和弦注释。准备过程中，我们将调性/调式统一为 C 大调，按起始小节记录段落边界，并将与节拍对齐的和弦标签存储为 MIDI 标记。前奏和尾奏被视为固定

在此标注协议中无需必备和弦标签的八小节区域。

预处理会将音符量化到八分音符网格，拆分跨越小节边界的音符，并在音域约束下，通过在12个调之间移调来扩增处理后的旋律。该流程基于98条处理后的源旋律，生成了847个分词后的适配样例。

C 分词、建模与训练

如图3所示，训练流程分为事件分词、合成预训练，以及结合合成回放的真实流行音乐适配。

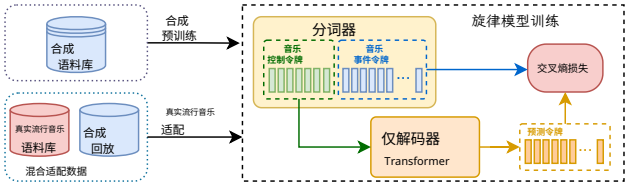


图3：旋律模型训练流程，包括使用合成回放进行合成预训练和真实流行音乐适配。

C.1 事件表示与条件生成

如图4所示，旋律及其元数据被线性化为事件词元序列。序列以调性/调式、速度类别、人声音域和歌曲结构的全局控制词元开头。随后，每个小节包含段落和和弦信息，之后是编码音高、起始位置及时值的音符事件。这种表示方式支持通过仅解码器 Transformer 进行条件自回归生成。

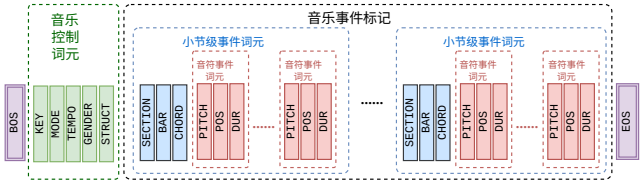


图4：基于事件的分词，先是控制词元前缀，随后为小节级和音符级事件。

设 $x_i = (x_{i,0}, \ldots, x_{i,T_i})$ 为经过分词的旋律，$x_{i,0}$ = 序列起始符。模型通过教师强制下一词元交叉熵进行预训练：

$$\mathcal{L}_{\text{pretrain}}(\theta) = -\sum_{i}\sum_{t=1}^{T_i} \log p_\theta(x_{i,t} \mid x_{i,<t}). \qquad (1)$$

由此得到的合成预训练检查点称为 Synth-Base。合成预训练涵盖多样的调性、速度、结构、人声音域和和弦进行，旨在学习专家合成器的可控行为。

C.2 借助合成 R 进行真实流行音乐适配 eplay

合成预训练提供了广泛的音乐条件覆盖范围，但基于规则的系统的输出分布仍可能与自然创作的流行旋律有所不同。直接在小型真实流行音乐语料库上微调合成预训练模型可能会导致过拟合，并削弱其遵循条件的能力。为保留预训练期间学到的部分控制能力，适配会将所有真实流行音乐样例与从合成语料库中抽样得到的

重放子集结合起来。设 $\mathcal{D}_{\text{real}}$ 和 $\mathcal{D}_{\text{syn}}$ 分别为真实集和合成集。对于重放比例 ρ,

$$\left|\mathcal{D}_{\text{syn}}^{(\rho)}\right| = \rho\,|\mathcal{D}_{\text{real}}|\,, \qquad \mathcal{D}_{\text{adapt}}^{(\rho)} = \mathcal{D}_{\text{real}} \cup \mathcal{D}_{\text{syn}}^{(\rho)}. \tag{2}$$

最终报告的 Real-FT 使用了 真实流行音乐:合成重放比例为 *2:1*，对应于 $\rho = 1/2$。两个训练阶段使用相同的分词器和事件词汇表。

C.3 架构与超参数

Item	预训练	适配
最大上下文长度	1,300	1,300
隐藏层维度	512	512
解码器层数	6	6
注意力头数	8	8
丢弃率	0.1	0.1
优化器	AdamW	AdamW
学习率	3×10^{-4}	3×10^{-5}
批量大小	16	16
训练轮数	20	20
训练集/验证集/测试集	8:1:1	8:1:1
训练样例	仅合成数据	真实流行音乐 + 回放

