

# (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



(10) 国际公布号

WO 2020/239001 A1

(43) 国际公布日  
2020 年 12 月 3 日 (03.12.2020)

(51) 国际专利分类号:  
G10L 15/02 (2006.01)

(21) 国际申请号: PCT/CN2020/092802

(22) 国际申请日: 2020 年 5 月 28 日 (28.05.2020)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:  
201910472410.9 2019年5月31日 (31.05.2019) 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: 叶波 (YE, Bo); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 吴小进 (WU, Xiaojin); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 周昕宇 (ZHOU, Xinyu); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (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, IR, IS, 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,

(54) Title: HUMMING RECOGNITION METHOD AND RELATED DEVICE

(54) 发明名称: 一种哼唱识别方法及相关设备

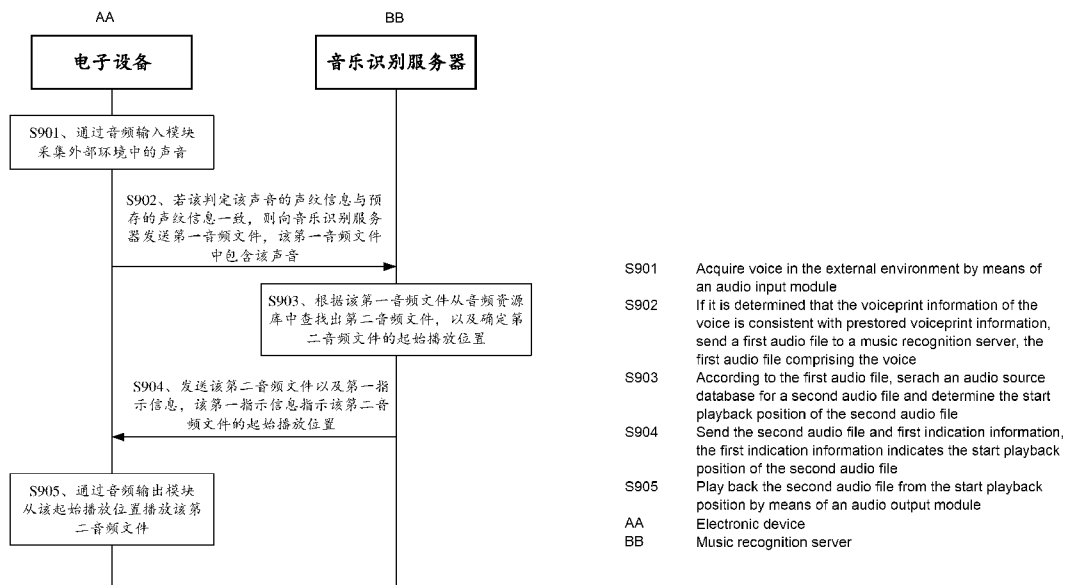


图 9

(57) Abstract: Disclosed in the present application are a humming recognition method and a related device. In the humming recognition method, an electronic device can continuously acquire voice in the external environment, and upon determining that the voice is produced by a preset user, the electronic device sends to a music recognition server a first audio file comprising the voice to perform humming recognition. Upon receiving a recognized second audio file and its start playback position sent by the music recognition server, the electronic device can start playing back the second audio file from the end position of the voice, wherein the start playback

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))。

position of the second audio file corresponds to the end position of the first audio file. In this way, the steps of operations conducted by a user to trigger a terminal to perform humming recognition can be reduced, and the efficiency of humming recognition can be improved; in addition, the audio can be played back subsequently to the humming of the user, so that the user experience can be improved.

(57) 摘要: 本申请公开了一种哼唱识别方法及相关设备, 在哼唱识别的方法中, 电子设备可以持续获取外部环境中的声音, 在判定该声音为预设用户发出的声音时, 该电子设备向音乐识别服务器发送包含该声音的第一音频文件, 以进行哼唱识别。在电子设备接收了音乐识别服务器发送的, 识别出的第二音频文件以及它的起始播放位置之后, 能够从该声音的结束位置开始播放该第二音频文件。其中, 该第二音频文件的起始播放位置与第一音频文件的结束位置相对应。通过这种方式, 可以减少用户触发终端进行哼唱识别的操作步骤, 提升哼唱识别的效率, 同时, 可以实现跟随用户的哼唱播放音频的效果, 提升用户体验。

### 一种哼唱识别方法及相关设备

本申请要求在 2019 年 5 月 31 日提交中国国家知识产权局、申请号为 201910472410.9 的中国专利申请的优先权，发明名称为“一种哼唱识别方法及相关设备”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

本申请涉及计算机技术领域，尤其涉及一种哼唱识别方法及相关设备。

#### 背景技术

哼唱识别是目前音频检索领域的研究热点。不同于利用文本（例如，歌曲名、演唱者或者歌词等文本）来检索音频的方式，也不同于利用一段正在播放的音乐来检索音频的方式，哼唱识别可以通过用户哼唱的音乐片段来检索音频。

现阶段，用户触发终端进行哼唱识别的方式主要有以下两种：第一种方式，用户首先需要查找具有哼唱识别功能的应用程序，再在该应用程序中查找哼唱识别对应的功能控件，之后对该功能控件执行操作，以触发终端进行哼唱识别。第二种方式，用户首先需要通过唤醒词唤醒智能语音助手（例如，siri，天猫精灵等），再输入语音指令以触发终端进行哼唱识别。可以看出，现有技术中，用户触发终端进行哼唱识别的方式较为繁琐。

#### 发明内容

本申请提供了一种哼唱识别方法及相关设备，可以减少用户触发终端进行哼唱识别的操作步骤，提升哼唱识别的效率，同时，可以实现跟随用户的哼唱播放音频的效果，提升用户体验。

上述目标和其他目标将通过独立权利要求中的特征来达成。进一步的实现方式在从属权利要求、说明书和附图中体现。

第一方面，本申请实施例提供了一种哼唱识别方法，该方法可包括：电子设备通过音频输入模块采集外部环境中的声音；若所述电子设备判定所述声音的声纹信息与预存的声纹信息一致，则所述电子设备向音乐识别服务器发送第一音频文件，所述第一音频文件中包含所述声音，所述音乐识别服务器用于根据所述第一音频文件从音频资源库中查找出第二音频文件，以及确定所述第二音频文件的起始播放位置；其中，所述第二音频文件的特征与所述第一音频文件的特征的相似度，高于第三音频文件的特征与所述声音的特征的相似度，所述第三音频文件为上述音频资源库中除所述第二音频文件的音频文件，所述第二音频文件的起始播放位置与所述第一音频文件的结束位置相对应；所述电子设备接收所述音乐识别服务器发送的所述第二音频文件以及第一指示信息，所述第一指示信息指示所述第二音频文件的起始播放位置；所述电子设备通过音频输出模块从所述起始播放位置播放所述第二音频文件。通过这种方式，可以减少用户触发终端进行哼唱识别的操作步骤，提升哼唱识别的效率，同时，可以实现跟随用户的哼唱播放音频的效果，提升了用户体验。

结合第一方面，在一种可能的实现方式中，所述方法还包括：所述电子设备通过摄像头获取用户的口型信息；若所述声音的声纹信息与预存的声纹信息一致，则所述电子设备向音

乐识别服务器发送所述口型信息；其中，所述音乐识别服务器还用于将所述口型信息转化为文本信息，所述根据所述第一音频文件从音频资源库中查找出第二音频文件，包括：根据所述第一音频文件和所述口型信息对应的文本信息从音频资源库中查找出第二音频文件，其中，所述第二音频文件对应的文本信息与所述口型信息对应的文本信息的相似度，高于所述第三音频文件对应的文本信息与所述口型信息对应的文本信息的相似度。

结合第一方面，在一种可能的实现方式中，所述电子设备通过摄像头获取用户的口型信息，包括：若所述电子设备判定所述声音为人声，则通过摄像头获取用户的口型信息。

结合第一方面，在一种可能的实现方式中，所述电子设备通过音频输入模块采集外部环境中的声音，包括：若所述电子设备判定所述音频输入模块和/或所述音频输出模块未被占用，则所述电子设备通过所述音频输入模块采集外部环境中的声音。

结合第一方面，在一种可能的实现方式中，所述第二音频文件的标签包含于第一用户的用户标签。

结合第一方面，在一种可能的实现方式中，在所述电子设备通过音频输出模块从所述起始播放位置播放所述第二音频文件之后，所述方法还包括：所述电子设备显示第二音频文件的标识信息，以及播放控件；其中，所述播放控件的显示状态为第一状态，所述第一状态表示所述第二音频文件正在被播放；若所述电子设备检测到作用于处于所述第一状态的所述播放控件的第一用户操作，响应于所述第一用户操作，所述电子设备暂停播放所述第二音频文件，并将所述播放控件的显示状态设为第二状态，所述第二状态表示所述第二音频文件暂停播放。

结合第一方面，在一种可能的实现方式中，所述方法还包括：当检测到所述电子设备处于锁定状态时，所述电子设备停止通过所述音频输入模块采集外部环境中的声音。

结合第一方面，在一种可能的实现方式中，所述方法还包括：当检测到所述电子设备处于预设地点时，所述电子设备停止通过所述音频输入模块采集外部环境中的声音。

结合第一方面，在一种可能的实现方式中，所述电子设备通过音频输出模块从所述起始播放位置播放所述第二音频文件，包括：若所述电子设备判定所述电子设备的位置与预设地点不一致，则所述电子设备通过所述音频输出模块从所述起始播放位置播放所述第二音频文件。

结合第一方面，在一种可能的实现方式中，所述方法还包括：所述电子设备在第一时间段内停止通过所述音频输入模块采集外部环境中的声音。

结合第一方面，在一种可能的实现方式中，该电子设备通过音频输入模块采集外部环境中的声音，包括：若电子设备判定自身的哼唱识别功能开启，则电子设备通过音频输入模块采集外部环境中的声音。

结合第一方面，在一种可能的实现方式中，该方法还包括：当检测到环境光亮度小于预设值的持续时间，大于预设时间时，所述电子设备停止通过音频输入模块采集外部环境中的声音。

结合第一方面，在一种可能的实现方式中，该音乐识别服务器还用于当该音乐识别服务器判定该声音信号为音乐片段时，根据所述第一音频文件从音频资源库中查找出第二音频文件。

结合第一方面，在一种可能的实现方式中，电子设备从开始播放第二音频文件的时刻到预设时刻（例如，第5秒，第6秒等时间值）的时间段内，将使播放第二音频文件的音量由

低到高逐渐增大。

结合第一方面，在一种可能的实现方式中，在所述电子设备通过音频输出模块从所述起始播放位置播放所述第二音频文件之后，该电子设备还可以检测该第二音频文件是否存储在预存的音乐文件夹中，若是，该电子设备可以在播放完该第二音频文件之后，播放该音乐文件夹中的其他音频文件。

第二方面，本申请实施例提供了一种电子设备，这种电子设备包括音频输入模块，音频输出模块，处理器，存储器，其中：所述存储器用于存储程序指令；所述处理器用于根据所述程序指令执行以下操作：通过音频输入模块采集外部环境中的声音；若判定所述声音的声纹信息与预存的声纹信息一致，则向音乐识别服务器发送第一音频文件，所述第一音频文件中包含所述声音，所述音乐识别服务器用于根据所述第一音频文件从音频资源库中查找出第二音频文件，以及确定第二音频文件的起始播放位置；其中，所述第二音频文件的特征与所述第一音频文件的特征的相似度，高于第三音频文件的特征与所述声音的特征的相似度，所述第三音频文件为上述音频资源库中除所述第二音频文件的音频文件，所述第二音频文件的起始播放位置与所述第一音频文件的结束位置相对应；接收所述音乐识别服务器发送的所述第二音频文件以及第一指示信息，所述第一指示信息指示所述第二音频文件的起始播放位置；通过音频输出模块从所述起始播放位置播放所述第二音频文件。通过这种电子设备，可以减少用户触哼唱识别的操作步骤，提升哼唱识别的效率，同时，可以实现跟随用户的哼唱播放音频的效果，提升了用户体验。

结合第二方面，在一种可能的实现方式中，所述电子设备还包括摄像头，所述处理器还用于根据所述程序指令执行以下操作：通过摄像头获取用户的口型信息；若所述声音的声纹信息与预存的声纹信息一致，则向音乐识别服务器发送所述口型信息；其中，所述音乐识别服务器还用于将所述口型信息转化为文本信息；所述音乐识别服务器还具体用于：根据所述第一音频文件和所述口型信息对应的文本信息从音频资源库中查找出第二音频文件，其中，所述第二音频文件对应的文本信息与所述口型信息对应的文本信息的相似度，高于所述第三音频文件对应的文本信息与所述口型信息对应的文本信息的相似度。

结合第二方面，在一种可能的实现方式中，所述处理器具体用于根据所述程序指令执行以下操作：若判定所述声音为人声，则通过摄像头获取用户的口型信息。

结合第二方面，在一种可能的实现方式中，所述处理器具体用于根据所述程序指令执行以下操作：若判定所述音频输入模块和/或音频输出模块未被占用，则通过音频输入模块采集外部环境中的声音。

结合第二方面，在一种可能的实现方式中，所述第二音频文件的标签包含于第一用户的用户标签。

结合第二方面，在一种可能的实现方式中，所述电子设备还包括显示屏，所述处理器还用于根据所述程序指令执行以下操作：通过显示屏显示第二音频文件的标识信息，以及播放控件；其中，所述播放控件的显示状态为第一状态，所述第一状态表示所述第二音频文件正在被播放；若检测到作用于处于所述第一状态的所述播放控件的第一用户操作，响应于所述第一用户操作，暂停播放所述第二音频文件，并将所述播放控件的显示状态设为第二状态，所述第二状态表示所述第二音频文件暂停播放。

结合第二方面，在一种可能的实现方式中，所述处理器还用于根据所述程序指令执行以

下操作：当检测到所述电子设备处于锁定状态时，停止通过音频输入模块采集外部环境中的声音。

结合第二方面，在一种可能的实现方式中，所述处理器还用于根据所述程序指令执行以下操作：当检测到所述电子设备处于预设地点时，停止通过音频输入模块采集外部环境中的声音。

结合第二方面，在一种可能的实现方式中，所述处理器具体用于根据所述程序指令执行以下操作：若判定所述电子设备的位置与预设地点不一致，则通过音频输出模块从所述起始播放位置播放所述第二音频文件。

结合第二方面，在一种可能的实现方式中，所述处理器还用于根据所述程序指令执行以下操作：在第一时间段内停止通过音频输入模块采集外部环境中的声音。

结合第二方面，在一种可能的实现方式中，所述处理器具体用于根据所述程序指令执行以下操作：若判定自身的哼唱识别功能开启，则通过音频输入模块采集外部环境中的声音。

结合第二方面，在一种可能的实现方式中，所述处理器还用于根据所述程序指令执行以下操作：当检测到环境光亮度小于预设值的持续时间，大于预设时间时，停止通过音频输入模块采集外部环境中的声音。

结合第二方面，在一种可能的实现方式中，该音乐识别服务器还用于当该音乐识别服务器判定该声音信号为音乐片段时，根据所述第一音频文件从音频资源库中查找出第二音频文件。

结合第二方面，在一种可能的实现方式中，所述处理器还用于根据所述程序指令执行以下操作：从开始播放第二音频文件的时刻到预设时刻（例如，第5秒，第6秒等时间值）的时间段内，将使播放第二音频文件的音量由低到高逐渐增大。

