(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 110009716 A
(43)申请公布日 2019.07.12

(21)申请号 201910241196.6

(22)申请日 2019.03.28

(71)申请人 网易(杭州)网络有限公司
地址 310052 浙江省杭州市滨江区网商路599号网易大厦

(72)发明人 袁燚 田冠中

(74)专利代理机构 北京同立钧成知识产权代理有限公司 11205
代理人 张宁 刘芳

(51)Int.Cl.
G06T 13/40(2011.01)
G10L 15/02(2006.01)

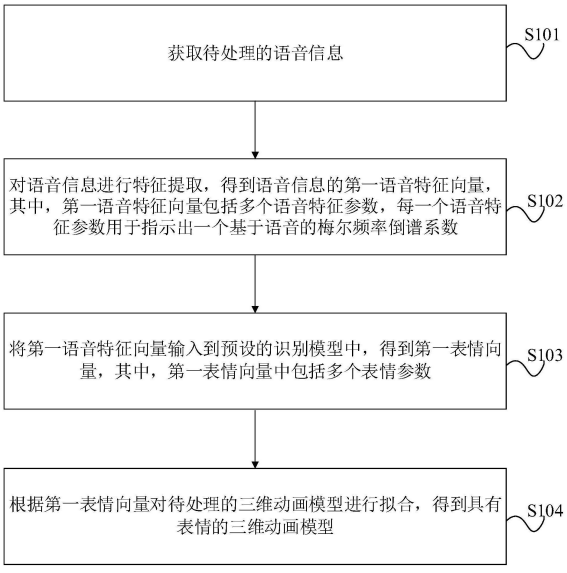
权利要求书2页 说明书12页 附图4页

(54)发明名称

面部表情的生成方法、装置、电子设备及存储介质

(57)摘要

本申请提供一种面部表情的生成方法、装置、电子设备及存储介质，该方法包括：对获取的语音信息进行特征提取，得到第一语音特征向量，第一语音特征向量包括多个语音特征参数，每一个语音特征参数指示出一个基于语音的梅尔频率倒谱系数；将第一语音特征向量输入到预设的识别模型中得到第一表情向量，第一表情向量中包括多个表情参数；根据第一表情向量对三维动画模型进行拟合，得到具有表情的三维动画模型。基于语音的梅尔频率倒谱系数可以保留语音的内容信息、音高、音色等等语音特征，可以更好的还原出表情动作；使得三维动画模型的面部表情更加正确，可以较好的对面部表情进行模拟，得到的表情动作是符合语音信息和语音数据的。



CN 110009716 A

1.一种面部表情的生成方法,其特征在于,包括:

获取待处理的语音信息,并对所述语音信息进行特征提取,得到所述语音信息的第一语音特征向量,其中,所述第一语音特征向量包括多个语音特征参数,每一个所述语音特征参数用于指示出一个基于语音的梅尔频率倒谱系数;

将所述第一语音特征向量输入到预设的识别模型中,得到第一表情向量,其中,所述第一表情向量中包括多个表情参数;

根据所述第一表情向量对待处理的三维动画模型进行拟合,得到具有表情的三维动画模型。

2.根据权利要求1所述的方法,其特征在于,对所述语音信息进行特征提取,得到所述语音信息的第一语音特征向量,包括:

对所述语音信息进行快速傅氏变换处理,得到频谱值;

对所述频谱值进行梅尔滤波处理,得到梅尔频谱信息;

对所述梅尔频谱信息进行倒谱分析处理,得到多个所述基于语音的梅尔频率倒谱系数,其中,多个所述基于语音的梅尔频率倒谱系数构成所述第一语音特征向量。

3.根据权利要求1所述的方法,其特征在于,根据所述第一表情向量对待处理的三维动画模型进行拟合,得到具有表情的三维动画模型,包括:

对所述第一表情向量进行平滑处理,得到平滑处理后的第一表情向量;

根据所述平滑处理后的第一表情向量对待处理的三维动画模型进行拟合,得到具有表情的三维动画模型。

4.根据权利要求1-3任一项所述的方法,其特征在于,在所述获取待处理的语音信息之前,还包括:

获取待训练的至少一个第二语音特征向量,其中,每一个所述第二语音特征向量中包括多个待训练的基于语音的梅尔频率倒谱系数;

获取每一个所述第二语音特征向量对应的第二表情向量,其中,每一个所述第二表情向量中包括多个待训练的表情参数,每一个所述第二表情向量用于指示出一种面部表情;

将所述至少一个第二语音特征向量和每一个所述第二语音特征向量对应的第二表情向量,输入到初始的识别模型中,得到所述预设的识别模型。

5.根据权利要求4所述的方法,其特征在于,获取待训练的至少一个第二语音特征向量,包括:

获取用户的动态视频,并获取所述动态视频中的用户在各时间段下的语音信息;

对每一个时间段下的语音信息进行特征提取,得到每一个时间段下的第二语音特征向量。

6.根据权利要求5所述的方法,其特征在于,获取每一个所述第二语音特征向量对应的第二表情向量,包括:

捕捉所述用户在所述每一个时间段下的面部表情,以得到每一个时间段下的第二表情向量,其中,第二语音特征向量与第二表情向量是一一对应的。

7.根据权利要求1-3任一项所述的方法,其特征在于,在根据所述第一表情向量对待处理的三维动画模型进行拟合,得到具有表情的三维动画模型之前,还包括:

采集用户的多种面部特征信息,其中,每一种所述面部特征信息中包括至少一种视觉

特征信息和至少一种面部深度信息；

根据所述多种面部特征信息，对所述三维动画模型进行调整，得到调整后的三维动画模型。

8.根据权利要求1-3任一项所述的方法，其特征在于，所述识别模型为双向长短期记忆Bi-LSTM神经网络模型。

9.一种面部表情的生成装置，其特征在于，包括：

第一获取模块，用于获取待处理的语音信息；

提取模块，用于对所述语音信息进行特征提取，得到所述语音信息的第一语音特征向量，其中，所述第一语音特征向量包括多个语音特征参数，每一个所述语音特征参数用于指示出一个基于语音的梅尔频率倒谱系数；

识别模块，用于将所述第一语音特征向量输入到预设的识别模型中，得到第一表情向量，其中，所述第一表情向量中包括多个表情参数；

拟合模块，用于根据所述第一表情向量对待处理的三维动画模型进行拟合，得到具有表情的三维动画模型。

10.一种电子设备，其特征在于，包括：存储器和处理器，存储器中存储有所述处理器的可执行指令；其中，所述处理器配置为经由执行所述可执行指令来执行权利要求1-8中任一项所述的方法。

11.一种计算机可读存储介质，其上存储有计算机程序，其特征在于，该程序被处理器执行时实现权利要求1-8中任一项所述方法。

面部表情的生成方法、装置、电子设备及存储介质

技术领域

[0001] 本申请涉及动画技术领域，尤其涉及一种面部表情的生成方法、装置、电子设备及存储介质。

背景技术

[0002] 随着三维动画技术的发展，三维动画受到越来越多用户的喜爱，并开始得到广泛的发展和应用。三维动画模型需要实现各种动作，其中，三维动画模型的面部需要面部表情的动作。从而，需要对三维动画模型的面部进行处理，以得到面部表情的动作。

