

表情交互的方法、装置、设备和存储介质

5 本申请要求 2024 年 01 月 05 日递交的申请号为 202410023197.4、标题为“表情交互的方法、装置、设备和存储介质”的中国发明专利申请的优先权，该申请的全部内容通过引用结合在本申请中。

技术领域

10 本公开的示例实施例总体涉及计算机领域，特别地涉及表情交互的方法、装置、设备和计算机可读存储介质。

背景技术

15 随着计算机技术的发展，互联网已经成为人们信息交互的重要平台。在人们通过互联网进行信息交互的过程中，例如在聊天会话中，各种类型的表情符号（emoji）已经成为人们社交表达和交换信息的重要媒介。

发明内容

20 在本公开的第一方面，提供了一种表情交互的方法。该方法包括：确定用于生成语音表情目标媒体资源；基于接收到的编辑操作，创建语音表情，语音表情包括视觉部分和语音部分，编辑操作包括以下至少一项：用于修改目标媒体资源所包括的第一视觉内容和/或第一语音内容的第一操作，用于输入第二视觉内容和/或第二语音内容的第二操作；以及发布语音表情，使得语音表情能够基于目标用户的选择而被发送至与目标用户相关联的会话中。

25 在本公开的第二方面，提供了一种表情交互的方法。该方法包括：接收对语音表情的选择；以及将语音表情发送至目标会话中，语音表情包括视觉部分和语音部分，其中，语音表情是基于与目标媒体资源相关联的编辑操作而被创建，编辑操作包括以下至少一项：用于修改目标媒体资源所包括的第一视觉内容和/或第一语音内容的第一操作，用于添加第二视觉内容和/或第二语音内容的第二操作。

30 在本公开的第三方面，提供了一种表情制作方法。该方法包括：基于目标语音内容，呈现基于目标语音内容所确定的第一组候选视觉内容；以及基于针对第一组候选视觉内容中的第一视觉内容的选择，创建语音表情，语音表情包括视觉部分和语音部分，其中视觉部分基于第一视觉内容而被确定，语音部分基于目标语音内容而被确定。

在本公开的第四方面，提供了一种表情制作方法。该方法包括：基于目标文本内容，呈现基于目标文本内容所确定的第二组候选视觉内容；以及基于针对第二组候选视觉内容中的第二视觉内容的选择，创建语音表情，语音表情包括视觉部分和语音部分，其中视觉部分基于目标视觉内容而被确定，语音部分基于目标文本内容而被生成。

5 在本公开的第五方面，提供了一种用于表情交互的装置。该装置包括：确定模块，被配置为确定用于生成语音表情的目标媒体资源；创建模块，被配置为基于接收到的编辑操作，创建语音表情，语音表情包括视觉部分和语音部分，编辑操作包括以下至少一项：用于修改目标媒体资源所包括的第一视觉内容和/或第一语音内容的第一操作，用于输入第二视觉内容和/或第二语音内容的第二操作；以及发布模块，被配置为发布
10 语音表情，使得语音表情能够基于目标用户的选择而被发送至与目标用户相关联的会话中。

在本公开的第六方面，提供了一种电子设备。该设备包括至少一个处理单元；以及至少一个存储器，至少一个存储器被耦合到至少一个处理单元并且存储用于由至少一个处理单元执行的指令。指令在由至少一个处理单元执行时使设备执行第一方面至
15 第四方面中任一方面的方法。

在本公开的第七方面，提供了一种计算机可读存储介质。该计算机可读存储介质上存储有计算机程序，计算机程序可由处理器执行以实现第一方面至第四方面中任一方面的方法。

应当理解，本内容部分中所描述的内容并非旨在限定本公开的实施例的关键特征
20 或重要特征，也不用于限制本公开的范围。本公开的其他特征将通过以下的描述而变得容易理解。

附图说明

结合附图并参考以下详细说明，本公开各实施例的上述和其他特征、优点及方面
25 将变得更加明显。在附图中，相同或相似的附图标记表示相同或相似的元素，其中：

- 图 1 示出了其中可以实施根据本公开的实施例的示例环境的示意图；
- 图 2 示出了根据本公开的一些实施例的表情交互的示例过程的流程图；
- 图 3A 至图 3E 示出了根据本公开的一些实施例的示例界面；
- 图 4 示出了根据本公开的一些实施例的表情交互的示例过程的流程图；
- 30 图 5A 至图 5B 示出了根据本公开的一些实施例的示例界面；
- 图 6 示出了根据本公开的一些实施例的表情制作的示例过程的流程图；
- 图 7 示出了根据本公开的一些实施例的表情制作的示例过程的流程图；
- 图 8 示出了根据本公开的一些实施例的用于表情交互的示例装置的示意性结构框图；以及

图 9 示出了能够实施本公开的多个实施例的电子设备的框图。

具体实施方式

下面将参照附图更详细地描述本公开的实施例。虽然附图中示出了本公开的某些
5 实施例，然而应当理解的是，本公开可以通过各种形式来实现，而且不应该被解释为
限于这里阐述的实施例，相反，提供这些实施例是为了更加透彻和完整地理解本公开。
应当理解的是，本公开的附图及实施例仅用于示例性作用，并非用于限制本公开的保护
范围。

需要注意的是，本文中所提供的任何节/子节的标题并不是限制性的。本文通篇描
10 述了各种实施例，并且任何类型的实施例都可以包括在任何节/子节下。此外，在任一
节/子节中描述的实施例可以以任何方式与同一节/子节和/或不同节/子节中描述的任何
其他实施例相结合。

