

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2023年2月9日 (09.02.2023)



(10) 国际公布号
WO 2023/010949 A1

- (51) 国际专利分类号:
G10H 1/00 (2006.01) *G06F 16/68* (2019.01)
- (21) 国际申请号: PCT/CN2022/093923
- (22) 国际申请日: 2022年5月19日 (19.05.2022)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202110876809.0 2021年7月31日 (31.07.2021) CN
- (71) 申请人: 华为技术有限公司 (HUAWEI TECHNOLOGIES CO., LTD.) [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。
- (72) 发明人: 王卓(WANG, Zhuo); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129

(CN)。王萌(WANG, Meng); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。杜春晖(DU, Chunhui); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。范泛(FAN, Fan); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。刘经纬(LIU, Jingwei); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。李贤胜(LI, Xiansheng); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。徐德著(XU, Dezh); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(74) 代理人: 北京中博世达专利商标代理有限公司 (BEIJING ZBSD PATENT & TRADEMARK AGENT)

(54) Title: METHOD AND APPARATUS FOR PROCESSING AUDIO DATA

(54) 发明名称: 一种音频数据的处理方法及装置

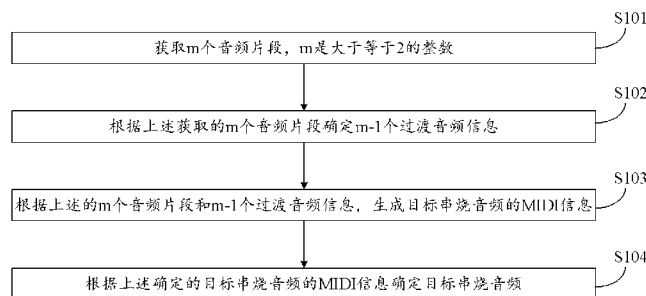


图 3

- S101 Acquire m audio clips, wherein m is an integer greater than or equal to 2
- S102 Determine m-1 pieces of transition audio information according to the obtained m audio clips
- S103 Generate MIDI information of target burn audio according to the m audio clips and the m-1 pieces of transition audio information
- S104 Determine the target burn audio according to determined MIDI information of the target burn audio

(57) Abstract: A method and apparatus (120) for processing audio data, which relate to the technical field of multimedia. The method comprises: obtaining m (m is an integer greater than or equal to 2) audio clips (S101); and determining m-1 pieces of transition audio information according to the m audio clips (S102); generating MIDI information of target burn audio according to the m audio clips and the m-1 pieces of transition audio information (S103); and determining the target burn audio according to determined MIDI information of the target burn audio (S104). The m-1 pieces of transition audio information are used to connect the m audio clips. For first transition audio information among the m-1 pieces of transition audio information, the first transition audio information is used to connect a first audio clip and a second audio clip that are ordered in sequence among the m audio clips. Ordering refers to the burning order of the m audio clips.



WO 2023/010949 A1

LTD.); 中国北京市海淀区交大东路31号11
号楼8层, Beijing 100044 (CN)。

- (81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。
- (84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告(条约第21条(3))。

(57) 摘要: 一种音频数据的处理方法及装置(120), 涉及多媒体技术领域。方法包括: 获取 m (m 是大于等于2的整数)个音频片段(S101); 根据 m 个音频片段确定 $m-1$ 个过渡音频信息(S102); 根据 m 个音频片段和 $m-1$ 个过渡音频信息, 生成目标串烧音频的MIDI信息(S103); 根据确定的目标串烧音频的MIDI信息确定目标串烧音频(S104)。其中, $m-1$ 个过渡音频信息用于衔接 m 个音频片段。对于 $m-1$ 个过渡音频信息中的第一过渡音频信息而言, 第一过渡音频信息用于衔接 m 个音频片段中排序连续的第一音频片段和第二音频片段。排序是指 m 个音频片段的串烧顺序。

一种音频数据的处理方法及装置

5 本申请要求于2021年07月31日提交国家知识产权局、申请号为202110876809.0、申请名称为“一种音频数据的处理方法及装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及多媒体技术领域，尤其涉及一种音频数据的处理方法及装置。

背景技术

10 随着数字流媒体音乐的广泛传播和应用，以及随着手机、平板、耳机等无线终端设备的普及和发展，聆听音乐成为大多数人在不同环境下的生活必需品，并且人们对音乐的多样化需求也日趋增长。例如，除了从音频开头至结尾全部聆听的听歌体验之外，人们对于由多个音频片段组合而成的音频，即串烧音频的需求也日益增多。

15 目前在实现音频串烧时，通常仅能实现对相似度较高的至少两个音频片段进行拼接串烧。然而这种方式实现的串烧音频的风格往往都比较单一。

因此，如何获得更丰富、更具多样性的串烧音频，是现有技术中亟待解决的技术问题。

发明内容

20 本申请提供了一种音频数据的处理方法及装置，基于该方法可以获得更丰富、更具多样性的串烧音频。

为达上述目的，本申请提供如下技术方案：

25 第一方面，本申请提供一种音频数据的处理方法，该方法包括：获取 m 个音频片段， m 是大于等于 2 的整数。根据该 m 个音频片段确定 $m-1$ 个过渡音频信息。根据该 m 个音频片段和该 $m-1$ 个过渡音频信息，生成目标串烧音频。其中，该 $m-1$ 个过渡音频信息用于衔接该 m 个音频片段。其中，对于该 $m-1$ 个过渡音频信息中的第一过渡音频信息而言，该第一过渡音频信息用于衔接 m 个音频片段中排序连续的第一音频片段和第二音频片段。这里， m 个音频片段的排序为该 m 个音频片段的串烧顺序。

30 可以看出，基于本申请提供的方法对多个音频片段进行串烧时，可以生成全新的用于衔接该多个音频片段的过渡音频信息。因此本申请提供的方法无需考虑用于串烧得到目标串烧音频的多个音频片段的相似度。也就是说，通过本申请实施例提供的方法能够获得更丰富、更具多样性的串烧音频。

35 在一种可能的设计方式中，上述根据该 m 个音频片段确定 $m-1$ 个过渡音频信息包括：根据上述第一音频片段的第一信息和上述第二音频片段的第二信息，确定第一过渡音频信息。其中，该第一信息包括第一音频片段的乐器数字接口（musical instrument digital interface, MIDI）信息和音频特征信息，该第二信息包括第二音频片段的 MIDI 信息和音频特征信息，该第一过渡音频信息包括第一过渡音频信息对应的第一过渡音频的乐器数字接口 MIDI 信息。

在另一种可能的设计方式中，上述的音频特征信息包括音频片段的主旋律轨位置

信息、风格标签、情感标签、节奏信息、节拍信息、或调号信息中的至少一种。

5 基于该两种可能的设计，本申请所提供方法生成的用于衔接多个音频片段的过渡音频信息是在 MIDI 域进行的。由于音频的 MIDI 信息是音频最原始的表现形式，其记录有音频的音符音高、音符力度、音符时长等信息。因此，相比在时域上直接对多个音频片段进行串烧，本申请所提供的方法在 MIDI 域中对音频片段的 MIDI 信息进行处理后所生成的用于衔接两个音频片段的过渡音频信息是基于音频乐理生成的。这样的话，基于该过渡音频信息获得的串烧音频在听觉上更加流畅自然。并且，在 MIDI 域处理数据也更有利于串烧音频在后期渲染时的灵活性和一致性。

10 在另一种可能的设计方式中，上述根据上述第一音频片段的第一信息和上述第二音频片段的第二信息，确定第一过渡音频信息包括：根据第一音频片段的第一信息、第二音频片段的第二信息以及预设神经网络模型，确定第一过渡音频信息。

15 在另一种可能的设计方式中，当在目标串烧音频中，第一音频片段位于第二音频片段之前，则：上述第一过渡音频信息是基于用于表征第一过渡音频信息的特征向量确定的，该第一过渡音频信息的特征向量是基于第一向量和第二向量确定的。其中，第一向量是根据第一信息在第一音频片段的时序末端生成的特征向量，第二向量是根据第二信息在第二音频片段的时序始端生成的特征向量。

20 基于该两种可能的设计，本申请所提供方法在 MIDI 域通过神经网络模型对多个音频片段的 MIDI 信息进行处理，从而得到了用于衔接多个音频片段的过渡音频的 MIDI 信息。这样的话，基于神经网络极强的学习能力，本申请在 MIDI 域基于对音频乐理的学习而得到的用于衔接多个音频片段的过渡音频信息，能够更自然流畅的衔接多个音频片段。

25 在另一种可能的设计方式中，上述获取 m 个音频片段包括：响应于用户的第一操作，确定 k 个目标音频。从该 k 个目标音频中提取 m 个音频片段。其中， $2 \leq k \leq m$ ，且 k 是整数。

30 基于该可能的设计，本申请可以基于用户的意愿对用户选择的多个目标音频中的音频片段进行串烧，从而提高了用户的体验。

在另一种可能的设计方式中，在上述根据该 m 个音频片段确定 $m-1$ 个过渡音频信息之前，上述方法还包括：确定该 m 个音频片段的串烧顺序。

35 在另一种可能的设计方式中，上述方法还包括：响应于用户的第二操作，重新确定上述 m 个音频片段的串烧顺序。根据重新确定的串烧顺序和 m 个音频片段，重新确定 $m-1$ 个过渡音频信息。根据重新确定的 $m-1$ 个过渡音频信息和 m 个音频片段，重新生成目标串烧音频。

根据该可能的设计，通过本申请所提供方法生成目标串烧音频后，当用户对该目标串烧音频不满意时，可以向终端设备输入第二操作，以使终端设备响应该第二操作，对生成该目标串烧音频的 m 个音频片段的串烧顺序进行调整，并重新生成新的目标串烧音频。这样的话，通过设备和用户之间的反复交互，可以令用户获得满意的目标串烧音频，从而提高了用户的体验。

在另一种可能的设计方式中，上述方法还包括：响应于用户的第三操作，对上述目标串烧音频进行渲染。

在另一种可能的设计方式中，上述方法还包括：输出上述目标串烧音频。

第二方面，本申请提供了一种音频数据的处理装置。

在一种可能的设计方式中，该处理装置用于执行上述第一方面提供的任一种方法。本申请可以根据上述第一方面提供的任一种方法，对该处理装置进行功能模块的划分。

5 例如，可以对应各个功能划分各个功能模块，也可以将两个或两个以上的功能集成在一个处理模块中。示例性的，本申请可以按照功能将该处理装置划分为获取单元、确定单元和生成单元等。上述划分的各个功能模块执行的可能的技术方案和有益效果的描述均可以参考上述第一方面或其相应的可能的设计提供的技术方案，此处不再赘述。

10 在另一种可能的设计中，该处理装置包括：一个或多个处理器和传输接口，该一个或多个处理器通过该传输接口接收或发送数据，该一个或多个处理器被配置为调用存储在存储器中的程序指令，以使得该处理装置执行如第一方面及其任一种可能的设计方式提供的任一种方法。

15 第三方面，本申请提供了一种计算机可读存储介质，该计算机可读存储介质包括程序指令，当程序指令在计算机或处理器上运行时，使得计算机或处理器执行第一方面中的任一种可能的实现方式提供的任一种方法。

第四方面，本申请提供了一种计算机程序产品，当其在音频数据的处理装置上运行时，使得第一方面中的任一种可能的实现方式提供的任一种方法被执行。

20 第五方面，本申请提供给了一种音频数据的处理系统，该系统包括终端设备和服务器。其中，该终端设备用于执行第一方面中的任一种可能的实现方式提供的任一种方法中与用户进行交互的方法部分，该服务器用于执行第一方面中的任一种可能的实现方式提供的任一种方法中生成目标串烧音频的方法部分。

可以理解的是，上述提供的任一种装置、计算机存储介质、计算机程序产品或系统等均可以应用于上文所提供的对应的方法，因此，其所能达到的有益效果可参考对应的方法中的有益效果，此处不再赘述。

25 在本申请中，上述音频数据的处理装置的名字对设备或功能模块本身不构成限定，在实际实现中，这些设备或功能模块可以以其他名称出现。只要各个设备或功能模块的功能和本申请类似，属于本申请权利要求及其等同技术的范围之内。

附图说明

图 1 为本申请实施例提供的一种手机硬件结构示意图；

30 图 2 为本申请实施例提供的一种音频数据的处理系统的示意图；

图 3 为本申请实施例提供的一种音频数据的处理方法的流程示意图；

图 4 为本申请实施例提供的一种用户在音频剪辑应用程序的音频剪辑界面上输入的第一操作的示意图；

35 图 5 为本申请实施例提供的另一种用户在音频剪辑应用程序的音频剪辑界面上输入的第一操作的示意图；

图 6 为本申请实施例提供的又一种用户在音频剪辑应用程序的音频剪辑界面上输入的第一操作的示意图；

图 7 为本申请实施例提供的又一种用户在音频剪辑应用程序的音频剪辑界面上输入的第一操作的示意图；

图 8 为本申请实施例提供的一种预设神经网络模型的结构示意图；

图 9 为本申请实施例提供的另一种预设神经网络模型的结构示意图；

图 10 为本申请实施例提供的第二操作的示意图；

5 图 11 为本申请实施例提供的一种对目标串烧音频的 MIDI 信息进行渲染以及输出的示意图；

图 12 为本申请实施例提供的一种音频数据的处理装置的结构示意图；

图 13 为本申请实施例提供的用于承载计算机程序产品的信号承载介质的结构示意图。

具体实施方式

10 为了更清楚的理解本申请实施例，下面对本申请实施例中涉及的部分术语或技术进行说明：

1) 、乐器数字接口 (musical instrument digital interface, MIDI)

MIDI 是编曲界应用最广泛的音乐标准格式，可以称为“计算机能理解的乐谱”。

15 MIDI 用音符的数字控制信号来记录音乐。即，MIDI 传输的不是声音信号本身，而是音符及控制参数等指令。这些指令可以指示 MIDI 设备演奏音乐，例如指示 MIDI 设备以指令中所指示的音量大小来演奏某个音符等。MIDI 所传输的指令可以被统一表示成 MIDI 消息或 MIDI 信息。

通常，MIDI 信息可以以图谱形式呈现，也可以以数据流的形式呈现。当 MIDI 信息以图谱形式呈现时，可以简称为 MIDI 谱。

20 对于一首以波形音频文件格式 (waveform audio file format, WAV) 存储的时域上的音乐波形信号，当该音乐波形信号被转录为 MIDI 信息时，即可以将该 MIDI 信息理解为该音乐波形信号在 MIDI 域的表达形式。其中，时域是指时间域。

可以理解，MIDI 信息一般可以包含多个声轨，且每一个声轨均标注有音符的起始位置、结束位置、音符的音高以及音符的力度信息等。其中，一个声轨用于表征一种 25 乐器声音/人声。可以理解，一首完整的通过 MIDI 信息表达的音乐的大小往往只有几十千字节 (kilobyte, KB)，但是却可以包含数十条声轨。

当前，几乎所有的现代音乐都是通过 MIDI 信息加上音色库来制作合成。其中，音色库 (或称为采样库) 包括人类所能听到和创造出来的各种声音，例如包括各种乐器的演奏、各种人声的演唱、念白，以及各种自然、人造声音的录音等。

30 2) 、隐空间

对于神经网络的某个中间层输出的特征，该特征所表示的原始数据经过若干神经网络层变换后的空间即可称为隐空间。通常，隐空间的维度一般小于原始数据的空间 35 维度。

隐空间也可以被理解为原始数据特征的某种抽象的提取和表示。

3) 、序列模型网络、双向序列模型网络

通常，输入或者输出中包含有序列数据的模型可以叫做序列模型。序列模型通常用于处理具有某种顺序关系数据。用于构建序列模型的神经网络可以称为序列模型网络。

其中，常见的序列模型网络包括循环神经网络 (recurrent neural network, RNN)、

长短期记忆网络(long short-term memory, LSTM)、门控循环单元(gated recurrent unit, GRU)、转换器(transformer)等。

应理解,序列模型网络在时间 t 预测得到的预测结果,通常是基于对输入数据在时间 t 之前的数据进行学习后得到的。

5 而在一些情况中,当序列模型网络在时间 t 预测得到的预测结果,是基于对输入数据在时间 t 之前的数据进行学习,以及对输入数据在时间 t 之后的数据进行学习后得到的,这种情况下,该序列网络模型称为双向序列模型网络。可以看出,双向序列网络模型在对输入数据进行预测时,结合了输入数据中时刻 t 的上下文信息来进行结果预测。

10 应理解,双向序列模型可以在输入数据的任意时刻预测得到预测结果。

其中,常见的双向序列模型网络包括(双向)循环神经网络((bidirectional) recurrent neural networks, (Bi-) RNN)、(双向)长短时记忆网络((bidirectional) long short-term memory, (Bi-) LSTM)、(双向)门控循环单元((bidirectional) gate recurrent unit, (Bi-) GRU)、转换器(transformer)等。

