



(12) 发明专利申请

(10) 申请公布号 CN 118379976 A

(43) 申请公布日 2024. 07. 23

(21) 申请号 202410318748.X

(22) 申请日 2024.03.20

(71) 申请人 浙江大学

地址 310058 浙江省杭州市西湖区余杭塘路866号

申请人 OPPO广东移动通信有限公司

(72) 发明人 张克俊 刘浩轩 王子豪 洪浩荣
冯佑玮 俞佳兴 许云飞

(74) 专利代理机构 杭州天勤知识产权代理有限公司 33224

专利代理师 彭剑

(51) Int. Cl.

G10H 1/00 (2006.01)

G06F 16/438 (2019.01)

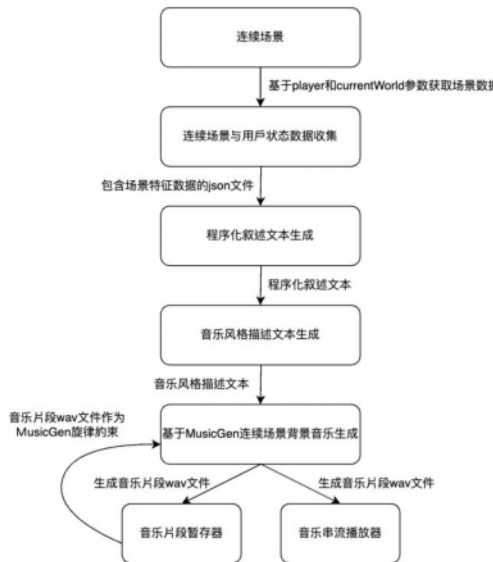
权利要求书2页 说明书6页 附图2页

(54) 发明名称

一种融入环境感知和个性化体验的连续场景背景音乐生成方法

(57) 摘要

本发明公开了一种融入环境感知和个性化体验的连续场景背景音乐生成方法,包括:(1)在连续场景中,获取用户当前时段所处的环境数据和用户状态数据;(2)将包含环境和用户状态信息的特征数据转化为音乐风格描述文本;(3)将音乐风格描述文本输入音频大模型MusicGen生成背景音乐片段,作为旋律约束;(4)再次通过步骤(1)和(2)生成新一时段的音乐风格描述文本输入MusicGen,同时在步骤(3)中存储的旋律约束下生成新一时段的背景音乐片段;(5)循环重复步骤(4),最终将每一时段的背景音乐片段连接得到对应连续场景和用户状态变换的背景音乐。利用本发明,能够根据连续的场景信息生成符合场景风格的背景音乐。



1. 一种融入环境感知和个性化体验的连续场景背景音乐生成方法, 其特征在于, 包括以下步骤:

(1) 在连续场景中, 获取用户当前时段所处的环境数据和用户状态数据, 输出包含环境和用户状态信息的特征数据;

(2) 将包含环境和用户状态信息的特征数据转化为音乐风格描述文本;

(3) 将得到的音乐风格描述文本输入音频大模型MusicGen, 通过构造提示词模版接收音乐风格描述文本生成背景音乐片段, 以暂存器模块存储该背景音乐片段作为旋律约束;

(4) 再次通过步骤(1)和(2)生成新一时段的音乐风格描述文本并输入音频大模型MusicGen, 同时将步骤(3)中暂存器模块存储的背景音乐片段作为新一时段MusicGen的旋律约束, 在旋律约束下生成新一时段的背景音乐片段;

(5) 循环重复步骤(4), 最终将每一时段的背景音乐片段连接得到对应连续场景和用户状态变换的背景音乐。

2. 根据权利要求1所述的融入环境感知和个性化体验的连续场景背景音乐生成方法, 其特征在于, 步骤(1)中, 连续场景为游戏中场景, 环境数据包含生态环境、时间、天气、温度; 用户状态数据包含生命值、饱食值、触碰水、跑步、潜行、碰触岩浆、着火、接触地面、在铁路上。

3. 根据权利要求1所述的融入环境感知和个性化体验的连续场景背景音乐生成方法, 其特征在于, 步骤(2)的具体过程为:

将环境数据和用户状态数据通过拼接方式转化为LLM提示词, 输入到LLaMA 2中进行文本润色和优化, 得到一段包括环境数据和用户状态数据的场景描述文本; 再将场景描述文本输入通过全量微调后的开源大语言模型, 得到音乐风格描述文本。

4. 根据权利要求1所述的融入环境感知和个性化体验的连续场景背景音乐生成方法, 其特征在于, 步骤(3)中, 先通过剪枝技术对音频大模型MusicGen进行模型轻量化, 再将得到的音乐风格描述文本输入轻量化的MusicGen, 通过构造针对音乐风格描述文本的提示词模板使MusicGen理解从音乐风格描述文本到生成当前时间片段对应场景的背景音乐的基本规则。

5. 根据权利要求1所述的融入环境感知和个性化体验的连续场景背景音乐生成方法, 其特征在于, 步骤(3)中, 针对用户个性化生成背景音乐, 增加用户设置参数模块, 用户通过选择特征统计方式和多久时间进行一次统计输出到MusicGen。

6. 根据权利要求1所述的融入环境感知和个性化体验的连续场景背景音乐生成方法, 其特征在于, 步骤(4)中, 增加旋律约束模块, 在旋律约束下生成新一时段的背景音乐片段时, 调整Top-p、Top-k、Temperature参数调节生成新一时段的伴奏与上个时段音乐的相似性和连贯性;

其中, Top-p用于动态设置音频Token候选列表的大小, 将可能性之和不超过特定值的top Token列入音乐生成候选名单;

Top-k用于允许其他高分音频Token有机会被选中, 从而使采样加入了随机性;

Temperature用于调整随机从音频大模型MusicGen中抽样的程度, 控制每次音乐生成时相同的音乐风格描述文本是否会产生不同的音乐输出。

7. 根据权利要求6所述的融入环境感知和个性化体验的连续场景背景音乐生成方法,

其特征在于,Top-p被预设0.75。

8.根据权利要求6所述的融入环境感知和个性化体验的连续场景背景音乐生成方法,其特征在于,Temperature为0时,每次音乐生成时相同的音乐风格描述文本将始终产生相同的输出;Temperature越高生成的伴奏随机性越大。

一种融入环境感知和个性化体验的连续场景背景音乐生成方法

技术领域

[0001] 本发明属于程序化音乐生成领域,尤其是涉及一种融入环境感知和个性化体验的连续场景背景音乐生成方法。

背景技术

[0002] 程序化音乐生成是指利用计算机程序或算法来创造音乐的过程,从早期简单音效的创作,到无重复地拼接人类音乐作品,再到如今创作具有复杂和声和旋律结构的音乐。根据程序化音乐生成的发展历程,可以将方法分为转换算法和生成算法两大类。

[0003] 在早期阶段,转换算法是主要的方法,它需要提供复杂的结构模版作为音乐生成的条件。例如,将录制好的音乐划分成多个不同的层,每个层包含音乐的不同功能。当需要改变音乐的听感时,可以在特定的时间点增加或移除对应的层以达到想要的效果。这种方法的背后是对已有音乐材料进行变换,创造出富有创意和独特性的新音乐。算法能够处理音频的不同层次,调整音乐的结构和听感。

[0004] 随着机器学习算法的发展和大规模音乐数据的出现,生成算法逐渐成为主流。生成算法可以在生成音乐的同时,赋予音乐以一定的旋律结构。常见的生成算法可以仅依靠模型自身的性能实现,或是融入知识进行知识增强的音乐生成。前者通常对模型的结构进行改进,让模型能够学习长期依赖关系,如修改注意力机制等。后者则在数据的表征中引入音乐特征,让模型显式地学习音乐元素之间的关系。生成算法的关键特点是通过算法本身创造独特的音乐,而不仅仅是在已有音乐基础上进行修改。

[0005] 如公开号为CN116072098A的中国专利文献公开了一种音频信号生成方法,包括:根据输入文本信息,确定输入文本信息的文本特征;根据文本特征和噪声相关信息,确定输入文本信息的音频表示;以及根据音频表示,生成与输入文本信息相对应的目标音频信号。