结合第二方面，在一种可能的实现方式中，所述处理器还用于根据所述程序指令执行以下操作：检测该第二音频文件是否存储在预存的音乐文件夹中，若是，则在播放完该第二音频文件之后，播放该音乐文件夹中的其他音频文件。

第三方面，本申请实施例提供了又一种哼唱识别方法，该方法包括：开放平台获取第一音频文件，所述第一音频文件中包括外部环境中的声音；若所述开放平台判定所述第一音频文件的声纹信息与预存的声纹信息一致，则所述开放平台根据所述第一音频文件从音频资源库中查找第二音频文件，以及确定所述第二音频文件的起始播放位置；其中，所述第二音频文件的特征与所述第一音频文件的特征的相似度，高于第三音频文件的特征与所述声音的特征的相似度，所述第三音频文件为上述音频资源库中除所述第二音频文件的音频文件，所述第二音频文件的起始播放位置与所述第一音频文件的结束位置相对应；所述开放平台从所述起始播放位置播放所述第二音频文件，或所述开发平台控制电子设备的其他应用程序从所述起始播放位置播放所述第二音频文件。通过这种方式，可以减少用户触发哼唱识别的操作步骤，提升哼唱识别的效率，同时，可以实现跟随用户的哼唱播放音频的效果，提升了用户体验。

结合第三方面，在一种可能的实现方式中，所述方法还包括：所述开放平台通过所述电子设备获取用户的口型信息；若所述开放平台判定所述第一音频文件的声纹信息与预存的声纹信息一致，所述开放平台将所述口型信息转化为文本信息；所述根据所述第一音频文件从音频资源库中查找出第二音频文件，包括：根据所述第一音频文件和所述口型信息对应的文

本信息从音频资源库中查找出第二音频文件，其中，所述第二音频文件对应的文本信息与所述口型信息对应的文本信息的相似度，高于所述第三音频文件对应的文本信息与所述口型信息对应的文本信息的相似度。

结合第三方面，在一种可能的实现方式中，所述开放平台获取用户的口型信息，包括：若所述开放平台判定所述第一音频文件中包括的声音为人声，则通过所述电子设备获取用户的口型信息。

结合第三方面，在一种可能的实现方式中，开放平台获取第一音频文件，包括：若所述开放平台判定音频输入模块和/或音频输出模块未被其他应用占用，则所述开放平台获取第一音频文件。

结合第三方面，在一种可能的实现方式中，所述第二音频文件的标签包含于第一用户的用户标签。

结合第三方面，在一种可能的实现方式中，在所述开放平台从所述起始播放位置播放所述第二音频文件之后，所述方法还包括：所述开放平台通过电子设备显示第二音频文件的标识信息，以及播放控件；其中，所述播放控件的显示状态为第一状态，所述第一状态表示所述第二音频文件正在被播放；若所述开放平台检测到作用于处于所述第一状态的所述播放控件的第一用户操作，响应于所述第一用户操作，所述开放平台暂停播放所述第二音频文件，或控制电子设备的其他应用程序暂停播放所述第二音频文件，并将所述播放控件的显示状态设为第二状态，所述第二状态表示所述第二音频文件暂停播放。

结合第三方面，在一种可能的实现方式中，所述方法还包括：当检测到所述电子设备处于锁定状态时，所述开放平台停止获取第一音频文件。

结合第三方面，在一种可能的实现方式中，所述方法还包括：当检测到所述电子设备处于预设地点时，所述开放平台停止获取第一音频文件。

结合第三方面，在一种可能的实现方式中，所述开放平台从所述起始播放位置播放所述第二音频文件，或所述开发平台控制电子设备的其他应用程序从所述起始播放位置播放所述第二音频文件，包括：若所述开放平台判定所述电子设备的位置与预设地点不一致，则所述开放平台从所述起始播放位置播放所述第二音频文件，或所述开发平台控制电子设备的其他应用程序从所述起始播放位置播放所述第二音频文件。

结合第三方面，在一种可能的实现方式中，所述方法还包括：所述开放平台在第一时间段内停止获取第一音频文件。

结合第三方面，在一种可能的实现方式中，该开放平台获取第一音频文件，包括：若所述电子设备的哼唱识别功能开启，则开放平台获取第一音频文件。

结合第三方面，在一种可能的实现方式中，该方法还包括：当检测到所述电子设备的环境光亮度小于预设值的持续时间，大于预设时间时，所述开放平台停止获取第一音频文件。

结合第三方面，在一种可能的实现方式中，该开放平台还用于当判定该第一音频文件为音乐片段时，根据所述第一音频文件从音频资源库中查找出第二音频文件。

结合第三方面，在一种可能的实现方式中，开放平台从开始播放第二音频文件的时刻到预设时刻（例如，第5秒，第6秒等时间值）的时间段内，将使播放第二音频文件的音量由低到高逐渐增大。

结合第三方面，在一种可能的实现方式中，在所述开发平台控制电子设备的其他应用程序从所述起始播放位置播放所述第二音频文件之后，该开放平台还可以检测该第二音频文件

是否存储在所述电子设备的预存的音乐文件夹中，若是，该开放平台可以在控制电子设备的其他应用程序播放完该第二音频文件之后，控制电子设备的其他应用程序播放该音乐文件夹中的其他音频文件。

第四方面，本申请实施例提供一种包含指令的计算机程序产品，当所述计算机程序产品在电子设备上运行时，使得所述电子设备执行如使得上述电子设备执行如第一方面中任一可能的实现方式，或者当所述计算机程序产品在开放平台上运行时，使得上述开放平台执行如上述第三方面中任一可能的实现方式。

第五方面，本申请实施例提供一种计算机可读存储介质，包括指令，其特征在于，当所述指令在电子设备上运行时，使得所述电子设备执行如使得上述电子设备执行如第一方面中任一可能的实现方式，或者当所述指令在开放平台上运行时，使得上述开放平台执行如上述第三方面中任一可能的实现方式。

在本申请提供的哼唱识别的方法中，电子设备可以持续获取外部环境中的声音，在判定该声音为预设用户发出的声音时，该电子设备向音乐识别服务器发送包含该声音的第一音频文件，以进行哼唱识别。在电子设备接收了音乐识别服务器发送的，识别出的第二音频文件以及它的起始播放位置之后，能够从该声音的结束位置开始播放该第二音频文件。其中，该第二音频文件的起始播放位置与第一音频文件的结束位置相对应。通过这种方式，可以减少用户触发终端进行哼唱识别的操作步骤，提升哼唱识别的效率，同时，可以实现跟随用户的哼唱播放音频的效果，提升用户体验。

#### 附图说明

为了更清楚地说明本申请实施例或背景技术中的技术方案，下面将对本申请实施例或背景技术中所需要使用的附图进行说明。

图 1A 是本申请实施例提供的一种智能终端的结构示意图；

图 1B 是本申请实施例提供的一种智能终端的软件结构框图；

图 1C 是本申请实施例提供的一种智能家居设备的结构示意图；

图 1D 是本申请实施例提供的一种车载设备的结构示意图；

图 2 是本申请实施例提供的一种智能终端上的用于显示应用程序菜单的用户界面；

图 3A-图 3B 是本申请实施例提供的一些显示识别结果的用户界面；

图 3C 是本申请实施例提供的一种智能终端处于锁定状态下显示的用户界面；

图 3D-图 3F 是本申请实施例提供的又一些显示识别结果的用户界面；

图 3G 是本申请实施例提供的一种用于哼唱识别的用户界面；

图 4A-图 4B 是本申请实施例提供的一些用于设置哼唱识别功能的用户界面；

图 5A-图 5C 是本申请实施例提供的又一些用于设置哼唱识别功能的用户界面；

图 5D-图 5F 是本申请实施例提供的一些用于设置哼唱识别功能的访问权限的用户界面；

图 5G 是本申请实施例提供的一种用于录入声纹信息的用户界面；

图 6A-图 6B 是本申请实施例提供的又一些用于设置哼唱识别功能的用户界面；

图 6C 是本申请实施例提供的又一种用于录入声纹信息的用户界面；

图 7A-图 7B 是本申请实施例提供的一些车载设备上用于设置哼唱识别功能的用户界面；

图 7C 是本申请实施例提供的又一种用于录入声纹信息的用户界面；

图 8A-图 8B 是本申请实施例提供的一些车载设备上用于显示识别结果的用户界面；  
图 9 是本申请实施例提供的一种哼唱识别方法的流程图。

### 具体实施方式

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描述。

其中，在本申请实施例的描述中，除非另有说明，“/”表示或的意思，例如，A/B 可以表示 A 或 B；本文中的“和/或”仅仅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，在本申请实施例的描述中，“多个”是指两个或两个以上。

以下，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。在本申请实施例的描述中，除非另有说明，“多个”的含义是两个或两个以上。

首先，对本申请涉及到的一些概念进行具体的介绍。

哼唱识别，是一种通过用户哼唱的音乐片段来进行音频检索的方式。哼唱识别的工作原理是：电子设备获取用户哼唱的一个音乐片段，然后再将该音乐片段发送给服务器，服务器通过相似度匹配出和用户哼唱片段最相似的音频文件，之后，该服务器将该音频文件反馈给电子设备。可选的，服务器通过从音乐片段中提取特征（例如，基频序列），然后利用该特征进行检索，从预存的音频资源库中匹配出和用户哼唱片段最相似的音频文件。由于用户哼唱的片段和库中实际音频文件的片段不可能完全相似，所以哼唱识别是一种模糊匹配。针对模糊匹配，可以利用字符串编辑距离，以及动态时间规整（dynamic time warping, DTW）算法等提升识别的准确性。

用户界面（user interface, UI），是应用程序或操作系统与用户之间进行交互和信息交换的介质接口，它实现信息的内部形式与用户可以接受形式之间的转换。应用程序的用户界面是通过 java、可扩展标记语言（extensible markup language, XML）等特定计算机语言编写的源代码，界面源代码在电子设备 300 设备上经过解析，渲染，最终呈现为用户可以识别的内容，比如图片、文字、按钮等控件。控件（control），是用户界面的基本元素，典型的控件有按钮（button）、小工具(widget)、工具栏（toolbar）、菜单栏（menu bar）、文本框（text box）、滚动条（scrollbar）、图片和文本。界面中的控件的属性和内容是通过标签或者节点来定义的，比如 XML 通过<TextView>、<ImageView>、<VideoView>等节点来规定界面所包含的控件。一个节点对应界面中一个控件或属性，节点经过解析和渲染之后呈现为用户可视的内容。此外，很多应用程序，比如混合应用（hybrid application）的界面中通常还包含有网页。网页，也称为页面，可以理解为内嵌在应用程序界面中的一个特殊的控件，网页是通过特定计算机语言编写的源代码，例如超文本标记语言（hyper text markup language, HTML），层叠样式表（cascading style sheets, CSS），java 脚本（JavaScript, JS）等，网页源代码可以由浏览器或与浏览器功能类似的网页显示组件加载和显示为用户可识别的内容。网页所包含的具体内容也是通过网页源代码中的标签或者节点来定义的，比如 HTML 通过<p>、<img>、<video>、<canvas>来定义网页的元素和属性。

用户界面常用的表现形式是图形用户界面（graphic user interface, GUI），是指采用图形方式显示的与计算机操作相关的用户界面。它可以是在电子设备的显示屏中显示的一个图标、

窗口、控件等界面元素。

本申请以下实施例提供了一种哼唱识别方法及电子设备，可以使得电子设备在用户哼唱音乐片段的过程中，能够跟随用户的哼唱播放该音乐片段对应的音频文件，减少用户触发终端进行哼唱识别的操作步骤，提升哼唱识别的效率。

本申请以下实施例中，电子设备（例如，智能终端、智能家居、车载设备等等）在具有哼唱识别的权限的情况下，会执行本申请实施例提供的哼唱识别操作。该哼唱识别操作的实施流程可参照以下步骤：首先，电子设备通过音频输入模块（例如，麦克风）采集外部环境中的声音；接着，若所述电子设备判定所述声音的声纹信息与预存的声纹信息一致，则电子设备将包含这段声音的第一音频文件发送给音乐识别服务器进行哼唱识别，以识别出和用户哼唱的音乐片段相匹配的音频文件，以及确定音频文件的起始播放位置。其中，该识别出的音频文件的起始播放位置与所述第一音频文件的结束位置相对应。在电子设备接收音乐识别服务器反馈的音频文件以及包含该起始播放位置的信息之后，电子设备可以从该起始播放位置播放该音频文件，从而达到跟随用户的哼唱播放音频文件的目的。后续内容将对执行该哼唱识别操作的系统架构以及实施流程作进一步的介绍，此处不具体展开。

本申请以下实施例中，在电子设备执行本申请实施例提供的哼唱识别操作之前，需要判定自身的音频输入模块和/或音频输出模块是否被占用，若自身的音频输入模块和/或音频输出模块被占用，例如，播放音频/视频、拨打电话、进行语音导航等等，则电子设备不执行本申请实施例提供的哼唱识别操作；若该电子设备自身的音频输入模块和/或音频输出模块未被占用，则电子设备执行本申请实施例提供的哼唱识别操作。可选的，在电子设备的音频输入模块和/或音频输出模块被释放之后，例如，音频/视频播放结束、电话挂断、语音导航结束等等，该电子设备可执行本申请实施例提供的哼唱识别操作。也可以理解为，本申请实施例提供的哼唱识别操作的优先级，低于该电子设备中除该哼唱识别操作的其他需占用音频输入模块和/或音频输出模块的操作的优先级。

本申请以下实施例中，在该电子设备执行本申请实施例提供的哼唱识别操作的过程中，若检测到其他需占用音频输入模块和/或音频输出模块的操作的对音频资源的请求，该电子设备调用该音频输入模块和/或音频输出模块执行该请求对应的操作。特殊的，若该请求需占用音频输出模块的时间小于预设值（例如，1秒），举例而言，该请求可以为发出通知提示音（例如，短消息提示音、应用程序推送提示音），该哼唱识别操作可占用音频输入模块，同时，该请求对应的操作占用音频输出模块。

本申请以下实施例中，用于执行本申请实施例提供的哼唱识别操作的可以是电子设备的一个系统应用程序或者第三方应用程序。在一种可能的实现方式中，该系统应用程序或者该第三方应用程序可以专用于执行本申请实施例提供的哼唱识别操作；在又一种可能的实现方式中，该系统应用程序或者该第三方应用程序还可以执行其他的服务（或功能），本申请实施例提供的哼唱识别操作仅作为一种服务（或功能）集成在该系统应用程序或者该第三方应用程序之中。

可以理解的是，“哼唱识别”只是本实施例中所使用的名称，其代表的含义在本实施例中已经记载，其名称并不能对本实施例构成任何限制。例如，在一些可能的实施方式中，“哼唱识别”还可以被称为“听歌识曲”“哼唱检索”等名称。