15 4)、其他术语

在本申请实施例中,“示例性的”或者“例如”等词用于表示作例子、例证或说明。本申请实施例中描述为“示例性的”或者“例如”的任何实施例或设计方案不应被解释为比其它实施例或设计方案更优选或更具优势。确切而言,使用“示例性的”或者“例如”等词旨在以具体方式呈现相关概念。

20 在本申请的实施例中,术语“第一”、“第二”仅用于描述目的,而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量。由此,限定有“第一”、“第二”的特征可以明示或者隐含地包括一个或者更多个该特征。本申请中的术语“至少一个”的含义是指一个或多个。在本申请的描述中,除非另有说明,术语“多个”的含义是两个或两个以上。

25 还应理解,本文中所使用的术语“和/或”是指并且涵盖相关联的所列出的项目中的一个或多个项目的任何和全部可能的组合。术语“和/或”,是一种描述关联对象的关联关系,表示可以存在三种关系,例如,A 和/或 B,可以表示:单独存在 A,同时存在 A 和 B,单独存在 B 这三种情况。另外,本申请中的字符“/”,一般表示前后关联对象是一种“或”的关系。

30 还应理解,在本申请的各个实施例中,各个过程的序号的大小并不意味着执行顺序的先后,各过程的执行顺序应以其功能和内在逻辑确定,而不应对本申请实施例的实施过程构成任何限定。

应理解,根据 A 确定 B 并不意味着仅仅根据 A 确定 B,还可以根据 A 和/或其它信息确定 B。

35 还应理解,术语“包括”(也称“includes”、“including”、“comprises”和/或“comprising”)当在本说明书中使用指定存在所陈述的特征、整数、步骤、操作、元素、和/或部件,但是并不排除存在或添加一个或多个其他特征、整数、步骤、操作、元素、部件、和/或其分组。

应理解,本申请实施例下文中所述的串烧,是指从不同音频中提取出多段音频段,

并将该多个音频段按照预设顺序串联组合在一起的过程，该预设顺序即为该多段音频段的串烧顺序。

5 本申请实施例提供一种音频数据的处理方法，该方法先根据预先获取的 m 个音频片段确定出 $m-1$ 个过渡音频信息，然后基于该 $m-1$ 个过渡音频信息对该 m 个音频片段进行衔接，从而生成了 m 个音频片段被串烧后的目标串烧音频。其中，一个过渡音频信息用于衔接串烧顺序相邻的两个音频片段。

由于通过该方法对多个音频片段进行串烧时，无需考虑该多个音频片段之间的特征相似度。因此，通过本申请实施例方法可以获得丰富的、风格多样的串烧音频。

10 本申请实施例还提供一种音频数据的处理装置，该处理装置可以是终端设备。其中，该终端设备可以是手机、平板电脑、笔记本电脑、个人数字助理（personal digital assistant, PDA）、上网本、可穿戴电子设备（例如智能手表、智能眼镜）等便携式设备，也可以是台式计算机、智能电视、车载等设备，还可以是其他任一能够实现本申请实施例的终端设备，本申请对此不作限定。

15 以上述处理装置是手机为例，参考图 1，图 1 示出了本申请实施例提供的一种手机 10 硬件结构示意图。如图 1 所示，手机 10 可以包括处理器 110，内部存储器 120，外部存储器接口 130，摄像头 140，触摸屏 150，音频模块 160 以及通信模块 170 等。

20 处理器 110 可以包括一个或多个处理单元，例如：处理器 110 可以包括应用处理器（application processor, AP），调制解调处理器，图形处理器（graphics processing unit, GPU），图像信号处理器（image signal processor, ISP），控制器，存储器，视频编解码器，数字信号处理器（digital signal processor, DSP），基带处理器，和/或神经网络处理器（neural-network processing unit, NPU）等。其中，不同的处理单元可以是独立的器件，也可以集成在一个或多个处理器中。

其中，控制器可以是手机 10 的神经中枢和指挥中心。控制器可以根据指令操作码和时序信号，产生操作控制信号，完成取指令和执行指令的控制。

25 NPU 为神经网络（neural-network, NN）计算处理器，通过借鉴生物神经网络结构，例如借鉴人脑神经元之间传递模式，对输入信息快速处理，还可以不断的自学习。通过 NPU 可以实现手机 10 的智能认知等应用，例如：文字识别、图像识别、人脸识别等。

30 处理器 110 中还可以设置存储器，用于存储指令和数据。在一些实施例中，处理器 110 中的存储器为高速缓冲存储器。该存储器可以保存处理器 110 刚用过或循环使用的指令或数据。如果处理器 110 需要再次使用该指令或数据，可从所述存储器中直接调用。避免了重复存取，减少了处理器 110 的等待时间，因而提高了系统的效率。

35 在一些实施例中，处理器 110 可以包括一个或多个接口。接口可以包括集成电路（inter-integrated circuit, I2C）接口，集成电路内置音频（inter-integrated circuit sound, I2S）接口，脉冲编码调制（pulse code modulation, PCM）接口，通用异步收发传输器（universal asynchronous receiver/transmitter, UART）接口，移动产业处理器接口（mobile industry processor interface, MIPI），通用输入输出（general-purpose input/output, GPIO）接口，用户标识模块（subscriber identity module, SIM）接口，和/或通用串行总线（universal serial bus, USB）接口等。

I2C 接口是一种双向同步串行总线，包括一根串行数据线（serial data line, SDA）和一根串行时钟线（derail clock line, SCL）。I2S 接口可以用于音频通信。PCM 接口也可以用于音频通信，将模拟信号抽样，量化和编码。UART 接口是一种通用串行数据总线，用于异步通信。该总线可以为双向通信总线。它将要传输的数据在串行通信与并行通信之间转换。MIPI 接口可以被用于连接处理器 110 与摄像头 140、触摸屏 150 等外围器件。MIPI 接口包括摄像头串行接口（camera serial interface, CSI），触摸屏串行接口（display serial interface, DSI）等。GPIO 接口可以通过软件配置。GPIO 接口可以被配置为控制信号，也可被配置为数据信号。

内部存储器 120，可以用于存储计算机可执行程序代码，所述可执行程序代码包括指令。处理器 110 通过运行存储在内部存储器 120 的指令，从而执行手机 10 的各种功能应用以及数据处理，例如执行本申请实施例提供的音频数据的处理方法。

外部存储器接口 130 可以用于连接外部存储卡，例如 Micro SD 卡，实现扩展手机 10 的存储能力。外部存储卡通过外部存储器接口 130 与处理器 110 通信，实现数据存储功能。例如将音乐，视频、图片等文件保存在外部存储卡中。

摄像头 140，用于获取静态图像或视频。物体通过镜头生成光学图像投射到感光元件。数字信号处理器用于处理数字信号，除了可以处理数字图像信号，还可以处理其他数字信号。应理解，手机 10 可以包括 n 个摄像头 140，n 是正整数。

触摸屏 150，用于手机 10 和用户之间的交互。触摸屏 150 包括显示面板 151 和触摸板 152。其中，显示面板 151 用于显示文字、图像、视频等。触摸板 152 用于输入用户的指令。

音频模块 160 用于将数字音频信息转换成模拟音频信号输出，也用于将模拟音频输入转换为数字音频信号。音频模块 160 可以包括扬声器 161、受话器 162、麦克风 163 以及耳机接口 164 中的至少一个。

其中，扬声器 161，也称“喇叭”，用于将音频电信号转换为声音信号。受话器 162，也称“听筒”，用于将音频电信号转换成声音信号。麦克风 163，也称“话筒”，“传声器”，用于将声音信号转换为电信号。耳机接口 164 用于连接有线耳机。耳机接口 164 可以是 USB 接口，也可以是 3.2mm 的开放移动电子设备平台（open mobile terminal platform, OMTP）标准接口，美国蜂窝电信工业协会（cellular telecommunications industry association of the USA, CTIA）标准接口。

这样，手机 10 即可以通过音频模块 160 中的扬声器 161，受话器 162，麦克风 163，耳机接口 164，以及应用处理器等实现音频功能。例如用户的语音输入、语音/音乐的播放等。

通信模块 170，用于实现手机 10 的通信功能。具体的，通信模块 170 可以通过天线、移动通信模块，无线通信模块，调制解调处理器以及基带处理器等实现。

天线用于发射和接收电磁波信号。手机 10 中的每个天线可用于覆盖单个或多个通信频带。不同的天线还可以复用，以提高天线的利用率。例如：可以将用于移动通信模块的天线 1 可以复用为无线局域网的分集天线。在另外一些实施例中，天线可以和调谐开关结合使用。

移动通信模块可以提供应用在手机 10 上的包括 2G/3G/4G/5G 等无线通信的解决

方案。移动通信模块可以包括至少一个滤波器，开关，功率放大器，低噪声放大器（low noise amplifier, LNA）等。移动通信模块可以由天线接收电磁波，并对接收的电磁波进行滤波，放大等处理，传送至调制解调处理器进行解调。移动通信模块还可以对经调制解调处理器调制后的信号放大，经天线转为电磁波辐射出去。在一些实施例中，移动通信模块的至少部分功能模块可以被设置于处理器 110 中。在一些实施例中，移动通信模块的至少部分功能模块可以与处理器 110 的至少部分模块被设置在同一个器件中。调制解调处理器可以包括调制器和解调器。

无线通信模块可以提供应用在手机 10 上的包括无线局域网（wireless local area networks, WLAN）（如无线保真（wireless fidelity, Wi-Fi）网络），蓝牙（bluetooth, BT），GNSS，调频（frequency modulation, FM），近距离无线通信技术（near field communication, NFC），红外技术（infrared, IR）等无线通信的解决方案。无线通信模块可以是集成至少一个通信处理模块的一个或多个器件。无线通信模块经由天线接收电磁波，将电磁波信号调频以及滤波处理，将处理后的信号发送到处理器 110。无线通信模块还可以从处理器 110 接收待发送的信号，对其进行调频，放大，经天线转为电磁波辐射出去。

示例性的，本申请实施例中的 GNSS 可以包括：全球卫星定位系统（global positioning system, GPS），全球导航卫星系统（global navigation satellite system, GLONASS），北斗卫星导航系统（beidou navigation satellite system, BDS），准天顶卫星系统（quasi-zenith satellite system, QZSS）和/或星基增强系统（satellite based augmentation systems, SBAS）等。

可以理解的是，本申请实施例示意的结构并不构成对手机 10 的具体限定。在本申请另一些实施例中，手机 10 可以包括比图示更多或更少的部件，或者组合某些部件，或者拆分某些部件，或者不同的部件布置。图示的部件可以以硬件，软件或软件和硬件的组合实现。

需要说明的是，当上述的处理装置为终端设备，上述音频数据的处理方法可以通过安装在终端设备上的应用程序（application, App）实现。其中，该 App 具有剪辑音频的功能。作为示例，该 App 可以是音乐剪辑 App 等。

其中，该 App 可以是具有人工介入功能的 App。这里，人工介入是指 App 可以接收用户输入的指令，并能够响应用户输入的指令。也就是说，该 App 可以和用户进行交互。该 App 可以包括用于与用户进行交互的交互界面，该交互界面通过终端设备的显示屏（例如图 1 所示的显示面板 151）显示。

应理解，如果终端设备包括有触摸屏（例如图 1 所示的触摸屏 150），则用户可以通过操作终端设备的触摸屏（例如操作图 1 所示的触摸板 152）实现和 App 的交互。如果终端设备不包括触摸屏（例如终端设备是普通的台式计算机），则用户可以通过终端设备的鼠标、键盘等输入输出器件和 App 进行交互。

还应理解，上述的 App 可以是安装在终端设备中的嵌入式应用程序（即终端设备的系统应用），也可以是可下载应用程序。

其中，嵌入式应用程序是设备（如手机）的操作系统提供的应用程序。例如，该嵌入式应用程序可以时手机出厂是提供的音乐应用程序等。可下载应用程序是一个可

以提供自己的通信连接的应用程序，该可下载应用程序是可以预先安装在设备中的 App，或者可以是由用户下载并安装在设备中的第三方 App，例如，该可下载应用程序可以是音乐剪辑 App，本申请实施例对此不作具体限定。

5 还需说明的是，上述的处理装置也可以是服务器。这种情况下，本申请实施例还提供一种音频数据的处理系统。其中，该处理系统包括服务器和终端设备，该服务器和终端设备之间可以通过有线或无线的方式连接通信。

如图 2 所示，图 2 示出了本申请实施例提供的一种处理系统 20 的示意图。处理系统 20 包括终端设备 21 和服务器 22。其中，终端设备 21 可以通过客户端 App（例如音频剪辑的客户端 App）与用户进行交互，例如接收用户输入的指令，并将接收到的指令传输至服务器 22。然后，服务器 22 用于基于从终端设备 21 接收到的指令执行本申请实施例提供的音频数据的处理方法，并将生成的目标串烧音频的 MIDI 信息和/或目标串烧音频发送至终端设备 21。这样，终端设备 21 即可接收到服务器 22 发送的目标串烧音频的 MIDI 信息和/或目标串烧音频，并通过音频模块向用户播放该目标串烧音频，和/或，通过显示屏向用户显示该目标串烧音频的 MIDI 谱。对此不作限定。

15 下面结合附图，对本申请实施例提供的音频数据的处理方法进行详细描述。

参考图 3，图 3 示出了本申请实施例提供的一种音频数据的处理方法的流程示意图。该方法由上文所述的音频数据的处理装置执行。下面，以上文所述的音频数据的处理装置是终端设备为例，该方法可以包括以下步骤：

S101、获取 m 个音频片段， m 是大于等于 2 的整数。

20 其中，该 m 个音频片段包括不同音频中的片段。

具体的，终端设备可以先确定出 k 个目标音频，然后再从 k 个目标音频中提取 m 个音频片段。其中， $2 \leq k \leq m$ ，且 k 是整数。

也就是说，终端设备可以从一个目标音频中提取至少一个音频片段。

S102、根据上述获取的 m 个音频片段确定 $m-1$ 个过渡音频信息。

25 其中，该 $m-1$ 个过渡音频信息用于衔接该 m 个音频片段。对于该 $m-1$ 个过渡音频信息中的第一过渡音频信息，该第一过渡音频信息是用于衔接终端设备所获取的 m 个音频片段中排序连续的第一音频片段和第二音频片段的过渡音频信息。应理解，这里所述排序是 m 个音频片段的串烧顺序的排序。应理解，该串烧顺序是终端设备预先确定的。

30 具体的，终端设备可以根据预设神经网络模型和在 S101 获取的 m 个音频片段的信息，确定出 $m-1$ 个过渡音频信息。这里，该过渡音频信息即为过渡音频段的 MIDI 信息。

S103、根据上述的 m 个音频片段和 $m-1$ 个过渡音频信息，生成目标串烧音频的 MIDI 信息。

35 终端设备通过上述的 $m-1$ 个过渡音频信息（即过渡音频段的 MIDI 信息）将 m 个音频片段的 MIDI 信息连接起来，即生成了目标串烧音频的 MIDI 信息。

S104、根据上述确定的目标串烧音频的 MIDI 信息确定目标串烧音频。

可选的，终端设备还可以基于目标串烧音频的 MIDI 信息确定出目标串烧音频后，输出该目标串烧音频。

通过上述本申请实施例提供的方法，终端设备可以基于 m 个音频片段，生成用于衔接该 m 个音频片段的 $m-1$ 个过渡音频信息。这样，通过该 $m-1$ 个过渡音频信息即可将 m 个音频片段的 MIDI 信息衔接起来，从而得到 m 个音频片段被串烧后的目标串烧音频的 MIDI 信息。这样的话，当终端设备将目标串烧音频的 MIDI 信息转换为音频格式，即得到了目标串烧音频。可以看出，通过这种方式对多个音频片段进行串烧时，终端可以生成全新的用于衔接该多个音频片段的过渡音频段，因此本申请实施例提供的方法无需考虑用于得到目标串烧音频的音频片段的相似度。也就是说，通过本申请实施例提供的方法能够获得更丰富、更具多样性的串烧音频。

并且，本申请实施例提供的方法生成过渡音频信息的过程是在 MIDI 域中进行的，由于音频的 MIDI 信息是音频最原始的表现形式，其记录有音频的音符音高、音符力度、音符时长等信息。因此，相比在时域上直接对多个音频片段进行串烧，本申请实施例提供的方法能够基于音频乐理生成用于衔接两个音频片段的过渡音频信息，且基于该过渡音频信息获得的串烧音频在听觉上更加流畅自然。并且，在 MIDI 域处理数据更有利于串烧音频在后期渲染时的灵活性和一致性。