在本公开的实施例的描述中，术语“包括”及其类似用语应当理解为开放性包含，
即“包括但不限于”。术语“基于”应当理解为“至少部分地基于”。术语“一个实施
15 例”或“该实施例”应当理解为“至少一个实施例”。术语“一些实施例”应当理解
为“至少一些实施例”。下文还可能包括其他明确的和隐含的定义。术语“第一”、
“第二”等可以指代不同的或相同的对象。下文还可能包括其他明确的和隐含的定义。

本公开的实施例中可能涉及用户的数据、数据的获取和/或使用等。这些方面均遵
循相应的法律法规及相关规定。在本公开的实施例中，所有数据的采集、获取、处理、
20 加工、转发、使用等，都是在用户知晓并且确认的前提下进行的。相应地，在实现本
公开的各实施例时，均应根据相关法律法规通过适当的方式，将可能所涉及的数据或
信息的类型、使用范围、使用场景等告知用户并获得用户的授权。具体的告知和/或授
权方式可以根据实际情况和应用场景而变化，本公开的范围在此方面不受限制。

本说明书及实施例中方案，如涉及个人信息处理，则均会在具备合法性基础（例
25 如征得个人信息主体同意，或者为履行合同所必需等）的前提下进行处理，且仅会在
规定或者约定的范围内进行处理。用户拒绝处理基本功能所需必要信息以外的个人信
息，不会影响用户使用基本功能。

在人们通过互联网进行信息交互的过程中，人们期望使用优质的表情来便捷地表
达想要的信息。传统的表情资源难以全面表达人们想要传达的信息。

30 本公开的实施例提出了一种表情交互方案。根据该方案，可以确定用于生成语音
表情的目标媒体资源。进一步地，可以基于接收到的编辑操作，创建语音表情，其中
语音表情包括视觉部分和语音部分。

这样的编辑操作可以包括以下至少一项：用于修改目标媒体资源所包括的第一视
觉内容和/或第一语音内容的第一操作，用于输入第二视觉内容和/或第二语音内容的第

二操作。

进一步地，可以发布语音表情，使得语音表情能够基于目标用户的选择而被发送至与目标用户相关联的会话中。

以此方式，本公开的实施例能够支持用户对于媒体资源的编辑来创建语音表情，从而提高了语音表情的制作效率。

以下进一步结合附图来详细描述该方案的各种示例实现。

示例环境

图 1 示出了本公开的实施例能够在其中实现的示例环境 100 的示意图。如图 1 所示，示例环境 100 可以包括电子设备 110。

在该示例环境 100 中，电子设备 110 可以运行有支持界面交互的应用 120。应用 120 可以是用于界面交互的任何适当类型应用，其示例可以包括但不限于：视频应用、社交应用或其他适当的应用。用户 140 可以经由电子设备 110 和/或其附接设备来与应用 120 进行交互。

在图 1 的环境 100 中，如果应用 120 处于活动状态，电子设备 110 可以通过应用 120 呈现用于支持界面交互的界面 150。

在一些实施例中，电子设备 110 与服务器 130 通信，以实现对应用 120 的服务的供应。电子设备 110 可以是任意类型的移动终端、固定终端或便携式终端，包括移动手机、台式计算机、膝上型计算机、笔记本计算机、上网本计算机、平板计算机、媒体计算机、多媒体平板、掌上电脑、便携式游戏终端、VR/AR 设备、个人通信系统（Personal Communication System，PCS）设备、个人导航设备、个人数字助理（Personal Digital Assistant，PDA）、音频/视频播放器、数码相机/摄像机、定位设备、电视接收器、无线电广播接收器、电子书设备、游戏设备或者前述各项的任意组合，包括这些设备的配件和外设或者其任意组合。在一些实施例中，电子设备 110 也能够支持任意类型的针对用户的接口（诸如“可佩戴”电路等）。

服务器 130 可以是独立的物理服务器，也可以是多个物理服务器构成的服务器集群或者分布式系统，还可以是提供云服务、云数据库、云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全服务、内容分发网络、以及大数据和人工智能平台等基础云计算服务的云服务器。服务器 130 例如可以包括计算系统/服务器，诸如大型机、边缘计算节点、云环境中的计算设备，等等。服务器 130 可以为电子设备 110 中支持虚拟场景的应用 120 提供后台服务。

服务器 130 与电子设备 110 之间可以建立有通信连接。通信连接可以通过有线方式或无线方式建立。通信连接可以包括但不限于蓝牙连接、移动网络连接、通用串行总线（Universal Serial Bus，USB）连接、无线保真（Wireless Fidelity，WiFi）连接等，

本公开的实施例在此方面不受限制。在本公开的实施例中，服务器 130 与电子设备 110 可以通过二者之间的通信连接实现信令交互。

应当理解，仅出于示例性的目的描述环境 100 中各个元素的结构和功能，而不暗示对于本公开的范围的任何限制。

5 以下将继续参考附图描述本公开的一些示例实施例。

第一示例过程

图 2 示出了根据本公开的一些实施例的示例交互过程 200 的流程图。过程 200 可以被实现在电子设备 110 处。下面参考图 1 来描述过程 200。

10 如图 2 所示，在框 210，电子设备 110 确定用于生成语音表情目标媒体资源。

在一些实施例中，这样的目标媒体资源可以包括图片、视频、音频、表情等包括视觉内容和/或音频内容的任何适当类型资源。

15 在一些实施例中，目标媒体资源可以包括用户输入的媒体资源。例如，用户可以拍摄一段视频、一张图片或录制一段语音以作为目标媒体资源，以用于语音表情的生成。

在一些实施例中，电子设备 110 还可以基于用户对于候选媒体资源中目标媒体资源的选择，来确定用于生成语音表情目标媒体资源。

20 以下将参考图 3A 至图 3E 来描述过程 200。图 3A 至图 3E 示出了根据本公开的一些实施例的示例界面 300A 至 300E。界面 300A 至界面 300E 例如可以由图 1 所示的电子设备 110 所提供。