[0006] 公开号为CN109260693A的中国专利文献公开了一种运动音乐的生成方法,包括:获取用户运动数据,并根据所述用户运动数据的运动向量确定当前的运动节奏;将已确定的所述运动节奏输入预设的音频生成器,以通过所述音频生成器的音乐学习框架生成与所述运动节奏对应的运动音乐;播放生成的运动音乐。

[0007] 近年来,生成算法被广泛应用于连续场景背景音乐的生成。在这类音乐的应用中,生成算法能够根据场景中的实时事件和用户行为动态生成音乐,为整体体验增添更多交互性和个性化。然而,目前的工作只能根据不同场景连续播放对应的音乐,忽视了这些音乐片段之间的过渡,导致生成的音乐难以保持一致性和协调性。

发明内容

[0008] 本发明提供了一种融入环境感知和个性化体验的连续场景背景音乐生成方法,能够根据连续的场景信息,生成符合场景风格的背景音乐。

[0009] 一种融入环境感知和个性化体验的连续场景背景音乐生成方法,包括以下步骤:

[0010] (1) 在连续场景中,获取用户当前时段所处的环境数据和用户状态数据,输出包含环境和用户状态信息的特征数据;

[0011] (2) 将包含环境和用户状态信息的特征数据转化为音乐风格描述文本;

[0012] (3) 将得到的音乐风格描述文本输入音频大模型MusicGen,通过构造提示词模版接收音乐风格描述文本生成背景音乐片段,以暂存器模块存储该背景音乐片段作为旋律约束;

[0013] (4) 再次通过步骤(1)和(2)生成新一时段的音乐风格描述文本并输入音频大模型MusicGen,同时将步骤(3)中暂存器模块存储的背景音乐片段作为新一时段MusicGen的旋律约束,在旋律约束下生成新一时段的背景音乐片段;

[0014] (5) 循环重复步骤(4),最终将每一时段的背景音乐片段连接得到对应连续场景和用户状态变换的背景音乐。

[0015] 进一步地,步骤(1)中,连续场景为游戏中场景,环境数据包含生态环境、时间、天气、温度等;用户状态数据包含生命值、饱食值、触碰水、跑步、潜行、碰触岩浆、着火、接触地面、在铁路上等。

[0016] 步骤(2)的具体过程为:

[0017] 将环境数据和用户状态数据通过拼接方式转化为LLM提示词,输入到LLaMA 2中进行文本润色和优化,得到一段包括环境数据和用户状态数据的场景描述文本;再将场景描述文本输入通过全量微调后的开源大语言模型,得到音乐风格描述文本。

[0018] 步骤(3)中,先通过剪枝技术对音频大模型MusicGen进行模型轻量化,再将得到的音乐风格描述文本输入轻量化的MusicGen,通过构造针对音乐风格描述文本的提示词模板使MusicGen理解从音乐风格描述文本到生成当前时间片段对应场景的背景音乐的基本规则。

[0019] 步骤(3)中,针对用户个性化生成背景音乐,增加用户设置参数模块,用户可通过选择特征统计方式和多久时间进行一次统计输出到MusicGen。

[0020] 步骤(4)中,增加旋律约束模块,在旋律约束下生成新一时段的背景音乐片段时,调整Top-p、Top-k、Temperature参数调节生成新一时段的伴奏与上个时段音乐的相似性和连贯性;

[0021] 其中,Top-p用于动态设置音频Token候选列表的大小,将可能性之和不超过特定值的top Token列入音乐生成候选名单;Top-p被预设为较高的值(0.75),目的是限制可能被采样的低概率的长度。

[0022] Top-k用于允许其他高分音频Token有机会被选中,从而使采样加入了随机性。

[0023] Temperature用于调整随机从音频大模型MusicGen中抽样的程度,控制每次音乐生成时相同的音乐风格描述文本是否会产生不同的音乐输出。

[0024] Temperature为0时,每次音乐生成时相同的音乐风格描述文本将始终产生相同的输出;Temperature越高生成的伴奏随机性越大。

[0025] 与现有技术相比,本发明具有以下有益效果:

[0026] 1. 本发明让融入知识的大语言模型理解从场景数据转换到音乐风格描述文本的基本规则,从而让生成的音乐更符合实际连续场景变换以及人类对音乐审美的基本需求。

[0027] 2. 本发明使用“音频大模型”框架MusicGen生成的连续场景背景音乐更具鲁棒性。

[0028] 3.本发明使用分段生成连续场景背景音乐,并加入旋律约束,解决了MusicGen长时音乐生成和连贯性问题。

附图说明

[0029] 图1为本发明实施例中一种融入环境感知和个性化体验的连续场景背景音乐生成方法流程图;

[0030] 图2是本发明实施例中场景数据收集模块运行流程图;

[0031] 图3是本发明实施例中音乐风格文本生成流程图;

[0032] 图4是本发明实施例中连续场景背景音乐生成整体架构图。

具体实施方式

[0033] 下面结合附图和实施例对本发明做进一步详细描述,需要指出的是,以下所述实施例旨在便于对本发明的理解,而对其不起任何限定作用。

[0034] 如图1所示,一种融入环境感知和个性化体验的连续场景背景音乐生成方法,包括以下步骤:

[0035] 步骤1,场景数据收集。

[0036] 在预设的时间片段内收集连续场景地图中的场景元素和用户信息,计算该时间段的数据特征并以规范的格式输出。

[0037] 以连续场景游戏Minecraft为例,借助Minecraft在开源社区完整模块开发工具库(Fabric),能够方便的调用Minecraft游戏场景源代码的众多函数。

[0038] 通过游戏场景数据分析,可以获取的数据有两个大方向(环境数据、用户状态数据),从各个大方向中也可以细分为几点:环境数据(生态环境、时间、天气、温度)、用户状态数据(生命值、饱食值、触碰水、跑步、潜行、碰触岩浆、着火、接触地面、在铁路上)。一切关于游戏场景数据获取的代码都基于两个游戏源代码中提供的参数player以及currentWorld获得,前者存有当前用户状态的数值,后者提供当前用户所处的世界(地图)数据。

[0039] 在环境数据中,当前的生态环境可以根据用户在场景中的位置,获取用户当前所属生态环境。在Minecraft中,存在多种生态环境(Biome),其描述方式如同: Birch Forest, Deep Frozen Ocean, Bamboo Jungle等。在代码实现上,经由一个BioMap将类型RegistryKey<Biome>转换为可读的字符串,在本程序中也实现记录了所有RegistryKey<Biome>对应的生态名称Map。

[0040] Minecraft的世界中,场景的当前时间是用long作为数据类型在currentWorld存储,其值位于0-24000这个区间,且不同的区间分别代表了不同的场景时间概念。经过分析发现,我们可以通过if判断区间对应到正确的当前时间状态。

[0041] 具体来说,(0-450)为日出、(450-5000)为白天、(5000-7000)为中午、(7000-11617)为下午、(11617-13702)为日落、(13702-18000)为晚上、(18000-22200)为午夜、(22200-24000)为日出。

[0042] 场景中天气的生成是根据用户坐标所在地的降雨类型以及当前世界的降雨类型两项参数得到,在Minecraft中一共会出现的天气类型共五种:晴朗、阴天、下雨、下雪、雷雨。其中下雨类型表示当前生态若真实世界正在下雨,表现在场景中是下雨还是下雪。

[0043] 代码设计中的降雨枚举类型一共有三类:RAIN(下雨)、SNOW(下雪)和NONE(不下雨也不下雪)。

[0044] 以(降雨枚举类型、是否下雨、是否打雷)举例,(NONE、NO、YES/NO)对应天气类型晴朗、(NONE、YES、YES/NO)对应天气类型阴天、(RAIN、YES/NO、YES)对应天气类型雷雨、(RAIN、YES、NO)对应天气类型下雨、(RAIN、NO、NO)对应天气类型晴朗、(SNOW、YES、YES/NO)对应天气类型下雪、(SNOW、YES、YES/NO)对应天气类型晴朗。