在本申请实施例中，执行哼唱识别操作的电子设备可以是智能终端，还可以是智能家居

设备，还可以是车载设备。接下来首先介绍本申请以下实施例中提供的示例性的智能终端100。

图1A示出了智能终端100的结构示意图。

智能终端100可以包括处理器110，外部存储器接口120，内部存储器121，通用串行总线(universal serial bus, USB)接口130，充电管理模块140，电源管理模块141，电池142，天线1，天线2，移动通信模块150，无线通信模块160，音频模块170，扬声器170A，受话器170B，麦克风170C，耳机接口170D，传感器模块180，按键190，马达191，指示器192，摄像模组193，显示屏194，以及用户标识模块(subscriber identification module, SIM)卡接口195等。其中传感器模块180可以包括压力传感器180A，陀螺仪传感器180B，加速度传感器180C，距离传感器180D，接近光传感器180E，指纹传感器180F，触摸传感器180G，环境光传感器180H等。

可以理解的是，本申请实施例示意的结构并不构成对智能终端100的具体限定。在本申请另一些实施例中，智能终端100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。

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)，中央处理器(central processing unit, CPU)，图形处理器(graphics processing unit, GPU)，神经网络处理器(neural-network processing unit, NPU)，调制解调处理器，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。在一些实施例中，智能终端100也可以包括一个或多个处理器110。

其中，控制器可以是智能终端100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。

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，因而提高了智能终端100的效率。

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)。在一些实施例中，处理器110可以包含多组I2C总线。处理器110可以通过不同的I2C总线接口分别耦合触摸传感器180K，充电器，闪光灯，摄像模组193等。例如：处理器110可以通过I2C接口耦合触摸传感器180K，使处理器110与触摸传感器180K通过I2C总线接口通信，实现智能终端100的触摸功能。

I2S接口可以用于音频通信。在一些实施例中，处理器110可以包含多组I2S总线。处理

器 110 可以通过 I2S 总线与音频模块 170 耦合, 实现处理器 110 与音频模块 170 之间的通信。在一些实施例中, 音频模块 170 可以通过 I2S 接口向无线通信模块 160 传递音频信号, 从而实现通过蓝牙耳机播放识别出的音频文件的功能。

PCM 接口也可以用于音频通信, 将模拟信号抽样, 量化和编码。在一些实施例中, 音频模块 170 与无线通信模块 160 可以通过 PCM 总线接口耦合。在一些实施例中, 音频模块 170 也可以通过 PCM 接口向无线通信模块 160 传递音频信号, 从而实现通过蓝牙耳机播放识别出的音频文件的功能。所述 I2S 接口和所述 PCM 接口都可以用于音频通信。

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。在一些实施例中, UART 接口通常被用于连接处理器 110 与无线通信模块 160。例如: 处理器 110 通过 UART 接口与无线通信模块 160 中的蓝牙模块通信, 实现蓝牙功能。在一些实施例中, 音频模块 170 可以通过 UART 接口向无线通信模块 160 传递音频信号, 实现通过蓝牙耳机播放识别出的音频文件的功能。

MIPI 接口可以被用于连接处理器 110 与显示屏 194, 摄像模组 193 等外围器件。MIPI 接口包括摄像头串行接口(camera serial interface, CSI), 显示屏串行接口(display serial interface, DSI)等。在一些实施例中, 处理器 110 和摄像模组 193 通过 CSI 接口通信, 实现智能终端 100 的摄像功能, 从而获取用户的口型信息。处理器 110 和显示屏 194 通过 DSI 接口通信, 实现智能终端 100 的显示功能。

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。在一些实施例中, GPIO 接口可以用于连接处理器 110 与摄像模组 193, 显示屏 194, 无线通信模块 160, 音频模块 170, 传感器模块 180 等。GPIO 接口还可以被配置为 I2C 接口, I2S 接口, UART 接口, MIPI 接口等。

USB 接口 130 是符合 USB 标准规范的接口, 具体可以是 Mini USB 接口, Micro USB 接口, USB Type C 接口等。USB 接口 130 可以用于连接充电器为智能终端 100 充电, 也可以用于智能终端 100 与外围设备之间传输数据。也可以用于连接耳机, 通过耳机播放音频。该接口还可以用于连接其他智能终端, 例如 AR 设备等。

可以理解的是, 本申请实施例示意的各模块间的接口连接关系, 只是示意性说明, 并不构成对智能终端 100 的结构限定。在另一些实施例中, 智能终端 100 也可以采用上述实施例中不同的接口连接方式, 或多种接口连接方式的组合。

充电管理模块 140 用于从充电器接收充电输入。其中, 充电器可以是无线充电器, 也可以是有线充电器。在一些有线充电的实施例中, 充电管理模块 140 可以通过 USB 接口 130 接收有线充电器的充电输入。在一些无线充电的实施例中, 充电管理模块 140 可以通过智能终端 100 的无线充电线圈接收无线充电输入。充电管理模块 140 为电池 142 充电的同时, 还可以通过电源管理模块 141 为智能终端供电。

电源管理模块 141 用于连接电池 142, 充电管理模块 140 与处理器 110。电源管理模块 141 接收电池 142 和/或充电管理模块 140 的输入, 为处理器 110, 内部存储器 121, 外部存储器, 显示屏 194, 摄像模组 193, 和无线通信模块 160 等供电。电源管理模块 141 还可以用于监测电池容量, 电池循环次数, 电池健康状态(漏电, 阻抗)等参数。在其他一些实施例中, 电源管理模块 141 也可以设置于处理器 110 中。在另一些实施例中, 电源管理模块 141 和充电管理模块 140 也可以设置于同一个器件中。

智能终端 100 的无线通信功能可以通过天线 1, 天线 2, 移动通信模块 150, 无线通信模

块 160, 调制解调处理器以及基带处理器等实现。

天线 1 和天线 2 用于发射和接收电磁波信号。智能终端 100 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: 可以将天线 1 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 150 可以提供应用在智能终端 100 上的包括 2G/3G/4G/5G 等无线通信的解决方案。移动通信模块 150 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 150 可以由天线 1 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 150 还可以对经调制解调处理器调制后的信号放大, 经天线 1 转为电磁波辐射出去。在一些实施例中, 移动通信模块 150 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 150 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。其中, 调制器用于将待发送的低频基带信号调制成中高频信号。解调器用于将接收的电磁波信号解调为低频基带信号。随后解调器将解调得到的低频基带信号传送至基带处理器处理。低频基带信号经基带处理器处理后, 被传递给应用处理器。应用处理器通过音频设备(不限于扬声器 170A, 受话器 170B 等)输出声音信号, 或通过显示屏 194 显示图像或视频。在一些实施例中, 调制解调处理器可以是独立的器件。在另一些实施例中, 调制解调处理器可以独立于处理器 110, 与移动通信模块 150 或其他功能模块设置在同一个器件中。

无线通信模块 160 可以提供应用在智能终端 100 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), 全球导航卫星系统(global navigation satellite system, 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 160 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 160 经由天线 2 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 160 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 2 转为电磁波辐射出去。示例性地, 无线通信模块 160 可以包括蓝牙模块、Wi-Fi 模块等。在一种可能的实现方式中, 智能终端可以通过无线通信模块 160 确定自身所在的位置。

在一些实施例中, 智能终端 100 的天线 1 和移动通信模块 150 耦合, 天线 2 和无线通信模块 160 耦合, 使得智能终端 100 可以通过无线通信技术与网络以及其他设备通信。所述无线通信技术可以包括全球移动通讯系统(global system for mobile communications, GSM), 通用分组无线服务(general packet radio service, GPRS), 码分多址接入(code division multiple access, CDMA), 宽带码分多址(wideband code division multiple access, WCDMA), 时分码分多址(time-division code division multiple access, TD-SCDMA), 长期演进(long term evolution, LTE), BT, GNSS, WLAN, NFC, FM, 和/或 IR 技术等。所述 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)。

智能终端 100 通过 GPU, 显示屏 194, 以及应用处理器等可以实现显示功能。GPU 为图像处理的微处理器, 连接显示屏 194 和应用处理器。GPU 用于执行数学和几何计算, 用于图

形渲染。处理器 110 可包括一个或多个 GPU，其执行指令以生成或改变显示信息。

显示屏 194 用于显示图像，视频等。显示屏 194 包括显示面板。显示面板可以采用液晶显示屏(liquid crystal display, LCD)，有机发光二极管(organic light-emitting diode, OLED)，有源矩阵有机发光二极体或主动矩阵有机发光二极体(active-matrix organic light emitting diode 的, AMOLED)，柔性发光二极管(flex light-emitting diode, FLED)，Miniled，MicroLed，Micro-oLed，量子点发光二极管(quantum dot light emitting diodes, QLED)等。在一些实施例中，智能终端 100 可以包括 1 个或 N 个显示屏 194，N 为大于 1 的正整数。

智能终端 100 可以通过摄像模组 193，ISP，视频编解码器，GPU，显示屏 194 以及应用处理器 AP、神经网络处理器 NPU 等实现摄像功能。

摄像模组 193 可用于采集拍摄对象的彩色图像数据。ISP 可用于处理摄像模组 193 采集的彩色图像数据。例如，拍照时，打开快门，光线通过镜头被传递到摄像头感光元件上，光信号转换为电信号，摄像头感光元件将所述电信号传递给 ISP 处理，转化为肉眼可见的图像。ISP 还可以对图像的噪点，亮度，肤色进行算法优化。ISP 还可以对拍摄场景的曝光，色温等参数优化。在一些实施例中，ISP 可以设置在摄像模组 193 中。

在一些实施例中，彩色摄像模组的摄像头的感光元件可以是电荷耦合器件(charge coupled device, CCD)或互补金属氧化物半导体(complementary metal-oxide-semiconductor, CMOS)光电子晶体管。感光元件把光信号转换成电信号，之后将电信号传递给 ISP 转换成数字图像信号。ISP 将数字图像信号输出到 DSP 加工处理。DSP 将数字图像信号转换成标准的 RGB，YUV 等格式的图像信号。

在一些实施例中，智能终端 100 可以包括 1 个或 N 个摄像模组 193，N 为大于 1 的正整数。具体的，智能终端 100 可以包括 1 个前置摄像模组 193 以及 1 个后置摄像模组 193。其中，前置摄像模组 193 通常可用于采集面对显示屏 194 的拍摄者自己的彩色图像数据，后置摄像模组 193 可用于采集拍摄者所面对的拍摄对象（如人物、风景等）的彩色图像数据。

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。例如，当智能终端 100 在频点选择时，数字信号处理器用于对频点能量进行傅里叶变换等。

视频编解码器用于对数字视频压缩或解压缩。智能终端 100 可以支持一种或多种视频编解码器。这样，智能终端 100 可以播放或录制多种编码格式的视频，例如：动态图像专家组(moving picture experts group, MPEG)-1, MPEG-2, MPEG-3, MPEG-4 等。

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 可以实现智能终端 100 的智能认知等应用，例如：图像识别，人脸识别，语音识别，文本理解等。

外部存储器接口 120 可以用于连接外部存储卡，例如 Micro SD 卡，实现扩展智能终端 100 的存储能力。外部存储卡通过外部存储器接口 120 与处理器 110 通信，实现数据存储功能。例如将音频文件、照片、视频等数据保存在外部存储卡中。

内部存储器 121 可以用于存储一个或多个计算机程序，该一个或多个计算机程序包括指令。处理器 110 可以通过运行存储在内部存储器 121 的上述指令，从而使得智能终端 100 执行本申请一些实施例中所提供的智能终端的拍照预览方法，以及各种功能应用以及数据处理等。内部存储器 121 可以包括存储程序区和存储数据区。其中，存储程序区可存储操作系统；该存储程序区还可以存储一个或多个应用程序(比如图库、联系人等)等。存储数据区可存储

智能终端 100 使用过程中所创建的数据(比如照片, 联系人等)。此外, 内部存储器 121 可以包括高速随机存取存储器, 还可以包括非易失性存储器, 例如至少一个磁盘存储器件, 闪存器件, 通用闪存存储器(universal flash storage, UFS)等。

智能终端 100 可以通过音频模块 170, 扬声器 170A, 受话器 170B, 麦克风 170C, 耳机接口 170D, 以及应用处理器等实现音频功能。例如音乐播放, 录音等。

音频模块 170 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 170 还可以用于对音频信号编码和解码。在一些实施例中, 音频模块 170 可以设置于处理器 110 中, 或将音频模块 170 的部分功能模块设置于处理器 110 中。

音频输出模块 170A, 也称为“扬声器”“喇叭”, 用于将音频电信号转换为声音信号。智能终端 100 可以通过扬声器 170A 收听音乐, 或收听免提通话。

音频输出模块 170B, 也称为“受话器”“听筒”, 用于将音频电信号转换成声音信号。当智能终端 100 接听电话或语音信息时, 可以通过将受话器 170B 靠近人耳接听语音。

音频输入模块 170C, 也称为“话筒”“传声器”, 用于将声音信号转换为电信号。当拨打电话或发送语音信息时, 用户可以通过人嘴靠近麦克风 170C 发声, 将声音信号输入到麦克风 170C。智能终端 100 可以设置至少一个麦克风 170C。在另一些实施例中, 智能终端 100 可以设置两个麦克风 170C, 除了采集声音信号, 还可以实现降噪功能。在另一些实施例中, 智能终端 100 还可以设置三个, 四个或更多麦克风 170C, 实现采集声音信号, 降噪, 还可以识别声音来源, 实现定向录音功能等。在本申请实施例中, 在智能终端 100 开启了“哼唱识别”功能的情况下, 麦克风 170C 可以采集智能终端 100 附近的的声音信号。

在一些实施例中, 处理器 110 中的 CPU 或数字处理器或音频处理器可以对麦克风 170C 所采集的声音进行处理。在一种实施例中, 当处理器 110 判定预设时间内收集到的声音为人声时, 处理器 110 从该声音中提取出声纹信息, 若所述声音的声纹信息与预存的声纹信息一致, 再将包含该声音的第一音频文件通过移动通信模块 150 或者无线通信模块 160 发送给音乐识别服务器。

在一些实施例中, 处理器 110 中包含用户画像模块, 该用户画像模块可以采集使用该智能终端的用户的用户信息, 该用户信息可以包括用户的属性(年龄、性别、职业等)生活习惯、用户行为等信息。在一种可能的实现方式中, 智能终端根据这些用户信息可抽象出来形成用户标签, 并发送给服务器进行存储。在又一种可能的实现方式中, 智能终端可以将这些用户信息发送给服务器, 服务器分析这些用户信息, 形成用户标签并存储。其中, 该用户标签与使用该智能终端的用户的用户账号(或称为用户 ID)具有对应关系。在本申请实施例中, 可以根据用户播放音频文件的习惯或偏好, 对用户抽象形成标签, 例如, 摇滚、民谣、流行等等, 还可以对喜欢的歌手进行记录形成标签, 例如, 李宗盛、梁静茹、陈奕迅等等。在一种可能的实现方式中, 音频文件识别出的第二音频文件的标签包含于第一用户的用户标签, 该第一用户可以为使用该智能终端的用户, 也可以为登录在该智能终端上的用户账号对应的用户。

耳机接口 170D 用于连接有线耳机。耳机接口 170D 可以是 USB 接口 130, 也可以是 3.5mm 的开放移动智能终端平台(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。