下面对 S101-S104 进行详细描述：

在 S101，可以理解，S101 所述的一个音频可以是一首完整/不完整的歌曲/音乐，一个音频片段则是从一个音频中截取的一段音频。还应理解，音频或音频片段均具有时序性。

可选的，终端设备可以随机的将媒体数据库或本地存储的音乐数据库中的 k 个音频确定为 k 个目标音频。

可选的，终端设备可以先接收用户输入的第一操作，并响应该第一操作，从而确定出 k 个目标音频。可以理解，该终端设备中安装有具有音频剪辑功能的应用程序，该第一操作即为用户在该应用程序中的音频剪辑界面上的操作。

一种可能的实现方式，上述的第一操作可以包括用户在音频剪辑界面上的目标音频的选择操作。可选的，选择操作可以包括音乐数据库的选择操作，以及在选中的音乐数据库中选择目标音频的操作。

其中，音乐数据库可以是存储在本地的音乐数据库，也可以是基于系统根据音频的场景标签、情感标签、风格标签等对音频进行分类后的音乐数据库，也可以是系统自动推荐的音乐数据库，还可以是用户对基于系统推荐或分类的音乐数据库中的音频进行删除或增加后配置得到的自定义的音乐数据库等，本申请实施例对此不作限定。这里，系统可以是与终端设备上安装的具有音频剪辑功能的应用程序连接通信的任意媒体系统，本申请实施例对此不作限定。

作为示例，系统推荐的音乐数据库可以是系统根据终端设备的传感器检测的当前用户所处场景/状态推荐的音乐数据库。例如，当终端设备传感器检测到当前用户的状态为跑步状态，则系统可以向用户推荐包括动感音乐的音乐数据库。或者，系统推荐的音乐数据库也可以是随机展示的流媒体的音乐数据库，例如各大音乐榜单，或者包括流行经典音乐的音乐数据库。等等。

这里需要说明的是，每个音频在制作时，都可以标注有包括但不限于场景、风格、情感等标签。其中，场景是指适合聆听该音频的场景，例如可以是工作场景、学习场

景、跑步场景等。风格是指音频的音乐风格，例如可以是摇滚、电子、轻音乐等。情感是指该音频表达的情感，例如可以是伤感、思恋、孤独等。对此不做详述。

以终端设备是手机 10，手机 10 中安装有音频剪辑应用程序为例，在一个示例中，参考图 4，图 4 示出了本申请实施例提供的一种用户在音频剪辑应用程序的音频剪辑界面上输入的第一操作的示意图。

如图 4 中的 (a) 所示，手机 10 的触摸屏上显示有音频剪辑应用程序的音频剪辑界面 401。可以看出，音频剪辑界面 401 为“分类曲库”标签下的音乐数据库选择界面。这样，用户即可在音频剪辑界面 401 上进行音乐数据库的选择操作。

如图 4 中的 (a) 所示，在音频剪辑界面 401 上，显示有系统基于音频的不同分类标准对音频进行分类时的类型标签。如图所示，音频剪辑界面 401 上显示有基于适于聆听音频的场景对音频进行分类时的类型标签，如“工作”标签、“跑步”标签等。音频剪辑界面 401 上还显示有基于音频所表达情感对音频进行分类时的类型标签，例如“快乐”标签、“兴奋”标签等。音频剪辑界面 401 上还显示有基于音频的音乐风格对音频进行分类时的类型标签，例如“流行”标签、“节奏蓝调”标签等。容易理解，图 4 中的 (a) 所示出的音频的类型标签及其显示格式仅为示例性说明，并不作为对本申请实施例保护范围的限定。

这样，用户可以基于自身的需求/喜好，操作（例如用手指/触摸笔点击）音频剪辑界面 401 上显示的类型标签，例如用户可以用手指分别点击“跑步”标签、“快乐”标签、“兴奋”标签、以及“节奏蓝调”标签，并在操作（例如用手指/触摸笔点击）音频剪辑界面 401 上的“确定”按钮后，作为响应，手机 10 即可显示系统基于用户选择的类型标签推荐的所有具有“跑步”标签、“快乐”标签、“兴奋”标签、以及“节奏蓝调”标签的音频的界面，例如图 4 中的 (b) 所示的目标音频选择界面 402。可以理解，目标音频选择界面 402 上显示的所有音频，即构成了用户所选择的音乐数据库。

可以理解，当用户在音频剪辑界面 401 上选择的类型标签为“自动”标签，则手机 10 在目标音频选择界面 402 上所显示的音频，即为手机 10 根据该手机 10 配置的传感器（例如陀螺仪传感器、噪声传感器等）检测到的当前操作手机 10 的用户所处的环境/状态，为用户自动推荐的适合用户在当前环境聆听的音频。对此不作赘述。

进一步的，用户可以在目标音频选择界面 402 上进行目标音频的选择操作，例如用户可以基于自身的需求/喜好，在目标音频选择界面 402 选中 k 个目标音频。作为响应，手机 10 即确定出 k 个目标音频。

在另一个示例中，参考图 5，图 5 示出了本申请实施例提供的另一种用户在音频剪辑应用程序的音频剪辑界面上输入的第一操作的示意图。

如图 5 中的 (a) 所示，手机 10 的触摸屏上显示有音频剪辑应用程序的音频剪辑界面 501。可以看出，音频剪辑界面 501 为用户在音频剪辑界面 401 上选中“推荐曲库”标签后显示的音乐数据库选择界面。这样，用户即可在音频剪辑界面 501 上进行音乐数据库的选择操作。

如图 5 中的 (a) 所示，在音频剪辑界面 501 上，显示有系统展示的多个音乐数据库的标识。例如“流行经典”音乐库标识、“网络甜歌”音乐库标识、“轻音乐集”音乐库标识、“金曲榜”音乐库标识等。容易理解，图 5 中的 (a) 所示出的音乐数据库及其标

识的显示格式仅为示例性说明，并不作为对本申请实施例保护范围的限定。例如，手机 10 还可以分多个界面来显示不同类型的音乐数据库的标识，对此不作限定。

5 这样，用户即可基于自身需求或喜好，操作（例如用手指/触摸笔点击）音频剪辑界面 501 上所显示的一个音乐数据库（例如“流行经典”音乐库），并在操作（例如用手指/触摸笔点击）音频剪辑界面 501 上的“确定”按钮后，作为响应，手机 10 即可显示“流行经典”音乐库中音频的界面，例如图 5 中的 (b) 所示的目标音频选择界面 502。

进一步的，用户可以在目标音频选择界面 502 上进行目标音频的选择操作，例如用户可以基于自身的需求/喜好，在目标音频选择界面 502 选中 k 个目标音频。作为响应，手机 10 即确定出 k 个目标音频。

10 在又一个示例中，参考图 6，图 6 示出了本申请实施例提供的又一种用户在音频剪辑应用程序的音频剪辑界面上输入的第一操作的示意图。

如图 6 所示，手机 10 的触摸屏上显示有音频剪辑应用程序的音频剪辑界面 601。可以看出，音频剪辑界面 601 为用户在音频剪辑界面 401 或音频剪辑界面 501 上选中“本地曲库”标签后显示的目标音频选择界面。这样，用户即可在音频剪辑界面 501 上
15 进行目标音频的选择操作。

可以理解，用户在音频剪辑界面 401 或音频剪辑界面 501 上选中“本地曲库”标签后显示音频剪辑界面 601 的操作，相当于是用户在音频剪辑界面 401 或音频剪辑界面 501 选择本地音乐数据库的操作。

20 如图 6 所示，在音频剪辑界面 601 上显示有本地存储的多个音频，例如以列表的形式显示多个音频。这样，用户可以在音频剪辑界面 601 上进行目标音频的选择操作，例如用户可以基于自身的需求/喜好，在音频剪辑界面 601 选中 k 个目标音频。作为响应，手机 10 即确定出 k 个目标音频。

容易理解，图 6 中所示出的本地存储的多个音频的显示的格式仅为示例性说明，并不作为对本申请实施例的限定。例如，手机 10 还可以将本地存储的多个音频划分为
25 多个组，并以多个层级界面来显示不同组别中的音频列表，对此不作限定。

另一种可能的实现方式，上述的第一操作可以包括用户在音频剪辑界面上输入目标音频的数量的输入操作，以及音乐数据库的选择操作。

作为示例，参考图 7，图 7 示出了本申请实施例提供的又一种用户在音频剪辑应用程序的音频剪辑界面上输入的第一操作的示意图。

30 如图 7 所示，图 7 所示的音频剪辑界面 701 上包括用于输入串烧音频数目的接口（即输入框 702），这样，用户即可通过输入框 702 输入串烧音频的数量，以 k 的取值是 3 为例，即用户可以在输入框 702 输入数值“3”。

此外，用户可以通过操作音频剪辑界面 701 上的“曲库”按钮，并根据自身的需求/喜好选择音乐数据库。这里，用户操作音频剪辑界面 701 上的“曲库”按钮后所显示
35 的选择音乐数据库的过程，可以参考上文中图 4 中的 (a)、图 5 中的 (b) 以及图 6 选择音乐数据库的描述，这里不作赘述。

当用户通过操作音频剪辑界面 701 上的“曲库”按钮选定音乐数据库后，手机 10 即可以根据用户在输入框 702 输入的 k 值，在用户选定的音乐数据库中选择 k 个音频作为目标音频。

可选的，手机 10 可以基于预设规则，并根据用户在输入框 702 输入的 k 值，在用户选定的音乐数据库中选择 k 个音频作为目标音频。例如，手机 10 可以在音乐数据库中随机选择 k 个音频作为目标音频，或者，手机 10 可以将音乐数据库中的前 k 个音频作为目标音频，等等，本申请实施例对此不作限定。

5 这样，当终端设备确定出 k 个目标音频后，终端设备可以通过预设算法从 k 个目标音频中提取 m 个音频片段。作为示例，该预设算法可以是用于提取歌曲中副歌/高潮部分的算法，本申请实施例对此不作限定。

可选的，终端设备预置有 m 个音频的串烧顺序，或者，终端设备可以进一步通过与用户进行交互，以确定出该 m 个音频片段在串烧时的串烧顺序。

10 一个示例，结合图 4、图 5 以及图 6，当用户在图 4 中的 (b) 所示的目标音频选择界面 402 上选中 k 个目标音频后，并对目标音频选择界面 402 上的“确定”按钮进行操作（例如用手指/触摸笔点击）后，或者，当用户在图 5 中的 (b) 所示的目标音频选择界面 502 上选中 k 个目标音频后，并对目标音频选择界面 502 上的“确定”按钮进行操作（例如用手指/触摸笔点击）后，或者，当用户在图 6 所示的音频剪辑界面 601
15 选中 k 个目标音频后，并对音频剪辑界面 601 上的“确定”按钮进行操作（例如用手指/触摸笔点击）后，终端设备（即手机 10）即可显示如图 4 中的 (c) 所示的串烧顺序选择界面 403。如图 4 中的 (c) 所示，串烧顺序选择界面 403 可以包括“顺序”、“随机”、以及“自定义”三个选项。

其中，当用户选择“顺序”选项时，作为响应，手机 10 可以按照 k 个目标音频在所属音乐数据库中的顺序，对从 k 个目标音频中提取的 m 个音频片段进行串烧。可选的，
20 k 个目标音频在所属音乐数据库中的顺序，可以以该 k 个目标音频在该音乐数据库中的编号体现。

当用户选择“随机”选项时，作为响应，手机 10 可以随机的对从 k 个目标音频中提取的 m 个音频片段进行串烧。

25 当用户选择“自定义”选项时，用户可以进一步的在“自定义”的选项框 4031 中以预设顺序依次输入 k 个目标音频的标识（例如编号）。这样，作为响应，手机 10 即可以该预设顺序对从 k 个目标音频中提取的 m 个音频片段进行串烧。其中，该预设顺序即为用户自定义的顺序。

30 另一个示例，在图 7 所示的音频剪辑界面 701 上，还可以包括三个用于输入串烧歌曲顺序的选项，其具体说明可以参考上文对图 4 中的 (c) 的说明，这里不再赘述。

在 S102，终端设备在确定出 m 个音频片段后，可以确定该 m 个音频片段的音频特征信息。

35 其中，音频片段的音频特征信息可以包括音频片段的主旋律轨位置信息、风格标签、情感标签、节奏信息、节拍信息、或调号信息等中的至少一种。可以理解，这里所述的节拍即为音乐的节拍，调号即为音调号。其中，本申请实施例对终端设备获得音频片段的音频特征信息的具体实现方式不作具体限定。本申请实施例对终端设备确定该 m 个音频片段的音频特征信息的过程不作详述。

对于终端设备确定出的 m 个音频片段中的每个音频片段而言，终端设备还可以通过音乐人声检测技术，对每个音频片段进行音乐和人声分离处理。

以 m 个音频片段中的第一音频片段为例，终端设备可以通过音乐人声检测技术，将第一音频片段中的人声和各种乐器（例如钢琴、贝斯、鼓、小提琴等）声音分离，并将分离得到的多轨乐器和人声转换为 MIDI 格式的数据，该 MIDI 格式的数据即为第一音频片段的 MIDI 信息。这里，本申请实施例对音乐人声检测技术不作详述。

5 需要说明的是，音频片段中也可以不包括人声。这种情况下，终端设备可以通过音乐人声检测技术，分离出第一音频片段中的多轨乐器声音，并将该多轨乐器声音转换为 MIDI 格式的数据。

进一步的，终端设备可以根据 m 个音频片段的音频特征信息和 MIDI 信息，以及预设神经网络模型，确定出 $m-1$ 个过渡音频信息。其中， $m-1$ 个过渡音频信息中的一个过渡音频信息，是用于衔接终端设备所获取的 m 个音频片段中排序连续的两个音频片段的过渡音频信息，其中，该 m 个音频片段的排序是指该 m 个音频片段的串烧顺序，过渡音频信息为过渡音频段的 MIDI 信息。这里，终端设备确定 m 个音频片段的串烧顺序的详细描述可以参考 S101 中的相关描述，这里不再赘述。

10 其中，预设神经网络模型可以预置在终端设备中，也可以预置在与终端设备具有通信连接的服务器上，本申请实施例对此不作限定。

其中，该预设神经网络模型包括编码器、信息提取模块、信息生成模块以及解码器。这里，编码器、信息提取模块、信息生成模块、以及解码器均为序列模型网络结构，信息提取模块是双向序列模型网络结构。

20 作为示例，其中，编码器、信息生成模块、以及解码器可以是 RNN、LSTM、GRU、Transformer 等网络。信息提取模块可以是 Bi-RNN、Bi-LSTM、Bi-GRU、Transformer 等网络。

应理解，该预设神经网络模型包括至少两个编码器、至少两个信息提取模块、至少一个信息生成模块以及至少一个解码器。其中，该至少两个编码器的网络结构相同，该至少两个信息提取模块的网络结构相同，该至少一个信息生成模块的网络结构相同，该至少一个解码器的网络结构相同。此外，在一个预设神经网络模型中，编码器和解码器的网络结构也可以是相同的。需要说明的是，编码器和解码器的数据流向是相反的。作为示例，编码器的输入可以作为解码器的输出，编码器的输出可以作为解码器的输入。

30 应理解，上述至少两个编码器、至少两个信息提取模块、至少一个信息生成模块以及至少一个解码器的网络参数均是在训练该预设神经网络模型时确定的。这里，训练预设神经网络模型的详细说明可以参考下文中训练图 8 所示预设神经网络模型 80 的描述，这里不作赘述。

35 在一种可能的实现方式中，当需要该预设神经网络模型处理的音频片段的数量为 m ，则该预设神经网络模型包括 m 个输入和 $m-1$ 个输出。这样的话，该预设神经网络模型中编码器的数量为 m ，信息提取模块的数量为 $2 \times (m-1)$ ，信息生成模块和解码器的数量均为 $m-1$ 。

这种情况下，该预设神经网络模型可以对输入的 m 个音频片段的信息同时进行处理，并输出 $m-1$ 个过渡音频的 MIDI 信息。

一个示例，以 m 的取值是 2 为例，参考图 8，图 8 示出了本申请实施例提供的一

种预设神经网络模型的结构示意图。

如图 8 所示，预设神经网络模型 80 包括用于接收 2 个输入的 2 个编码器，分别为编码器 811 和编码器 812。预设神经网络模型 80 包括 2 个（即 $2 \times (2-1)$ 个）信息提取模块，分别为信息提取模块 821 和信息提取模块 822。预设神经网络模型 80 还包括 1 个（即 $(2-1)$ 个）信息生成模块（即信息生成模块 83）和 1 个（即 $(2-1)$ 个）解码器（即解码器 84）。