如图 3A 所示，界面 300A 例如可以是会话的会话界面。电子设备 110 可以在界面 300A 中显示表情面板 305。作为示例，该表情面板 305 可以显示已添加的语音表情，例如，语音表情 315。

25 附加地，表情面板 305 例如还可以包括创建入口 310。在接收到对创建入口 310 的选择的情况下，电子设备 110 可以呈现如图 3B 所示的界面 300B。

在界面 300B 中，电子设备 110 可以显示一组候选图像，例如，候选图像 320-1、候选图像 320-2、候选图像 320-3 和候选图像 320-4（单独或统一成为候选图像 320）。在一些实施例中，这样的候选图像可以包括静态图像和/或静态图像。

30 在一些实施例中，这样的候选图像 320 例如可以与当前用户相关联。例如，候选图像 320 可以包括当前用户所收藏的图像。作为另一示例，候选图像 320 还可以包括与当前用户发布的作品相关联的图像，例如，当前用户发布的图像作品或当前用户发布的视频作品的视频帧。

在一些实施例中，在用户知晓并授权的情况下，候选图像 320 还可以包括电子设备 110 本地的图像。

进一步地，电子设备 110 例如可以接收用户对于候选图像 320-3 的选择，并可以相应地将候选图像 320-3 确定为用于生成语音表情的目标媒体资源。

电子设备 110 还可以基于其他适当的方式来确定用于生成语音表情的目标媒体资源。在一些实施例中，电子设备 110 例如还可以支持用户将可获取的表情选择作为目标媒体资源。

在一些示例中，用户可获取的表情可以包括用户已添加的图像表情、语音表情等；或者，其他用户所分享的图像表情、语音表情等。

在一些实施例中，电子设备 110 可以在表情面板 305 中呈现一组候选图像表情。作为示例，这样的候选图像表情例如可以仅具有视觉部分，而不包括语音部分。

进一步地，电子设备 110 可以接收用户对于该组候选图像表情中的目标图像表情的选择，并将该目标图像表情确定为目标媒体资源。

在又一些实施例中，电子设备 110 例如还可以接收与第一作品相关联的预设操作，并可以相应地将该第一作品确定为目标媒体资源。

例如，电子设备 110 例如可以关联于第一作品的查看界面来提供语音表情创建入口，并可以基于用户对该创建入口的选择，而将该第一作品确定为目标媒体资源。

继续参考图 2，在框 220，电子设备 110 基于接收到的编辑操作，创建语音表情，语音表情包括视觉部分和语音部分，编辑操作包括以下至少一项：用于修改目标媒体资源所包括的第一视觉内容和/或第一语音内容的第一操作，用于输入第二视觉内容和/或第二语音内容的第二操作。

在一些实施例中，电子设备 110 可以为用户提供用于创建语音表情的编辑界面。

在一些实施例中，电子设备 110 可以基于用户在编辑界面中的第一操作来相应地修改目标媒体资源所包括的第一视觉内容和/或第一语音内容。

以图 3B 作为示例，在目标媒体资源为候选图像 320-3 的情况下，电子设备 110 例如可以提供用于编辑候选图像 320-3 的图像编辑界面。应当理解的是，这样的图像编辑界面例如可以支持任何适当类型的图像编辑操作，例如，裁剪图像、添加文字、添加贴纸、应用滤镜等等。

示例性地，如图 3C 所示，电子设备 110 例如可以在界面 300C 中显示编辑后的图像 325。应当理解的，这样的图像编辑过程不是必须的。图像 325 例如也可以是原始的候选图像 320-3。

继续参考图 3C，电子设备 110 可以提供用于添加语音内容的控件 330。例如，在用户触发控件 330 的情况下，电子设备 110 可以获取用户输入的语音内容。

应当理解的是，这样的语音内容可以包括任何适当的音频内容，而不论这样的音频内容是否表述对应的文本信息。例如，语音内容可以包括所录制的用户说的一段话、环境中的特定声音等。又例如，语音内容例如也可以包括有歌词或无歌词的音乐内容。

继续参考图 3D,在语音内容输入完成后,电子设备 110 例如可以呈现界面 300D。在界面 300D 中,电子设备 110 可以提供预览入口 335,以用于播放所输入的语音内容。

附加地,电子设备 110 例如还可以将特定的语音风格应用于输入的语音内容。如图 3D 所示,电子设备 110 例如可以提供一组候选语音风格 340,并可以基于用户的选择而将指定的目标语音风格应用于输入的语音内容。这样的语音风格例如可以对应于不同音色、语调等。

以图 3D 作为示例,在用户选择了目标语音风格(例如,“风格 1”)的情况下,预览入口 335 所触发播放的语音内容例如可以与该目标语音风格匹配。

在又一些实施例中,电子设备 110 例如也可以支持用户在输入语音内容之前就选择待应用的目标语音风格。相应地,在语音内容输入完成后,预览入口 335 所触发播放的语音内容例如可以是已经应用了该目标语音风格的语音内容。