[0003] 现有技术中，可以采用语音对三维动画模型的面部进行拟合，进而得到具有表情的三维动画模型。首先，对原始的语音数据进行特征提取，得到音素共振峰特征；然后，采用音素共振峰特征对三维动画模型的面部进行驱动，进而得到表情动作，从而得到具有表情的三维动画模型。

[0004] 然而现有技术中，对三维动画模型的面部进行驱动以得到表情动作的时候，只能采用音素共振峰特征；音素共振峰特征只包含语音的内容信息，缺失了音高和音色；从而得到的表情动作并不完整，并且，得到的表情动作与当前的语音数据所对应的真实表情是不符合的。进而，无法较好的对面部表情进行模拟，得到的三维动画模型的面部表情不正确；并且，得到的三维动画模型的面部表情与对应的语音数据并不相符合。

发明内容

[0005] 本申请提供一种面部表情的生成方法、装置、电子设备及存储介质，以解决现有技术中无法较好的对面部表情进行模拟，得到的三维动画模型的面部表情不正确，得到的三维动画模型的面部表情与对应的语音数据并不相符合的问题。

[0006] 第一方面，本申请实施例提供一种面部表情的生成方法，包括：

[0007] 获取待处理的语音信息，并对所述语音信息进行特征提取，得到所述语音信息的第一语音特征向量，其中，所述第一语音特征向量包括多个语音特征参数，每一个所述语音特征参数用于指示出一个基于语音的梅尔频率倒谱系数；

[0008] 将所述第一语音特征向量输入到预设的识别模型中，得到第一表情向量，其中，所述第一表情向量中包括多个表情参数；

[0009] 根据所述第一表情向量对待处理的三维动画模型进行拟合，得到具有表情的三维动画模型。

[0010] 可选的，对所述语音信息进行特征提取，得到所述语音信息的第一语音特征向量，包括：

[0011] 对所述语音信息进行快速傅氏变换处理，得到频谱值；

[0012] 对所述频谱值进行梅尔滤波处理，得到梅尔频谱信息；

[0013] 对所述梅尔频谱信息进行倒谱分析处理，得到多个所述基于语音的梅尔频率倒谱系数，其中，多个所述基于语音的梅尔频率倒谱系数构成所述第一语音特征向量。

[0014] 可选的，根据所述第一表情向量对待处理的三维动画模型进行拟合，得到具有表情的三维动画模型，包括：

[0015] 对所述第一表情向量进行平滑处理，得到平滑处理后的第一表情向量；

[0016] 根据所述平滑处理后的第一表情向量对待处理的三维动画模型进行拟合，得到具有表情的三维动画模型。

[0017] 可选的，在所述获取待处理的语音信息之前，还包括：

[0018] 获取待训练的至少一个第二语音特征向量，其中，每一个所述第二语音特征向量中包括多个待训练的基于语音的梅尔频率倒谱系数；

[0019] 获取每一个所述第二语音特征向量对应的第二表情向量，其中，每一个所述第二表情向量中包括多个待训练的表情参数，每一个所述第二表情向量用于指示出一种面部表情；

[0020] 将所述至少一个第二语音特征向量和每一个所述第二语音特征向量对应的第二表情向量，输入到初始的识别模型中，得到所述预设的识别模型。

[0021] 可选的，获取待训练的至少一个第二语音特征向量，包括：

[0022] 获取用户的动态视频，并获取所述动态视频中的用户在各时间段下的语音信息；

[0023] 对每一个时间段下的语音信息进行特征提取，得到每一个时间段下的第二语音特征向量。

[0024] 可选的，获取每一个所述第二语音特征向量对应的第二表情向量，包括：

[0025] 捕捉所述用户在所述每一个时间段下的面部表情，以得到每一个时间段下的第二表情向量，其中，第二语音特征向量与第二表情向量是一一对应的。

[0026] 可选的，在根据所述第一表情向量对待处理的三维动画模型进行拟合，得到具有表情的三维动画模型之前，还包括：

[0027] 采集用户的多种面部特征信息，其中，每一种所述面部特征信息中包括至少一种视觉特征信息和至少一种面部深度信息；

[0028] 根据所述多种面部特征信息，对所述三维动画模型进行调整，得到调整后的三维动画模型。

[0029] 可选的，所述识别模型为双向长短期记忆Bi-LSTM神经网络模型。

[0030] 第二方面，本申请实施例提供一种面部表情的生成装置，包括：

[0031] 第一获取模块，用于获取待处理的语音信息；

[0032] 提取模块，用于对所述语音信息进行特征提取，得到所述语音信息的第一语音特征向量，其中，所述第一语音特征向量包括多个语音特征参数，每一个所述语音特征参数用于指示出一个基于语音的梅尔频率倒谱系数；

[0033] 识别模块，用于将所述第一语音特征向量输入到预设的识别模型中，得到第一表情向量，其中，所述第一表情向量中包括多个表情参数；

[0034] 拟合模块，用于根据所述第一表情向量对待处理的三维动画模型进行拟合，得到具有表情的三维动画模型。

[0035] 可选的，所述提取模块，包括：

[0036] 变换子模块，用于对所述语音信息进行快速傅氏变换处理，得到频谱值；

[0037] 滤波子模块，用于对所述频谱值进行梅尔滤波处理，得到梅尔频谱信息；

[0038] 处理子模块，用于对所述梅尔频谱信息进行倒谱分析处理，得到多个所述基于语音的梅尔频率倒谱系数，其中，多个所述基于语音的梅尔频率倒谱系数构成所述第一语音特征向量。

[0039] 可选的，所述拟合模块，包括：

[0040] 平滑子模块，用于对所述第一表情向量进行平滑处理，得到平滑处理后的第一表情向量；

[0041] 拟合子模块，用于根据所述平滑处理后的第一表情向量对待处理的三维动画模型进行拟合，得到具有表情的三维动画模型。

[0042] 可选的，所述装置，还包括：

[0043] 第二获取模块，用于在所述第一获取模块获取待处理的语音信息之前，获取待训练的至少一个第二语音特征向量，其中，每一个所述第二语音特征向量中包括多个待训练的基于语音的梅尔频率倒谱系数；

[0044] 第三获取模块，用于获取每一个所述第二语音特征向量对应的第二表情向量，其中，每一个所述第二表情向量中包括多个待训练的表情参数，每一个所述第二表情向量用于指示出一种面部表情；

[0045] 训练模块，用于将所述至少一个第二语音特征向量和每一个所述第二语音特征向量对应的第二表情向量，输入到初始的识别模型中，得到所述预设的识别模型。

[0046] 可选的，所述第二获取模块，包括：

[0047] 获取子模块，用于获取用户的动态视频，并获取所述动态视频中的用户在各时间段下的语音信息；

[0048] 提取子模块，用于对每一个时间段下的语音信息进行特征提取，得到每一个时间段下的第二语音特征向量。

[0049] 可选的，所述第三获取模块，具体用于：

[0050] 捕捉所述用户在所述每一个时间段下的面部表情，以得到每一个时间段下的第二表情向量，其中，第二语音特征向量与第二表情向量是一一对应的。

[0051] 可选的，所述装置，还包括：

[0052] 采集模块，用于在所述拟合模块根据所述第一表情向量对待处理的三维动画模型进行拟合，得到具有表情的三维动画模型之前，采集用户的多种面部特征信息，其中，每一种所述面部特征信息中包括至少一种视觉特征信息和至少一种面部深度信息；