另一个示例，以 m 的取值是 4 为例，参考图 9，图 9 示出了本申请实施例提供的另一种预设神经网络模型的结构示意图。

如图 9 所示，预设神经网络模型 90 包括用于接收 4 个输入的 4 个编码器，分别为编码器 911、编码器 912、编码器 913、以及编码器 914。预设神经网络模型 90 包括 6 个（即 $2 \times (4-1)$ 个）信息提取模块，分别为信息提取模块 921、信息提取模块 922、信息提取模块 923、信息提取模块 924、信息提取模块 925、以及信息提取模块 926。预设神经网络模型 90 包括 3 个（即 $(4-1)$ 个）信息生成模块，分别为信息生成模块 931、信息生成模块 932、以及信息生成模块 933。预设神经网络模型 90 还包括 3 个（即 $(4-1)$ 个）解码器，分别为解码器 941、解码器 942、以及解码器 943。

在另一种可能的实现方式中，预设神经网络模型中编码器和信息提取模块的数量均为 2，信息生成模块和解码器的数量均为 1（例如图 8 所示的预设神经网络模型）。也就是说，该预设神经网络模型包括 2 个输入、1 个输出。

这种情况下，该预设神经网络模型一次可以对 2 个输入（即 2 个音频片段的信息）同时进行处理，并输出 1 个过渡音频信息。当需要该预设神经网络模型处理的音频片段的数量为 m ，且 m 大于 2 时，则该预设神经网络模型可以对 m 个音频片段的信息串行的进行 $m-1$ 次处理，即可得到 $m-1$ 个过渡音频信息。应理解，该预设神经网络模型每次处理的两个音频片段是串烧顺序相邻的两个音频片段。

作为示例，当需要该预设神经网络模型处理的音频片段包括音频片段 1、音频片段 2、音频片段 3 以及音频片段 4，即 m 的取值为 4 时，假设该 4 个音频片段的串烧顺序为：音频片段 1→音频片段 4→音频片段 3→音频片段 2，则终端设备可以对该 4 个音频片段的信息串行的进行 3（即 $(4-1)$ ）次处理，即可得到 3 个过渡音频信息。具体的，终端设备可以将音频片段 1 和音频片段 4 的信息作为该预设神经网络的两个输入信息，这样即可得到用于衔接音频片段 1 和音频片段 4 的过渡音频信息 1。终端设备可以将音频片段 4 和音频片段 3 的信息作为该预设神经网络的两个输入信息，这样即可得到用于衔接音频片段 4 和音频片段 3 的过渡音频信息 2。终端设备还可以将音频片段 3 和音频片段 2 的信息作为该预设神经网络的两个输入信息，这样即可得到用于衔接音频片段 3 和音频片段 2 的过渡音频信息 3。

下面以 m 的取值是 2，即以终端设备获取到 2 个音频片段（例如该 2 个音频片段包括第一音频片段和第二音频片段），且串烧顺序为第一音频片段→第二音频片段为例，结合图 8 对本申请实施例提供的预设神经网络模型中的各个模块及其对音频片段的信息进行处理的过程予以详细说明。从串烧顺序可以看出，第一音频片段是目标串烧音频中在前的乐段（简称前乐段），第二音频片段是目标串烧音频中在后的乐段（简称后乐段）。

这里，第一音频片段作为前乐段，其音频特征信息和 MIDI 信息可以被称为第一信息。第二音频片段作为后乐段，其音频特征信息和 MIDI 信息可以被称为第二信息。其中，第一信息即可作为预设神经网络模型 80 的一个输入，第二信息即可作为预设神经网络模型 80 的另一个输入。

5 需要说明的是，第一信息中的 MIDI 信息（即第一音频片段的 MIDI 信息）中的多个声轨，和第二信息中的 MIDI 信息（即第二音频片段的 MIDI 信息）中的多个声轨是相同的。具体的，第一信息中的 MIDI 信息所包括的声轨数量及其类型，与第二信息中的 MIDI 信息所包括的声轨数量及其类型均相同。

10 例如，第一信息中的 MIDI 信息包括 3 个声轨，分别为人声声轨、钢琴声轨以及小提琴声轨。那么，第二信息中的 MIDI 信息也包括这 3 个声轨。

应理解，假设第一信息中的 MIDI 信息所包括的声轨数量和第二信息中的 MIDI 信息所包括的声轨数量不同时，例如第一信息中的 MIDI 信息所包括的声轨数量，大于第二信息中的 MIDI 信息所包括的声轨数量，且第一信息中的 MIDI 信息包括第二信息中的 MIDI 信息所包括的所有声轨类型，则终端设备可以为第二信息中的 MIDI 信息增加空轨，以使第一信息中的 MIDI 信息所包括的声轨数量和第二信息中的 MIDI 信息所包括的声轨数量相同。

15 这样，终端设备可以将第一信息作为预设神经网络模型 80 的一个输入（例如输入 1）输入到编码器 811 中。编码器 811 即可对接收到的第一信息进行处理后输出第一音频片段对应的第一序列。这里，第一序列是编码器 811 对第一信息中的 MIDI 信息和音频特征信息进行特征提取后得到的序列。这里，第一序列可以被理解为编码器 811 对第一信息进行降维后得到的序列，或者，第一序列可以被理解为编码器 811 将第一信息压缩至隐空间后得到的序列。

20 应理解，第一序列是时序上的一维序列，且第一序列的长度由第一音频片段的长度确定。可以理解，第一音频片段越长，第一序列越长；第一音频片段越短，第一序列越短。

作为示例，第一序列可以表示为“ $\{P_1, P_2, \dots, P_s\}$ ”，其中， P 表示特征向量， s 表示特征向量的个数。应理解，由于音频片段本身具有时序性，因此第一序列也具有时序性。这样的话， P_1 可以是音频片段始端时刻对应的特征向量， P_s 可以是音频片段末端时刻对应的特征向量。

30 类似的，终端设备可以将第二信息作为预设神经网络模型 80 的另一个输入（例如输入 2）输入到编码器 812 中，这样，编码器 812 即可对接收到的第二信息进行处理后输出第二音频片段对应的第二序列。这里，第二序列的描述可以参考第一序列，不再赘述。

35 作为示例，第二序列可以表示为“ $\{F_1, F_2, \dots, F_t\}$ ”，其中， F 表示特征向量， t 表示特征向量的个数。其中， F_1 可以是音频片段始端时刻对应的特征向量， F_t 可以是音频片段末端时刻对应的特征向量。

接着，信息提取模块 821 接收编码器 811 输出的第一序列。由于音频片段本身具有时序性，且第一音频片段是前乐段，因此信息提取模块 821 对第一序列进行学习后，可以输出与第一音频片段的末端时刻对应的第一向量。该过程也可以理解为信息提取

模块 821 对第一序列进一步进行了降维。需要说明，第一向量携带有第一序列的特征，且与第一序列的末端时刻对应。

类似的，信息提取模块 822 接收编码器 812 输出的第二序列。信息提取模块 821 对第二序列进行学习后，可以输出与第二音频片段的始端时刻对应的第二向量。该过程也可以理解为信息提取模块 822 对第二序列进一步进行了降维。需要说明，第二向量携带有第二序列的特征，且与第二序列的始端时刻对应。

然后，预设神经网络模型 80 对第一向量和第二向量求和，得到第三向量。即第三向量=第一向量+第二向量。这里，第一音频片段末端时刻对应的隐空间向量（即第一向量）和第二音频片段始端时刻对应的隐空间向量（即第二向量）的和（即第三向量），可以作为衔接第一音频片段和第二音频片段的过渡音频段对应的隐空间向量。

这样，基于该过渡向量，预设神经网络模型 80 可以确定出用于衔接第一音频片段和第二音频片段的过渡音频段。具体的，预设神经网络模型 80 将第三向量输入到信息生成网络模块 83，信息生成网络模块 83 即可对接收到的第三向量进行学习，并输出第三序列。应理解，第三序列即为用于衔接第一音频片段和第二音频片段的过渡音频段的特征向量所构成的序列。

作为示例，第三序列可以表示为“ $\{M_1, M_2, \dots, M_j\}$ ”，其中，M 表示特征向量，j 表示特征向量的个数。其中， M_1 可以是用于衔接第一音频片段和第二音频片段的过渡音频段的始端时刻对应的特征向量， M_j 可以是该过渡音频段的末端时刻对应的特征向量。

然后，解码器 84 接收到信息生成模块 83 输出的第三序列，并对第三序列进行学习，并输出用于衔接第一音频片段和第二音频片段的过渡音频信息，即过渡音频段的 MIDI 信息。

应理解，上述图 8 所示的预设神经网络模型 80 可以是预先基于多个训练样本训练得到的神经网络模型。其中，一个训练样本包括两个音频片段的 MIDI 信息和音频特征信息，该训练样本的标签值为领域专家根据这两个音频片段构建的过渡音频段的 MIDI 信息。这样，基于多个训练样本对神经网络进行反复的迭代训练，即可得到本申请实施例图中图 8 所示的预设神经网络模型。

在 S103，终端设备将在 S102 步骤生成的 m-1 个过渡音频信息（即过渡音频段的 MIDI 信息）分别插入 m 个音频片段的 MIDI 信息中，以实现通过该 m-1 个过渡音频信息来衔接该 m 个音频片段的 MIDI 信息的目的，也即生成了该 m 个目标音频片段串烧后得到的目标串烧音频。

作为示例，假设 m 取值为 3，m 个音频片段包括音频片段 1、音频片段 2 以及音频片段 3，m-1 个过渡音频信息包括过渡音频信息 1 和过渡音频信息 2。则当该 3 个音频片段的串烧顺序为音频片段 1→音频片段 3→音频片段 2，且过渡音频信息 1 是用于衔接音频片段 1 和音频片段 3 过渡音频信息，过渡音频信息 2 是用于衔接音频片段 3 和音频片段 2 过渡音频信息时，则终端设备可以将过渡音频信息 1 插入音频片段 1 和音频片段 3 的 MIDI 信息之间，以及可以将过渡音频信息 2 插入音频片段 3 和音频片段 2 的 MIDI 信息之间。这样，终端设备即生成了音频片段 1、音频片段 2 以及音频片段 3 串烧后目标串烧音频的 MIDI 信息。

在实际应用中，终端设备在生成目标串烧音频的 MIDI 信息后，可以向用户播放该目标串烧音频。当用户认为该目标串烧音频不是自己想要的串烧音频时，用户可以向终端设备输入第二操作。这样，终端设备响应该第二操作，对用于串烧得到该目标串烧音频的 m 个音频片段的串烧顺序进行调整，并基于调整后的串烧顺序和 m 个音频片段，重新生成 m-1 个过渡音频信息，并重新生成目标串烧音频的 MIDI 信息。

5 然后，终端设备可以向用户播放重新生成的目标串烧音频。当用户对该目标串烧音频满意，则可以执行 S104；当用户对该目标串烧音频不满意，则可以再次向终端设备输入第二操作，以使终端设备对 m 个音频片段的串烧顺序进行再次调整，并再次重新生成目标串烧音频的 MIDI 信息。可以看出，通过终端设备和用户之间的反复交互，
10 可以获得令用户满意的目标串烧音频，提高了用户的体验。

在一种可能的实现方式中，终端设备在生成目标串烧音频的 MIDI 信息后，通过图谱的形式在显示面板上显示目标串烧音频的 MIDI 谱。这样，用户的第二操作即可以是对该 MIDI 谱的拖拽操作。这样，终端设备在接收到用户的第二操作后，可以响应该第二操作，以重新确定出用于串烧得到前述目标串烧音频的 m 个音频片段的串烧
15 顺序。

这样的话，终端设备基于重新确定出的 m 个音频片段的串烧顺序，以及该 m 个音频片段，并通过执行 S102，即可重新确定出 m-1 个过渡音频信息。进一步的，终端设备根据重新确定出的 m-1 个过渡音频信息和该 m 个音频片段，即可重新生成目标串烧音频。这里，终端设备重新生成目标串烧音频的过程，可以参考是上文中终端设备生成目标串烧音频的过程的详细描述，这里不作赘述。
20

作为示例，以终端设备是手机 10，且手机 10 生成的目标串烧音频的 MIDI 信息中包括 3 个声轨，且目标串烧音频是对音频片段 1、音频片段 2 以及音频片段 3，以音频片段 1→音频片段 3→音频片段 2 的串烧顺序进行串烧后得到的串烧音频，参考图 10 中的 (a)，图 10 中的 (a) 示出了本申请实施例提供的一种第二操作的示意图。

25 如图 10 中的 (a) 所示，当手机 10 第一次生成目标串烧音频的 MIDI 信息后，可以在显示面板上显示该目标串烧音频的 MIDI 谱，例如图 10 中的 (a) 中所示的串烧音频剪辑界面 1001 上所显示的 MIDI 谱。

其中，串烧音频剪辑界面 1001 上显示的 MIDI 谱包括三个声轨，分别为黑色条带所示的声轨 1、白色条带所示的声轨 2、以及条纹条带所示的声轨 3。

30 其中，串烧音频剪辑界面 1001 上显示的 MIDI 谱上的起始线用于标记目标串烧音频的起始。串烧音频剪辑界面 1001 上显示的 MIDI 谱上还包括多个分割线，该多个分割线用于区分目标串烧音频中的不同音频片段以及过渡音频段。

例如，基于目标串烧音频中音频片段的串烧顺序，位于起始线和分割线 1 之间的音频段为音频片段 1，位于分割线 1 和分割线 2 之间的音频段为衔接音频片段 1 和音频片段 3 的过渡音频段 1，位于分割线 2 和分割线 3 之间的音频段为音频片段 3，位于分割线 3 和分割线 4 之间的音频段为衔接音频片段 3 和音频片段 2 的过渡音频段 2，位于分割 4 右侧的音频段为音频片段 2（图 10 中的 (a) 中 MIDI 谱未示出用于标记目标串烧音频结束的终止线）。可以理解，在串烧音频剪辑界面 1001 上显示的 MIDI 谱上也可以显示各个音频段的名称，本申请实施例对此不作限定。
35

当用户可以对串烧音频剪辑界面 1001 上的播放图标 1002 进行操作（例如用手指或触摸笔点击）后，作为响应，手机 10 向用户播放目标串烧音频。当用户对该目标串烧音频不满意时，可以向手机 10 输入第二操作。这里，第二操作可以是用户对串烧音频剪辑界面 1001 上显示的 MIDI 谱进行拖拽的操作（例如通过手指或触摸笔在显示面板上滑动的操作），例如用户通过手指按住音频片段 1 的 MIDI 谱（即图 10 中的（a）所示 MIDI 谱中起始线和分割线 1 之间的区域），并沿着图 10 中的（a）中箭头所示方向滑动至音频片段 2 的 MIDI 谱（即图 10 中的（a）所示 MIDI 谱中分割线 4 右侧的区域）位置处。作为响应，手机 10 交换音频片段 1 和音频片段 2 的串烧顺序，即手机 10 重新确定出音频片段 1、音频片段 2 以及音频片段 3 的串烧顺序为音频片段 2→音频片段 3→音频片段 1。

进一步的，手机 10 可以根据音频片段 1、音频片段 2、音频片段 3 以及重新确定的串烧顺序，重新生成目标串烧音频。

在另一种实现方式中，第二操作可以是用户在终端设备生成目标串烧音频的 MIDI 信息后所显示的音频剪辑界面上，输入目标串烧顺序的操作。作为响应，终端设备即接收到用户输入的目标串烧顺序。也就是说，终端设备重新确定出 m 个音频片段的目标串烧顺序。

这样的话，终端设备基于接收到的 m 个音频片段的目标串烧顺序，以及该 m 个音频片段，并通过执行 S102，即可重新确定出 m-1 个过渡音频信息。进一步的，终端设备根据重新确定出的 m-1 个过渡音频信息和该 m 个音频片段，即可重新生成目标串烧音频。

作为示例，以终端设备是手机 10，且目标串烧音频是手机 10 对音频片段 1、音频片段 2 以及音频片段 3，以音频片段 1→音频片段 3→音频片段 2 的串烧顺序进行串烧后得到的串烧音频为例，参考图 10 中的（b），图 10 中的（b）示出了本申请实施例提供的另一种第二操作的示意图。

如图 10 中的（b）所示，当手机 10 第一次生成目标串烧音频的 MIDI 信息后，可以在显示面板上显示如图 10 中的（b）所示的串烧音频剪辑界面 1001。当用户可以对串烧音频剪辑界面 1001 上的播放图标 1002 进行操作（例如用手指或触摸笔点击）后，作为响应，手机 10 可以播放该目标串烧音频。当用户对该目标串烧音频不满意时，可以向手机 10 输入第二操作。