[0045] 当前温度可根据用户当前位置,加上当前生态性质,判定是否为寒冷。例如,在Desert(沙漠生态)中,就不存在任一个坐标会判定为寒冷。

[0046] 表述用户当前状态的数值统一出九项进行表述,包括:生命值、饱食值、碰触水(雨淋、碰到水、没碰到水)、跑步、潜行、碰触岩浆、着火、接触地面、在铁路上。

[0047] 在场景中,将用户的饱食值、饱食值定义为一个在0-20区间的数值,可经由player参数直接得出,程序中将数值翻译为可阅读的字符串。

[0048] 是否碰触到水,表示的是用户当前与水相关的状态。在Minecraft中,用户与水相关的状态有三种:被雨淋、触碰到地图上的水源(接触到河流、海洋等)、没有碰触到水。

[0049] 如图2所示,在这一步骤设计了四大模块。

[0050] 循环获取数据主线程模块:在连续场景的场景及用户状态数据收集程序担任主线程,在程序运行期间进入循环,在每一次的循环中调用其它模块进行场景数据的读取,写入磁盘中。

[0051] 数据获取模块:为连续场景的场景及用户状态数据收集程序的核心模块,其目标为获取当前场景内用户所处坐标的环境数据以及用户当前自身的状态数据。

[0052] 数据计算/写入模块:模块实现的功能包括,统计一定时间区间内所有获取到的数值的分布占比,将统计后的数据输出至终端,以及写入磁盘中。

[0053] 用户设置模块:为了方便进行实验的调试,本程序设计了一个用户设置参数的模块,可由用户决定多久时间进行一次统计输出,以及一些其余的设置。

[0054] 以下举例描述如何计算获取的数据。假设在间隔1秒的区间中,经过循环获取到了三个数据,包括环境数据(生态环境、时间、天气、温度)、用户状态数据(生命值、饱食值、触碰到水、跑步、潜行、碰触岩浆、着火、接触地面、在铁路上)。

[0055] 统计以上三个数据各自的十二项子数据占比状况,可以得到下表1的最后数据:

[0056] 表1

[0057]

属性	选项一	选项一占比	选项二	选项二占比
生态环境	Forest	2/3	River	1/3
时间	Noon	2/3	Day	1/3
天气	Clear	2/3	Snow	1/3
温度	Cold	1/1	/	/
生命值	Full	1/1	/	/
饱食值	Full	1/1	/	/
触碰到水	Not Touching	2/3	Touching	1/3
跑步	Not Sprinting	1/1	/	/
潜行	Not Sneaking	2/3	Sneaking	1/3

碰触岩浆	Not In Lava	1/1	/	/
着火	Not On Fire	1/1	/	/
接触地面	On Ground	2/3	Not On Ground	1/3
在铁路上	Not On Rail	1/1	/	/

[0058] 步骤2,音乐风格描述文本生成。

[0059] 将步骤1生成的含有连续场景及用户状态特征数据转化为音乐风格描述文本,具体流程如图3所示。

[0060] (2-1) 程序化叙事生成 (Procedural Narrative Generation),这是程序化内容生成 (Procedural Content Generation for Games,PCG-G) 的一个子概念。

[0061] 将场景数据以及用户角色数据通过拼接等方式转化为LLM提示词,输入到LLaMA 2中进行文本润色和优化,得到一段包括用户状态信息、场景信息的较为丰富的文本描述。

[0062] 其中场景信息包括生态环境、时间、天气、温度等,用户信息包括生命值、饱食值、触碰水、跑步、潜行、碰触岩浆、着火、接触地面、在铁路上等。这些信息被拼接成提示词,由LLaMA 2进行整合梳理,通过进行细节的补充,最终形成一段场景及用户信息的文本描述。

[0063] (2-2) 音乐风格描述文本生成 (Music styledescription text generation),将程序化叙事文本的连续场景及用户描述转换为音乐风格描述文本。

[0064] 通过全量微调开源的ChatGLM-6B、LLaMA 2等大语言模型构建数据量充足且高质量的场景描述-音乐描述训练集9000多条。使用LLaMA-13B生成多样化的训练data-pair,然后通过数据质检、数据筛选等方式构建高质量训练数据集,通过全量微调增进大语言模型在场景描述-音乐描述任务中的性能,从而实现这一项工作。