[0053] 调整模块，用于根据所述多种面部特征信息，对所述三维动画模型进行调整，得到调整后的三维动画模型。

[0054] 可选的，所述识别模型为双向长短期记忆Bi-LSTM神经网络模型。

[0055] 第三方面，本申请实施例提供一种电子设备，包括：存储器和处理器，存储器中存储有所述处理器的可执行指令；其中，所述处理器配置为经由执行所述可执行指令来执行第一方面中任一项所述的方法。

[0056] 第四方面，本申请实施例提供一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现第一方面任一项所述的方法。

[0057] 本申请提供一种面部表情的生成方法、装置、电子设备及存储介质，通过对语音信息进行特征提取，得到语音信息的第一语音特征向量，第一语音特征向量包括多个语音特

征参数，并且，每一个语音特征参数是一个基于语音的梅尔频率倒谱系数；将第一语音特征向量输入到预设的识别模型中，得到第一表情向量，第一表情向量中包括多个表情参数；根据第一表情向量对待处理的三维动画模型进行拟合，得到具有表情的三维动画模型。由于基于语音的梅尔频率倒谱系数可以保留语音的内容信息、音高、音色等等语音特征，从而最大程度的保证了语音的特点，基于这样的语音得到的表情参数可以更好的还原出表情动作；使得三维动画模型的面部表情更加正确，可以较好的对面部表情进行模拟，并且，得到的表情动作是符合语音信息和语音数据的。

附图说明

[0058] 为了更清楚地说明本申请实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作一简单地介绍，显而易见地，下面描述中的附图是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

[0059] 图1为本申请实施例提供的一种面部表情的生成方法的流程图；

[0060] 图2为本申请实施例提供的另一种面部表情的生成方法的流程图；

[0061] 图3为本申请实施例提供的一种面部表情的生成装置的结构示意图；

[0062] 图4为本申请实施例提供的另一种面部表情的生成装置的结构示意图；

[0063] 图5为本申请实施例提供的一种电子设备的结构示意图。

[0064] 通过上述附图，已示出本申请明确的实施例，后文中将有更详细的描述。这些附图和文字描述并不是为了通过任何方式限制本申请构思的范围，而是通过参考特定实施例为本领域技术人员说明本申请的概念。

具体实施方式

[0065] 为使本申请实施例的目的、技术方案和优点更加清楚，下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

[0066] 本申请的说明书和权利要求书及上述附图中的术语“第一”、“第二”、“第三”“第四”等(如果存在)是用于区别类似的对象，而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换，以便这里描述的本申请的实施例例如能够以除了在这里图示或描述的那些以外的顺序实施。此外，术语“包括”和“具有”以及他们的任何变形，意图在于覆盖不排他的包含，例如，包含了一系列步骤或单元的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元，而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

[0067] 下面以具体地实施例对本申请的技术方案进行详细说明。下面这几个具体的实施例可以相互结合，对于相同或相似的概念或过程可能在某些实施例不再赘述。

[0068] 随着三维动画技术的发展，三维动画受到越来越多用户的喜爱，并开始得到广泛的发展和应用。例如，将三维动画应用到虚拟现实领域、影视娱乐领域、辅助教学领域、等等。三维动画模型需要实现各种动作，其中，三维动画模型的面部需要面部表情的动作。如

何让三维动画模型具有逼真的表情、流畅自然的脸部动作变化，是一个较难解决的问题。

[0069] 通过对多种表情拟合技术的实践，人们发现通过语音来生成表情动作是一种比较合适技术方式，可以得到更生动自然的人脸动画。语音中的包含了重音、情感等因素，采用语音驱动人脸去自然生动地变化，可以极大地优化虚拟现实的展示与交互，使得三维动画模型的面部表情更加的生动；例如，采用语音得到面部表情，进而去驱动的三维动画模型，可以提高虚拟会议、游戏、个人虚拟助手、教育辅导等多个领域的用户体验。

[0070] 现在，采用语音得到面部表情的方法大致包括以下几种：一种是，基于声道信息或发音音素，提取口型特征的映射，进而得到与不同的语音场景对应的口型动作；另一种是，利用基于语音参数和生理的脸部模型，构建出三维动画模型的表情动作；还有一种是，利用脸部动作与表情向量融合，得到与不同的情感对应的面部表情。

[0071] 现有的基于语音的面部表情的生成算法包括以下过程：获取原始的语音数据，提出语音数据中的音素共振峰特征；然后，将整体的音素共振峰特征以2倍重叠分段截取，得到多段音素共振峰特征，其中，每一段音素共振峰特征是520毫秒(ms)的语音特征；然后，将每一段音素共振峰特征作为单个语音窗口，输入到一个5层的卷积神经网络模型中，得到分析后的每一段音素共振峰特征；然后，将每一段音素共振峰特征与每一种表情信息，人工的进行一一对应；将对应后的音素共振峰特征和表情信息输入到两层的全连接神经网络模型中，得到训练后的全连接神经网络模型；最后，对待处理的语音数据进行特征提取，得到待处理的音素共振峰特征；将待处理的音素共振峰特征输入到训练后的全连接神经网络模型中，得到表情参数；采用表情参数就可以驱动三维动画模型的面部表情了。

[0072] 然而上述方法，由于提取的音素共振峰特征只包含语音的内容信息，缺失了音高、音色等信息；从而得到的表情动作并不完整，并且，得到的表情动作与当前的语音数据所对应的真实表情是不符合的。并且，在对全连接神经网络模型进行训练的时候，采用的每一段音素共振峰特征是基于0.5秒左右的语音而得到的，那么每一段音素共振峰特征所对应的表情信息也是0.5秒左右的；但是一个表情的持续时间会超过0.5秒，从而基于这样的时段对语音特征和表情信息进行划分，每一段音素共振峰特征不足表征出对应的表情和情感是什么，进而训练得到的神经网络模型也并不正确；导致拟合得到的三维动画模型的面部表情不正确。

[0073] 本申请提供一种面部表情的生成方法、装置、电子设备及存储介质，可以最大程度的保证了语音的特点，基于这样的语音得到的表情参数可以更好的还原出表情动作；使得三维动画模型的面部表情更加正确，可以较好的对面部表情进行模拟，并且，得到的表情动作是符合语音信息和语音数据的。

[0074] 下面以具体地实施例对本申请的技术方案以及本申请的技术方案如何解决上述技术问题进行详细说明。下面这几个具体的实施例可以相互结合，对于相同或相似的概念或过程可能在某些实施例中不再赘述。下面将结合附图，对本申请的实施例进行描述。

[0075] 图1为本申请实施例提供的一种面部表情的生成方法的流程图，如图1所示，本实施例中的方法可以包括：

[0076] S101、获取待处理的语音信息。

[0077] 本实施例中，本申请实施例的执行主体可以是终端设备、或者服务器、或者面部表情的生成装置或设备、或者其他可以执行本申请提供的方法的装置或设备。

[0078] 首先获取语音信息，例如，用户可以发出语音，并采用录音设备采集用户发出的语音，进而就可以获取到语音信息。

[0079] S102、对语音信息进行特征提取，得到语音信息的第一语音特征向量，其中，第一语音特征向量包括多个语音特征参数，每一个语音特征参数用于指示出一个基于语音的梅尔频率倒谱系数。