压力传感器 180A 用于感受压力信号, 可以将压力信号转换成电信号。在一些实施例中,

压力传感器 180A 可以设置于显示屏 194。压力传感器 180A 的种类很多，如电阻式压力传感器，电感式压力传感器，电容式压力传感器等。电容式压力传感器可以是包括至少两个具有导电材料的平行板。当有力作用于压力传感器 180A，电极之间的电容改变。智能终端 100 根据电容的变化确定压力的强度。当有触摸操作作用于显示屏 194，智能终端 100 根据压力传感器 180A 检测所述触摸操作强度。智能终端 100 也可以根据压力传感器 180A 的检测信号计算触摸的位置。在一些实施例中，作用于相同触摸位置，但不同触摸操作强度的触摸操作，可以对应不同的操作指令。例如：当有触摸操作强度小于第一压力阈值的触摸操作作用于短消息应用图标时，执行查看短消息的指令。当有触摸操作强度大于或等于第一压力阈值的触摸操作作用于短消息应用图标时，执行新建短消息的指令。

陀螺仪传感器 180B 可以用于确定智能终端 100 的运动姿态。在一些实施例中，可以通过陀螺仪传感器 180B 确定智能终端 100 围绕三个轴(即，x，y 和 z 轴)的角速度。陀螺仪传感器 180B 可以用于拍摄防抖。示例性的，当按下快门，陀螺仪传感器 180B 检测智能终端 100 抖动的角度，根据角度计算出镜头模组需要补偿的距离，让镜头通过反向运动抵消智能终端 100 的抖动，实现防抖。陀螺仪传感器 180B 还可以用于导航，体感游戏场景。在一些实施例中，智能终端 100 可以通过陀螺仪传感器 180B 确定自身的移动方向，以提升确定自身的位置的准确性。

加速度传感器 180C 可检测智能终端 100 在各个方向上(一般为三轴)加速度的大小。当智能终端 100 静止时可检测出重力的大小及方向。还可以用于识别智能终端 100 的姿态，应用于横竖屏切换，计步器等应用。在一些可能的实现方式中，以下实施例中所示例的用户界面可以随着智能终端的姿态的变换，进行横竖屏的切换。

距离传感器 180D，用于测量距离。智能终端 100 可以通过红外或激光测量距离。在一些实施例中，拍摄场景，智能终端 100 可以利用距离传感器 180D 测距以实现快速对焦，提升获取到的口型信息的准确性。

接近光传感器 180E 可以包括例如发光二极管(LED)和光检测器，例如光电二极管。发光二极管可以是红外发光二极管。智能终端 100 通过发光二极管向外发射红外光。智能终端 100 使用光电二极管检测来自附近物体的红外反射光。当检测到充分的反射光时，可以确定智能终端 100 附近有物体。当检测到不充分的反射光时，智能终端 100 可以确定智能终端 100 附近没有物体。智能终端 100 可以利用接近光传感器 180E 检测用户手持智能终端 100 贴近耳朵通话，以便自动熄灭屏幕达到省电的目的。接近光传感器 180G 也可用于皮套模式，口袋模式自动解锁与锁屏。

环境光传感器 180F 用于感知环境光亮度。智能终端 100 可以根据感知的环境光亮度自适应调节显示屏 194 亮度。环境光传感器 180F 也可用于拍照时自动调节白平衡。环境光传感器 180F 还可以与接近光传感器 180G 配合，检测智能终端 100 是否在口袋里，以防误触。在一种可能的实现方式中，当智能终端通过环境光传感器 180F 检测到，环境光亮度小于预设值的持续时间大于预设时间时，所述智能终端停止通过音频输入模块采集外部环境中的声音。指纹传感器 180G 用于采集指纹。智能终端 100 可以利用采集的指纹特性实现指纹解锁，以解除该智能终端 100 的锁定状态。

触摸传感器 180H，也可称触控面板或触敏表面。触摸传感器 180H 可以设置于显示屏 194，由触摸传感器 180H 与显示屏 194 组成触摸屏，也称“触控屏”。触摸传感器 180H 用于检测作用于其上或附近的触摸操作。触摸传感器可以将检测到的触摸操作传递给应用处理器，以确

定触摸事件类型。可以通过显示屏 194 提供与触摸操作相关的视觉输出。在另一些实施例中，触摸传感器 180H 也可以设置于智能终端 100 的表面，与显示屏 194 所处的位置不同。

按键 190 包括开机键，音量键等。按键 190 可以是机械按键。也可以是触摸式按键。智能终端 100 可以接收按键输入，产生与智能终端 100 的用户设置以及功能控制有关的键信号输入。

马达 191 可以产生振动提示。马达 191 可以用于来电振动提示，也可以用于触摸振动反馈。例如，作用于不同应用(例如拍照，音频播放等)的触摸操作，可以对应不同的振动反馈效果。作用于显示屏 194 不同区域的触摸操作，马达 191 也可对应不同的振动反馈效果。不同的应用场景(例如：时间提醒，接收信息，闹钟，游戏等)也可以对应不同的振动反馈效果。触摸振动反馈效果还可以支持自定义。

指示器 192 可以是指示灯，可以用于指示充电状态，电量变化，也可以用于指示消息，未接来电，通知等。

SIM 卡接口 195 用于连接 SIM 卡。SIM 卡可以通过插入 SIM 卡接口 195，或从 SIM 卡接口 195 拔出，实现和智能终端 100 的接触和分离。智能终端 100 可以支持 1 个或 N 个 SIM 卡接口，N 为大于 1 的正整数。SIM 卡接口 195 可以支持 Nano SIM 卡，Micro SIM 卡，SIM 卡等。同一个 SIM 卡接口 195 可以同时插入多张卡。所述多张卡的类型可以相同，也可以不同。SIM 卡接口 195 也可以兼容不同类型的 SIM 卡。SIM 卡接口 195 也可以兼容外部存储卡。智能终端 100 通过 SIM 卡和网络交互，实现通话以及数据通信等功能。在一些实施例中，智能终端 100 采用 eSIM，即：嵌入式 SIM 卡。eSIM 卡可以嵌在智能终端 100 中，不能和智能终端 100 分离。

图 1A 示例性所示的智能终端 100 可以通过显示屏 194 显示以下各个实施例中所描述的各个用户界面。智能终端 100 可以通过触摸传感器 180H 在各个用户界面中检测触控操作，例如在各个用户界面中的点击操作(如在图标上的触摸操作、双击操作)，又例如在各个用户界面中的向上或向下的滑动操作，或执行画圆圈手势的操作，等等。在一些实施例中，智能终端 100 可以通过陀螺仪传感器 180B、加速度传感器 180C 等检测用户手持智能终端 100 执行的运动手势，例如晃动智能终端。在一些实施例中，智能终端 100 可以通过摄像模组 193(如 3D 摄像头、深度摄像头)检测非触控的手势操作。

智能终端 100 的软件系统可以采用分层架构，事件驱动架构，微核架构，微服务架构，或云架构。本申请实施例以分层架构的 Android 系统为例，示例性说明智能终端 100 的软件结构。

图 1B 是本申请实施例提供的一种智能终端 100 的软件结构框图。

分层架构将软件分成若干个层，每一层都有清晰的角色和分工。层与层之间通过软件接口通信。在一些实施例中，将 Android 系统分为四层，从上至下分别为应用程序层，应用程序框架层，安卓运行时(Android runtime)和系统库，以及内核层。

应用程序层可以包括一系列应用程序包。

如图 1B 所示，应用程序包可以包括相机，图库，日历，通话，地图，导航，WLAN，蓝牙，音乐，视频，短信息等应用程序。

应用程序框架层为应用程序层的应用程序提供应用编程接口(application programming interface, API)和编程框架。应用程序框架层包括一些预先定义的函数。

如图 1B 所示, 应用程序框架层可以包括窗口管理器, 内容提供器, 视图系统, 电话管理器, 资源管理器, 通知管理等。

窗口管理器用于管理窗口程序。窗口管理器可以获取显示屏大小, 判断是否有状态栏, 锁定屏幕, 截取屏幕等。

内容提供器用来存放和获取数据, 并使这些数据可以被应用程序访问。所述数据可以包括视频, 图像, 音频, 拨打和接听的电话, 浏览历史和书签, 电话簿等。

视图系统包括可视控件, 例如显示文字的控件, 显示图片的控件等。视图系统可用于构建应用程序。显示界面可以由一个或多个视图组成的。例如, 包括短信通知图标显示界面, 可以包括显示文字的视图以及显示图片的视图。

电话管理器用于提供智能终端 100 的通信功能。例如通话状态的管理(包括接通, 挂断等)。

资源管理器为应用程序提供各种资源, 比如本地化字符串, 图标, 图片, 布局文件, 视频文件等等。

通知管理器使应用程序可以在状态栏中显示通知信息, 可以用于传达告知类型的消息, 可以短暂停留后自动消失, 无需用户交互。比如通知管理器被用于告知下载完成, 消息提醒等。通知管理器还可以是以图表或者滚动条文本形式出现在系统顶部状态栏的通知, 例如后台运行的应用程序的通知, 还可以是以对话窗口形式出现在屏幕上的通知。例如在状态栏提示文本信息, 发出提示音, 智能终端振动, 指示灯闪烁等。

Android Runtime 包括核心库和虚拟机。Android runtime 负责安卓系统的调度和管理。

核心库包含两部分: 一部分是 java 语言需要调用的功能函数, 另一部分是安卓的核心库。

应用程序层和应用程序框架层运行在虚拟机中。虚拟机将应用程序层和应用程序框架层的 java 文件执行为二进制文件。虚拟机用于执行对象生命周期的管理, 堆栈管理, 线程管理, 安全和异常的管理, 以及垃圾回收等功能。

系统库可以包括多个功能模块。例如: 表面管理器(surface manager), 媒体库(Media Libraries), 三维图形处理库(例如: OpenGL ES), 2D 图形引擎(例如: SGL)等。

表面管理器用于对显示子系统进行管理, 并且为多个应用程序提供了 2D 和 3D 图层的融合。

媒体库支持多种常用的音频, 视频格式回放和录制, 以及静态图像文件等。媒体库可以支持多种音视频编码格式, 例如: MPEG4, G.264, MP3, AAC, AMR, JPG, PNG 等。

三维图形处理库用于实现三维图形绘图, 图像渲染, 合成, 和图层处理等。

2D 图形引擎是 2D 绘图的绘图引擎。

内核层是硬件和软件之间的层。内核层至少包含显示驱动, 摄像头驱动, 音频驱动, 传感器驱动。

图 1B 所示的软件系统涉及到使用分享能力的应用呈现(如图库, 文件管理器), 提供分享能力的即时分享模块, 提供打印能力的打印服务(print service)和打印后台服务(print spooler), 以及应用框架层提供打印框架 WLAN 服务 蓝牙服务, 以及内核和底层提供 WLAN 蓝牙能力和基本通信协议。

下面结合一种设置哼唱识别权限的场景, 示例性说明智能终端 100 软件以及硬件的工作流程。

当触摸传感器 180K 接收到触摸操作, 相应的硬件中断被发给内核层。内核层将触摸操

作加工成原始输入事件(包括触摸坐标, 触摸操作的时间戳等信息)。原始输入事件被存储在内核层。应用程序框架层从内核层获取原始输入事件, 识别该输入事件所对应的控件。以该触摸操作是触摸触摸操作, 该触摸操作所对应的控件为哼唱识别功能的开关控件为例, 哼唱识别应用调用应用框架层的接口, 启动哼唱识别应用, 进而通过调用内核层启动麦克风驱动, 通过麦克风 170C 采集外部环境中的声音。

图 1C 示例性示出了本申请实施例提供的智能家居设备 110 的结构示意图。

举例而言, 该智能家居设备可以是智能音箱或者智能电视等设备。如图 1C 所示, 智能家居设备 110 可包括处理器 102、存储器 103、无线通信处理模块 104、电源开关 105、RJ45 通信处理模块 106、USB 接口模块 107、音频输入模块 108 和音频输出模块 109。这些部件可以通过总线连接。其中:

处理器 102 可用于读取和执行计算机可读指令。具体实现中, 处理器 102 可主要包括控制器、运算器和寄存器。其中, 控制器主要负责指令译码, 并为指令对应的操作发出控制信号。运算器主要负责执行定点或浮点算数运算操作、移位操作以及逻辑操作等, 也可以执行地址运算和转换。寄存器主要负责保存指令执行过程中临时存放的寄存器操作数和中间操作结果等。具体实现中, 处理器 102 的硬件架构可以是专用集成电路 (Application Specific Integrated Circuits, ASIC) 架构、MIPS 架构、ARM 架构或者 NP 架构等等。

在一些实施例中, 处理器 102 可以用于解析无线通信处理模块 104 接收到的信号, 例如, 智能终端 100 发送的修改设置信息的请求, 音乐识别服务器发送的识别出的音频文件以及用于指示起始播放位置的指示信息等等。处理器 102 可以用于根据解析结果进行相应的处理操作, 如根据请求修改智能家居设备 110 的设置信息, 又如从播放位置播放识别出的音频文件等等。

在一些实施例中, 处理器 102 还可以用于处理智能家居设备 110 采集到的外部环境中的声音。例如, 处理器 102 可以提取该声音中的声纹信息, 若所述处理器 102 判定所述声音的声纹信息与预存的声纹信息一致, 则通过无线通信模块 104 将该包含该声音的第一音频文件发送给音乐识别服务器。

在一些实施例中, 处理器 102 还可以用于生成无线通信处理模块 104 向外发送的信号, 如向智能终端 100 发送的用于反馈识别状态 (如识别成功、识别失败等) 的信号。

存储器 103 与处理器 102 耦合, 用于存储各种软件程序和/或多组指令。具体实现中, 存储器 103 可包括高速随机存取的存储器, 并且也可包括非易失性存储器, 例如一个或多个磁盘存储设备、闪存设备或其他非易失性固态存储设备。存储器 103 可以存储操作系统, 例如 DuerOS、AliGenie 等嵌入式操作系统。存储器 103 还可以存储通信程序, 该通信程序可用于与智能终端 100, 一个或多个服务器 (例如, 音乐识别服务器), 或附加设备进行通信。

无线通信处理模块 104 可以包括蓝牙 (BT) 通信处理模块 104A、WLAN 通信处理模块 104B 中的一项或多项。

在一些实施例中, 蓝牙 (BT) 通信处理模块、WLAN 通信处理模块中的一项或多项可以监听到其他设备 (智能终端 100) 发射的信号, 如播放请求、更改设置信息的请求等等, 并可以发送响应信号, 如请求响应等, 使得其他设备 (如智能终端 100) 可以发现智能家居设备 110, 并与其他设备建立无线通信连接, 通过蓝牙或 WLAN 中的一种或多种无线通信技术与其他设备进行通信。

在另一些实施例中，蓝牙（BT）通信处理模块、WLAN 通信处理模块中的一项或多项也可以发射信号，如广播蓝牙信号、信标信号，使得其他设备（如智能终端 100）可以发现智能家居设备 110，并与其他设备（如智能终端 100）建立无线通信连接，通过蓝牙或 WLAN 中的一种或多种无线通信技术与其他设备（如智能终端 100）进行通信。