进一步地,电子设备 110 可以基于用户对于完成控件 345 的选择来创建对应的语音表情。该语音表情可以包括视觉部分和语音部分。

相应地,该视觉部分例如可以对应于图 3C 中所示的图像 325。语音部分可以对应于预览入口 335 所触发播放的语音内容。

在一些实施例中,在图像 325 为动态图像的情况下,语音表情的语音内容可以与动态图像在时间上对齐,使得该语音内容例如可以从动态图像的起始帧播放。

在一些实施例中,电子设备 110 还可以通过其他适当的方式来添加语音内容。作为示例,电子设备 110 例如可以提供一组候选语音内容,并可以基于用户对于候选语音内容的选择,而将该候选语音内容添加至该语音表情。

作为示例,电子设备 110 例如可以获取一组语音素材,这样的语音素材例如可以由其他用户所创建并分享的,或者可以包括平台所提供的语音素材。相应地,用户例如可以选择特定的语音素材并与已有的图像进行组合,以生成对应的语音表情。

在一些实施例中,语音内容例如也可以是基于用户输入的文本内容而被相应地添加。例如,电子设备 110 可以接收用户输入的文本内容,并可以相应地生成与该文本内容匹配的语音内容,以添加至该语音表情。备选地,电子设备 110 也可以从已有的语音素材中搜索与该文本内容匹配的语音内容,以添加至语音表情。

以上以目标媒体资源包括图像为例描述了语音表情的创建过程。如上文所提及的,目标媒体资源例如还可以包括用户所选择的图像表情。应当理解的是,可以基于上文所描述的类似过程来支持对该图像表情的视觉部分的编辑,并支持音频内容的添加,以创建语音表情。

在一些实施例中,目标媒体资源例如还可以包括语音表情。在这种情况下,电子设备 110 例如可以支持用户对于该语音表情的视觉部分的编辑或替换和/或语音部分的

编辑或替换。

例如，用户可以选择保留已有语音表情的视觉部分，并可以录入新的语音部分，以生成语音表情。或者，用户可以选择保留已有语音表情的视觉部分，并可以将特定语音风格应用于已有的语音部分，以生成语音表情。

5 在一些实施例中，目标媒体资源例如还可以包括音频内容，例如，一段语音。作为示例，电子设备 110 例如可以接收用户对于会话中一段语音的选择，以将该语音确定为目标媒体资源。或者，电子设备 110 也可以接收用户录制的一段语音，以作为目标媒体资源。

10 相应地，电子设备 110 例如可以支持用户添加对应的视觉内容。类似于上文关于图 3B 所讨论的内容，电子设备 110 例如可以提供一组候选图像以供用户选择。在一些示例中，这样的候选的图像例如可以包括与用户相关联的图像。或者，这样的候选图像还可以包括基于该段语音而生成的或搜索得到的图像。

 基于上文所讨论的过程，本公开的实施例能够支持用户对于媒体资源的编辑来创建语音表情，从而提高了语音表情的制作效率。

15 继续参考图 2，在框 230，电子设备 110 发布语音表情，使得语音表情能够基于目标用户的选择而被发送至与目标用户相关联的会话中。

 在一些实施例中，电子设备 110 可以基于用户的发布操作来相应地发布所创建的语音表情。

20 在一些示例中，这样的发布操作可以包括将语音表情添加至表情面板。如图 3E 所示，电子设备 110 可以相应地在表情面板 300E 中显示所发布的语音表情 350，以使得该语音表情 350 可以被发送至相关联的会话中。

 在一些示例中，这样的发布操作可以包括对于语音表情的分享操作。例如，电子设备 110 可以基于用户的分享请求而将该语音表情分享至其他用户或群组。作为示例，用户例如可以通过私信或者评论等方式来将语音表情分享至其它用户或群组。

25 或者，电子设备 110 也可以基于用户的分享请求而将该语音表情发布为其他用户可获取的资源。其他用户例如可以进一步添加该语音表情，以支持在会话中使用这样的语音表情。

30 在又一示例中，这样的发布操作可以包括将语音表情 350 发布为作品（也称为第二作品），使得用户能够经由该作品来添加该语音表情 350。例如，可以在该作品的查看界面提供用于添加表情的添加入口，当前用户或其他可以通过该添加入口来添加该语音表情。

 关于语音表情的使用过程将在下文参考图 4 详细描述。

第二示例过程

图 4 示出了根据本公开的一些实施例的表情交互的示例过程 400 的流程图。过程 400 可以被实现在电子设备 110 处。下面参考图 1 来描述过程 400。

在框 410，电子设备 110 接收对语音表情的选择。

以下将参考图 5A 和图 5B 来描述过程 400。图 5A 和图 5B 示出了根据本公开的一些实施例的示例界面 500A 和 500B。

如图 5A 所示，界面 500A 例如可以是会话（例如，与“用户 B”的会话）的会话界面。如图所示，电子设备 110 例如可以在界面 500A 中显示表情面板 505。表情面板 505 例如可以选择一组语音表情，例如，语音表情 510-1 和语音表情 510-2（单独或统一称为语音表情 510）。这样的语音表情 510 例如可以是基于上文参考图 2 所描述的过程所创建。