[0080] 可选的，步骤S102包括以下步骤：

[0081] 第一步、对语音信息进行快速傅氏变换处理，得到频谱值。

[0082] 第二步、对频谱值进行梅尔滤波处理，得到梅尔频谱信息。

[0083] 第三步、对梅尔频谱信息进行倒谱分析处理，得到多个基于语音的梅尔频率倒谱系数，其中，多个基于语音的梅尔频率倒谱系数构成第一语音特征向量。

[0084] 本实施例中，对语音信息进行特征提取的处理，以得到一个第一语音特征向量，第一语音特征向量包括了多个语音特征参数，其中，每一个语音特征参数可以是一个基于语音的梅尔频率倒谱系数；其中，基于语音的梅尔频率倒谱系数可以保留语音的内容信息、音高、音色等等语音特征。

[0085] 具体来说，首先，对语音信息依次进行预加重、分帧、加窗处理。由于语音信息可以配置有一个短时分析窗，从而，后续的处理过程是对每一个时间段内的语音信息进行处理，其中，该时间段为短时分析窗所指示的时间，优选的，该时间段大于现有的时间窗口的时间，例如，该时间段大于0.5秒。由于短时分析窗的时间段可以大于现有的时间窗口的时间，则后得到的第一语音特征向量可以对应出时间超长的表情。基于这样的时段对语音特征和表情信息进行划分，语音信息可以更好的表征出对应的表情和情感是什么，可以更好的对识别模型进行训练。

[0086] 然后，对于每一个时间段内的语音信息分别依次进行以下处理：可以采用快速傅氏变换(Fast Fourier Transformation，简称FFT)算法对每一个时间段内的语音信息进行处理，得到频谱值；然后，采用梅尔(Mel)滤波器组对频谱值进行梅尔滤波处理，得到梅尔频谱信息；接着，在梅尔频谱信息上进行倒谱分析，获得与一个时间段内的语音信息对应的一个梅尔频率倒谱系数(Mel Frequency Cepstrum Coefficient，MFCC)；由于梅尔频率倒谱系数是对语音分析得到，梅尔频率倒谱系数也可以称为基于语音的梅尔频率倒谱系数。

[0087] 由于上述过程是对每一个时间段内的语音信息进行的处理，从而所有时间段内的语音信息对应的各个梅尔频率倒谱系数可以构成一个第一语音特征向量，该第一语音特征向量是与步骤S101中获取到的语音信息相对应的。并且，梅尔频率倒谱系数可以表征出语音特征，语音特征包括了语音的内容信息、音高、音色等等。

[0088] S103、将第一语音特征向量输入到预设的识别模型中，得到第一表情向量，其中，第一表情向量中包括多个表情参数。

[0089] 本实施例中，预先设置了一个识别模型，例如，识别模型为双向长短期记忆(Bi-directional Long Short-Term Memory，简称Bi-LSTM)神经网络模型，该识别模型已经被预先训练，该识别模型是一个成熟的神经网络模型。

[0090] 将获取到的第一语音特征向量，输入到上述识别模型中进行处理，可以输出一个第一表情向量。其中，第一表情向量是多维的，例如是51维的；第一表情向量中包括了多个表情参数。

[0091] S104、根据第一表情向量对待处理的三维动画模型进行拟合，得到具有表情的三维动画模型。

[0092] 可选的，步骤S104包括以下步骤：

[0093] 第一步、对第一表情向量进行平滑处理，得到平滑处理后的第一表情向量。

[0094] 第二步、根据平滑处理后的第一表情向量对待处理的三维动画模型进行拟合，得到具有表情的三维动画模型。

[0095] 本实施例中，将第一表情向量中的各个表情参数，输送给三维动画模型；然后，采用现有的拟合方法对表情进行拟合；然后，对三维动画模型进行渲染，得到具有表情的三维动画模型。

[0096] 其中，为了使得三维动画模型的表情生动自然，可以首先对第一表情向量中的各个表情参数进行平滑处理，得到平滑处理后的第一表情向量；平滑处理的方法例如有，加权平滑方法、FIR(Finite Impulse Response，简称FIR)滤波器。然后再将平滑处理后的第一表情向量输送给三维动画模型；采用现有的拟合方法对表情进行模拟，得到具有表情的三维动画模型。

[0097] 从而，由于表情参数经过了平滑处理，可以防止出现细微表情抖动的情况，生成的表情动作更加生动、自然、顺畅。

[0098] 本实施例，通过对语音信息进行特征提取，得到语音信息的第一语音特征向量，第一语音特征向量包括多个语音特征参数，并且，每一个语音特征参数是一个基于语音的梅尔频率倒谱系数；将第一语音特征向量输入到预设的识别模型中，得到第一表情向量，第一表情向量中包括多个表情参数；根据第一表情向量对待处理的三维动画模型进行拟合，得到具有表情的三维动画模型。由于基于语音的梅尔频率倒谱系数可以保留语音的内容信息、音高、音色等等语音特征，从而最大程度的保证了语音的特点，基于这样的语音得到的表情参数可以更好的还原出表情动作；使得三维动画模型的面部表情更加正确，可以较好的对面部表情进行模拟，并且，得到的表情动作是符合语音信息和语音数据的。

[0099] 图2为本申请实施例提供的另一种面部表情的生成方法的流程图，如图2所示，本实施例中的方法可以包括：

[0100] S201、获取待训练的至少一个第二语音特征向量，其中，每一个第二语音特征向量中包括多个待训练的基于语音的梅尔频率倒谱系数。

[0101] 可选的，步骤S201包括以下步骤：

[0102] 第一步、获取用户的动态视频，并获取动态视频中的用户在各时间段下的语音信息。

[0103] 第二步、对每一个时间段下的语音信息进行特征提取，得到每一个时间段下的第二语音特征向量。

[0104] 本实施例中，本申请实施例的执行主体可以是终端设备、或者服务器、或者面部表情的生成装置或设备、或者其他可以执行本申请提供的方法的装置或设备。

[0105] 需要得到成熟的识别模型。首先，需要获取多个第二语音特征向量，每一个第二语音特征向量中包括了多个待训练的基于语音的梅尔频率倒谱系数。

[0106] 具体来说，在封闭无噪空间内，用户对剧本进行演绎；同时，采用专业的录制设备对用户进行录像，进而得到用户的动态视频。

[0107] 然后,对动态视频进行时间分段,进而得到各时间段下的语音信息。

[0108] 然后,对每一个时间段下的语音信息配置一个短时分析窗,每一个时间段下的语音信息包括了多个时间窗口下的语音数据,即,一个时间窗口下具有一个语音数据。

[0109] 然后,对于对每一个时间段下的语音信息来说,采用FFT算法对每一个时间窗口下的语音数据进行处理,得到频谱值;然后,采用梅尔滤波器组对频谱值进行梅尔滤波处理,得到梅尔频谱信息;接着,在梅尔频谱信息上进行倒谱分析,获得与一个时间窗口下的语音数据对应的一个梅尔频率倒谱系数;由于梅尔频率倒谱系数是对语音分析得到,梅尔频率倒谱系数也可以称为基于语音的梅尔频率倒谱系数;然后,各个时间窗口下的语音数据对应的各个梅尔频率倒谱系数可以构成一个第二语音特征向量;该第一语音特征向量是与一个时间段下的语音信息相对应的。