无线通信处理模块 104 还可以包括蜂窝移动通信处理模块（未示出）。蜂窝移动通信处理模块可以通过蜂窝移动通信技术与其他设备（如服务器）进行通信。

电源开关 105 可用于控制电源向智能家居设备 110 的供电。

RJ45 通信处理模块 106 可以用于处理通过 RJ45 接口接收或发送的数据。RJ45 接口主要用来联接 modem 调制解调器。

USB 接口 107 可用于通过数据线与其他设备（例如，计算机、笔记本电脑等）进行通信。

音频输入模块 108 可用于采集外部环境中的声音，并将该声音转换为电信号。在一种可能的实现方式中，智能家居设备 110 可以通过音频输入模块 108 接收用户输入的语音指令，响应于该语音指令，智能家居设备执行该语音指令对应的操作。

音频输出模块 109 用于将音频电信号转换为声音信号，智能家居设备 100 可以通过音频输出模块 109 播放声音信号。

在一种可能的实现方式中，该智能家居设备 110 还可以包括显示屏 110（未示出），显示屏 110 可用于显示图像，视频等。显示屏 110 包括显示面板。显示面板可以采用液晶显示屏（liquid crystal display, LCD），有机发光二极管（organic light-emitting diode, OLED），有源矩阵有机发光二极体或主动矩阵有机发光二极体（active-matrix organic light emitting diode 的，AMOLED），柔性发光二极管（flex light-emitting diode, FLED），Miniled, MicroLed, Micro-oLed, 量子点发光二极管（quantum dot light emitting diodes, QLED）等。在一些实施例中，智能家居设备 110 可以包括 1 个或 N 个显示屏 110，N 为大于 1 的正整数。

可以理解的是，图 1C 示意的结构并不构成对智能家居设备 110 的具体限定。在本申请另一些实施例中，智能家居设备 110 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。

图 1D 示例性示出了本申请提供的车载设备 120 的结构示意图。

举例而言，该车载设备可以是车载音箱或者车载电脑等设备。如图 1C 所示，车载设备 120 可包括处理器 102、存储器 103、无线通信处理模块 104、电源开关 105、显示屏 106、USB 接口模块 107、音频输入模块 108 和音频输出模块 109。这些部件可以通过总线连接。其中：

处理器 102 可用于读取和执行计算机可读指令。具体实现中，处理器 102 可主要包括控制器、运算器和寄存器。其中，控制器主要负责指令译码，并为指令对应的操作发出控制信号。运算器主要负责执行定点或浮点算数运算操作、移位操作以及逻辑操作等，也可以执行地址运算和转换。寄存器主要负责保存指令执行过程中临时存放的寄存器操作数和中间操作结果等。具体实现中，处理器 102 的硬件架构可以是专用集成电路（Application Specific Integrated Circuits, ASIC）架构、MIPS 架构、ARM 架构或者 NP 架构等等。

在一些实施例中，处理器 102 可以用于解析无线通信处理模块 104 接收到的信号，例如，智能终端 100 发送的修改设置信息的请求，音乐识别服务器发送的识别出的音频文件以及用

于指示起始播放位置的指示信息等等。处理器 102 可以用于根据解析结果进行相应的处理操作，如根据请求修改智能家居设备 110 的设置信息，又如从播放位置播放识别出的音频文件等等。

在一些实施例中，处理器 102 还可以用于处理车载设备 120 采集到的外部环境中的声音。例如，处理器 102 可以提取该声音中的声纹信息，若所述处理器 102 判定所述声音的声纹信息与预存的声纹信息一致，则通过无线通信模块 104 将该包含该声音的第一音频文件发送给音乐识别服务器。

在一些实施例中，处理器 102 还可以用于生成无线通信处理模块 104 向外发送的信号，如向智能终端 100 发送的用于反馈识别状态（如识别成功、识别失败等）的信号。

存储器 103 与处理器 102 耦合，用于存储各种软件程序和/或多组指令。具体实现中，存储器 103 可包括高速随机存取的存储器，并且也可包括非易失性存储器，例如一个或多个磁盘存储设备、闪存设备或其他非易失性固态存储设备。存储器 103 可以存储操作系统，例如 uCLinux、GENIVI、ecos 等嵌入式操作系统。存储器 103 还可以存储通信程序，该通信程序可用于与智能终端 100，一个或多个服务器（例如，音乐识别服务器），或附加设备进行通信。

无线通信处理模块 104 可以包括蓝牙（BT）通信处理模块 104A、WLAN 通信处理模块 104B 中的一项或多项。

在一些实施例中，蓝牙（BT）通信处理模块、WLAN 通信处理模块中的一项或多项可以监听到其他设备（智能终端 100）发射的信号，如播放请求、更改设置信息的请求等等，并可以发送响应信号，如请求响应等，使得其他设备（如智能终端 100）可以发现车载设备 120，并与其他设备建立无线通信连接，通过蓝牙或 WLAN 中的一种或多种无线通信技术与其他设备进行通信。

在另一些实施例中，蓝牙（BT）通信处理模块、WLAN 通信处理模块中的一项或多项也可以发射信号，如广播蓝牙信号、信标信号，使得其他设备（如智能终端 100）可以发现车载设备 120，并与其他设备（如智能终端 100）建立无线通信连接，通过蓝牙或 WLAN 中的一种或多种无线通信技术与其他设备（如智能终端 100）进行通信。

无线通信处理模块 104 还可以包括蜂窝移动通信处理模块（未示出）。蜂窝移动通信处理模块可以通过蜂窝移动通信技术与其他设备（如服务器）进行通信。

电源开关 105 可用于控制电源向车载设备 120 的供电。

显示屏 110 可用于显示图像，视频等。显示屏 110 包括显示面板。显示面板可以采用液晶显示屏(liquid crystal display, LCD)，有机发光二极管(organic light-emitting diode, OLED)，有源矩阵有机发光二极体或主动矩阵有机发光二极体(active-matrix organic light emitting diode 的, AMOLED)，柔性发光二极管(flex light-emitting diode, FLED)，Miniled, MicroLed, Micro-oLed, 量子点发光二极管(quantum dot light emitting diodes, QLED)等。在一些实施例中，智能家居设备 110 可以包括 1 个或 N 个显示屏 110，N 为大于 1 的正整数。

USB 接口 107 可用于通过数据线与其他设备例如显示器、智能终端 100 或者音频外放设备进行通信。

音频输入模块 108 可用于采集外部环境中的声音，并将该声音转换为电信号。在一种可能的实现方式中，车载设备 120 可以通过音频输入模块 108 接收用户输入的语音指令，响应于该语音指令，车载设备执行该语音指令对应的操作。

音频输出模块 109 用于将音频电信号转换为声音信号，车载设备 120 可以通过音频输出

模块 109 播放声音信号。

在一些实施例中，车载设备 120 还可以包括 RS-232 接口等串行接口。该串行接口可连接至其他设备，如音箱等音频外放设备，使得音频外放设备协作播放识别出的音频文件。

可以理解的是，图 1C 示意的结构并不构成对车载设备 120 的具体限定。在本申请另一些实施例中，车载设备 120 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。

下面介绍智能终端 100 上的用于显示应用程序菜单的示例性用户界面。

图 2 示例性示出了智能终端 100 用于显示应用程序菜单的用户界面 21。如图 2 所示，用户界面 21 可包括：状态栏 201，具有常用应用程序图标托盘 217，日历小工具 213，天气小工具 215，以及其他应用程序图标。其中：

状态栏 201 可包括：移动通信信号（又可称为蜂窝信号）的一个或多个信号强度指示符 203、无线高保真（wireless fidelity, Wi-Fi）信号的一个或多个信号强度指示符 205，电池状态指示符 209、时间指示符 211。

日历小工具 213 可用于指示当前时间，例如日期、星期几、时分信息等。

天气小工具 215 可用于指示天气类型，例如多云转晴、小雨等，还可以用于指示气温等信息。

具有常用应用程序图标托盘 217 可展示：电话图标 219、联系人图标 221、短信图标 223、相机图标 225。

其他应用程序图标可例如：微信（Wechat）的图标 227、QQ 的图标 229、推特（Twitter）的图标 231、脸书（Facebook）的图标 233、邮箱的图标 235、云共享的图标 237、备忘录的图标 239、支付宝的图标 221、图库的图标 225、设置的图标 227。用户界面 21 还可包括页面指示符 229。其他应用程序图标可分布在其他页面，页面指示符 229 可用于指示页面数量，以及用户当前浏览的是哪一个页面，比如页面指示符 229 显示 3 个小圆点，并且第 2 个小圆点是黑色的，另外两个小圆点是白色的，表明当前手机包括 3 个页面，且用户正在浏览第 2 个页面。此外，用户可以在当前页面左右滑动，来浏览其他页面中的应用程序图标。在一些实施例中，图 2 示例性所示的用户界面 21 可以为主界面（Home screen）中的一个用户界面。

在其他一些实施例中，智能终端 100 还可以包括主屏幕键。该主屏幕键可以是实体按键，也可以是虚拟按键。该主屏幕键可用于接收用户的指令，响应于该用户的指令，将当前显示的 UI 返回到主界面，这样可以方便用户随时查看主屏幕。上述指令具体可以是用户单次按下主屏幕键的操作指令，也可以是用户在短时间内连续两次按下主屏幕键的操作指令，还可以是用户在预定时间内长按主屏幕键的操作指令。在本申请其他一些实施例中，主屏幕键还可以集成指纹识别器，以便用于在按下主屏幕键的时候，随之进行指纹采集和识别。

可以理解的是，图 2 仅仅示例性示出了智能终端 100 上的用户界面，不应构成对本申请实施例的限定。

接下来，将介绍本申请实施例提供的一些哼唱识别的实施例。

在本申请实施例中，智能终端 100 可以跟随用户的哼唱进度播放该识别出的音频文件，另外，智能终端 100 可以通过显示屏 194 显示识别结果。该识别结果可能在智能终端 100 处于使用状态下显示，还可能在智能终端 100 处于锁定状态下显示，以下将这种两种应用场景

下的实施例作进一步的介绍。需要说明的是，在本申请实施例中，智能终端 100 处于使用状态表示智能终端 100 正在被用户使用，智能终端 100 的显示屏 194 处于常亮状态，该显示屏 194 可以显示桌面，应用程序界面，下拉通知栏，负一屏等用户界面。智能终端 100 处于锁定状态表示智能终端的屏幕被锁定，大多数情况下，智能终端 100 锁定后需要接收用户输入的密码或验证其他解锁方式（例如，指纹解锁，面部解锁等等）才能解除锁定。通常情况下，用户可以通过单击智能终端 100 的电源键，或者点击“锁定屏幕”的虚拟控件，使智能终端 100 熄屏，并且进入锁定状态。另外，锁屏界面是指智能终端 100 进入锁定状态后，未解除锁定之前，智能终端 100 显示的用户界面。智能终端 100 处于锁定状态的过程中，智能终端可以显示锁屏界面，还可以为熄屏（或称为黑屏）状态。

首先对智能终端 100 处于使用状态下显示识别结果的实施例进行介绍。

图 3A 示例性示出了在智能终端 100 的应用程序的使用界面中，显示识别结果的用户界面 31。需要说明的是，执行哼唱识别操作的应用程序与该处于使用过程中应用程序可以是同一个应用程序，也可以是不同的应用程序，本申请实施例不作限制。另外，本申请对该处于使用过程的应用程序不作限制，可以为微信、QQ、微博、邮箱等应用程序，示例性的，图 3A 中以微信使用过程中的聊天界面为例。如图 3A 所示，用户界面 31 可包括：显示区域 318，输入区域 319，通知窗口 315。其中：

显示区域 318，可以用于显示聊天内容，该聊天内容可包括使用本智能终端 100 的用户与另一社交账号的用户的文字/语音的交流内容。

输入区域 319，可以用于输入聊天内容，输入区域 319 可以包括第一控件 319A，第二控件 319B，第三控件 319C，第四控件 319D。其中，该第一控件 319A，用于接收用户的操作，响应于该用户的操作，智能终端 100 显示语音输入按钮，通常，用户可以通过长按该语音输入按钮以录入语音信息。需要说明的是，当该语音输入按钮接收到用户的操作时，智能终端 100 需采集用户输入的语音信息，音频输入模块将被该社交应用程序的语音输入服务占用，智能终端 100 不执行本申请实施例提供的哼唱识别操作。第二控件 319B，用于接收用户的操作，响应于该用户的操作，智能终端 100 显示键盘/手写板，通常，智能终端 100 可以通过键盘/手写板接收用户输入的文字信息。第三控件 319C，用于接收用户的操作，响应于该用户的操作，智能终端 100 显示多个表情图案/动图供用户进行选择。第四控件 319D，用于接收用户的操作，响应于该用户的操作，智能终端 100 显示多个输入类型的选择框，例如，图片，拍摄，文档，红包，视频通话等等，供用户进行选择。相似的，当“拍摄”或者“视频通话”选择框接收到用户的操作时，智能终端 100 需采集用户输入的音视频信息，音频输入模块和/或音频输出模块将被该社交应用程序的音视频输入服务占用，智能终端 100 不执行本申请实施例提供的哼唱识别操作。

通知窗口 315，用于显示对用户哼唱的音乐片段的识别结果，通知窗口 315 可包括：哼唱识别图标 316，第一显示区域 314，播放控件 310，控件 312。

其中，哼唱识别图标 316 用于指示通知窗口 315 的来源，为了方便用户快捷的了解到通知窗口 315 是哼唱识别服务（或者称为功能、应用程序）输出的识别结果。需要说明的是，哼唱识别图标 316 仅为示例图标，在具体的实施过程中，哼唱识别图标还可以为其他的图案，例如，音符或者其他样式的图标，本申请实施例不作限制。

第一显示区域 314，可以用于显示识别出的音频文件的标识信息，能够为用户提供更多的有关于该识别出的音频文件信息。其中，该音频文件的标识信息可以为音频文件的歌名，

歌词,演唱者名字,专辑名称,专辑封面图片,演唱者海报,等等。如图所示,第一显示区域 314 中包含了歌曲的名称《漂洋过海来看你》。第一显示区域 314 还可以包括操作指示信息(例如,第一显示区域 314 中包含的“可单击停止播放”),可为用户提供操作的提醒,提升用户操作的便利性。可选的,第一显示区域 314 还可以包含演唱者信息或者当前播放的音频文件的歌词信息。在又一种可能的情况下,该智能终端 100 还可以以悬浮窗的形式显示当前播放的音频文件的歌词信息。该悬浮窗是在智能终端 100 的显示界面中悬浮显示的一个可以移动的窗口。

播放控件 310,可以用于接收用户的操作,响应于该用户的操作,该智能终端 100 暂停播放或者继续播放音频文件。具体的,该智能终端 100 识别出用户哼唱的音乐片段对应的音频文件后,将跟随用户的演唱进度播放该音频文件,此时的播放控件 310 显示第一状态,该第一状态表示该音频文件正在被播放。可选的,在识别出的音频文件的播放过程中,智能终端 100 不再执行本申请实施例提供的哼唱识别的操作。在播放控件 310 显示为第一状态的情况下,若播放控件 310 接收到用户的操作,则该智能终端 100 暂停播放该音频文件,并将播放控件 310 显示为第二状态,该第二状态表示该音频文件暂停播放。可以理解的是,在播放 310 显示为第二状态的情况下,若播放控件 310 接收到用户的操作,则该智能终端 100 继续播放该音频文件,并将播放控件 310 显示为第一状态。