如图 5A 所示，电子设备 110 可以关联于语音表情 510-1 和语音表情 510-2 来提供对应的预览入口 515-1 和预览入口 515-2。

在例如接收到对于预览入口 515-1 的选择的情况下，电子设备 110 可以播放语音表情 510-1 的语音部分。

继续参考图 4，在框 420，电子设备 110 可以将语音表情发送至目标会话中，语音表情包括视觉部分和语音部分。

以图 5B 作为示例，在接收到对于图 5A 中所示出的语音表情 515-2 的选择的情况下，电子设备 110 可以将语音表情 515-2 发送至该会话中。

具体地，电子设备 110 可以在该会话的消息窗口中显示语音表情 510-2 的视觉部分 520。在该视觉部分 520 包括动态图像的情况下，该视觉部分 520 例如可以自动地播放一次或多次。

在一些实施例中，电子设备 110 还可以在消息窗口中显示语音表情 510-2 的播放入口 525。进一步地，在接收到该播放入口 525 的选择的情况下，电子设备 110 可以播放该语音表情 510-2 的语音部分。

在一些实施例中，在该视觉部分 520 包括动态图像的情况下，在播放入口 525 被触发的情况下，该动态图像例如可以重新播放，以使得动态图像例如可以与语音部分的播放同步。

应当理解的是，对于语音表情的接收方而言，其可以类似地查看语音表情的视觉部分 520 并可以基于对播放入口 525 的选择来播放语音表情的语音部分。

第三示例过程

图 6 示出了根据本公开的一些实施例的表情制作的示例过程 600 的流程图。过程 600 可以被实现在电子设备 110 处。下面参考图 1 来描述过程 600。

如图所示，在框 610，电子设备 110 基于目标语音内容，呈现基于目标语音内容

所确定的第一组候选视觉内容。

在一些实施例中，目标语音内容可以包括用户输入的语音内容。例如，电子设备 110 可以通过语音采集设备来获取用户输入的语音内容。

5 在一些实施例中，目标语音内容还可以包括用户以其他方式选择的语音内容。例如，电子设备 110 可以接收对会话中一段语音的选择，以将该段语音作为目标语音内容。或者，电子设备 110 可以接收对其他用户或平台提供的语音素材的选择，以将该语音素材作为目标语音内容。

10 在一些实施例中，电子设备 110 所提供的第一组候选视觉内容是基于与目标语音内容对应的文本内容而被生成。例如，电子设备 110 可以利用任何适当的机器学习模型来生成与文本内容对应的一组候选视觉内容，例如，静态图像、动态图像或视频等。

在框 620，电子设备 110 基于针对第一组候选视觉内容中的第一视觉内容的选择，创建语音表情，语音表情包括视觉部分和语音部分，其中视觉部分基于第一视觉内容而被确定，语音部分基于目标语音内容而被确定。

15 在一些实施例中，电子设备 110 可以基于目标语音内容来确定语音表情的语音部分。例如，目标语音内容本身可以被用作语音表情的语音部分。或者，电子设备 110 也可以将特定语音风格应用于该目标语音内容，以生成语音表情的语音部分。

20 在一些实施例中，电子设备 110 可以基于所选择的第一视觉内容来确定语音表情的视觉部分。例如，第一视觉内容本身可以被用作语音表情的视觉部分。或者，电子设备 110 也可以接收用户对于第一视觉内容的编辑操作，并相应地生成语音表情的视觉部分。

基于这样的方式，本公开的实施例能够支持用户通过录入语音的方式来高效地创建匹配的语音表情。

第四示例过程

25 图 7 示出了根据本公开的一些实施例的表情制作的示例过程 700 的流程图。过程 700 可以被实现在电子设备 110 处。下面参考图 1 来描述过程 700。

如图所示，在框 710，电子设备 110 基于目标文本内容，呈现基于目标文本内容所确定的第二组候选视觉内容。

30 在一些实施例中，电子设备 110 例如可以接收用户所输入或选择的目标文本内容，并可以相应地提供基于该目标文本内容所确定的第二组候选视觉内容。

在一些实施例中，第二组候选视觉内容例如可以是基于目标文本内容而从视觉内容库中搜索得到。备选地或附加地，第二组候选视觉内容例如可以是基于目标文本内容所生成。例如，电子设备 110 可以利用任何适当的机器学习模型来生成与文本内容对应的一组候选视觉内容，例如，静态图像、动态图像或视频等。

在框 720, 电子设备 110 基于针对第二组候选视觉内容中的第二视觉内容的选择, 创建语音表情, 语音表情包括视觉部分和语音部分, 其中视觉部分基于目标视觉内容而被确定, 语音部分基于目标文本内容而被生成。

5 在一些实施例中, 电子设备 110 可以基于所选择的第一视觉内容来确定语音表情的视觉部分。例如, 第一视觉内容本身可以被用作语音表情的视觉部分。或者, 电子设备 110 也可以接收用户对于第一视觉内容的编辑操作, 并相应地生成语音表情的视觉部分。

10 在一些实施例中, 电子设备 110 可以基于目标文本内容来生成语音表情的语音部分。示例性地, 电子设备 110 可以提供一组候选语音风格。进一步地, 电子设备 110 可以接收针对一组候选语音风格中的目标语音风格的选择, 使得语音表情的语音部分基于目标语音风格而被生成。

基于这样的方式, 本公开的实施例能够支持用户通过输入文本的方式来高效地创建匹配的语音表情。

15 **示例装置和设备**

本公开的实施例还提供了用于实现上述方法或过程的相应装置。图 8 示出了根据本公开的某些实施例的用于表情交互的示例装置 800 的示意性结构框图。装置 800 可以被实现为或者被包括在电子设备 110 中。装置 800 中的各个模块/组件可以由硬件、软件、固件或者它们的任意组合来实现。