[0110] 从而,通过以上过程,得到每一个时间段下的第二语音特征向量,即得到多个第二语音特征向量。

[0111] S202、获取每一个第二语音特征向量对应的第二表情向量,其中,每一个第二表情向量中包括多个待训练的表情参数,每一个第二表情向量用于指示出一种面部表情。

[0112] 可选的,步骤S202具体包括:捕捉用户在每一个时间段下的面部表情,以得到每一个时间段下的第二表情向量,其中,第二语音特征向量与第二表情向量是一一对应的。

[0113] 本实施例中,在获取上述动态视频的时候,是采用深度相机来采集的,进而可以直接的捕捉到用户在每一个时间段下的面部表情,进而可以得到每一个时间段下的第二表情向量。由于,在步骤S201中,每一个时间段下具有语音信息和面部表情,从而,每一个时间段下具有一个第二语音特征向量和一个第二表情向量,即第二语音特征向量与第二表情向量是一一对应的;并且,每一个第二表情向量用于指示出一种面部表情。

[0114] 面部表情,例如是生气、高兴、哭泣、等等。

[0115] 其中,第二表情向量中包括多个待训练的表情参数。表情参数可以是视觉特征、深度信息等等。表情参数,例如是,面部各个部位的视觉特征,面部各个部位是否凸起,面部各个部位的凸起高度。第二表情向量可以是52维的向量;表情参数的取值空间为[0,100]。

[0116] 根据以上步骤可知,第二语音特征向量与第二表情向量之间自动进行了关联和标注,第二语音特征向量与第二表情向量构成了训练集。

[0117] S203、将至少一个第二语音特征向量和每一个第二语音特征向量对应的第二表情向量,输入到初始的识别模型中,得到预设的识别模型。

[0118] 本实施例中,将步骤S201中的每一个第二语音特征向量、步骤S202中的每一个第二语音特征向量对应的第二表情向量,输入到初始的识别模型中,进而对初始的识别模型进行训练。

[0119] 其中,识别模型可以采用Bi-LSTM神经网络模型。第二语音特征向量与第二表情向量是一一对应的,Bi-LSTM神经网络模型可以将语音信息与表情参数之间进行映射。

[0120] Bi-LSTM神经网络模型的数学模型为Y=g(x),其中,x为第二语音特征向量,Y为第二表情向量,x表征了输入,Y表征了输出。Bi-LSTM神经网络模型的网络层数为两层,每一层有256个隐藏节点,随机失活(dropout参数)设置为0.5;Bi-LSTM神经网络模型的最后一层,可以通过全连接层来获取全局最优的输出序列。

[0121] 在训练的过程中,使用Pytorch深度学习框架,可以选择随机梯度下降法,选择均

方误差(Mean-Square Error,简称MSE)损失函数作为损失函数。

[0122] 其中,MSE损失函数为$MSE = \sum_{i=1}^{n}(y_i - y_i^p)^2$,其中,$y_i$为Bi-LSTM神经网络模型的输出值,$y_i^p$为第二表情向量,n为迭代次数;i大于等于1,且i小于等于n;i、n为正整数。

[0123] 可选的,经过500代的训练之后,可以得到性能良好的Bi-LSTM神经网络模型。

[0124] S204、获取待处理的语音信息。

[0125] 本实施例中,本步骤可以参见图1的步骤S101,不再赘述。

[0126] S205、对语音信息进行特征提取,得到语音信息的第一语音特征向量,其中,第一语音特征向量包括多个语音特征参数,每一个语音特征参数用于指示出一个基于语音的梅尔频率倒谱系数。

[0127] 本步骤可以参见图1的步骤S102,不再赘述。

[0128] S206、将第一语音特征向量输入到预设的识别模型中,得到第一表情向量,其中,第一表情向量中包括多个表情参数。

[0129] 本步骤可以参见图1的步骤S103,不再赘述。

[0130] S207、采集用户的多种面部特征信息,其中,每一种面部特征信息中包括至少一种视觉特征信息和至少一种面部深度信息。

[0131] 本实施例中,在对步骤S209中的对待处理的三维动画模型进行拟合之前,可以对待处理的三维动画模型进行调整,使得三维动画模型与真实人脸更加贴合和对应。

[0132] 首先,利用深度相机采集用户的多种面部特征信息,面部特征信息例如是视觉特征信息、面部深度信息、等等。其中,视觉特征信息可以是额头区域、眼部区域、鼻子区域、嘴巴区域、双颌部区域的视觉特征,面部深度信息可以是凸起信息、高度信息等等。

[0133] S208、根据多种面部特征信息,对三维动画模型进行调整,得到调整后的三维动画模型。

[0134] 本实施例中,在步骤S208之后,可以根据多种面部特征信息,对三维动画模型进行调整。例如,采用软件Faceshift的各个功能,根据深度相机采集的面部特征信息,建立起三维动画模型的面部轮廓和基础表情,进而对三维动画模型进行调整。使得三维动画模型可以更符合真实人物的轮廓动作幅度,以达到更加自然的效果。

[0135] S209、根据第一表情向量对待处理的三维动画模型进行拟合,得到具有表情的三维动画模型。

[0136] 本步骤可以参见图1的步骤S104,不再赘述。

[0137] 并且,本申请在实际应用过程中,涉及的各个信息和参数,是游戏程序可以识别的,进而游戏客户端可以识别第一表情向量,即可以识别到面部动画参数。

[0138] 本实施例,通过对语音信息进行特征提取,得到语音信息的第一语音特征向量,第一语音特征向量包括多个语音特征参数,并且,每一个语音特征参数是一个基于语音的梅尔频率倒谱系数;将第一语音特征向量输入到预设的识别模型中,得到第一表情向量,第一表情向量中包括多个表情参数;根据第一表情向量对待处理的三维动画模型进行拟合,得到具有表情的三维动画模型。由于基于语音的梅尔频率倒谱系数可以保留语音的内容信息、音高、音色等等语音特征,从而最大程度的保证了语音的特点,基于这样的语音得到的表情参数可以更好的还原出表情动作;使得三维动画模型的面部表情更加正确,可以较好

的对面部表情进行模拟，并且，得到的表情动作是符合语音信息和语音数据的。并且，由于训练数据来自于无噪空间内的真实语音信息和真实表情信息，并建立起语音信息和表情信息之间的映射关系，即，建立起语音特征向量和表情向量之间的映射关系，可以对识别模型进行正确的训练。

[0139] 图3为本申请实施例提供的一种面部表情的生成装置的结构示意图，如图3所示，本实施例的装置，可以包括：

[0140] 第一获取模块31，用于获取待处理的语音信息。

[0141] 提取模块32，用于对语音信息进行特征提取，得到语音信息的第一语音特征向量，其中，第一语音特征向量包括多个语音特征参数，每一个语音特征参数用于指示出一个基于语音的梅尔频率倒谱系数。

[0142] 识别模块33，用于将第一语音特征向量输入到预设的识别模型中，得到第一表情向量，其中，第一表情向量中包括多个表情参数。

[0143] 拟合模块34，用于根据第一表情向量对待处理的三维动画模型进行拟合，得到具有表情的三维动画模型。

[0144] 本实施例的装置，可以执行图1所示方法中的技术方案，其具体实现过程和技术原理参见图1所示方法中的相关描述，此处不再赘述。

[0145] 图4为本申请实施例提供的另一种面部表情的生成装置的结构示意图，在图3所示实施例的基础上，如图4所示，本实施例的装置中，提取模块32，包括：

[0146] 变换子模块321，用于对语音信息进行快速傅氏变换处理，得到频谱值。

[0147] 滤波子模块322，用于对频谱值进行梅尔滤波处理，得到梅尔频谱信息。

[0148] 处理子模块323，用于对梅尔频谱信息进行倒谱分析处理，得到多个基于语音的梅尔频率倒谱系数，其中，多个基于语音的梅尔频率倒谱系数构成第一语音特征向量。