控件 312,可以用于接收用户的操作,响应于该用户的操作,智能终端 100 暂停播放音频文件重新获取用户的语音信号,并对重新获取的语音信号进行哼唱识别。在一种可能的情况下,在重新获取语音信号的过程中,智能终端 100 可以显示提示信息(例如,“正在识别中...”),该提示信息用于指示该智能终端 100 正在重新获取语音信号以进行哼唱识别。在又一种可能的情况下,响应于该用户针对控件 312 的操作,智能终端 100 暂停播放音频文件,并跳转到用于显示哼唱识别的用户界面 35。该用户界面 35 在后续会作具体的介绍,此处不展开说明。

在一种可能的实现方式中,通知窗口 315 在显示预设时间后消失,该预设时间可以为 4 秒,5 秒等时间值。或者,当通知窗口 315 接收到用户的向上滑动的操作时,响应于该操作,智能终端 100 不再在用户界面 31 中显示通知窗口 315。

在一种可能的实现方式中,通知窗口 315 还可以在下拉通知栏中显示,这种显示方式可参见图 3B。如图 3B 所示,当检测到在状态栏 201 上的向下滑动手势时,响应于该手势,智能终端 100 可以在用户界面 21 上显示下拉通知栏 318,下拉通知栏 318 中包括通知窗口 315 和控制窗口 313,其中:

通知窗口 315 可参照图 3A 中的描述,此处不再赘述。

控制窗口 313,可以显示多个开关控件,例如显示“哼唱识别”的开关控件 317,还可以显示有其他功能(如 Wi-Fi、蓝牙、手电筒等等)的开关控件。控制窗口 313 将在后续介绍哼唱识别的设置界面中作详细的介绍,此处不具体展开。

在一种可能的情况下,在哼唱识别功能开启的过程中,状态栏 201 中显示哼唱识别的图标 311。可以理解的是,在智能终端 100 的多个显示界面中均可包括状态栏 201。通过这种方式,可以方便用户通过智能终端 100 的多个显示界面知晓哼唱识别功能的开启状态。

接下来对智能终端 100 处于锁定状态下显示识别结果的实施例进行介绍。

图 3C 示例性示出了智能终端 100 处于锁定状态下显示的用户界面 32,用户界面 32 也可

以称为锁屏界面。如图 3C 所示, 用户界面 32 包括状态栏 201, 日历小工具 213 以及锁屏壁纸 523。其中:

状态栏 201 可参照图 2 中的描述, 此处不再赘述。特别的, 此处的状态栏 201 包括了哼唱识别图标 311 以及锁定图标 323, 哼唱识别图标 311 用于指示哼唱识别功能处于开启状态, 锁定图标 323 用于指示智能终端 100 处于锁定状态。

日历小工具 213 可参照图 2 中的描述, 此处不再赘述。可选的, 用户界面 32 还可以包括天气小工具 215。

锁屏壁纸 523 可以为用户设定的图片, 也可以为智能终端 100 预设的图片, 或者为智能终端 100 从网络中下载的图片。

图 3D 示例性示出了又一种显示识别结果的用户界面 32。

如图 3D 所示, 在智能终端处于锁定状态的情况下, 当智能终端 100 识别出用户哼唱的音乐片段的音频文件时, 智能终端 100 在用户界面 32 上方显示通知窗口 324, 通知窗口 324 可包括: 哼唱识别图标 316, 第二显示区域 322, 播放控件 310, 控件 312, 音量控件 328。

其中, 哼唱识别图标 316, 播放控件 310 以及控件 312 可参照图 3A 中的相关描述, 此处不再赘述。

第二显示区域 322 与图 3A 中的第一显示区域 314 的作用相同, 均可以显示识别出的音频文件的标识信息。不同的是, 此处的第二显示区域 322 中不仅包括了音频文件的名称《漂洋过海来看你》, 还包括了音频文件的演唱者“李宗盛”, 以及当前播放的音频文件的歌词信息“陌生的城市啊, 熟悉的角落里……”, 其中歌词信息中加粗的部分“不管将会面对”为用户当前演唱的歌词部分。可以理解的是, 该歌词信息会随着音频文件的播放进度而改变, 以使得歌词信息与音频文件的播放保持同步。

在一种可能的实现方式中, 第二音频文件的标签包含于第一用户的用户标签, 该第一用户的含义可参照上述的介绍。不同的智能终端对同一首歌曲的哼唱片段的识别结果可能不同, 例如, 用户 1 在演唱《漂洋过海来看你》的过程中, 用户 1 的智能终端识别出的音频文件可能是李宗盛演唱的版本; 用户 2 在演唱《漂洋过海来看你》的过程中, 用户 2 的智能终端识别出的音频文件可能是梁静茹演唱的版本。可以理解的是, 识别结果不同是由于用户 1 和用户 2 的用户标签不同。

音量控件 328, 可以用于调整播放音频文件的音量。音量控件 328 可以用于接收用户的操作, 响应于该操作, 智能终端 100 调整播放音频文件的音量。可选的, 当接收到的用户操作为向左的滑动时, 智能终端 100 降低播放音频文件的音量; 当接收到的用户操作为向右的滑动时, 智能终端 100 提高播放音频文件的音量。在一种可能的实现方式中, 音频文件控件 328 到所在线段的左端点的距离与所在线段的线段长的比值, 和当前音量与系统播放的最大音量的比值具有对应关系。

在一种可能的实现方式中, 智能终端 100 从开始播放音频文件的时刻到预设时刻(例如, 第 5 秒, 第 6 秒等时间值)的时间段内, 将使播放音频文件的音量由低到高逐渐增大。例如, 从音量的最小值逐渐增大到用户设定的音量值, 或者, 从用户设定的音量值的 30% 逐渐增大到用户设定的音量值的 100%, 还可以存在其他的音量增大方式, 本申请实施例不作限制。需要说明的是, 用户设定的音量值为音频文件控件 328 指示的音量值, 可选的, 离当前最近一次的, 用户调整音量的结果为用户设定的音量值。

在一种可能的实现方式中, 通知窗口 324 在识别出的音频文件播放完毕后消失。

在一种可能的实现方式中，智能终端 100 还可以把通知窗口 324 的内容以一个用户界面的方式显示在用户界面 32 上方，该用户界面可如图 3E 的用户界面 33 所示。在图 3E 中，用户界面 33 中包含通知窗口 324 中包含的内容，例如，哼唱识别图标 316，第二显示区域 322，播放控件 310，控件 312，音量控件 328。可选的，用户界面 32 还可以显示背景图片，例如，该背景图片可以是歌曲演唱者的海报，识别出的音频文件所收录的专辑的图片，等等。可选的，当用户界面 33 接收到用户的操作（例如向右的滑动操作），智能终端 100 显示用户界面 33 下方的用户界面 32（也即锁屏界面）。

图 3F 示例性示出了又一种显示识别结果的用户界面 34。

在一个实施例中，在用户界面 31 或者用户界面 21 中，若检测到用户输入的针对通知窗口 315 的操作（例如，点击操作，长按操作，按压操作，等等），或者，在用户界面 32 中，若检测到用户输入的针对通知窗口 324 的操作（例如，点击操作，长按操作，按压操作，等等），则智能终端 100 显示用户界面 34。可选的，在用户界面 32 跳转到用户界面 34 之前，智能终端 100 接收用户输入的解锁操作（例如，指纹解锁，密码解锁，人脸解锁，等等），在解锁成功的情况下，智能终端 100 执行从用户界面 32 跳转到用户界面 34 的操作。

用户界面 34 中包括：哼唱识别图标 316，第二显示区域 322，播放控件 310，控件 312，音量控件 328，控件 330，控件 332，控件 334，其中：

其中，哼唱识别图标 316，第二显示区域 322，播放控件 310，控件 312，音量控件 328 可参照上述描述，此处不再赘述。

控件 330，可以用于收藏识别出的音频文件。控件 330 可以接收用户的操作，响应于该用户的操作，智能终端 100 将识别出的音频文件的标识，添加到预设的收藏夹（或者称为“喜欢的音乐”的文件夹，本申请不作限制）中，便于用户下一次查找或播放该识别出的音频文件。

控件 332，可以用于下载识别出的音频文件。控件 332 可以接收用户的操作，响应于该用户的操作，智能终端 100 从网络中下载该识别出的音频文件的音频资源。可选的，响应于该用户的操作，智能终端 100 显示选择框，该选择框中包含“标准品质”“高品质”“无损品质”等音质选项。该选择框用于接收用户对一个选项的选择操作，响应于该用户对一个选项的选择操作，智能终端 100 下载该选项对应的音质的音频资源。

控件 334，可以用于分享识别出的音频文件。控件 334 可以接收用户的操作，响应于该用户的操作，智能终端 100 显示分享框，该分享框中包含多个分享对象，例如，QQ，微信，微博，推特，等等。该分享框用于接收用户对一个分享对象的选择操作，响应于该用户对一个分享对象的选择操作，智能终端 100 将该音频文件的标识或者音频资源发送给该选择操作对应的分享对象。

图 3G 示例性示出了一种用于哼唱识别的用户界面 35。

在一个实施例中，在用户界面 21，用户界面 31，用户界面 32，用户界面 33 以及用户界面 34 中，若检测到用户针对控件 312 的操作（例如，点击操作），则智能终端 100 显示用于哼唱识别的用户界面 35。

用户界面 35 中包括哼唱识别图标 316，指示符 350，控件 352 以及控件 354。其中：

哼唱识别图标 316 可参照上述描述，此处不再赘述。

指示符 350，可以指示用户已哼唱的音乐片段的时间信息，该时间信息随着用户哼唱音频文件的时间的增长而改变，与用户哼唱的时间长度保持同步。指示符 350 还可以指示用户

录入声音信号的操作提示信息（例如，指示符 350 中包含的“多哼唱几句识别更准确”），可以为用户提供操作的提醒，以便于提升哼唱识别的准确性。该操作提示信息还可以是其他的内容，举例而言，当检测到用户的声音的音量较小时，可显示例如“增大音量（或者靠近设备发声）识别更准确”的操作提示信息。

控件 352，可以用于接收用户的操作（例如，长按操作），响应于该用户的操作，智能终端 100 通过麦克风 170C 采集用户的输入的声音信号。当检测到用户的手指离开显示屏 194 时，智能终端 100 根据采集到的声音信号进行哼唱识别。可选的，当智能终端 100 从音乐识别服务器中接收到识别出音频文件时，智能终端 100 可显示用于显示识别结果的用户界面 34。

以上介绍了一些在智能终端 100 中用于显示识别结果以及用于进行哼唱识别的用户界面，在本申请实施例中，在智能终端 100 可以实施哼唱识别功能之前，用户可以通过智能终端 100 的设置界面对哼唱识别的功能进行开启或关闭。以下将对一些哼唱识别的设置界面进行介绍。

图 4A 示例性示出了一种用于设置哼唱识别功能的用户界面 41。

与上述介绍的显示下拉通知栏 318 的方式相似，当检测到在状态栏 201 上的向下滑动手势时，响应于该手势，智能终端 100 可以在用户界面 41 上显示下拉通知栏 401，下拉通知栏 401 中包括控制窗口 313，其中：

控制窗口 313，可以显示多个开关控件，例如显示“哼唱识别”的开关控件 317，还可以显示有其他功能（如 Wi-Fi、蓝牙、手电筒等等）的开关控件。开关控件 317 存在两种显示状态，第一显示状态（也可以称为“开（ON）”状态）表明哼唱识别功能开启，第二显示状态（也可称为“关（OFF）”状态）表明哼唱识别功能关闭。在开关控件 317 的显示状态为第二显示状态的情况下，当智能终端 100 检测到在控制窗口 318 中的开关控件 317 上的操作（如在开关控件 317 上的触摸操作）时，响应于该操作，智能终端 100 可以开启“哼唱识别”，并将开关控件 317 的显示状态调整为第一显示状态。在开关控件 317 的显示状态为第一显示状态的情况下，当智能终端 100 检测到在控制窗口 318 中的开关控件 317 上的操作时，响应于该操作，智能终端 100 可以关闭“哼唱识别”，并将开关控件 317 的显示状态调整为第二显示状态。通过这种方式，可以方便用户开启/关闭哼唱识别的功能。

在一种可能的情况下，在哼唱识别功能开启的过程中，状态栏 201 中显示哼唱识别的图标 311。可以理解的是，在智能终端 100 的多个显示界面中均可包括状态栏 201。通过这种方式，可以方便用户通过智能终端 100 的多个显示界面知晓哼唱识别功能的开启状态。

图 4B 示例性示出了又一种用于设置哼唱识别功能的用户界面 42。

如图 4B 所示，用户界面 42 包括显示区域 410，显示区域 410 用于显示多个可设置的选项，例如“飞行模式”“Wi-Fi”“蓝牙”等等。显示区域 410 还包括多个开关控件以及多个跳转控件，以开关控件 412 和跳转控件 416 介绍两种控件的作用，其中：

开关控件 412，可用于接收用户的操作（例如，点击操作，滑动操作，等等），响应于该用户的操作，智能终端 100 改变开关控件 412 对应的功能/服务/应用程序（即，哼唱识别功能）的开启状态。例如，在接收用户的操作之前，开关控件 412 的显示状态为“ON”，则表明此时哼唱识别功能处于开启状态。若开关控件 412 接收到用户的操作，响应于该用户的操作，智能终端 100 将开关控件 412 的显示状态调整为“OFF”，并关闭哼唱识别功能。

跳转控件 416，可用于接收用户的操作，响应于该用户的操作，智能终端 100 跳转到跳转控件 416 对应的功能/服务/应用程序（即，勿扰模式）的设置界面，需要说明的是，该设备

界面可以包括“勿扰模式”这个功能的多个设置选项，例如，勿扰模式的开启状态的调整，勿扰模式开启时间的设定，勿扰模式下自动回复的设置，等等。

图 5A-图 5C 示例性示出了又一些用于设置哼唱识别功能的用户界面。

如图 5A 所示，用户界面 51 包括显示区域 522，显示区域与用户界面 41 中包含的显示区域 410 相似，显示区域 522 用于显示多个可设置的选项，例如“飞行模式”“Wi-Fi”“蓝牙”等等。

与显示区域 410 不同的是，该哼唱识别功能对应的控件为跳转控件 520，跳转控件 520 可用于跳转用户界面至“哼唱识别”设置界面。如图 5A-图 5B 示例性所示，跳转控件 520 接收用户的操作（例如，点击操作），响应于该用户的操作，智能终端 100 从用户界面 51 跳转到“哼唱识别”的设置界面（即，用户界面 52）。

如图 5B 所示，用户界面 52 包括返回键 530，开关控件 532，文本信息 534，开关控件 536，控件 538，控件 540，控件 552，多个跳转控件（例如，跳转控件 554），开关控件 556。其中：

返回键 530，可用于接收用户的操作，响应于该用户的操作，智能终端 100 返回当前页面上一个界面，即图 5A 所示的用户界面 41。本领域技术人员应知，一个界面上一个界面在应用程序设定时便已确定。

开关控件 532 的功能可参照图 4B 中的开关控件 412 的功能，此处不再赘述。

文本信息 534，可用于对哼唱识别功能开启之后，智能终端 100 所获得的权限进行说明，方便用户根据该说明判定是否赋予智能终端 100 哼唱识别的权限。文本信息 534 的表述可按照需求进行更改，此处不做限定。