表 3：两个训练阶段的训练配置。

C.4 推理与小节级和弦条件控制

在神经网络推理过程中，用户指定的全局条件用于初始化自回归解码。为确保生成所选和弦进行，在生成音符事件之前，会在每个小节边界注入对应的和弦标记，如图 5 所示。相比之下，基于规则的 Expert 后端采用自身的和声与音高生成规则。

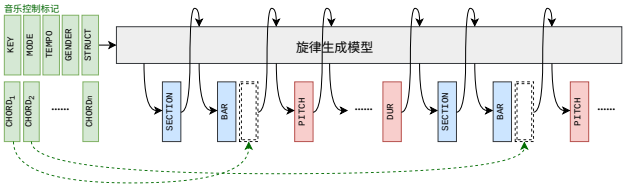


图 5：在线注入和弦：在每个小节边界插入和弦标记，以引导后续的音符事件解码。

D 评估方案与客观结果 D.1 提示网格与指标

比较涉及三种生成系统：基于规则的 Expert、合成数据预训练的 Synth-Base，以及经过适配的 Real-FT。

每个模型根据评估网格生成 4,320 条旋律。该网格包含 12 个调、3 种速度类别、2 种人声音域、2 种结构类别、6 种和弦进行，每种条件采样 5 次；三个系统合计生成 12,960 条输出。

采用四项音乐学上可解释的指标：

- **和声遵循率（HAR）：** 和声对齐度，结合强拍位置的和弦音占比与弱拍位置的音阶音占比；
- **CAD（终止式遵循度）：** 乐句末尾或段落末尾的音符是否符合指定的终止式目标；
- **RGL（副歌–主歌音域提升）：** 副歌与主歌之间的音域对比；
- **SMR（级进旋律运动率）：** 主歌与副歌段落中局部级进旋律运动所占的比例。

HAR 和 CAD 是控制遵循指标，数值越高表示越符合专家定义的约束。RGL 和 SMR 则用于刻画旋律风格，应结合处理后的真实流行音乐参考集进行解读。要完整复现结果，需采用评估代码中包含的具体阈值和聚合流程。

D.2 合成预训练规模

合成数据规模研究比较了 1万、3万、10 万和 30万个样例。表4 重现了规模消融实验中记录的 HAR 和 CAD 数值。从 1 万 增至 10 万时，这两项指标均显著提升；而 30 万 的设置并未进一步提升 HAR，CAD 则有所下降。图 6 展示了这一趋势，因此选择 Synth-100k 作为 Synth-Base。

系统	HAR ↑	CAD ↑
专家参考	1.000	1.000
合成-1万	0.914	0.543
合成-3万	0.991	0.834
Synth-100k	**0.998**	**0.893**
Synth-300k	0.998	0.828

表4：合成预训练规模消融。

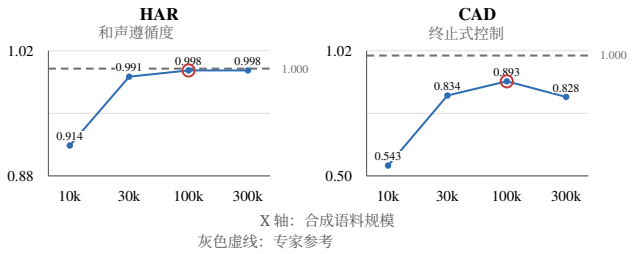


图 6：合成预训练规模消融实验在 HAR 和 CAD 指标上的结果。选中的 10 万设置被选作 Synth-Base 检查点。

D.3 回放混合选择

适应实验改变了真实流行数据与合成回放比例。表 5 重现了混合消融实验中的数值。仅使用真实数据进行微调会显著削弱 CAD，而加入合成回放则能恢复 CAD 性能。更高的回放比例会进一步增强 HAR 和 CAD，但也会使

模型更接近合成预训练行为。图7展示了这一权衡；由于 2:1 的比例在保留较强和声与节奏可控性的同时，最接近真实流行音乐参考集，因此我们为最终的 Real-FT 设置选择了该比例。

真实流行音乐：合成回放	HAR ↑	CAD ↑
1:0（仅真实数据）	0.917	0.320
2:1（最终 Real-FT）	**0.983**	**0.834**
1:1	0.987	0.844
1:2	0.994	0.890
专家参考	1.000	1.000
真实流行音乐参考集	0.898	0.798