具体的，用户可以在串烧音频剪辑界面 1001 上的目标串烧顺序的输入框 1003 内输入期望的目标串烧顺序，例如用户在输入框 1003 输入“2, 3, 1”，其中，“2”可以用于表示音频片段 2 的标识，“3”可以用于表示音频片段 3 的标识，“1”可以用于表示音频片段 1 的标识，“2, 3, 1”可以用于表示音频片段 1、音频片段 2 以及音频片段 3 的串烧顺序为音频片段 2→音频片段 3→音频片段 1。作为响应，手机 10 即接收到用户输入的目标串烧顺序。这样，手机 10 即确定出音频片段 1、音频片段 2 以及音频片段 3 的目标串烧顺序。

然后，手机 10 可以根据音频片段 1、音频片段 2、音频片段 3、以及接收到的目标串烧顺序，重新生成目标串烧音频。

可选的，图 10 中的（b）所示的串烧音频剪辑界面 1001 上也可以是显示有当前串

烧顺序，例如“当前串烧顺序：1，3，2”。应理解，该当前串烧顺序可以在用户提供输入目标串烧顺序时作为参考。

应理解，上述可能的实现第二操作的方式仅为示例性说明，并不构成对本申请实施例保护范围的限定。

5 在 S104，可选的，终端设备可以直接将 S103 生成的最新的目标串烧音频的 MIDI 信息直接进行保存并输出。

可选的，终端设备还可以根据 S103 生成的最新的目标串烧音频的 MIDI 信息合成目标串烧音频的时域波形，从而得到目标串烧音频。可选的，终端设备还可以保存/输出该目标串烧音频。

10 其中，本申请实施例对终端设备根据目标串烧音频的 MIDI 信息合成目标串烧音频的时域波形的具体方式不作限定。作为示例，终端设备可以通过为目标串烧音频的 MIDI 信息加载音色库来合成目标串烧音频的时域波形，或者可以根据目标串烧音频的 MIDI 信息和波表（波表是预先将各种真实乐器所能发出的所有声音（包括各个音域、声调等）录制下来所存贮的文件）合成目标串烧音频的时域波形，或者还可以根据目
15 标串烧音频的 MIDI 信息，和物理模型/神经网络模型来合成目标串烧音频的时域波形，当然不限于此。这里，物理模型/神经网络模型是预先构建的用于合成音频波形的模型，本申请实施例对此不作详述。

应理解，在终端设备根据目标串烧音频的 MIDI 信息合成目标串烧音频的时域波形时，仅根据目标串烧音频的 MIDI 信息中除人声声轨之外的所有乐器声轨来合成目
20 标串烧音频的时域波形。

可选的，终端设备可以接收用户的第三操作，并响应该第三操作，对 S103 生成的最新的目标串烧音频的 MIDI 信息进行渲染，并根据渲染后的 MIDI 信息合成目标串烧音频的时域波形。然后，终端设备对合成的目标串烧音频的时域波形进行进一步渲染后，得到渲染后的目标串烧音频。可选的，终端设备还可以保存/输出该渲染后的目标
25 串烧音频。

其中，终端设备根据渲染后的 MIDI 信息合成目标串烧音频的时域波形的详细描述，可以参考上文中终端设备根据目标串烧音频的 MIDI 信息合成目标串烧音频的时域波形的说明，这里不再赘述。

其中，用户的第三操作，可以包括用户在音频渲染界面输入的音频渲染处理方式
30 的选择操作，以及用户在音频渲染界面输入的用于合成目标串烧音频时域波形的处理方式的选择操作。其中，对目标串烧音频的 MIDI 信息进行渲染，例如可以包括音源分离。用于合成目标串烧音频时域波形的处理方式，例如可以包括加载音色库、波表合成、物理模型合成等。对目标串烧音频的时域波形进行渲染可以包括混音，人声风格迁移等，本申请实施例对此不作限定。

35 可选的，终端设备可以将渲染后的目标串烧音频的时域波形保存为任一种音频格式音频文件，本申请实施例对此不作具体限定。作为示例，终端设备可以将渲染后的目标串烧音频波形保存为 WAV 格式的音频文件、无损音频压缩编码（free lossless audio codec, FLAC）格式的音频文件、动态影像专家压缩标准音频层面 3（moving picture experts group audio layer III, MP3）格式的音频文件、音频压缩格式（OGGVobis, ogg）

的音频文件等，当然不限于此。

可选的，终端设备还可以保存生成目标串烧音频的工程。这样的话，终端可以根据所保存的工程文件重新设置用于串烧得到该目标串烧音频的 m 个音频片段的串烧顺序，并重新进行串烧处理，这样可以提高未来对该 m 个音频片段再次进行串烧时的效率。

作为示例，以终端设备是手机 10 为例，参考图 11，图 11 示出了本申请实施例提供的一种对目标串烧音频的 MIDI 信息进行渲染以及输出的示意图。

如图 11 中的 (a) 所示，在手机 10 最终确定目标串烧音频的 MIDI 信息后，手机 10 可以显示音频渲染界面 1101。用户可以在音频渲染界面 1101 上输入第三操作，该第三操作可以包括：用户在“音源分离”选项下选择开启“去除人声”（图 11 中的 (a) “去除人声”标签的黑色方块表示开启，白色方框表示关闭）的选择操作，用户在“音频波形合成”选项下选择“加载音色库”的选择操作，用户在“混音”选项下选择“录制人声”的选择操作，以及用户在“人声风格迁移”选项下选择“歌星 A”作为迁移目标的选择操作。

作为响应，手机 10 接收到用户的第三操作，以及接收到用户对音频渲染界面 1101 上的“确定”按钮的操作（例如点击）后，手机 10 可以将目标串烧音频的 MIDI 信息中的人声声轨删除或置为无效，并对目标串烧音频的 MIDI 信息加载音色库以合成目标串烧音频的时域波形，然后开启录音界面为目标串烧音频录制人声，以及将目标串烧音频中的人声迁移为歌星 A 的声音。

进一步的，手机 10 可以在接收到用户的第三操作，并接收到用户对音频渲染界面 1101 上的“确定”按钮的操作（例如点击）后，显示如图 10 中的 (b) 所显示的音频发布界面 1102。这样，终端设备 10 可以通过音频发布界面 1102 与用户进行交互，并按照用户输入的指示导出目标串烧音频。

如图 11 中的 (b) 所示，手机 10 可以在音频发布界面 1102 的“导出格式”选项下接收用户输入的选择导出音频格式的选择操作，例如选择“音频格式 1”的操作。手机 10 可以在音频发布界面 1102 的“导出路径”选项下接收用户输入的目标串烧音频的名称（如名称 A）以及路径的操作。手机 10 还可以在音频发布界面 1102 的“保存工程”选项下接收用户输入的开启“保存工程”功能的操作。

这样，当用户对音频发布界面 1102 的“导出”按钮进行操作后（例如点击），作为响应，手机 10 即将目标串烧音频按照用户的指示进行保存。

在一些实施例中，本申请实施例所提供方法中确定过渡音频信息和生成目标音频的方法部分（即步骤 S102-S104），也可以在终端设备向用户实时播放音频的过程中进行。这种情况下，本申请实施例所提供方法中确定过渡音频信息和生成目标音频的方法部分，可以通过能够提供音频聆听的 App 的一个功能模块实现。

作为示例，能够提供音频聆听的 App 例如可以是云音乐 App。为简单叙述，下文中以能够提供音频聆听的 App 是云音乐 App 为例进行说明。

具体的，云音乐 App 在向用户提供音乐的聆听模式时，可以提供串烧模式。该串烧模式即可通过运行该云音乐 App 的终端设备、或者该云音乐 App 连接通信的服务器执行本申请实施例所提供方法中的步骤 S102-S104 来实现。为简单描述，下文中以通过运行该云音乐 App 的终端设备执行本申请实施例所提供方法中的步骤 S102-S104 来

实现该串烧模式为例进行说明。

这样的话，可选的，当终端设备向用户通过该终端设备中运行的云音乐 App 播放音乐时，终端设备所播放的音乐可以是云音乐媒体库自动推荐的音乐，也可以是本地媒体库中的音乐，本申请实施例对此不作限定。

5 这样，当终端设备通过与用户的交互，确定通过上述串烧模式向用户播放音乐时，终端设备可以将正在向用户播放的当前音乐和将要向用户播放的下一首音乐作为 2 个目标音频。并基于该 2 个目标音频和预设串烧顺序，并执行上文所述的 S102-S104，以生成对该 2 个目标音频进行串烧后得到的第一目标串烧音频。其中，该预设串烧顺序为：终端设备正在向用户播放的当前音乐→终端设备将要向用户播放的下一首音乐。

10 这里应注意，当终端设备当前播放的音乐是云音乐媒体库自动推荐的音乐，则终端设备在向用户播放当前音乐的过程中，可以确定出为用户自动推荐的下一首音乐，该下一首音乐即为终端将要向用户播放的下一首音乐。

需要说明的是，在终端设备向用户播放完当前音乐之前，终端设备可以完成对确定出的 2 个目标音频的串烧并获得第一目标串烧音频。

15 可选的，终端设备可以在向用户播放完当前音乐后，为用户播放第一目标串烧音频。进一步的，终端设备在向用户播放完第一目标串烧音乐后，为用户播放原先的下一首音乐。

20 作为示例，如果终端设备正在向用户播放的当前音乐是音乐 1，原先将要播放的下一首音乐为音乐 2，则终端设备可以在向用户播放完音乐 1 后，向用户播放第一目标串烧音频。然后，终端设备在向用户播放完第一目标串烧音乐后，为用户播放音乐 2。

25 类似的，当终端设备为用户播放原先的下一首音乐时，该原先的下一首音乐即变为终端设备正在向用户播放的新的当前音乐。这样，终端设备即可重复上述过程，生成对该新的当前音乐和该新的当前音乐的下一首音乐进行串烧后得到的第二目标串烧音频。并且，终端设备可以在向用户播放完该新的当前音乐后，为用户播放该第二串烧音频。

可以看出，在终端设备通过串烧模式向用户播放音乐时，终端设备可以动态的生成当前音乐和下一首音乐的串烧音频，并为用户播放该串烧音频，从而提高了用户体验。

30 还可以理解的是，在终端设备通过串烧模式向用户播放音乐时，终端设备为用户播放当前音乐和下一首音乐时，可以仅播放当前音乐和下一首音乐中用于生成串烧音频的音频片段，例如仅播放当前音乐和下一首音乐中的副歌/高潮部分，本申请实施例对此不作限定。

35 其中，还需说明的是，当终端设备基于确定出的 2 个目标音频和预设串烧顺序，执行上文所述步骤 S102-S104 以生成目标串烧音频（例如第一目标串烧音频、第二目标串烧音频、...、第 q 个目标串烧音频等，q 是正整数）的过程中，在 S103，终端设备只需执行一次生成过渡音频信息的过程，而无需接收用户输入的第二操作。

另外，云音乐 App 的串烧模式中预置有目标串烧音频的预设渲染模式和预设导出模式，该预设渲染模式中包括音源分离处理方式、混音处理方式，声音迁移方式等中

的至少一种。该预设导出模式包括串烧音频的导出格式，以及指示是否保存目标串烧音频的工程等。因此，在 S104，终端设备无需通过与用户进行交互来获取对目标串烧音频的渲染模式和导出模式。

5 应理解，云音乐 App 的串烧模式中所预置的目标串烧音频的预设渲染模式和预设导出模式，可以通过终端设备和用户的交互预先配置得到，也可以在终端设备向用户播放音乐的过程中，通过与用户的交互配置得到，这里不作限定。还应理解，在对目标串烧音频的预设渲染模式和预设导出模式配置完成后，终端设备还可以在向用户播放音乐的过程中，通过与用户的交互来更新该预先配置的预设渲染模式和预设导出模式，本申请实施例对此不作限定。

10 当然，上述预设导出模式中也可以不包括指示是否保存目标串烧音频的工程。这种情况下，终端设备可以在通过与用户进行交互后停止为用户播放音乐时，通过与用户进行交互，接收用户输入的是否保存目标串烧音频的工程的指示，并基于用户输入的指示执行对目标串烧音频的工程的保存。可以理解，在此之前，终端设备可以将动态生成的所有目标串烧音频的工程进行缓存。

15 还应理解，上述“串烧模式”的名称仅为示例性说明，并不作为对本申请实施例的限定。

20 综上，本申请实施例提供了一种音频数据的处理方法，通过该方法，本申请实施例可以在 MIDI 域基于 m 个音频片段，生成用于衔接该 m 个音频片段的 $m-1$ 个过渡音频信息。这样，通过该 $m-1$ 个过渡音频信息即可将 m 个音频片段的 MIDI 信息衔接起来，从而得到该 m 个音频片段被串烧后的目标串烧音频。可以看出，通过本申请方法对多个音频片段进行串烧时，终端设备可以生成全新的用于衔接该多个音频片段的过渡音频段，因此本申请实施例提供的方法无需考虑用于得到目标串烧音频的音频片段的相似度。也就是说，通过本申请实施例提供的方法能够获得更丰富、更具多样性的串烧音频。

25 并且，由于音频的 MIDI 信息是音频最原始的表现形式，其记录有音频的音符音高、音符力度、音符时长等信息。因此，相比在时域上直接对多个音频片段进行串烧，本申请实施例提供的方法中所生成的用于衔接两个音频片段的过渡音频信息是基于音频乐理生成的，这样的话，基于该过渡音频信息获得的串烧音频在听觉上更加流畅自然。并且，在 MIDI 域处理数据也更有利于串烧音频在后期渲染时的灵活性和一致性。

30 此外，通过本申请实施例提供的方法对 m 个音频片段进行串烧时，用户可以参与度高，这样即可以获得令用户满意的串烧音频，即用户体验度高。

35 上述主要从方法的角度对本申请实施例提供的方案进行了介绍。为了实现上述功能，其包含了执行各个功能相应的硬件结构和/或软件模块。本领域技术人员应该很容易意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，本申请能够以硬件或硬件和计算机软件的结合形式来实现。某个功能究竟以硬件还是计算机软件驱动硬件的方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本申请的范围。

本申请实施例可以根据上述方法示例对音频数据的处理装置进行功能模块的划分，

例如，可以对应各个功能划分各个功能模块，也可以将两个或两个以上的功能集成在一个处理模块中。上述集成的模块既可以采用硬件的形式实现，也可以采用软件功能模块的形式实现。需要说明的是，本申请实施例中对模块的划分是示意性的，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式。

5 如图 12 所示，图 12 示出了本申请实施例提供的一种音频数据的处理装置 120 的结构示意图。处理装置 120 可以用于执行上述的音频数据的处理方法，例如用于执行图 3 所示的方法。其中，处理装置 120 可以包括获取单元 121、确定单元 122 以及生成单元 123。

获取单元 121，用于获取 m 个音频片段。确定单元 122，用于根据该 m 个音频片段确定 $m-1$ 个过渡音频信息。生成单元 123，用于根据该 m 个音频片段和该 $m-1$ 个过渡音频信息，生成目标串烧音频。其中，该 $m-1$ 个过渡音频信息用于衔接该 m 个音频片段。对于该 $m-1$ 个过渡音频信息中的第一过渡音频信息而言，该第一过渡音频信息用于衔接 m 个音频片段中排序连续的第一音频片段和第二音频片段。这里， m 个音频片段的排序是指 m 个音频片段的串烧顺序。

15 作为示例，结合图 3，获取单元 121 可以用于执行 S101，确定单元 122 以用于执行 S102，生成单元 123 可以用于执行 S103-S104。

可选的，确定单元 122，具体用于根据第一音频片段的第一信息和第二音频片段的第二信息，确定第一过渡音频信息。其中，该第一信息包括第一音频片段的 MIDI 信息和音频特征信息，该第二信息包括第二音频片段的 MIDI 信息和音频特征信息，20 该第一过渡音频信息包括第一过渡音频信息对应的第一过渡音频的 MIDI 信息。

作为示例，结合图 3，确定单元 122 可以用于执行 S102。

可选的，上述的音频特征信息包括音频片段的主旋律轨位置信息、风格标签、情感标签、节奏信息、节拍信息、或调号信息中的至少一种。

可选的，确定单元 122，具体用于根据第一音频片段的第一信息、第二音频片段的第二信息以及预设神经网络模型，确定第一过渡音频信息。

作为示例，结合图 3，确定单元 122 可以用于执行 S102。

可选的，当在目标串烧音频中，第一音频片段位于第二音频片段之前，则：上述第一过渡音频信息是基于用于表征第一过渡音频信息的特征向量确定的，该第一过渡音频信息的特征向量是基于第一向量和第二向量确定的。其中，该第一向量是根据第一信息在第一音频片段的时序末端生成的特征向量，该第二向量是根据第二信息在第二音频片段的时序始端生成的特征向量。

可选的，确定单元 122，还用于响应于用户的第一操作，确定 k 个目标音频。获取单元 121，具体用于从该 k 个目标音频中提取 m 个音频片段。其中， $2 \leq k \leq m$ ，且 k 是整数。