20 如图 8 所示, 装置 800 包括确定模块 810, 被配置为确定用于生成语音表情目标媒体资源; 创建模块 820, 被配置为基于接收到的编辑操作, 创建语音表情, 语音表情包括视觉部分和语音部分, 编辑操作包括以下至少一项: 用于修改目标媒体资源所包括的第一视觉内容和/或第一语音内容的第一操作, 用于输入第二视觉内容和/或第二语音内容的第二操作; 以及发布模块 830, 被配置为发布语音表情, 使得语音表情能够基
25 于目标用户的选择而被发送至与目标用户相关联的会话中。

在一些实施例中, 确定模块 810 还被配置为: 呈现一组候选图像, 一组候选图像包括静态图像和/或动态图像; 以及基于对一组候选图像中的目标图像的选择, 将目标图像确定为目标媒体资源。

30 在一些实施例中, 一组候选图像包括以下至少一项: 当前用户收藏的图像; 与当前用户发布的作品相关联的图像。

在一些实施例中, 语音表情的视觉部分基于目标图像而被确定。

在一些实施例中, 装置 800 还包括第一接收模块, 被配置为: 接收用于输入第二语音内容的第二操作, 其中语音表情的语音部分基于第二语音内容而被确定。

在一些实施例中, 装置 800 还包括第二接收模块, 被配置为: 接收用于编辑目标

图像的第一操作，其中语音表情的视觉部分基于经编辑的目标图像而被确定。

在一些实施例中，装置 800 还包括风格确定模块，被配置为：提供一组候选语音风格；以及接收针对一组候选语音风格中的目标语音风格的选择，使得语音表情的语音部分与目标语音风格匹配。

5 在一些实施例中，目标图像为目标动态图像，并且语音部分被配置为从目标动态图像的起始帧播放。

在一些实施例中，确定模块 810 还被配置为：基于对表情面板中的创建入口的选择，呈现一组候选图像。

10 在一些实施例中，确定模块 810 还被配置为：呈现一组候选图像表情；以及基于对一组候选图像表情中的目标图像表情的选择，将目标图像表情确定为目标媒体资源。

在一些实施例中，装置 800 还包括第二接收模块，被配置为：经由编辑界面，接收用于输入第二语音内容的第二操作，其中语音表情的视觉部分基于目标图像表情而被确定，语音部分基于第二语音内容而被确定。

15 在一些实施例中，确定模块 810 还被配置为：基于与第一作品相关联的预设操作，将第一作品确定为目标媒体资源；或获取用户拍摄的图片或视频，以作为目标媒体内容。

在一些实施例中，装置 800 还包括发送模块，被配置为：在表情面板中显示经发布的语音表情；以及基于针对语音表情的选择，将语音表情发送至会话中。

20 在一些实施例中，装置 800 还包括播放模块，被配置为：在会话的消息窗口中显示所发送的语音表情的视觉部分和语音表情的播放入口；以及基于对播放入口的选择，播放语音表情的语音部分。

在一些实施例中，装置 800 还包括预览模块，被配置为：在表情面板中，关联于语音表情提供预览入口；以及基于对预览入口的选择，播放语音表情的语音部分。

25 在一些实施例中，发布模块 830 还被配置为执行以下至少一项：将语音表情发布为第二作品，其中使得至少一个用户能够经由第二作品添加语音表情；添加语音表情，使得语音表情能够用于当前用户；向至少一个用户分享语音表情。

30 图 9 示出了其中可以实施本公开的一个或多个实施例的电子设备 900 的框图。应当理解，图 9 所示出的电子设备 900 仅仅是示例性的，而不应当构成对本文所描述的实施例的功能和范围的任何限制。图 9 所示出的电子设备 900 可以用于实现图 1 的电子

如图 9 所示，电子设备 900 是通用电子设备的形式。电子设备 900 的组件可以包括但不限于一个或多个处理器或处理单元 910、存储器 920、存储设备 930、一个或多个通信单元 940、一个或多个输入设备 950 以及一个或多个输出设备 960。处理单元 910 可以是实际或虚拟处理器并且能够根据存储器 920 中存储的程序来执行各种处理。

在多处处理器系统中，多个处理单元并行执行计算机可执行指令，以提高电子设备 900 的并行处理能力。

电子设备 900 通常包括多个计算机存储介质。这样的介质可以是电子设备 900 可访问的任何可以获取的介质，包括但不限于易失性和非易失性介质、可拆卸和不可拆卸介质。存储器 920 可以是易失性存储器（例如寄存器、高速缓存、随机访问存储器（RAM））、非易失性存储器（例如，只读存储器（ROM）、电可擦除可编程只读存储器（EEPROM）、闪存）或它们的某种组合。存储设备 930 可以是可拆卸或不可拆卸的介质，并且可以包括机器可读介质，诸如闪存驱动、磁盘或者任何其他介质，其可以能够用于存储信息和/或数据并且可以在电子设备 900 内被访问。

电子设备 900 可以进一步包括另外的可拆卸/不可拆卸、易失性/非易失性存储介质。尽管未在图 9 中示出，可以提供用于从可拆卸、非易失性磁盘（例如“软盘”）进行读取或写入的磁盘驱动和用于从可拆卸、非易失性光盘进行读取或写入的光盘驱动。在这些情况中，每个驱动可以由一个或多个数据介质接口被连接至总线（未示出）。存储器 920 可以包括计算机程序产品 925，其具有一个或多个程序模块，这些程序模块被配置为执行本公开的各种实施例的各种方法或动作。