表 5：通过合成回放进行真实流行音乐适配。2:1 混合是报告中的最终检查点；其他行是消融设置。

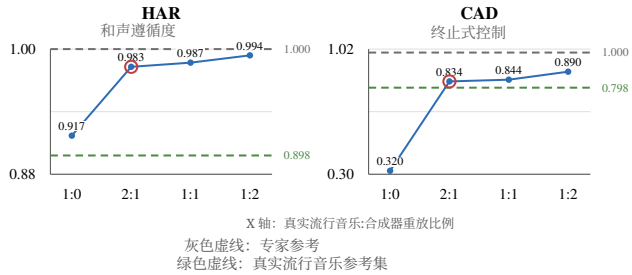


图7：HAR 和 CAD 指标上的真实流行音乐适配混合消融。高亮显示的 2:1 混合被选作最终 Real-FT 设置。

D.4 ThreeFinalBackends 的比较

客观评估的最终检查点数值见表6和图8。结果展示了两种预期行为。首先，Synth-Base 在全部四项指标上都与专家参考高度一致，表明 Transformer 通过大规模合成预训练，成功学习了专家合成器所编码的音乐特征。其次，经过真实流行音乐适配后，Real-FT 在全部四项指标上都持续向真实流行音乐参考集靠拢，这表明微调有效地使合成预训练模型适配真实流行音乐旋律特征，同时保留了合成预训练过程中获得的大部分可控性。

E 人工评估

在人工评估中，30名参与者使用5点MOS量表和三选一偏好任务，对来自6个提示组和3个系统的18首生成歌曲进行了评估。如表7所示，Real-FT取得了最高的平均意见分（3.548）和偏好率（48.3%）。

系统	HAR ↑	CAD ↑	RGL	SMR
专家	1.000	1.000	5.992	0.634
Synth-Base（10 万）	0.998	0.893	5.881	0.612
Real-FT（2:1）	0.983	0.834	4.946	0.663
RWC 真实流行音乐参考集	0.898	0.798	3.231	0.724

表6：三个最终系统和描述性真实流行音乐参考的客观评估。RGL 和 SMR 并非单调质量评分。

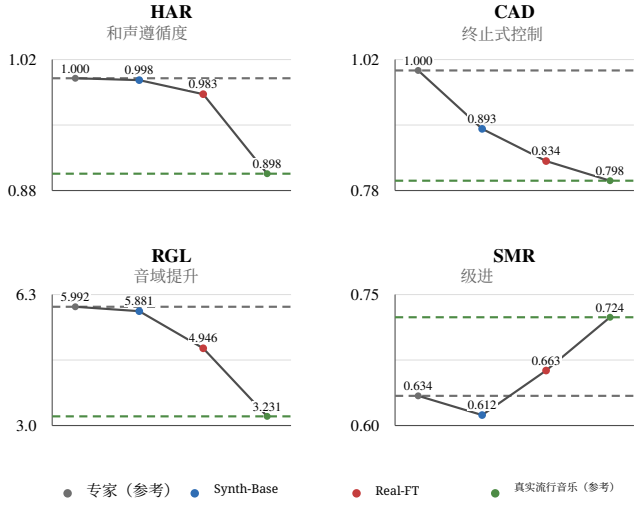


图8：专家、Synth-Base 和 Real-FT 在四项符号指标上的客观比较。虚线表示专家和真实流行音乐参考值。

系统	平均意见分 ↑	三选最佳 ↑
专家	3.231 ± 0.059	25.6%
Synth-Base	3.244 ± 0.064	26.1%
Real-FT	$\mathbf{3.548 \pm 0.057}$	**48.3%**

表7：对30位参与者和18首生成歌曲开展的人体研究。MOS表示五分制下的平均意见分±SE；三选最佳表示各提示组内的偏好胜率。

F 相似性检查详情

对于每个生成的旋律，我们提取移调不变的音程 n-gram，并从选定的RWC参考集中检索其最近匹配项。图9展示了三个系统生成的12,960条旋律的最近邻相似度分布。对得分最高的配对进行人工检查后，未发现明显的近重复旋律。

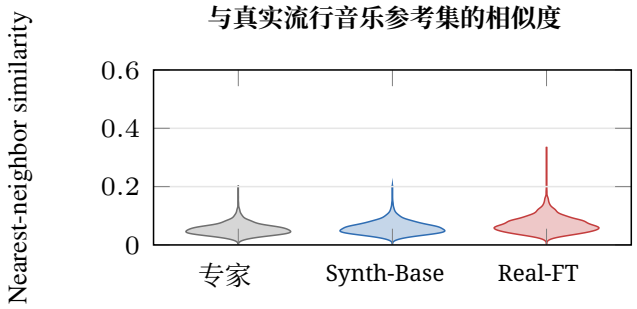


图 9：三个系统生成的旋律与 RWC 参考集之间的最近邻相似度。

尽管该分析基于规模有限的真实流行音乐参考语料库，但它为评估旋律相似性提供了实证证据，并为评估真实应用中的潜在近似复制风险提供了有用参考。

G　演示与局限性

G.1 创作者与研究者工作流

图 10 和图 11 展示了已部署平台的两种工作流。创作者模式支持交互式旋律生成、试听和导出，而研究者模式则支持批量生成带标签的MIDI–元数据对。Node.js 前端通过推理 API 与生成后端进行通信。