[0149] 可选的，拟合模块34，包括：

[0150] 平滑子模块341，用于对第一表情向量进行平滑处理，得到平滑处理后的第一表情向量。

[0151] 拟合子模块342，用于根据平滑处理后的第一表情向量对待处理的三维动画模型进行拟合，得到具有表情的三维动画模型。

[0152] 可选的，本实施例提供的装置，还包括：

[0153] 第二获取模块41，用于在第一获取模块31获取待处理的语音信息之前，获取待训练的至少一个第二语音特征向量，其中，每一个第二语音特征向量中包括多个待训练的基于语音的梅尔频率倒谱系数；

[0154] 第三获取模块42，用于获取每一个第二语音特征向量对应的第二表情向量，其中，每一个第二表情向量中包括多个待训练的表情参数，每一个第二表情向量用于指示出一种面部表情。

[0155] 训练模块43，用于将至少一个第二语音特征向量和每一个第二语音特征向量对应的第二表情向量，输入到初始的识别模型中，得到预设的识别模型。

[0156] 可选的，第二获取模块41，包括：

[0157] 获取子模块411，用于获取用户的动态视频，并获取动态视频中的用户在各时间段下的语音信息。

[0158] 提取子模块412,用于对每一个时间段下的语音信息进行特征提取,得到每一个时间段下的第二语音特征向量。

[0159] 可选的,第三获取模块42,具体用于:捕捉用户在每一个时间段下的面部表情,以得到每一个时间段下的第二表情向量,其中,第二语音特征向量与第二表情向量是一一对应的。

[0160] 可选的,本实施例提供的装置,还包括:

[0161] 采集模块44,用于在拟合模块34根据第一表情向量对待处理的三维动画模型进行拟合,得到具有表情的三维动画模型之前,采集用户的多种面部特征信息,其中,每一种面部特征信息中包括至少一种视觉特征信息和至少一种面部深度信息。

[0162] 调整模块45,用于根据多种面部特征信息,对三维动画模型进行调整,得到调整后的三维动画模型。

[0163] 识别模型为Bi-LSTM神经网络模型。

[0164] 本实施例的装置,可以执行图2所示方法中的技术方案,其具体实现过程和技术原理参见图2所示方法中的相关描述,此处不再赘述。

[0165] 图5为本申请实施例提供的一种电子设备的结构示意图,如图5所示,本实施例的电子设备50可以包括:处理器51和存储器52。

[0166] 存储器52,用于存储计算机程序(如实现上述方法的应用程序、功能模块等)、计算机指令等;

[0167] 上述的计算机程序、计算机指令等可以分区存储在一个或多个存储器42中。并且上述的计算机程序、计算机指令、数据等可以被处理器51调用。

[0168] 处理器51,用于执行存储器52存储的计算机程序,以实现上述实施例涉及的方法中的各个步骤。

[0169] 具体可以参见前面方法实施例中的相关描述。

[0170] 处理器51和存储器52可以是独立结构,也可以是集成在一起的集成结构。当处理器51和存储器52是独立结构时,存储器52、处理器51可以通过总线53耦合连接。

[0171] 本实施例的电子设备可以执行图1、图2所示方法中的技术方案,其具体实现过程和技术原理参见图1、图2所示方法中的相关描述,此处不再赘述。

[0172] 此外,本申请实施例还提供一种计算机可读存储介质,计算机可读存储介质中存储有计算机执行指令,当用户设备的至少一个处理器执行该计算机执行指令时,用户设备执行上述各种可能的方法。

[0173] 其中,计算机可读介质包括计算机存储介质和通信介质,其中通信介质包括便于从一个地方向另一个地方传送计算机程序的任何介质。存储介质可以是通用或专用计算机能够存取的任何可用介质。一种示例性的存储介质耦合至处理器,从而使处理器能够从该存储介质读取信息,且可向该存储介质写入信息。当然,存储介质也可以是处理器的组成部分。处理器和存储介质可以位于ASIC中。另外,该ASIC可以位于用户设备中。当然,处理器和存储介质也可以作为分立组件存在于通信设备中。

[0174] 本领域普通技术人员可以理解:实现上述各方法实施例的全部或部分步骤可以通过程序指令相关的硬件来完成。前述的程序可以存储于一计算机可读取存储介质中。该程序在执行时,执行包括上述各方法实施例的步骤;而前述的存储介质包括:ROM、RAM、磁碟或

者光盘等各种可以存储程序代码的介质。

[0175] 本申请还提供一种程序产品，程序产品包括计算机程序，计算机程序存储在可读存储介质中，服务器的至少一个处理器可以从可读存储介质读取计算机程序，至少一个处理器执行计算机程序使得服务器实施上述本申请实施例任一的方法。

[0176] 本领域普通技术人员可以理解：实现上述各方法实施例的全部或部分步骤可以通过程序指令相关的硬件来完成。前述的程序可以存储于一计算机可读取存储介质中。该程序在执行时，执行包括上述各方法实施例的步骤；而前述的存储介质包括：ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

[0177] 最后应说明的是：以上各实施例仅用以说明本申请的技术方案，而非对其限制；尽管参照前述各实施例对本申请进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分或者全部技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本申请各实施例技术方案的范围。