[0065] 通过步骤2,搭建了一个场景数据-场景描述文本-音乐描述文本的基线,完成了将连续场景数据,转化为音频大模型MusicGen可以识别的音乐风格描述文本。

[0066] 步骤3,连续场景背景音乐生成

[0067] 将步骤2中生成的音乐风格描述文本输入音频大模型MusicGen,通过构造提示词模版接收音乐风格描述文本生成背景音乐,以暂存器模块存取音乐片段作为旋律约束。

[0068] 为进一步考虑背景音乐在实时性生成上的需求,通过剪枝技术对MusicGen进行模型轻量化,并通过构造针对音乐风格描述文本的提示词模板使音频大模型理解从音乐风格描述文本到生成当前时间片端对应场景的背景音乐的基本规则。

[0069] 针对用户个性化生成背景音乐,加入了用户设置参数模块,通过用户配置文件outputDataConfig.json进行控制,用户可选择特征统计方式和多久时间进行一次统计输出到MusicGen。

[0070] 步骤4,重复步骤1和步骤2生成新的音乐风格描述,并将步骤3中生成的背景音乐片段作为新一时段的旋律约束,即可生成时且对应连续场景和用户状态变换的背景音乐。

[0071] 通过步骤1收集新一时间片段的场景数据,并通过步骤2生成新的音乐风格文本做为新的提示词,将步骤3中存储的背景音乐作为新时间片段音频大模型的旋律约束输入新的音乐风格描述文本生成新的背景音乐。循环重复步骤4,生成新的音乐风格文本和旋律约束,便可生成时且符合连续场景场景信息风格的背景音乐。

[0072] 在旋律约束方面本发明加入了相关模块,可通过调整Top-p、Top-k、Temperature

等参数调节生成对应时间片段的伴奏与上个时间片段音乐的相似性和连贯性。

[0073] Top-p动态设置音频Token候选列表的大小。将可能性之和不超过特定值的top Token列入音乐生成候选名单。Top-p被预设为较高的值(0.75),目的是限制可能被采样的低概率的长度。

[0074] Top-k允许其他高分音频Token有机会被选中。这种采样引入的随机性有助于生成音乐生成的质量。

[0075] Temperature用于调整随机从音频生成模型中抽样的程度,因此每次音乐生成时相同的音乐风格描述文本可能会产生不同的音乐输出,Temperature为0时将始终产生相同的输出,Temperature越高生成的伴奏随机性越大。

[0076] 音频大模型多采用Transformer架构,生成长时音乐出现音乐风格前后不一致,产生相关的可控性问题。MusicGen使用文本作为提示词,其MusicGen-melody版本除文本输入外,也支持旋律输入作为生成条件,约束每个时间片段音乐生成上的风格差异。针对音乐风格描述文本的完善,构造了提示词模版,补全音乐配齐选择和调性、和弦进行。

[0077] 使用旋律输入作为约束条件实现了不同时间片段前后音乐的柔和过度问题。具体实施方式为将上一时间片段生成的背景音乐,作为当前时间段的旋律约束。如此反复生成长时、具有明显结构且柔和过度的背景音乐并通过串流播放器播放,如图4所示。

[0078] 以上所述的实施例对本发明的技术方案和有益效果进行了详细说明,应理解的是以上所述仅为本发明的具体实施例,并不用于限制本发明,凡在本发明的原则范围内所做的任何修改、补充和等同替换,均应包含在本发明的保护范围之内。