通信单元 940 实现通过通信介质与其他电子设备进行通信。附加地，电子设备 900 的组件的功能可以以单个计算集群或多个计算机器来实现，这些计算机器能够通过通信连接进行通信。因此，电子设备 900 可以使用与一个或多个其他服务器、网络个人计算机（PC）或者另一个网络节点的逻辑连接来在联网环境中进行操作。

输入设备 950 可以是一个或多个输入设备，例如鼠标、键盘、追踪球等。输出设备 960 可以是一个或多个输出设备，例如显示器、扬声器、打印机等。电子设备 900 还可以根据需要通过通信单元 940 与一个或多个外部设备（未示出）进行通信，外部设备诸如存储设备、显示设备等，与一个或多个使得用户与电子设备 900 交互的设备进行通信，或者与使得电子设备 900 与一个或多个其他电子设备通信的任何设备（例如，网卡、调制解调器等）进行通信。这样的通信可以经由输入/输出（I/O）接口（未示出）来执行。

根据本公开的示例性实现方式，提供了一种计算机可读存储介质，其上存储有计算机可执行指令，其中计算机可执行指令被处理器执行以实现上文描述的方法。根据本公开的示例性实现方式，还提供了一种计算机程序产品，计算机程序产品被有形地存储在非瞬态计算机可读介质上并且包括计算机可执行指令，而计算机可执行指令被处理器执行以实现上文描述的方法。

这里参照根据本公开实现的方法、装置、设备和计算机程序产品的流程图和/或框图描述了本公开的各个方面。应当理解，流程图和/或框图的每个方框以及流程图和/或框图中各方框的组合，都可以由计算机可读程序指令实现。

5 这些计算机可读程序指令可以提供给通用计算机、专用计算机或其他可编程数据处理装置的处理单元，从而生产出一种机器，使得这些指令在通过计算机或其他可编程数据处理装置的处理单元执行时，产生了实现流程图和/或框图中的一个或多个方框中规定的功能/动作的装置。也可以把这些计算机可读程序指令存储在计算机可读存储介质中，这些指令使得计算机、可编程数据处理装置和/或其他设备以特定方式工作，从而，存储有指令的计算机可读介质则包括一个制造品，其包括实现流程图和/或框图中的一个或多个方框中规定的功能/动作的各个方面的指令。

10 可以把计算机可读程序指令加载到计算机、其他可编程数据处理装置、或其他设备上，使得在计算机、其他可编程数据处理装置或其他设备上执行一系列操作步骤，以产生计算机实现的过程，从而使得在计算机、其他可编程数据处理装置、或其他设备上执行的指令实现流程图和/或框图中的一个或多个方框中规定的功能/动作。

15 附图中的流程图和框图显示了根据本公开的多个实现的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上，流程图或框图中的每个方框可以代表一个模块、程序段或指令的一部分，模块、程序段或指令的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。在有些作为替换的实现中，方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如，两个连续的方框实际上可以基本并行地执行，它们有时也可以按相反的顺序执行，这依所涉及的功能而定。也要注意的，框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合，可以用执行规定的功能或动作的专用的基于硬件的系统来实现，或者可以用专用硬件与计算机指令的组合来实现。

20 以上已经描述了本公开的各实现，上述说明是示例性的，并非穷尽性的，并且也不限于所公开的各实现。在不偏离所说明的各实现的范围和精神的情况下，对于本技术领域的普通技术人员来说许多修改和变更都是显而易见的。本文中所用术语的选择，旨在最好地解释各实现的原理、实际应用或对市场中的技术的改进，或者使本技术领域的其他普通技术人员能理解本文公开的各个实现方式。

权利要求书

1. 一种表情交互的方法，包括：
确定用于生成语音表情的目标媒体资源；
- 5 基于接收到的编辑操作，创建语音表情，所述语音表情包括视觉部分和语音部分，
所述编辑操作包括以下至少一项：用于修改所述目标媒体资源所包括的第一视觉内容和/或第一语音内容的第一操作，用于输入第二视觉内容和/或第二语音内容的第二操作；以及
发布所述语音表情，使得所述语音表情能够基于目标用户的选择而被发送至与所
10 述目标用户相关联的会话中。
2. 根据权利要求1所述的方法，其中确定用于生成语音表情的目标媒体资源包括：
呈现一组候选图像，所述一组候选图像包括静态图像和/或动态图像；以及
基于对所述一组候选图像中的目标图像的选择，将所述目标图像确定为所述目标
媒体资源。
- 15 3. 根据权利要求2所述的方法，其中所述一组候选图像包括以下至少一项：
当前用户收藏的图像；
与当前用户发布的作品相关联的图像。
4. 根据权利要求2所述的方法，其中所述语音表情的所述视觉部分基于所述目标
图像而被确定。
- 20 5. 根据权利要求2所述的方法，还包括：
接收用于输入所述第二语音内容的所述第二操作，其中所述语音表情的所述语音
部分基于所述第二语音内容而被确定。
6. 根据权利要求5所述的方法，还包括：
接收用于编辑所述目标图像的所述第一操作，其中所述语音表情的所述视觉部分
25 基于经编辑的所述目标图像而被确定。
7. 根据权利要求5所述的方法，还包括：
提供一组候选语音风格；以及
接收针对所述一组候选语音风格中的目标语音风格的选择，使得所述语音表情的
所述语音部分与所述目标语音风格匹配。
- 30 8. 根据权利要求5所述的方法，其中所述目标图像为目标动态图像，并且所述语
音部分被配置为从所述目标动态图像的起始帧播放。
9. 根据权利要求2所述的方法，其中呈现一组候选图像包括：
基于对表情面板中的创建入口的选择，呈现所述一组候选图像。