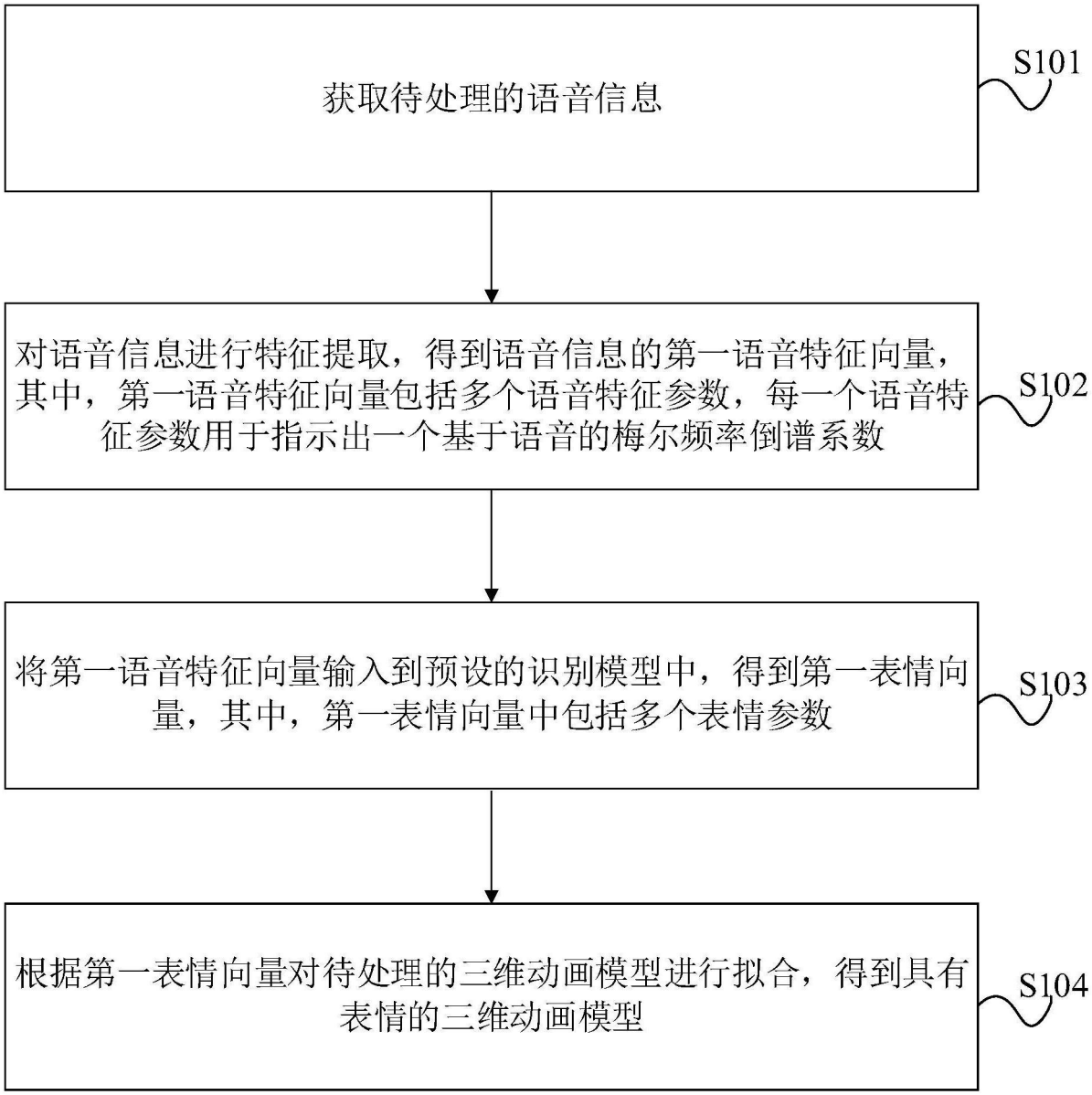

获取待处理的语音信息
S101
对语音信息进行特征提取，得到语音信息的第一语音特征向量，其中，第一语音特征向量包括多个语音特征参数，每一个语音特征参数用于指示出一个基于语音的梅尔频率倒谱系数
S102
将第一语音特征向量输入到预设的识别模型中，得到第一表情向量，其中，第一表情向量中包括多个表情参数
S103
根据第一表情向量对待处理的三维动画模型进行拟合，得到具有表情的三维动画模型
S104

图1

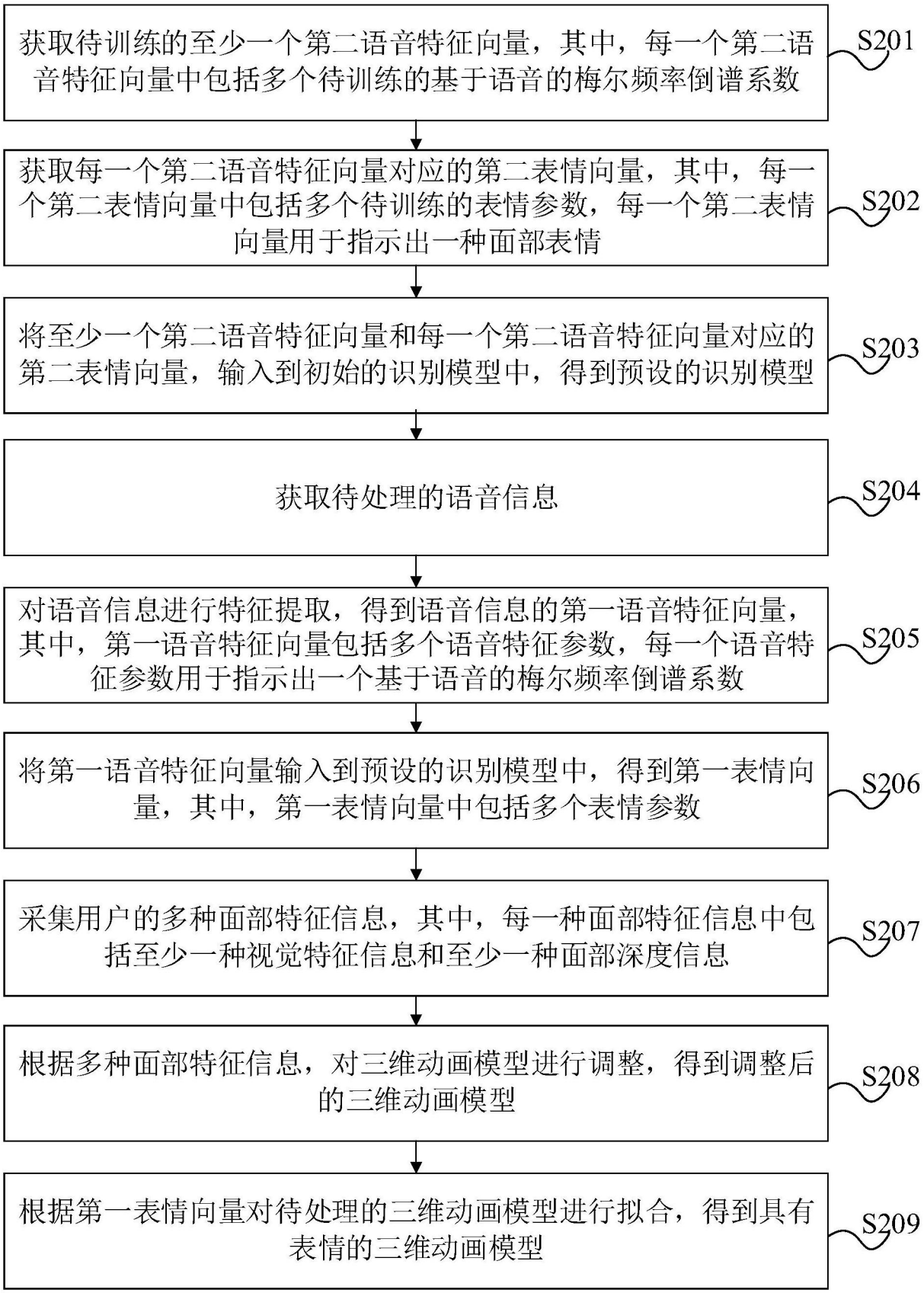

获取待训练的至少一个第二语音特征向量，其中，每一个第二语音特征向量中包括多个待训练的基于语音的梅尔频率倒谱系数
S201
获取每一个第二语音特征向量对应的第二表情向量，其中，每一个第二表情向量中包括多个待训练的表情参数，每一个第二表情向量用于指示出一种面部表情
S202
将至少一个第二语音特征向量和每一个第二语音特征向量对应的第二表情向量，输入到初始的识别模型中，得到预设的识别模型
S203
获取待处理的语音信息
S204
对语音信息进行特征提取，得到语音信息的第一语音特征向量，其中，第一语音特征向量包括多个语音特征参数，每一个语音特征参数用于指示出一个基于语音的梅尔频率倒谱系数
S205
将第一语音特征向量输入到预设的识别模型中，得到第一表情向量，其中，第一表情向量中包括多个表情参数
S206
采集用户的多种面部特征信息，其中，每一种面部特征信息中包括至少一种视觉特征信息和至少一种面部深度信息
S207
根据多种面部特征信息，对三维动画模型进行调整，得到调整后的三维动画模型
S208
根据第一表情向量对待处理的三维动画模型进行拟合，得到具有表情的三维动画模型
S209