35 作为示例，结合图 3，确定单元 122 和获取单元 121 可以用于执行 S101，

可选的，确定单元 122，还用于在根据 m 个音频片段确定 $m-1$ 个过渡音频信息之前，确定该 m 个音频片段的串烧顺序。

可选的，确定单元 122 还用于响应于用户的第二操作，重新确定 m 个音频片段的串烧顺序。确定单元 122 还用于根据重新确定的串烧顺序和 m 个音频片段，重新确定

m-1 个过渡音频信息生成单元 123，还用于根据重新确定的 m-1 个过渡音频信息和 m 个音频片段，重新生成目标串烧音频。

可选的，处理装置 120 还包括：渲染单元 124，用于响应于用户的第三操作，对上述的目标串烧音频进行渲染。

5 作为示例，结合图 3，渲染单元 124 可以用于执行 S104。

可选的，处理装置 120 还包括：输出单元 125，用于输出上述的目标串烧音频。

关于上述可选方式的具体描述可以参见前述的方法实施例，此处不再赘述。此外，上述提供的任一种处理装置 120 的解释以及有益效果的描述均可参考上述对应的方法实施例，不再赘述。

10 作为示例，结合图 1，处理装置 120 中的获取单元 121 和输出单元 125，可以通过图 1 中的触摸屏 150 和处理器 110 实现其功能。确定单元 122、生成单元 123、渲染单元 124 可以通过图 1 中的处理 110 执行图 1 中的内部存储器 120 中的程序代码实现。

15 图 13 示出本申请实施例提供的用于承载计算机程序产品的信号承载介质的结构示意图，该信号承载介质用于存储计算机程序产品或用于存储计算设备上执行计算机进程的计算机程序。

如图 13 所示，信号承载介质 130 可以包括一个或多个程序指令，其当被一个或多个处理器运行时可以提供以上针对图 3 描述的功能或者部分功能。因此，例如，参考图 3 中 S101~S104 的一个或多个特征可以由与信号承载介质 130 相关联的一个或多个指令来承担。此外，图 13 中的程序指令也描述示例指令。

20 在一些示例中，信号承载介质 130 可以包含计算机可读介质 131，诸如但不限于，硬盘驱动器、紧光盘（CD）、数字视频光盘（DVD）、数字磁带、存储器、只读存储记忆体（read-only memory, ROM）或随机存储记忆体（random access memory, RAM）等等。

25 在一些实施方式中，信号承载介质 130 可以包含计算机可记录介质 132，诸如但不限于，存储器、读/写（R/W）CD、R/W DVD、等等。

在一些实施方式中，信号承载介质 130 可以包含通信介质 133，诸如但不限于，数字和/或模拟通信介质（例如，光纤电缆、波导、有线通信链路、无线通信链路、等等）。

30 信号承载介质 130 可以由无线形式的通信介质 133（例如，遵守 IEEE 1902.11 标准或者其它传输协议的无线通信介质）来传达。一个或多个程序指令可以是，例如，计算机可执行指令或者逻辑实施指令。

在一些示例中，诸如针对图 3 描述的音频数据的处理装置可以被配置为，响应于通过计算机可读介质 131、计算机可记录介质 132、和/或通信介质 133 中的一个或多个程序指令，提供各种操作、功能、或者动作。

35 应该理解，这里描述的布置仅仅是用于示例的目的。因而，本领域技术人员将理解，其它布置和其它元素（例如，机器、接口、功能、顺序、和功能组等等）能够被取而代之地使用，并且一些元素可以根据所期望的结果而一并省略。另外，所描述的元素中的许多是可以被实现为离散的或者分布式的组件的、或者以任何适当的组合和位置来结合其它组件实施的功能实体。

在上述实施例中，可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。当使用软件程序实现时，可以全部或部分地以计算机程序产品的形式来实现。该计算机程序产品包括一个或多个计算机指令。在计算机上和执行计算机执行指令时，全部或部分地产生按照本申请实施例的流程或功能。计算机可以是通用计算机、专用计算机、计算机网络、或者其他可编程装置。计算机指令可以存储在计算机可读存储介质中，或者从一个计算机可读存储介质向另一个计算机可读存储介质传输，例如，计算机指令可以从一个网站站点、计算机、服务器或者数据中心通过有线（例如同轴电缆、光纤、数字用户线（digital subscriber line, DSL））或无线（例如红外、无线、微波等）方式向另一个网站站点、计算机、服务器或数据中心进行传输。计算机可读存储介质可以是计算机能够存取的任何可用介质或者是包含一个或多个可以用介质集成的服务器、数据中心等数据存储设备。可用介质可以是磁性介质（例如，软盘、硬盘、磁带），光介质（例如，DVD）、或者半导体介质（例如固态硬盘（solid state disk, SSD））等。

以上所述，仅为本发明的具体实施方式，但本发明的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本发明揭露的技术范围内，可轻易想到变化或替换，都应涵盖在本发明的保护范围之内。因此，本发明的保护范围应以所述权利要求的保护范围为准。

权 利 要 求 书

1. 一种音频数据的处理方法，其特征在于，包括：

获取 m 个音频片段， m 是大于等于 2 的整数；

5 根据所述 m 个音频片段确定 $m-1$ 个过渡音频信息，所述 $m-1$ 个过渡音频信息用于衔接所述 m 个音频片段；其中，所述 $m-1$ 个过渡音频信息中的第一过渡音频信息，用于衔接所述 m 个音频片段中排序连续的第一音频片段和第二音频片段，所述排序为所述 m 个音频片段的串烧顺序；

根据所述 m 个音频片段和所述 $m-1$ 个过渡音频信息，生成目标串烧音频。

10 2. 根据权利要求 1 所述的方法，其特征在于，所述根据所述 m 个音频片段确定 $m-1$ 个过渡音频信息，包括：

根据所述第一音频片段的第一信息和所述第二音频片段的第二信息，确定所述第一过渡音频信息，所述第一过渡音频信息包括所述第一过渡音频信息对应的第一过渡音频的乐器数字接口 MIDI 信息；

15 其中，所述第一信息包括所述第一音频片段的 MIDI 信息和音频特征信息，所述第二信息包括所述第二音频片段的 MIDI 信息和音频特征信息。

3. 根据权利要求 2 所述的方法，其特征在于，所述音频特征信息包括音频片段的主旋律轨位置信息、风格标签、情感标签、节奏信息、节拍信息、或调号信息中的至少一种。

20 4. 根据权利要求 2 或 3 所述的方法，其特征在于，所述根据所述第一音频片段的第一信息和所述第二音频片段的第二信息，确定第一过渡音频信息，包括：

根据所述第一音频片段的第一信息、所述第二音频片段的第二信息以及预设神经网络模型，确定第一过渡音频信息。

5. 根据权利要求 4 所述的方法，其特征在于，当在所述目标串烧音频中，所述第一音频片段位于所述第二音频片段之前，则：

25 所述第一过渡音频信息是基于用于表征所述第一过渡音频信息的特征向量确定的，所述第一过渡音频信息的特征向量是基于第一向量和第二向量确定的；其中，所述第一向量是根据所述第一信息在所述第一音频片段的时序末端生成的特征向量，所述第二向量是根据所述第二信息在所述第二音频片段的时序始端生成的特征向量。

30 6. 根据权利要求 1-5 中任一项所述的方法，其特征在于，所述获取 m 个音频片段，包括：

响应于用户的第一操作，确定 k 个目标音频， $2 \leq k \leq m$ ，且 k 是整数；

从所述 k 个目标音频中提取所述 m 个音频片段。

7. 根据权利要求 1-6 中任一项所述的方法，其特征在于，在所述根据所述 m 个音频片段确定 $m-1$ 个过渡音频信息之前，所述方法还包括：

35 确定所述 m 个音频片段的串烧顺序。

8. 根据权利要求 1-7 中任一项所述的方法，其特征在于，所述方法还包括：

响应于用户的第二操作，重新确定所述 m 个音频片段的串烧顺序；

根据重新确定的串烧顺序和所述 m 个音频片段，重新确定 $m-1$ 个过渡音频信息；

根据重新确定的 $m-1$ 个过渡音频信息和所述 m 个音频片段，重新生成目标串烧音

频。

9. 根据权利要求 1-8 中任一项所述的方法，其特征在于，所述方法还包括：
响应于用户的第三操作，对所述目标串烧音频进行渲染。

5 10. 根据权利要求 1-9 中任一项所述的方法，其特征在于，所述方法还包括：
输出所述目标串烧音频。

11. 一种音频数据的处理装置，其特征在于，包括：

获取单元，用于获取 m 个音频片段， m 是大于等于 2 的整数；

10 确定单元，用于根据所述 m 个音频片段确定 $m-1$ 个过渡音频信息，所述 $m-1$ 个过渡音频信息用于衔接所述 m 个音频片段；其中，所述 $m-1$ 个过渡音频信息中的第一过渡音频信息，用于衔接所述 m 个音频片段中排序连续的第一音频片段和第二音频片段，所述排序为所述 m 个音频片段的串烧顺序；

生成单元，用于根据所述 m 个音频片段和所述 $m-1$ 个过渡音频信息，生成目标串烧音频。

12. 根据权利要求 11 所述的装置，其特征在于，

15 所述确定单元，具体用于根据所述第一音频片段的第一信息和所述第二音频片段的第二信息，确定所述第一过渡音频信息，所述第一过渡音频信息包括所述第一过渡音频信息对应的第一过渡音频的乐器数字接口 MIDI 信息；

其中，所述第一信息包括所述第一音频片段的 MIDI 信息和音频特征信息，所述第二信息包括所述第二音频片段的 MIDI 信息和音频特征信息。

20 13. 根据权利要求 12 所述的装置，其特征在于，所述音频特征信息包括音频片段的主旋律轨位置信息、风格标签、情感标签、节奏信息、节拍信息、或调号信息中的至少一种。

14. 根据权利要求 12 或 13 所述的装置，其特征在于，

25 所述确定单元，具体用于根据所述第一音频片段的第一信息、所述第二音频片段的第二信息以及预设神经网络模型，确定第一过渡音频信息。

15. 根据权利要求 14 所述的装置，其特征在于，当在所述目标串烧音频中，所述第一音频片段位于所述第二音频片段之前，则：

30 所述第一过渡音频信息是基于用于表征所述第一过渡音频信息的特征向量确定的，所述第一过渡音频信息的特征向量是基于第一向量和第二向量确定的；其中，所述第一向量是根据所述第一信息在所述第一音频片段的时序末端生成的特征向量，所述第二向量是根据所述第二信息在所述第二音频片段的时序始端生成的特征向量。

16. 根据权利要求 11-15 中任一项所述的装置，其特征在于，

所述确定单元，还用于响应于用户的第一操作，确定 k 个目标音频， $2 \leq k \leq m$ ，且 k 是整数；

35 所述获取单元，具体用于从所述 k 个目标音频中提取所述 m 个音频片段。

17. 根据权利要求 11-16 中任一项所述的装置，其特征在于，

所述确定单元，还用于在根据所述 m 个音频片段确定 $m-1$ 个过渡音频信息之前，确定所述 m 个音频片段的串烧顺序。

18. 根据权利要求 11-17 中任一项所述的装置，其特征在于，

所述确定单元,还用于响应于用户的第二操作,重新确定所述 m 个音频片段的串烧顺序;以及,还用于根据重新确定的串烧顺序和所述 m 个音频片段,重新确定 m-1 个过渡音频信息;

5 所述生成单元,还用于根据重新确定的 m-1 个过渡音频信息和所述 m 个音频片段,重新生成目标串烧音频。

19. 根据权利要求 11-18 中任一项所述的装置,其特征在于,所述装置还包括:渲染单元,用于响应于用户的第三操作,对所述目标串烧音频进行渲染。

20. 根据权利要求 11-19 中任一项所述的装置,其特征在于,所述装置还包括:输出单元,用于输出所述目标串烧音频。

10 21. 一种音频数据的处理装置,其特征在于,包括:一个或多个处理器和传输接口,所述一个或多个处理器通过所述传输接口接收或发送数据,所述一个或多个处理器被配置为调用存储在存储器中的程序指令,以执行如权利要求 1-10 中任一项所述的方法。

15 22. 一种计算机可读存储介质,其特征在于,所述计算机可读存储介质包括程序指令,当所述程序指令在计算机或处理器上运行时,使得所述计算机或所述处理器执行权利要求 1-10 中任一项所述的方法。