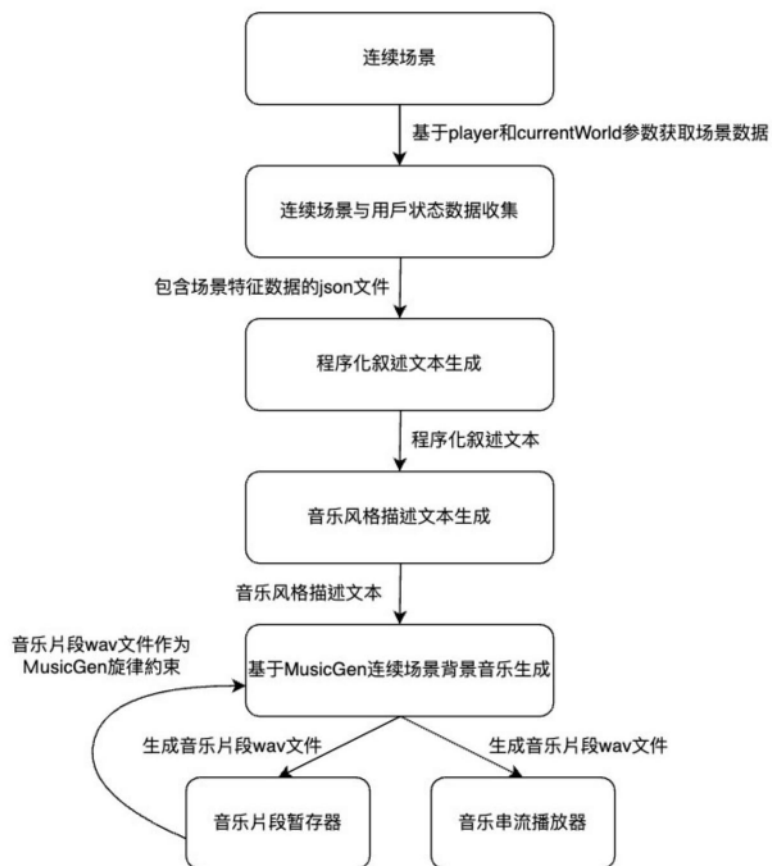


图1

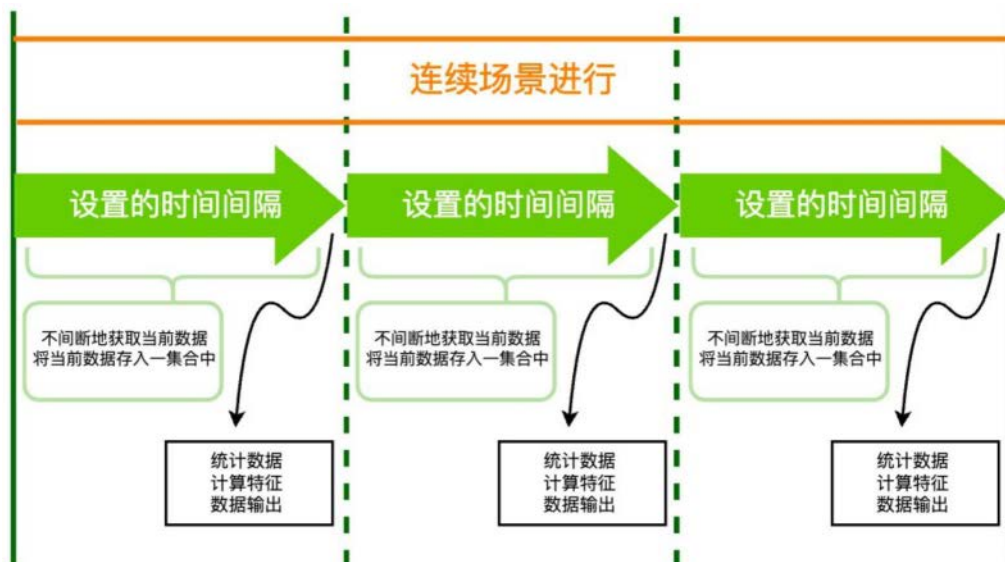


图2