开关控件 536，可用于接收用户的操作，响应于该用户的操作，智能终端 100 执行设定哼唱识别的启用时间段的操作。举例而言，若在接收用户的操作之前，开关控件 536 的显示状态为“OFF”，表明该哼唱识别操作没有设置启用时间，哼唱识别操作可以一直在运行过程中。可选的，在这种情况下，智能终端不显示控件 538 和控件 540。在接收用户对开关控件 536 的操作之后，响应于该用户的操作，智能终端将开关控件 536 的显示状态转变为“ON”，并显示显示控件 538 和控件 540。其中，控件 538 用于接收用户输入的哼唱识别操作的启用时间，响应于该用户的操作，智能终端 100 在该启用时间之后，执行本申请实施例提供的哼唱识别的操作；控件 540 用于接收用户输入的哼唱识别操作的结束时间，响应于该用户的操作，智能终端 100 在该结束时间之后，不再执行本申请实施例提供的哼唱识别的操作。需要说明的是，在智能终端 100 不再执行本申请实施例提供的哼唱识别的操作的过程中，用户依然可以按照现有技术中的方式主动触发哼唱识别。

控件 522，可以用于添加，可用于启用哼唱识别的声纹信息。

跳转 554，可用于接收用户的操作，响应于该用户的操作，智能终端 100 从用户界面 52 跳转至声纹 1 的设置界面。该声纹 1 的设备界面可以包括命名和删除功能，等等。

开关控件 556，可用于接收用户的操作，响应于该用户的操作，智能终端 100 设置哼唱识别功能的可使用状态。举例而言，若在接收用户的操作之前，开关控件 556 的显示状态为“OFF”，表明该哼唱识别操作在智能终端 100 锁定时不可使用，即智能终端 100 在锁定时不执行哼唱识别操作。在接收用户对开关控件 536 的操作之后，响应于该用户的操作，智能终端 100 将开关控件 556 的显示状态转变为“ON”，并调整哼唱识别功能的可使用状态，即在智能终端 100 锁定时也运行哼唱识别操作。

开关控件 557, 可用于接收用户的操作, 响应于该用户的操作, 智能终端 100 设置哼唱识别功能的可使用状态。在该开关控件 557 处于开启状态时, 在哼唱识别功能运行的过程中, 智能终端可以获取自身所处的位置。智能终端 100 通过判定自身所处的位置是否为预设地点, 来确定是否停止通过音频输入模块采集外部环境中的声音, 或者确定是否从起始播放位置播放识别出的音频文件。这种判定方式将在后续内容中进行介绍, 此处不具体展开。

如图 5B-图 5C 示例性所示, 显示屏 194 接收用户的操作 (例如, 上滑操作), 响应于该用户的操作, 智能终端 100 显示开关控件 556 下方的“哼唱识别”的设置内容。如图 5C 所示, 用户界面 52 中还包括设置哼唱识别功能的访问权限的内容, 具体的, 用户界面 52 中还包括跳转控件 558, 以及多个开关控件 (例如开关控件 560)。其中:

跳转控件 558, 可用于设置哼唱识别功能允许访问的无线数据的类型, 例如, 关闭、WLAN、WLAN 与蜂窝移动数据。

开关控件 560, 可用于设置哼唱识别功能允许访问的系统功能 (即, 定位服务)。举例而言, 若在接收用户的操作之前, 开关控件 560 的显示状态为“OFF”, 表明哼唱识别功能运行时, 不可以获取智能终端 100 的位置信息。在接收用户对开关控件 560 的操作之后, 响应于该用户的操作, 智能终端 100 将开关控件 560 的显示状态转变为“ON”, 并允许哼唱识别功能获取智能终端 100 的位置信息。相似的, 其他哼唱识别系统访问的系统功能也可以参照上述方式设置。

图 5D-图 5F 示例性示出了一些设置哼唱识别功能的访问权限的用户界面。

如图 5D-图 5E 示例性所示, 响应于用户对跳转控件 524 的操作, 智能终端 100 从用户界面 51 跳转至用户界面 53, 用户界面 53 用于显示多个系统功能, 例如, 蓝牙、定位服务、麦克风、图库, 等等。其中, 一个系统服务对应一个跳转控件 (例如, “麦克风”这个系统服务对应于跳转控件 562)。

如图 5E-图 5F 示例性所示, 响应于用户对跳转控件 562 的操作, 智能终端 100 从用户界面 53 跳转至用户界面 54, 用户界面 54 用于显示要求访问麦克风的多个应用程序。用户可以通过应用程序对应的开关控件, 控制应用程序访问麦克风的权限。举例而言, 若在接收用户的操作之前, 开关控件 572 的显示状态为“OFF”, 表明哼唱识别功能不可以访问麦克风。在接收用户对开关控件 572 的操作之后, 响应于该用户的操作, 智能终端 100 将开关控件 572 的显示状态转变为“ON”, 并允许哼唱识别功能访问麦克风。相似的, 其他应用程序访问系统功能的方式也可以参照上述方式。

图 5G 示例性示出了一种用于录入声纹信息的用户界面 55。

在一个实施例中, 响应于用户针对用户界面 52 中的控件 552 的操作, 智能终端 100 从用户界面 52 跳转至用户界面 55, 以录入用户想要添加的声纹信息。在用户界面 55 中, 包括指示符 570, 文本信息 572, 控件 574。其中:

指示符 570, 可以用于为用户提供提示信息, 以指示用户进行声纹信息的录入。

文本信息 572, 为用户需要朗读的文字内容。可选的, 智能终端可以多次显示不同的文本信息供用户朗读。这样, 可以录入更多用户的声音信号, 以提升声纹信息的准确性。

在一种可能的情况下, 智能终端也可以指示用户演唱几段音乐片段, 以进行声纹信息的录入。这种情况下, 指示符 570 的内容可以是“请长按按钮, 并演唱以下歌曲片段以录入声纹信息”, 对应的, 文本信息 572 为一段歌词。

控件 574, 可以用于接收用户的操作 (例如, 长按操作), 响应于该用户的操作, 智能终

端 100 通过麦克风 170C 采集用户的输入的声音信号。当检测到用户的手指离开显示屏 194 时, 智能终端 100 存储这段过程中采集到的声音信号, 并对该提取该采集到的声音信号作的声纹信息, 之后, 再将提取出的声纹信息进行存储。

本申请实施例提供的哼唱识别操作还可以应用于智能家居设备 (例如, 智能音箱, 电视机, 等等) 以及车载设备 (例如, 车载音箱), 该智能家居设备或者车载设备可以执行本申请实施例提供的哼唱识别操作。在一种可能的情况下, 该智能家居设备或者车载设备上并不配置有显示屏 (例如, 智能音箱, 车载音响), 用户可以通过智能终端 100 对智能家居设备或者车载设备上的哼唱识别功能进行设置。

图 6A-图 6B 示例性示出了又一些设置哼唱识别功能的用户界面。可选的, 这些用户界面可以是智能家居类的应用程序中的界面。

如图 6A 所示, 用户界面 61 包括显示区域 60, 显示区域 60 包括指示信息 600, 提醒信息 602, 选择框 610, 选择框 614, 控件 608, 显示区域 606, 其中:

指示信息 600, 可以用于指示用户设置的家庭信息, 还可以为 “安妮的家” “杰克的家” 等文字信息。

提醒信息 602, 可以用于提示用户需要注意的一些异常情况, 智能终端 100 可以根据各个智能家居设备的状态, 生成对应的提醒信息。例如, 若防盗门长时间未关闭, 则智能终端 100 可以显示提醒信息 602。或者, 若空气净化器的滤芯剩余量小于预设值, 则智能终端 100 可以显示提醒信息 “空气净化器的滤芯需要更换”, 等等。

选择框 610, 可以显示多个可选的家居状态供用户进行选择, 例如 “回家” “离家” “睡眠” “阅读” 以及 “更多”。对应于一个家居状态, 各个家具设备可存在预设的启用状态。举例而言, 若用户对 “回家” 这个选择框执行了选择操作, 响应于该选择操作, 智能终端 100 控制客厅的吊灯以及空调开启。可选的, 用户可以对各个家居状态下, 各个家居设备的启用状态进行设定, 还可以自定义更多的家居状态。

选择框 604, 可以显示多个家居空间供用户进行选择, 例如 “全部” “客厅” “主卧” “次卧”, 等等。该选择框 604 可接收用户的操作 (例如, 点击操作, 滑动操作, 等等), 响应于该操作, 智能终端 100 在显示框 606 中显示与选择的家居空间对应包含的智能家居设备。例如, 选择框 604B 接收到用户的点击操作, 则智能终端 100 在显示框 606 中显示 “客厅” 中包含的智能家居设备

控件 608, 可用于接收用户的操作, 响应于该用户的操作, 智能终端 100 显示智能家居设备添加界面。用户可通过该添加界面录入新的智能家居设备的信息。

显示区域 606, 可用于显示一个或多个智能家居设备的信息, 这些信息可以包括图片、名称、开启状态等基本信息。显示区域 606 还可以用于接收用户的操作, 响应于该用户的操作, 智能终端 100 显示该操作对应的智能家居设备的设置界面。

如图 6A-图 6B 示例性所示, 显示区域 606 接收用户的点击操作, 响应于该用户的点击操作, 智能终端 100 从用户界面 61 跳转到用户界面 62。用户界面 62 中包括返回键 620, 开关控件 622, 音量控件 626, 开关控件 628, 控件 630, 开关控件 620, 开关控件 634, 控件 636, 控件 638, 控件 640 和跳转控件 642。其中:

返回键 620, 可以用于接收用户的操作, 响应于该用户的操作, 该智能终端 100 返回当前页面的上一个页面 (即, 用户界面 61)。

开关控件 622, 可以用于接收用户的操作, 响应于该用户的操作, 智能终端 100 控制智能音箱的开启或者关闭状态。电子设备控制智能音箱的方式可以是发送控制指令, 以指示智能音箱执行该控制指令对应的操作。

音量控件 626, 可以用于调整播放音频文件的音量。音量控件 626 可以用于接收用户的操作, 响应于该操作, 智能终端 100 控制智能音箱调整播放音频文件的音量。可选的, 当接收到的用户操作为向左的滑动时, 智能终端 100 控制智能音箱降低播放音频文件的音量; 当接收到的用户操作为向右的滑动时, 智能终端 100 控制智能音箱提高播放音频文件的音量。在一种可能的实现方式中, 音频文件控件 626 到所在线段的左端点的距离与所在线段的线段长的比值, 和当前音量与智能音箱播放的最大音量的比值具有对应关系。

开关控件 628, 可以用于接收用户输入的操作, 响应于该用户的操作, 智能终端 100 控制智能音箱开启音效优化功能, 或者关闭音效优化功能。

控件 630, 可以用于接收用户输入时间的操作, 响应于该用户的操作, 智能终端 100 控制智能音箱设定关闭时间为用户输入的时间。

开关控件 620, 可用于接收用户的操作 (例如, 点击操作, 滑动操作, 等等), 响应于该用户的操作, 智能终端 100 控制智能音箱改变哼唱识别功能的开启状态。举例而言, 在接收用户的操作之前, 开关控件 412 的显示状态为 “ON”, 则表明此时智能音箱的哼唱识别功能处于开启状态。若开关控件 412 接收到用户的操作, 响应于该用户的操作, 智能终端 100 将开关控件 412 的显示状态调整为 “OFF”, 并控制智能音箱关闭哼唱识别功能。

开关控件 634, 可用于接收用户的操作, 响应于该用户的操作, 智能终端 100 控制智能音箱执行设定哼唱识别的启用时间段的操作。举例而言, 若在接收用户的操作之前, 开关控件 634 的显示状态为 “OFF”, 表明该智能音箱的哼唱识别操作没有设置启用时间, 哼唱识别操作可以一直在运行过程中。可选的, 在这种情况下, 智能终端 100 不显示控件 636 和控件 638。在接收用户对开关控件 634 的操作之后, 响应于该用户的操作, 电子设备将开关控件 634 的显示状态转变为 “ON”, 并显示显示控件 636 和控件 638。其中, 控件 636 用于接收用户输入的哼唱识别操作的启用时间, 响应于该用户的操作, 智能终端 100 控制智能音箱执行, 设定哼唱识别的启用时间为用户输入的启用时间的操作; 控件 638 用于接收用户输入的哼唱识别操作的结束时间, 响应于该用户的操作, 智能终端 100 控制智能音箱执行, 设定哼唱识别的结束时间为用户输入的结束时间的操作。

在一种可能的实现方式中, 智能音箱有可能自身并不能设置哼唱识别功能的启用时间, 在这种情况下, 响应于该用户的针对控件 636 的操作, 智能终端 100 在开启时间向智能音箱发送启用哼唱识别功能的指令, 以控制智能音箱开启哼唱识别功能; 响应于该用户针对控件 638 的操作, 智能终端 100 在结束时间向智能音箱发送停止哼唱识别功能的指令, 以控制智能音箱停止哼唱识别功能。

控件 640, 可以用于添加, 可用于启用哼唱识别的声纹信息。

跳转控件 554, 可用于接收用户的操作, 响应于该用户的操作, 智能终端 100 从用户界面 62 跳转至声纹 1 的设置界面。该声纹 1 的设备界面可以包括命名和删除功能, 等等。

在一种可能的情况下, 智能音箱中用于匹配的声纹信息为智能终端 100 中存储的声纹信息。可选的, 在智能终端 100 中接收了用户输入的声音信号之后, 将该提取该声音信号的声纹信息发送给, 可进行哼唱识别的智能音箱进行存储。这样, 具有哼唱识别功能的智能音箱可以使用电子设备中存储的声纹信息进行声音信号的匹配。这种情况下录入声纹信息的用户

界面可参照图 5G。

在又一种可能的情况下，智能音箱中用于匹配的声纹信息为，该智能音箱重新录入的用户的声音信号中提取出的声纹信息。图 6C 示例性示出了又一种用于录入声纹信息的用户界面 63。响应于用户对控件 640 的操作，智能终端 100 从用户界面 62 跳转到用户界面 63。用户界面 63 包括：

提示信息 650，可以用于为用户提供提示信息，以指示用户进行声纹信息的录入。

文本信息 652，为用户需要朗读的文字内容。可选的，电子设备可以多次显示不同的文本信息供用户朗读。这样，可以录入更多用户的声音信号，以提升声纹信息的准确性。

在一种可能的情况下，电子设备也可以指示用户演唱几段音乐片段，以进行声纹信息的录入。这种情况下，指示符 650 的内容可以是“请靠近智能音箱，长按播放键，并演唱以下歌曲片段以录入声纹信息”，对应的，文本信息 652 为一段歌词。需要说明的是，该播放键是指智能音箱的播放键，该播放键可以为物理按键，也可以为虚拟按键。

需要说明的是，不限于智能音箱，其他智能家居设备（不局限于没有显示屏的智能家居设备，也可以是配置有显示屏的智能家居设备）均可按照上述介绍的方式对哼唱识别功能进行设置。相似的，车载设备的功能设置也可以在智能终端 100 上进行，这种情况也可参照上述介绍的方式。

以上介绍了一些智能终端 100 对智能家居设备上的哼唱识别功能进行设置的用户界面。在一种可能的实现方式中，智能家居设备或者车载设备上设置有显示屏，可对自身的哼唱识别功能进行设置。以下对车载设备上用于哼唱识别的设置界面进行介绍。