23. 一种计算机程序产品,其特征在于,当所述计算机程序产品在音频数据的处理装置上运行时,使得所述装置执行如权利要求 1-10 中任一项所述的方法。

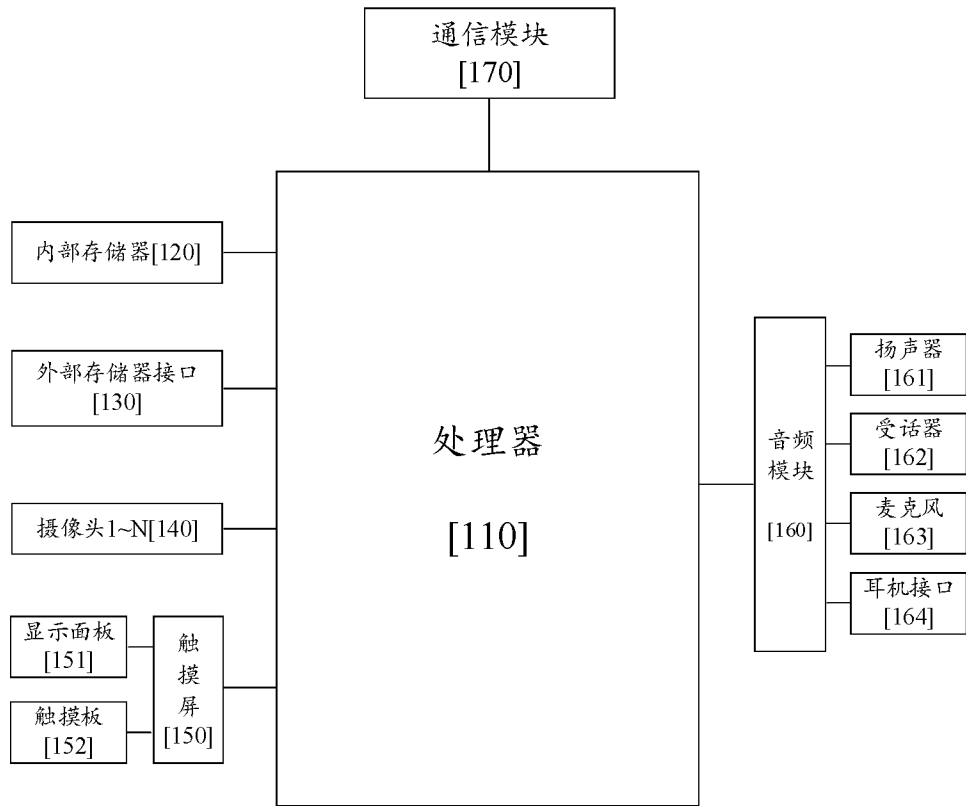


图 1

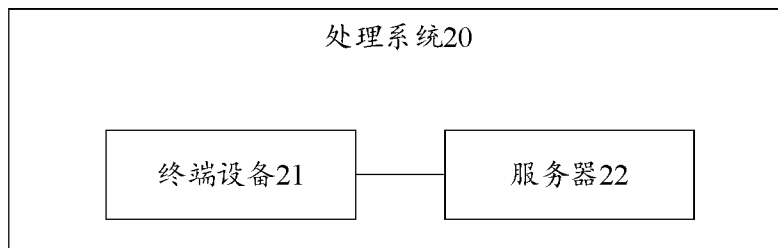


图 2

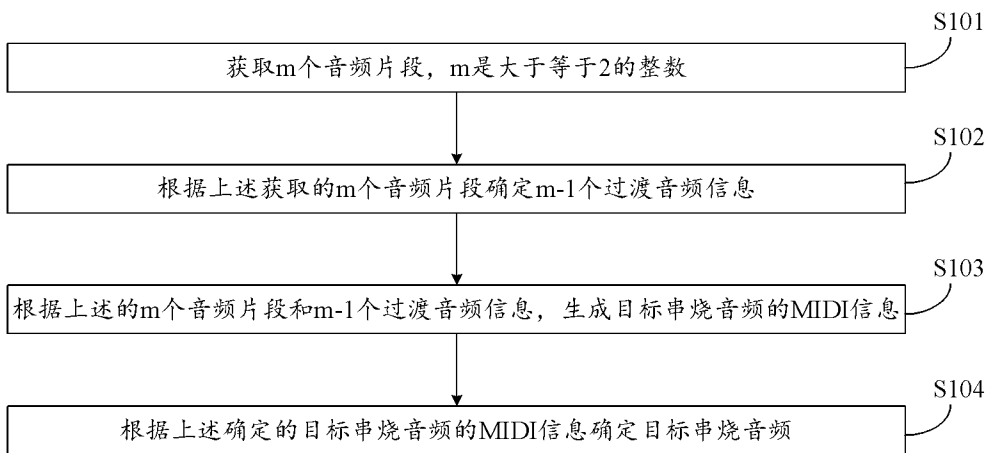


图 3

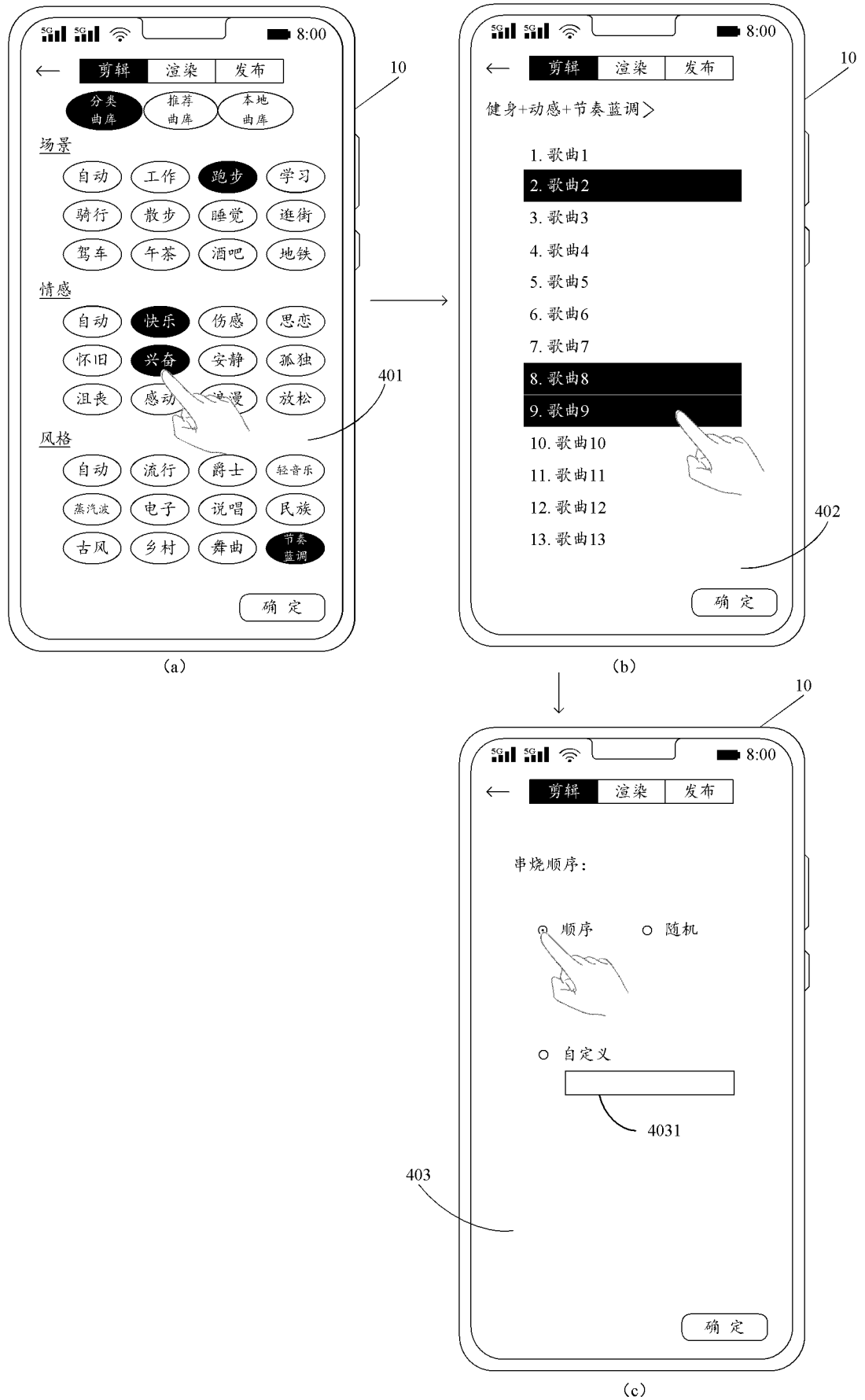


图 4

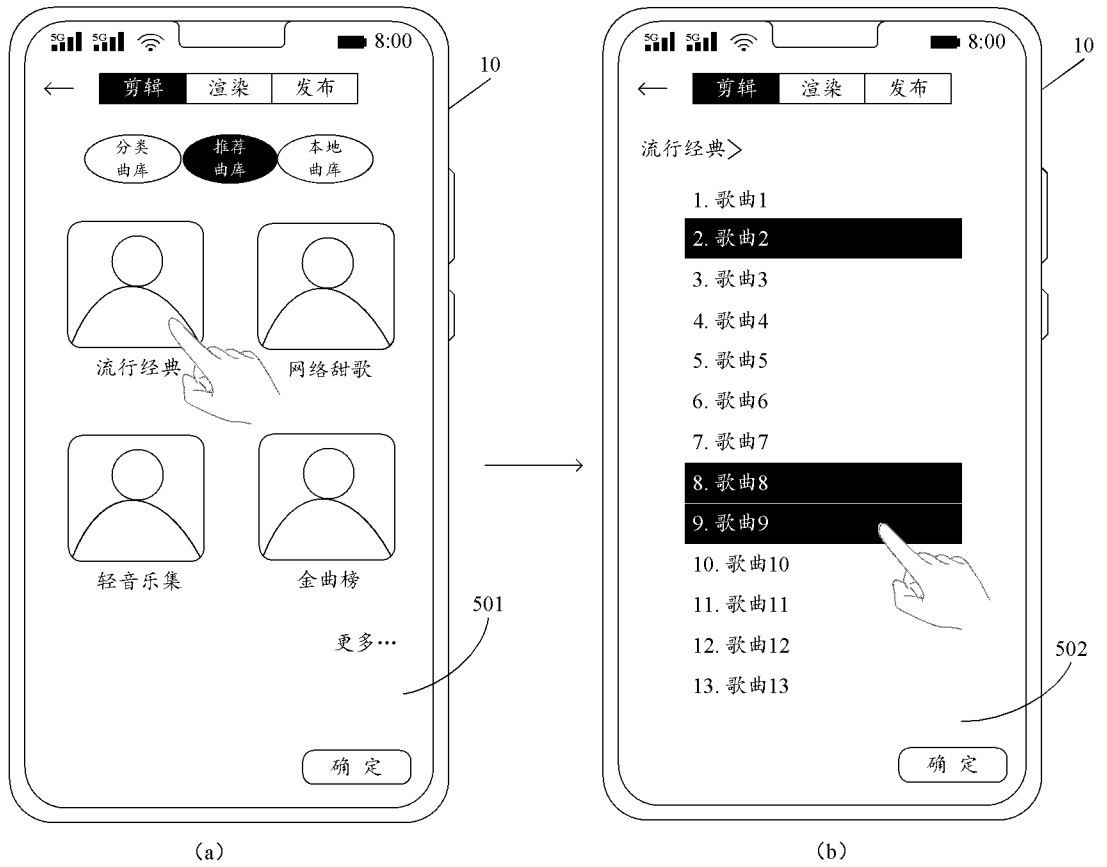


图 5

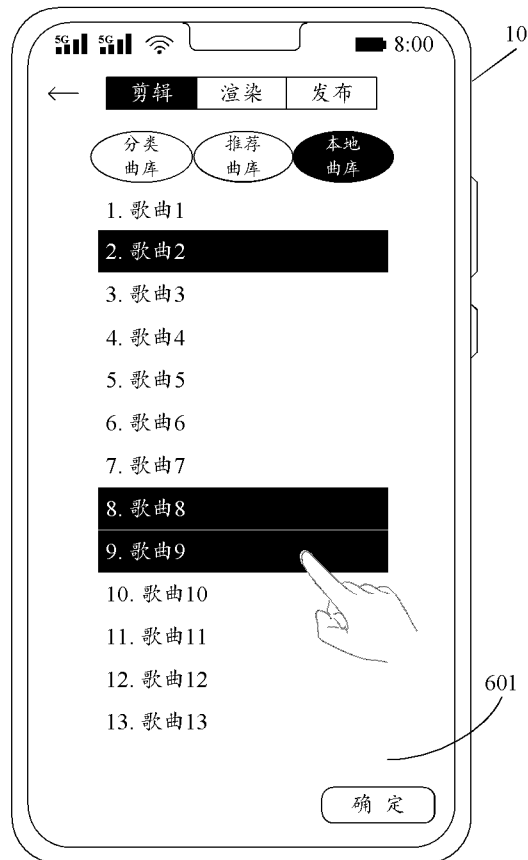


图 6

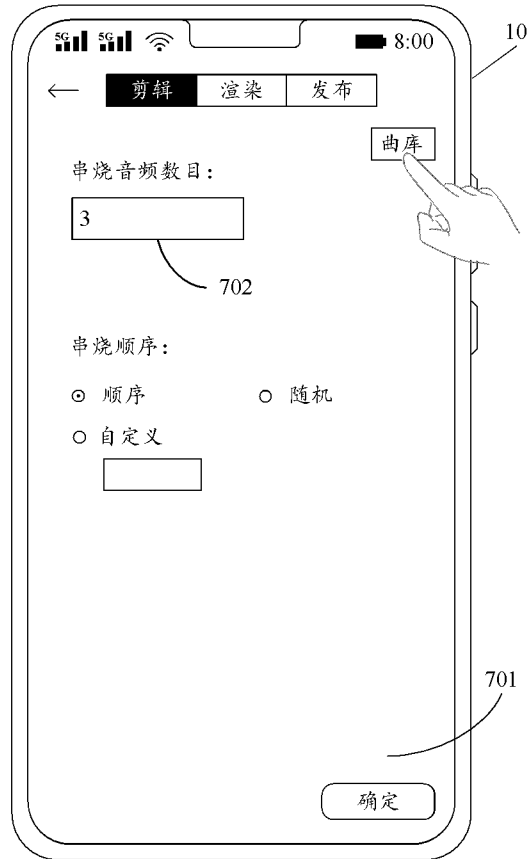


图 7

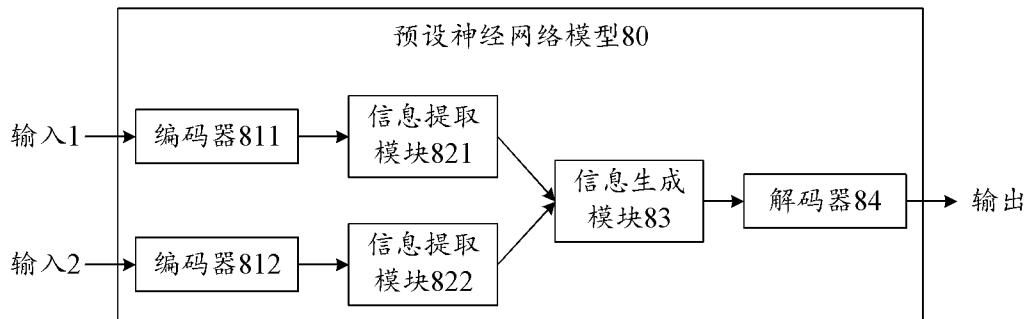


图 8

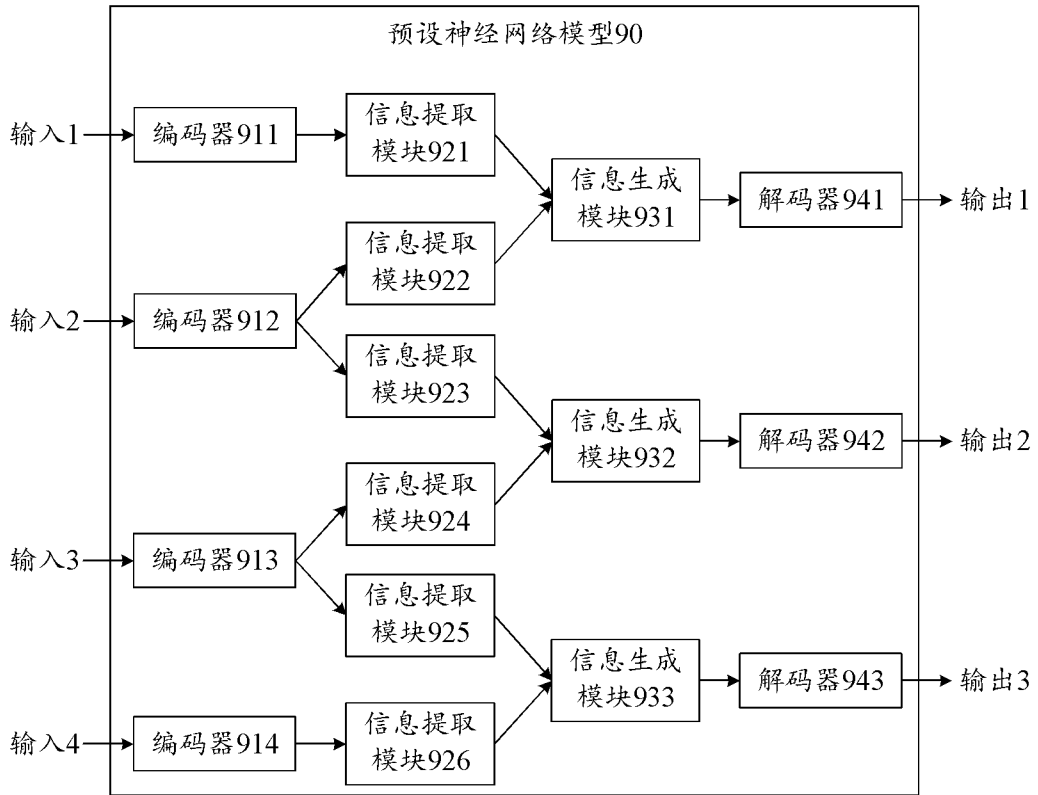


图 9

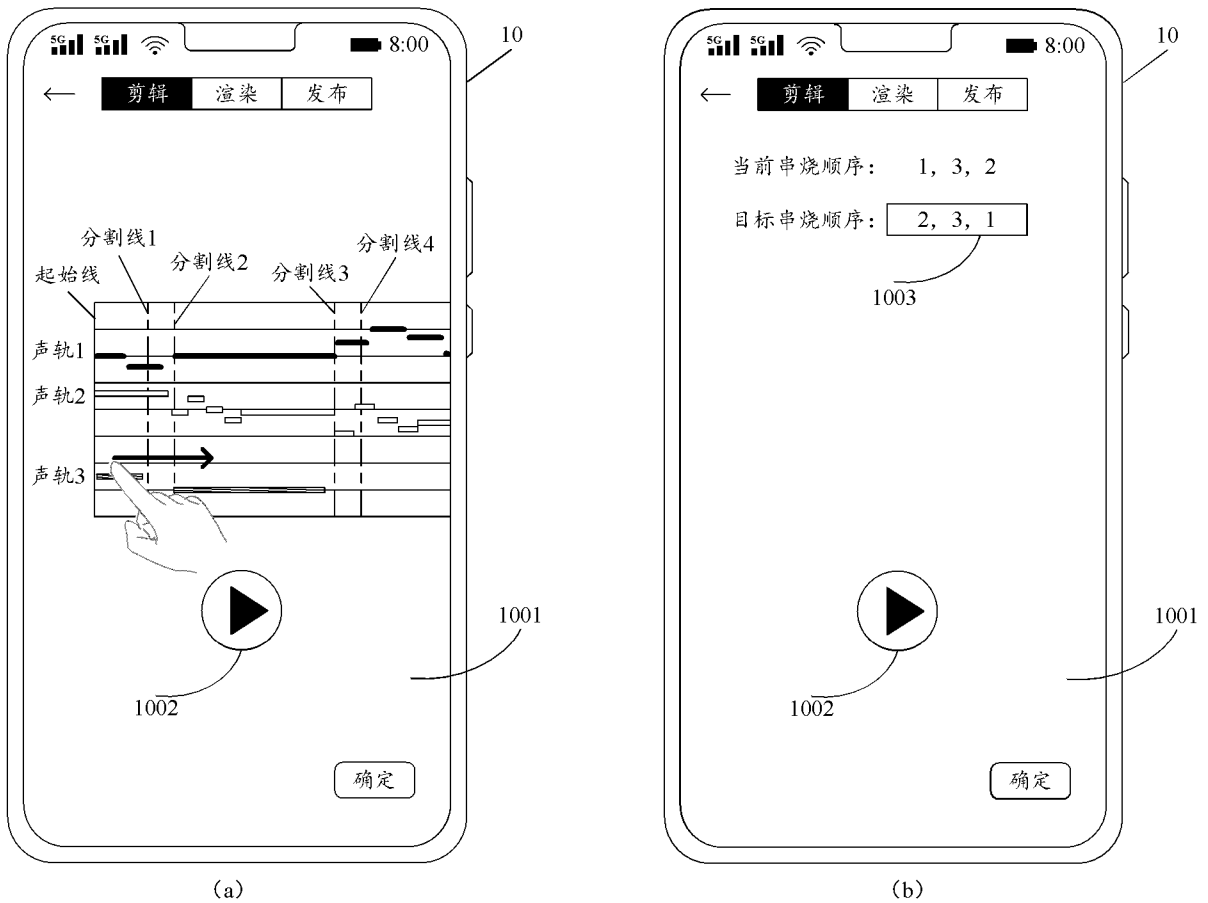


图 10

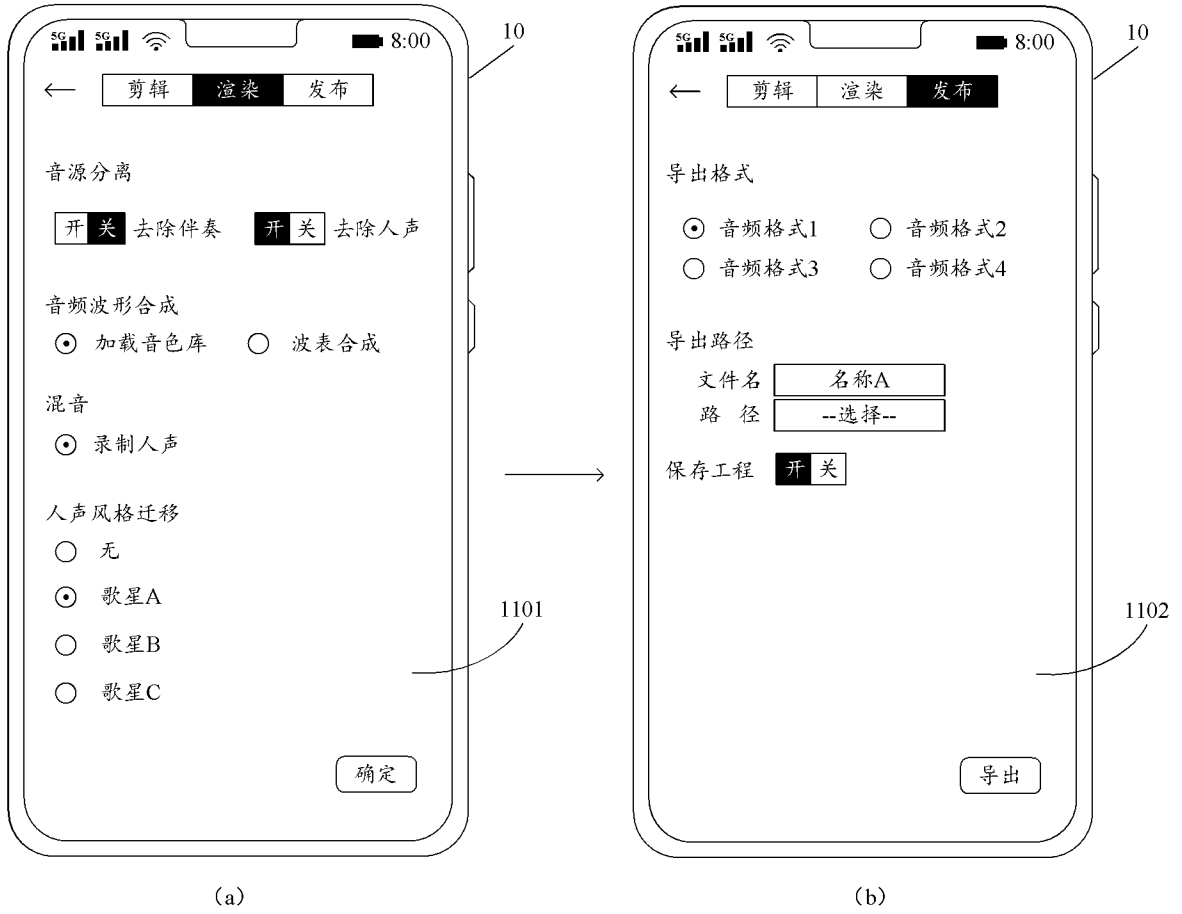


图 11

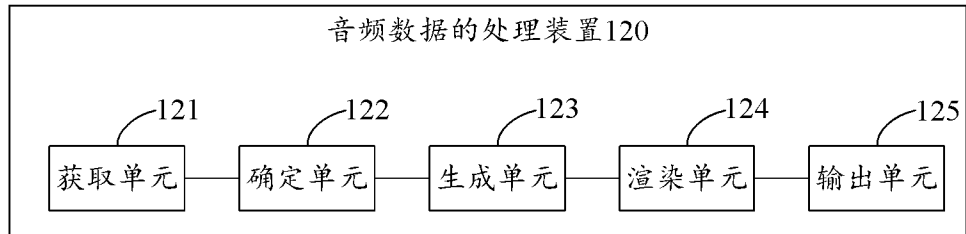


图 12

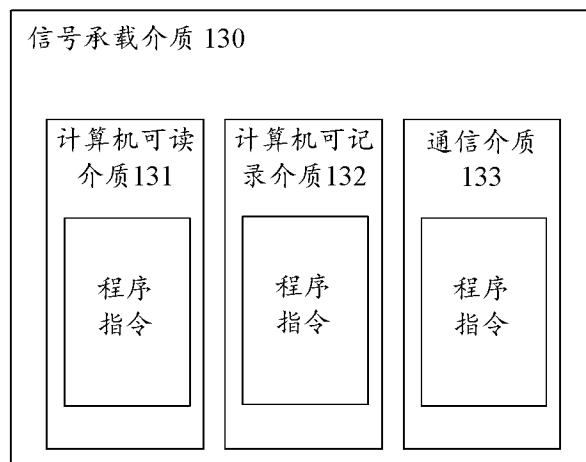


图 13

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/093923

A. CLASSIFICATION OF SUBJECT MATTER		
G10H 1/00(2006.01)i; G06F 16/68(2019.01)i		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols)		
G10H1;G06F16		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)		
CNPAT, CNKI, WPI, EPODOC: 音乐, 乐曲, 歌曲, 音频, 音轨, 乐轨, 声轨, 组合, 混音, 串烧, 拼接, 音符, 乐理, 过渡, 衔接, 渲染, 神经网络, MIDI, music, song, audio, soundtrack, track, combination, mixing, fade, splicing, note, transition, join, render, neural network		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 112037739 A (TENCENT MUSIC ENTERTAINMENT TECHNOLOGY (SHENZHEN) CO., LTD.) 04 December 2020 (2020-12-04) description, paragraphs [0037]-[0123]	1, 6-11, 16-23
Y	CN 112037739 A (TENCENT MUSIC ENTERTAINMENT TECHNOLOGY (SHENZHEN) CO., LTD.) 04 December 2020 (2020-12-04) description, paragraphs [0037]-[0123]	2-5, 12-15
Y	CN 108766407 A (TENCENT MUSIC ENTERTAINMENT TECHNOLOGY (SHENZHEN) CO., LTD.) 06 November 2018 (2018-11-06) description, paragraphs [0059]-[0137]	2-5, 12-15
X	WO 2009107137 A1 (TECHNION RESEARCH & DEVELOPMENT FOUNDATION LTD.) 03 September 2009 (2009-09-03) description, p. 8, line 1 to p. 15, line 32	1, 6-11, 16-23
X	US 2020410968 A1 (AI MUSIC LTD.) 31 December 2020 (2020-12-31) description, paragraphs [0024]-[0060]	1, 6-11, 16-23
A	US 2013282388 A1 (DOLBY INTERNATIONAL AB.) 24 October 2013 (2013-10-24) entire document	1-23
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search		Date of mailing of the international search report
26 July 2022		16 August 2022
Name and mailing address of the ISA/CN		Authorized officer
China National Intellectual Property Administration (ISA/CN) No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088, China		
Facsimile No. (86-10)62019451		Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/093923

C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 1938753 A (PANASONIC CORP.) 28 March 2007 (2007-03-28) entire document	1-23
A	US 2014254831 A1 (NIKE INC.) 11 September 2014 (2014-09-11) entire document	1-23
A	US 2019051272 A1 (COMMONEDITS, INC.) 14 February 2019 (2019-02-14) entire document	1-23

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2022/093923

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	112037739	A	04 December 2020	None			
CN	108766407	A	06 November 2018	None			
US	2020410968	A1	31 December 2020	WO	2019162703	A1	29 August 2019
				EP	3759706	A1	06 January 2021
US	2013282388	A1	24 October 2013	EP	2659483	A1	06 November 2013
				WO	2012089313	A1	05 July 2012
US	2014254831	A1	11 September 2014	KR	20170065685	A	13 June 2017
				KR	20150121178	A	28 October 2015
				JP	2017215611	A	07 December 2017
				US	2019244594	A1	08 August 2019
				WO	2014137377	A1	12 September 2014
				US	2017148425	A1	25 May 2017
				US	2021407478	A1	30 December 2021
				EP	2965310	A1	13 January 2016
				JP	2016511844	A	21 April 2016
WO	2009107137	A1	14 February 2019	None			
CN	1938753	A	28 March 2007	WO	2005096265	A1	13 October 2005
				EP	1734505	A1	20 December 2006
				KR	20060130692	A	19 December 2006
				US	2007209499	A1	13 September 2007
WO	2009107137	A1	03 September 2009	US	2017148425	A1	25 May 2017

国际检索报告

国际申请号

PCT/CN2022/093923

<p>A. 主题的分类</p> <p>G10H 1/00(2006.01)i; G06F 16/68(2019.01)i</p> <p>按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类</p>																										
<p>B. 检索领域</p> <p>检索的最低限度文献(标明分类系统和分类号)</p> <p>G10H1;G06F16</p> <p>包含在检索领域中的除最低限度文献以外的检索文献</p> <p>在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))</p> <p>CNPAT, CNKI, WPI, EPDOC: 音乐, 乐曲, 歌曲, 音频, 音轨, 乐轨, 声轨, 组合, 混音, 串烧, 拼接, 音符, 乐理, 过渡, 衔接, 渲染, 神经网络, MIDI, music, song, audio, soundtrack, track, combination, mixing, fade, splicing, note, transition, join, render, neural network</p>																										