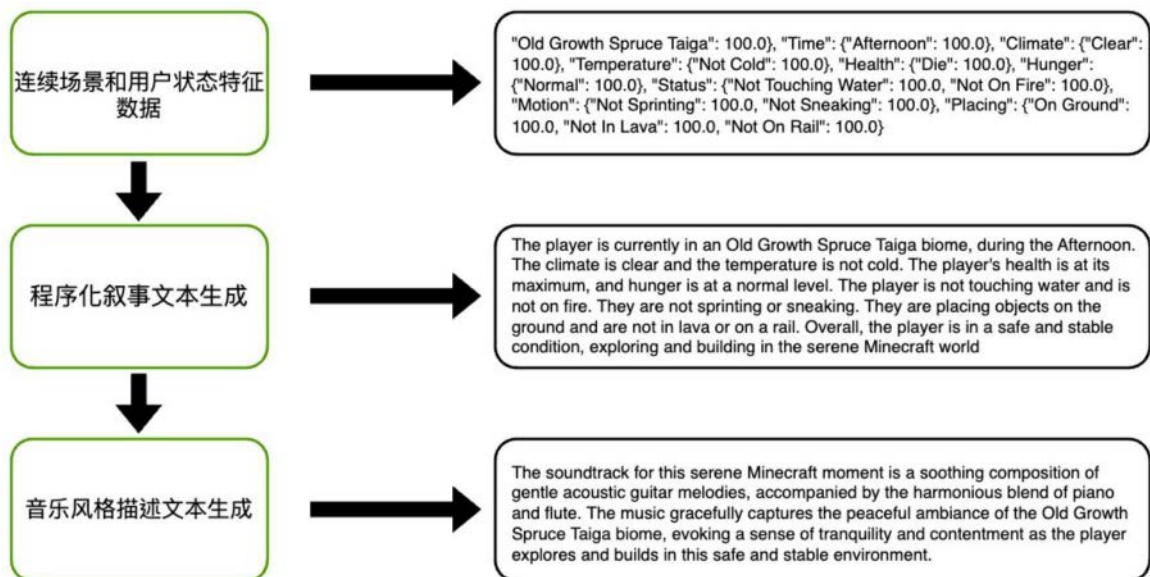


图3

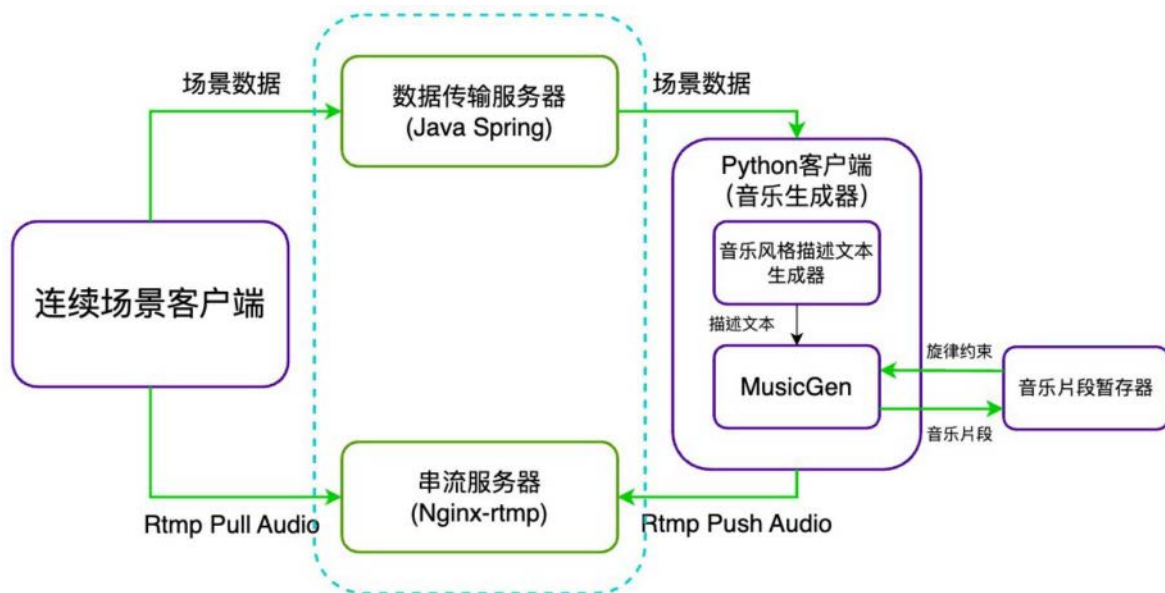


图4