图2

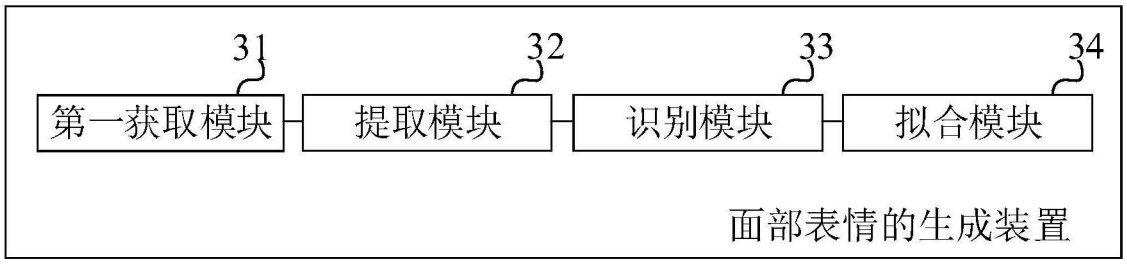

31
32
33
34
第一获取模块
提取模块
识别模块
拟合模块
面部表情的生成装置

图3

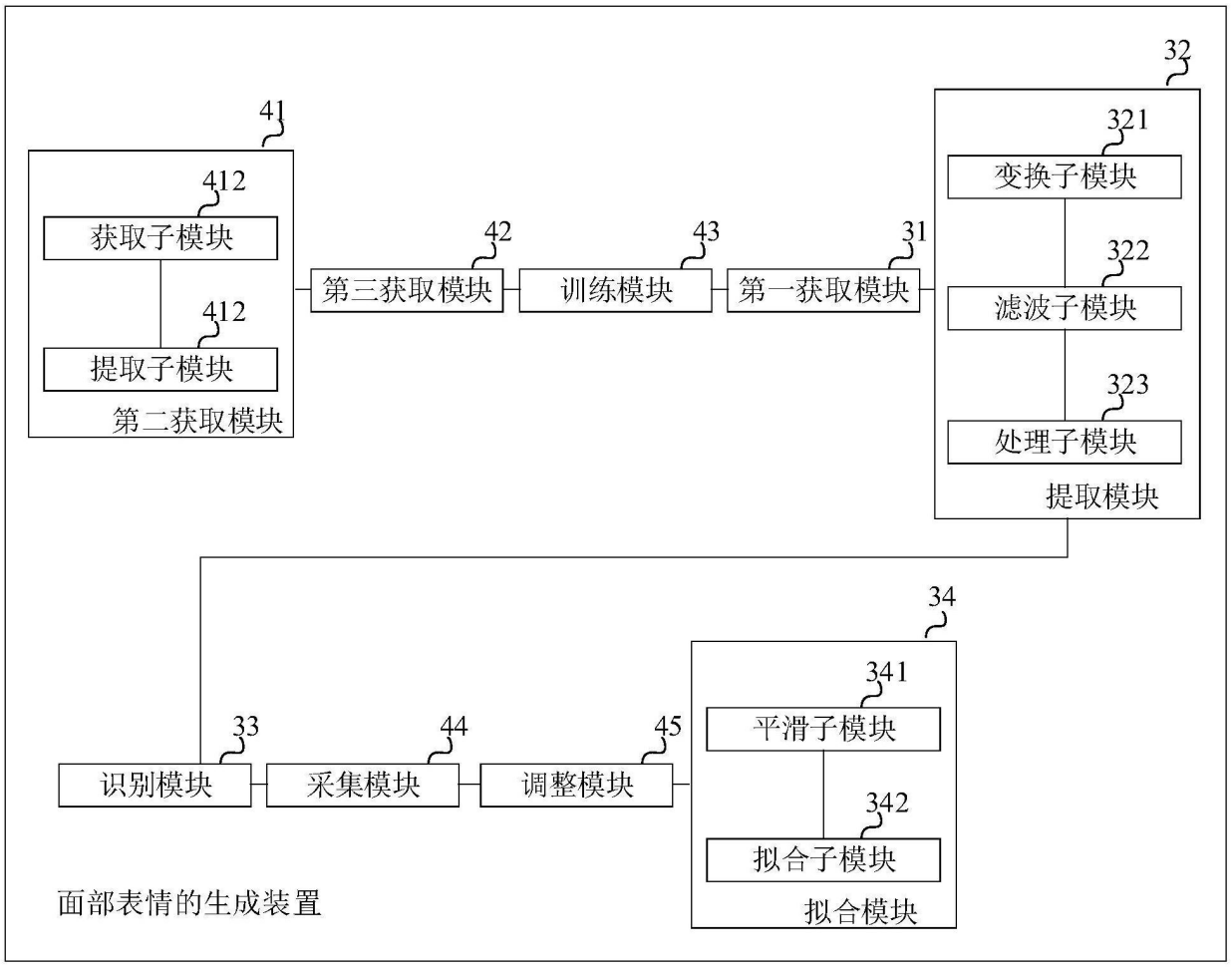

32
41
321
412
变换子模块
获取子模块
42
43
31
322
412
第三获取模块
训练模块
第一获取模块
滤波子模块
提取子模块
323
第二获取模块
处理子模块
提取模块
34
341
33
44
45
平滑子模块
识别模块
采集模块
调整模块
342
拟合子模块
面部表情的生成装置
拟合模块

图4

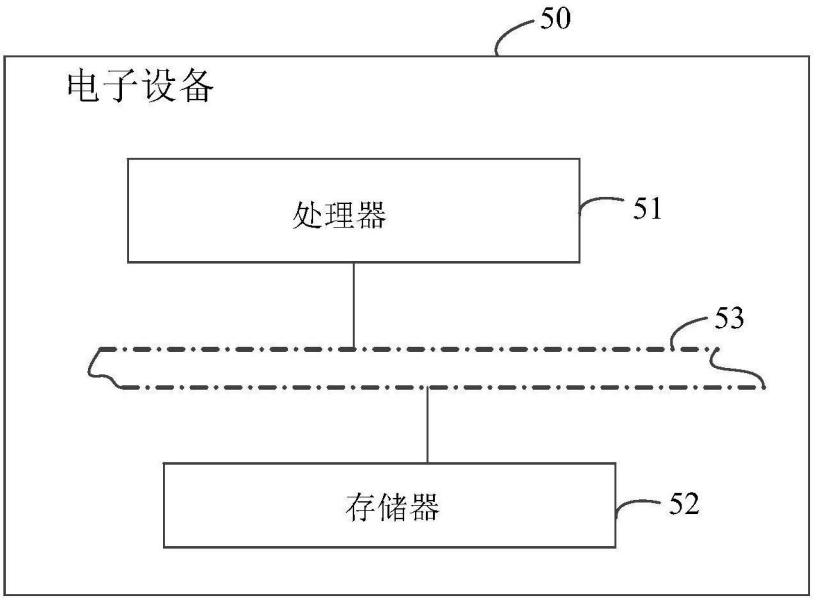


图5