G.2 范围与局限性

专家合成器依赖预定义的音乐结构、经过整理的和弦模板以及简化的乐句级规则。尽管这些约束有助于提升可控性和可复现性，但也限制了风格多样性。此外，真实流行音乐适配使用的语料库相对较小，因此对音乐风格和结构的覆盖范围有限。最后，神经模型侧重于符号旋律生成，而非完整歌曲制作。尽管该平台为试听提供伴奏，但歌词、演唱人声和经学习得到的多轨编曲仍不在当前建模范围内。

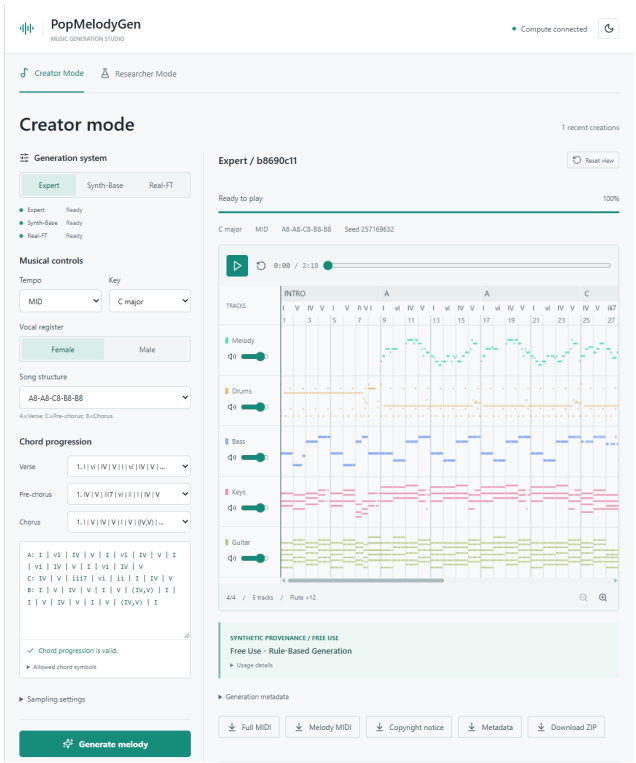


图10：用于可控旋律生成、试听和导出的创作者模式界面。

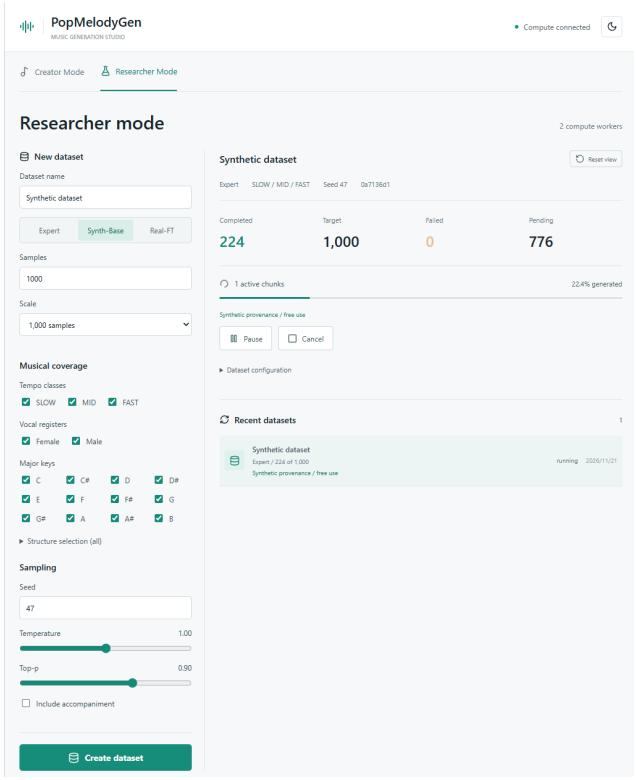


图11：用于批量生成带注释的符号旋律的研究者模式界面。