10. 根据权利要求 1 所述的方法，其中确定用于生成语音表情的目标媒体资源包括：
- 呈现一组候选图像表情；以及
- 基于对所述一组候选图像表情中的目标图像表情的选择，将所述目标图像表情确定为所述目标媒体资源。
11. 根据权利要求 10 所述的方法，还包括：
- 经由所述编辑界面，接收用于输入所述第二语音内容的所述第二操作，其中所述语音表情的所述视觉部分基于所述目标图像表情而被确定，所述语音部分基于所述第二语音内容而被确定。
12. 根据权利要求 1 所述的方法，其中确定用于生成语音表情的目标媒体资源包括：
- 基于与第一作品相关联的预设操作，将所述第一作品确定为所述目标媒体资源；
- 或
- 获取用户拍摄的图片或视频，以作为所述目标媒体内容。
13. 根据权利要求 1 所述的方法，还包括：
- 在表情面板中显示经发布的所述语音表情；以及
- 基于针对所述语音表情的选择，将所述语音表情发送至所述会话中。
14. 根据权利要求 13 所述的方法，还包括：
- 在所述会话的消息窗口中显示所发送的所述语音表情的所述视觉部分和所述语音表情的播放入口；以及
- 基于对所述播放入口的选择，播放所述语音表情的所述语音部分。
15. 根据权利要求 13 所述的方法，还包括：
- 在所述表情面板中，关联于所述语音表情提供预览入口；以及
- 基于对所述预览入口的选择，播放所述语音表情的所述语音部分。
16. 根据权利要求 1 所述的方法，其中发布目标表情包括以下至少一项：
- 将所述语音表情发布为第二作品，其中使得至少一个用户能够经由所述第二作品添加所述语音表情；
- 添加所述语音表情，使得所述语音表情能够用于当前用户；
- 向至少一个用户分享所述语音表情。
17. 一种表情交互的方法，包括：
- 接收对语音表情的选择；以及
- 将所述语音表情发送至目标会话中，所述语音表情包括视觉部分和语音部分，
- 其中，所述语音表情是基于与目标媒体资源相关联的编辑操作而被创建，所述编辑操作包括以下至少一项：用于修改所述目标媒体资源所包括的第一视觉内容和/或第

一语音内容的第一操作，用于添加第二视觉内容和/或第二语音内容的第二操作。

18. 根据权利要求 17 所述的方法，其中接收对语音表情的选择包括：

在所述目标会话的会话界面中呈现表情面板；以及

接收针对所述表情面板中的所述语音表情的选择。

5 19. 根据权利要求 18 所述的方法，还包括：

在所述表情面板中，关联于所述语音表情提供预览入口；以及

基于对所述预览入口的选择，播放所述语音表情的所述语音部分。

20. 根据权利要求 17 所述的方法，还包括：

10 在所述目标会话的消息窗口中显示所发送的所述语音表情的所述视觉部分和所述语音表情的播放入口；以及

基于对所述播放入口的选择，播放所述语音表情的所述语音部分。

21. 一种表情制作方法，包括：

基于目标语音内容，呈现基于所述目标语音内容所确定的第一组候选视觉内容；
以及

15 基于针对所述第一组候选视觉内容中的第一视觉内容的选择，创建语音表情，所述语音表情包括视觉部分和语音部分，其中所述视觉部分基于所述第一视觉内容而被确定，所述语音部分基于所述目标语音内容而被确定。

22. 根据权利要求 21 所述的方法，其中所述第一组候选视觉内容是基于与所述目标语音内容对应的文本内容而被生成。

20 23. 一种表情制作方法，包括：

基于目标文本内容，呈现基于所述目标文本内容所确定的第二组候选视觉内容；
以及

25 基于针对所述第二组候选视觉内容中的第二视觉内容的选择，创建语音表情，所述语音表情包括视觉部分和语音部分，其中所述视觉部分基于所述目标视觉内容而被确定，所述语音部分基于所述目标文本内容而被生成。

24. 根据权利要求 23 所述的方法，还包括：

提供一组候选语音风格；以及

接收针对所述一组候选语音风格中的目标语音风格的选择，使得所述语音表情的所述语音部分基于所述目标语音风格而被生成。

30 25. 一种用于表情交互的装置，包括：

确定模块，被配置为确定用于生成语音表情的目标媒体资源；

创建模块，被配置为基于接收到的编辑操作，创建语音表情，所述语音表情包括视觉部分和语音部分，所述编辑操作包括以下至少一项：用于修改所述目标媒体资源所包括的第一视觉内容和/或第一语音内容的第一操作，用于输入第二视觉内容和/或第

二语音内容的第二操作；以及

发布模块，被配置为发布所述语音表情，使得所述语音表情能够基于目标用户的选择而被发送至与所述目标用户相关联的会话中。

26. 一种电子设备，包括：

5 至少一个处理单元；以及

至少一个存储器，所述至少一个存储器被耦合到所述至少一个处理单元并且存储用于由所述至少一个处理单元执行的指令，所述指令在由所述至少一个处理单元执行时使所述电子设备执行根据权利要求 1 至 16、17 至 20、21 至 22 或 23 至 24 中任一项所述的方法。

10 27. 一种计算机可读存储介质，其上存储有计算机程序，所述计算机程序可由处理器执行以实现根据权利要求 1 至 16、17 至 20、21 至 22 或 23 至 24 中任一项所述的方法。