<p>C. 相关文件</p> <table border="1"> <thead> <tr> <th>类型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>CN 112037739 A (腾讯音乐娱乐科技深圳有限公司) 2020年12月4日 (2020 - 12 - 04) 说明书第[0037]-[0123]段</td> <td>1, 6-11, 16-23</td> </tr> <tr> <td>Y</td> <td>CN 112037739 A (腾讯音乐娱乐科技深圳有限公司) 2020年12月4日 (2020 - 12 - 04) 说明书第[0037]-[0123]段</td> <td>2-5, 12-15</td> </tr> <tr> <td>Y</td> <td>CN 108766407 A (腾讯音乐娱乐科技深圳有限公司) 2018年11月6日 (2018 - 11 - 06) 说明书第[0059]-[0137]段</td> <td>2-5, 12-15</td> </tr> <tr> <td>X</td> <td>WO 2009107137 A1 (TECHNION RESEARCH & DEVELOPMENT FOUNDATION LTD.) 2009年9月3日 (2009 - 09 - 03) 说明书第8页第1行-第15页第32行</td> <td>1, 6-11, 16-23</td> </tr> <tr> <td>X</td> <td>US 2020410968 A1 (AI MUSIC LTD.) 2020年12月31日 (2020 - 12 - 31) 说明书第[0024]-[0060]段</td> <td>1, 6-11, 16-23</td> </tr> <tr> <td>A</td> <td>US 2013282388 A1 (DOLBY INTERNATIONAL AB.) 2013年10月24日 (2013 - 10 - 24) 全文</td> <td>1-23</td> </tr> <tr> <td>A</td> <td>CN 1938753 A (松下电器产业株式会社) 2007年3月28日 (2007 - 03 - 28) 全文</td> <td>1-23</td> </tr> </tbody> </table>			类型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	CN 112037739 A (腾讯音乐娱乐科技深圳有限公司) 2020年12月4日 (2020 - 12 - 04) 说明书第[0037]-[0123]段	1, 6-11, 16-23	Y	CN 112037739 A (腾讯音乐娱乐科技深圳有限公司) 2020年12月4日 (2020 - 12 - 04) 说明书第[0037]-[0123]段	2-5, 12-15	Y	CN 108766407 A (腾讯音乐娱乐科技深圳有限公司) 2018年11月6日 (2018 - 11 - 06) 说明书第[0059]-[0137]段	2-5, 12-15	X	WO 2009107137 A1 (TECHNION RESEARCH & DEVELOPMENT FOUNDATION LTD.) 2009年9月3日 (2009 - 09 - 03) 说明书第8页第1行-第15页第32行	1, 6-11, 16-23	X	US 2020410968 A1 (AI MUSIC LTD.) 2020年12月31日 (2020 - 12 - 31) 说明书第[0024]-[0060]段	1, 6-11, 16-23	A	US 2013282388 A1 (DOLBY INTERNATIONAL AB.) 2013年10月24日 (2013 - 10 - 24) 全文	1-23	A	CN 1938753 A (松下电器产业株式会社) 2007年3月28日 (2007 - 03 - 28) 全文	1-23
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求																								
X	CN 112037739 A (腾讯音乐娱乐科技深圳有限公司) 2020年12月4日 (2020 - 12 - 04) 说明书第[0037]-[0123]段	1, 6-11, 16-23																								
Y	CN 112037739 A (腾讯音乐娱乐科技深圳有限公司) 2020年12月4日 (2020 - 12 - 04) 说明书第[0037]-[0123]段	2-5, 12-15																								
Y	CN 108766407 A (腾讯音乐娱乐科技深圳有限公司) 2018年11月6日 (2018 - 11 - 06) 说明书第[0059]-[0137]段	2-5, 12-15																								
X	WO 2009107137 A1 (TECHNION RESEARCH & DEVELOPMENT FOUNDATION LTD.) 2009年9月3日 (2009 - 09 - 03) 说明书第8页第1行-第15页第32行	1, 6-11, 16-23																								
X	US 2020410968 A1 (AI MUSIC LTD.) 2020年12月31日 (2020 - 12 - 31) 说明书第[0024]-[0060]段	1, 6-11, 16-23																								
A	US 2013282388 A1 (DOLBY INTERNATIONAL AB.) 2013年10月24日 (2013 - 10 - 24) 全文	1-23																								
A	CN 1938753 A (松下电器产业株式会社) 2007年3月28日 (2007 - 03 - 28) 全文	1-23																								
<p><input checked="" type="checkbox"/> 其余文件在C栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。</p> <p>* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件</p>																										
<p>国际检索实际完成的日期</p> <p>2022年7月26日</p>		<p>国际检索报告邮寄日期</p> <p>2022年8月16日</p>																								
<p>ISA/CN的名称和邮寄地址</p> <p>中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451</p>		<p>授权官员</p> <p>王玥</p> <p>电话号码 86-(10)-53962545</p>																								

C. 相关文件		
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求
A	US 2014254831 A1 (NIKE INC.) 2014年9月11日 (2014 - 09 - 11) 全文	1-23
A	US 2019051272 A1 (COMMONEDITS INC.) 2019年2月14日 (2019 - 02 - 14) 全文	1-23

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2022/093923

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	112037739	A	2020年12月4日	无			
CN	108766407	A	2018年11月6日	无			
US	2020410968	A1	2020年12月31日	WO	2019162703	A1	2019年8月29日
				EP	3759706	A1	2021年1月6日
US	2013282388	A1	2013年10月24日	EP	2659483	A1	2013年11月6日
				WO	2012089313	A1	2012年7月5日
US	2014254831	A1	2014年9月11日	KR	20170065685	A	2017年6月13日
				KR	20150121178	A	2015年10月28日
				JP	2017215611	A	2017年12月7日
				US	2019244594	A1	2019年8月8日
				WO	2014137377	A1	2014年9月12日
				US	2017148425	A1	2017年5月25日
				US	2021407478	A1	2021年12月30日
				EP	2965310	A1	2016年1月13日
				JP	2016511844	A	2016年4月21日
WO	2009107137	A1	2019年2月14日	无			
CN	1938753	A	2007年3月28日	WO	2005096265	A1	2005年10月13日
				EP	1734505	A1	2006年12月20日
				KR	20060130692	A	2006年12月19日
				US	2007209499	A1	2007年9月13日
WO	2009107137	A1	2009年9月3日	US	2017148425	A1	2017年5月25日