图 7A-图 7B 示例性示出了车载设备上用于显示设置哼唱识别功能的用户界面。

图 7A 示例性示出了车载设备上用于显示应用程序菜单的用户界面 71。用户界面 71 也可称为主菜单。如图 7A 所示，用户界面 71 可包括：日历小工具 700，状态栏 702，显示区域 708，控件 706，其中：

日历小工具 700 可用于指示当前时间，例如日期、星期几、时分信息等。

状态栏 201 可包括：蓝牙指示符 704、无线高保真（wireless fidelity，Wi-Fi）信号的一个或多个信号强度指示符 705，时间指示符 703。

显示区域 708，可用于显示多个应用程序图标，例如，导航的图标 708A、电话的图标 708B、音乐的图标 708C、视频的图标 708D、图库的图标 708E、收音机的图标 708F、机车记录仪的图标 708G、设置的图标 708H。

控件 706，可用于接收用户的操作，响应于用户的操作，车载设备从当前界面跳转回用户界面 71（即主菜单界面）。

如图 7A-图 7B 所示，当接收到用户对设置的图标 708H 的操作时，响应于该用户的操作，车载设备从用户界面 71 跳转至用户界面 72。用户界面 72 为用于显示设置菜单的用户界面，该用户界面中包括多个设置选项，例如，“系统设置 720”“用户设置 722”“音效设置 724”“网络设置 726”“时间设置 728”，等等。显示区域 716 中显示的内容为设置选项对应的设置内容。可选的，“系统设置 720”可以为默认选择的设置选项，这种情况在显示区域 716 中显示的内容为系统设置对应的设置内容。可选的，若一个设置选项接收到用户的操作，响应于该用户的操作，显示区域 716 显示该一个设置选项对应的设置内容。

在一种可能的实现方式中，显示区域 716 可接收用户的操作（例如，向上或向下的滑动

操作), 响应于该操作, 显示区域 716 可显示更多的设置内容。如图 7B 所示, 显示区域 716 中显示的内容为哼唱识别的设置内容。

显示区域 716 中可包括开关控件 710, 控件 712, 控件 714。

开关控件 710, 可用于开启或者关闭哼唱识别的功能。

控件 712, 可以用于添加, 可用于启用哼唱识别的声纹信息。可选的, 控件 712 可用于接收用户的操作, 响应于该用户的操作, 车载设备跳转到用于录入声纹信息的用户界面, 例如, 示例性示出的用户界面用户界面 73。之后将会对用户界面 73 作更详细的说明, 此处不具体展开。

跳转 714, 可用于接收用户的操作, 响应于该用户的操作, 车载设备从用户界面 72 跳转至声纹 1 的设置界面。该声纹 1 的设备界面可以包括命名和删除功能, 等等。

图 7C 示例性示出了一种用于录入声纹信息的用户界面 73。

在一个实施例中, 响应于用户针对用户界面 72 中的控件 712 的操作, 车载设备从用户界面 72 跳转至用户界面 73, 以录入用户想要添加的声纹信息。在用户界面 73 中, 包括指示信息 730 和文本信息 572。其中:

指示信息 730, 可以用于为用户提供提示信息, 以指示用户进行声纹信息的录入。需要说明的是, 播放键为音箱的播放键, 在一种可能的情况下, 音箱的播放键为车载设备显示屏周围的一个物理按键。

文本信息 732, 为用户需要朗读的文字内容。可选的, 车载设备可以多次显示不同的文本信息供用户朗读。这样, 可以录入更多用户的声音信号, 以提升声纹信息的准确性。

在一种可能的情况下, 车载设备也可以指示用户演唱几段音乐片段, 以进行声纹信息的录入。这种情况下, 指示符 730 的内容可以是“请靠近音箱, 长按播放键, 并演唱以下歌曲片段以录入声纹信息”, 对应的, 文本信息 732 为一段歌词。

在录入用户输入的声音信号后, 车载设备可以提取采集的声音信号的声纹信息, 并将该声纹信息进行存储。

以上介绍了车载设备上用于哼唱识别的设置界面, 需要说明的是, 不限于以上介绍的用户界面, 车载设备上用于哼唱识别的设置界面, 还可以参照上述介绍的智能终端中的用户界面 51、用户界面 52、用户界面 53 以及用户界面 54。接下来对车载设备中显示哼唱识别结果的用户界面进行进一步的介绍。

图 8A 示例性示出了一种车载设备上用于显示识别结果的用户界面 81。如图 8A 所示, 当车载设备根据用户哼唱的音乐片段识别出音频文件时, 车载设备跟随用户哼唱的进度播放音频文件, 并在自身当前界面上显示通知窗口 842, 用于显示对用户哼唱的音乐片段的识别结果, 通知窗口 842 可包括: 哼唱识别图标 840, 第三显示区域 841, 播放控件 843, 控件 844。

其中, 哼唱识别图标 840 用于指示通知窗口 842 的来源, 为了方便用户快捷的了解到通知窗口 842 是哼唱识别服务 (或者称为功能、应用程序) 输出的识别结果。需要说明的是, 哼唱识别图标 840 仅为示例图标, 在具体的实施过程中, 哼唱识别图标还可以为其他的图案, 例如, 音符或者其他样式的图标, 本申请实施例不作限制。

第三显示区域 841, 可以用于显示识别出的音频文件的标识信息。例如, 第三显示区域 841 中包含了歌曲的名称《漂洋过海来看你》。第三显示区域 841 还可以包括操作指示信息, 例如, 第三显示区域 841 中包含的“可单击停止播放”, 可为用户提供操作的提醒, 提升用户

操作的便利性。可选的，第三显示区域 841 还可以包含演唱者信息或者当前播放的音频文件的歌词信息。在又一种可能的情况下，该车载设备还可以以悬浮窗的形式显示当前播放的音频文件的歌词信息。该悬浮窗是在车载设备的显示界面中悬浮显示的一个可以移动的窗口。

播放控件 843，可以用于接收用户的操作，响应于该用户的操作，该车载设备暂停播放或者继续播放音频文件。具体的，该车载设备识别出用户哼唱的音乐片段对应的音频文件后，将跟随用户的演唱进度播放该音频文件，此时的播放控件 843 显示第一状态。在播放控件 843 显示为第一状态的情况下，若播放控件 843 接收到用户的操作，则该车载设备暂停播放该音频文件，并将播放控件 843 显示为第二状态。在播放 843 显示为第二状态的情况下，若播放控件 843 接收到用户的操作，则该车载设备继续播放该音频文件，并将播放控件 843 显示为第一状态。

控件 844，可以用于接收用户的操作，响应于该用户的操作，车载设备暂停播放音频文件重新获取用户的语音信号，并对重新获取的语音信号进行哼唱识别。

在一种可能的实施方式中，车载设备中不显示控件 844，可显示指示信息“语音呼出“重新识别”即可再次进行哼唱识别”。这种情况下，若车载设备检测到用户输入的语音信息“重新识别”，车载设备将暂停播放音频文件，并对用户紧接着再次哼唱的片段进行哼唱识别，之后，在根据重新识别的结果再次显示通知窗口。这种方式，无需用户进行手动操作，便于在用户驾驶过程中输出重新进行哼唱识别的指令。

在一种可能的实现方式中，通知窗口 842 在显示预设时间后消失，该预设时间可以为 4 秒，5 秒等时间值。或者，当通知窗口 842 接收到用户的向上滑动的操作时，响应于该操作，车载设备不再在用户界面 81 中显示通知窗口 842。或者，该通知窗口可以在当前歌曲播放完毕后消失。

图 8B 示例性示出了又一种车载设备上用于显示识别结果的用户界面 82。如图 8B 所示，当车载设备根据用户哼唱的音乐片段识别出音频文件时，车载设备跟随用户哼唱的进度播放音频文件，并在自身当前界面上显示用户界面 82，用于显示对用户哼唱的音乐片段的识别结果，用户界面 82 可包括：哼唱识别图标 840，第三显示区域 841，播放控件 843，控件 844，音量控件 851，控件 853 和控件 854。

其中，哼唱识别图标 840，第三显示区域 841，播放控件 843，控件 844 均可参照上述图 8A 中的描述，此处不再赘述。

音量控件 851，可以用于调整播放音频文件的音量。音量控件 851 可以用于接收用户的操作，响应于该操作，车载设备调整播放音频文件的音量。可选的，当接收到的用户操作为向左的滑动时，车载设备降低播放音频文件的音量；当接收到的用户操作为向右的滑动时，车载设备提高播放音频文件的音量。在一种可能的实现方式中，音频文件控件 851 到所在线段的左端点的距离与所在线段的线段长的比值，和当前音量与系统播放的最大音量的比值具有对应关系。

控件 853，可以用于收藏识别出的音频文件。控件 853 可以接收用户的操作，响应于该用户的操作，车载设备将识别出的音频文件的标识，添加到预设的收藏夹（或者称为“喜欢的音乐”的文件夹，本申请不作限制）中，便于用户下一次查找或播放该识别出的音频文件。

控件 854，可以用于下载识别出的音频文件。控件 854 可以接收用户的操作，响应于该用户的操作，车载设备从网络中下载该识别出的音频文件的音频资源。可选的，响应于该用户的操作，车载设备显示选择框，该选择框中包含“标准品质”“高品质”“无损品质”等音

质选项。该选择框用于接收用户对一个选项的选择操作，响应于该用户对一个选项的选择操作，车载设备下载该选项对应的音质的音频资源。

控件 855，可以用于分享识别出的音频文件。控件 855 可以接收用户的操作，响应于该用户的操作，车载设备显示分享框，该分享框中包含多个分享对象，例如，与车载设备蓝牙连接的一个或多个终端设备。该分享框用于接收用户对一个分享对象的选择操作，响应于该用户对一个分享对象的选择操作，车载设备将该音频文件的标识或者音频资源发送给该选择操作对应的分享对象。

在一种可能的实现方式中，通知窗口 842 可用于接收用户的操作，响应于该用户的操作，车载设备显示用户界面 82。

在一种可能的实现方式中，用户界面 82 可接收用户的滑动操作，响应于该滑动操作，车载设备显示用户界面 82 之前最近显示的用户界面。

需要说明的是，不限于车载设备，其他智能家居设备（配置有显示屏的智能家居设备）均可按照上述图 7A-图 7C 介绍的方式对哼唱识别功能进行设置，按照上述图 8A-图 8B 介绍的方式对识别结果进行显示。另外，车载设备以及智能家居设备中显示识别结果的用户界面，还可以参照上述介绍的智能终端中的用户界面 21、用户界面 31、用户界面 32、用户界面 33 以及用户界面 34。但由于各个设备具备的功能不完全相同，用户界面中包含的界面元素可以作相应的调整。

基于前述介绍的 UI 实施例，下面实施例介绍执行本申请提供的哼唱识别方法的系统架构。该系统架构中包括电子设备和音乐识别服务器。其中：

电子设备，可以为图 1A 示例性示出的智能终端 100，具体可以为手机、平板电脑等便携式电子设备，或者智能手表、智能手环等可穿戴设备，电子设备还可以为图 1C 示例性示出的智能家居设备 110 或者图 1D 示例性示出的车载设备 120。具体的，该电子设备可以具有音频输入模块，以及音频输出模块。电子设备可以通过音频输入模块采集外部环境中的声音，并将声音信号发送给音乐识别服务器进行哼唱识别，之后，电子设备从音乐识别服务器接收识别出的音频文件以及播放位置，再通过扬声器模块从该播放位置播放识别出的音频文件。在一种可能的实现方式中，该电子设备中还可以包括摄像头模块，该摄像头模块用于获取用户的口型信息，电子设备可以将获取到的口型信息发送给音乐识别服务器，用于音乐识别服务器结合该口型信息与该声音信号进行哼唱识别。

音乐识别服务器，可以对接收到的声音信号进行特征提取，并利用提取出的特征（例如，基频序列）进行检索，从预存的音频资源库（或者称为特征数据库）中匹配出和用户哼唱片段最相似的音频信息。可选的，该音乐识别服务器可以是一个单独的服务器，该音乐识别服务器还可以由多个服务器共同组成。可选的，该音频资源库可以保存在该音乐识别服务器中，该音频资源库还可以保存在与该音乐识别服务器建立有连接关系的其他的设备（例如，数据库服务器）中。

参见图 9，是本申请实施例提供的一种哼唱识别方法的流程图，本申请实施例提供的哼唱识别方式包括但不限于如下步骤。

S901、电子设备通过音频输入模块采集外部环境中的声音。

可选的，在电子设备通过音频输入模块采集外部环境中的声音之前，电子设备需要判定

自身的音频输入模块和/或音频输出模块是否被占用。若自身的音频输入模块和/或音频输出模块被占用，例如，播放音频/视频、拨打电话、进行语音导航等等，则电子设备不通过音频输入模块采集外部环境中的声音以用于哼唱识别操作，需要说明的是，电子设备不通过音频输入模块采集外部环境中的声音以用于哼唱识别操作，不代表在这种情况下，电子设备不进行通过音频输入模块采集外部环境中的声音的操作，而是代表电子设备获取声音的目的不是为了进行哼唱识别。举例而言，在通话过程中，电子设备（例如，手机）需要通过音频输入模块采集外部环境中的声音，是为了获取用户输入的语音信息，以及获取环境声音用于降噪。

若该电子设备自身的音频输入模块和/或音频输出模块未被占用，则通过音频输入模块采集外部环境中的声音。可选的，在电子设备的音频输入模块和/或音频输出模块被释放之后，例如，音频/视频播放结束、电话挂断、语音导航结束等等，该电子设备可通过音频输入模块采集外部环境中的声音。也可以理解为，本申请实施例提供的哼唱识别操作的优先级，低于该电子设备中除该哼唱识别操作的其他需占用音频资源的操作的优先级。

S902、若该判定该声音的声纹信息与预存的声纹信息一致，则该电子设备向音乐识别服务器发送第一音频文件，该第一音频文件中包含该声音。

具体的，在电子设备向音乐识别服务器发送第一音频文件之前，电子设备会对该声音的声纹信息与预存的声纹信息进行匹配。若匹配成功，即该声音的声纹信息与预存的声纹信息一致，则电子设备向音乐识别服务器发送该第一音频文件，以进行哼唱识别；若匹配失败，即该声音的声纹信息与预存的声纹信息不一致，则电子设备继续通过音频输入模块采集外部环境中的声音。需要说明的是，该预存的声纹信息为，预存的从用户输入的声音信号中提取的声纹信息。具体的，该电子设备可以通过上述实施例中的用户界面 55，用户界面 63，用户界面 73 示例性所示的用户界面接收用户输入的声音；之后，电子设备对采集到的声音作声纹提取处理，再将提取出的声纹信息进行存储。

另外，该声音的声纹信息与预存的声纹信息一致，不代表该声音的声纹信息与预存的声纹信息完全相同；在该声音的声纹信息与预存的声纹信息的相似度不小于预设值（例如，90%，95%）时，则可判定该声音的声纹信息与预存的声纹信息一致。具体的，该电子设备对该声音的声纹信息与预存的声纹信息进行匹配的方式可以为：电子设备从该声音信号中提取出声纹信息，电子设备计算该提取出的声纹信息与预测的声纹信息的相似度。若该相似度大于或等于预设值，则该电子设备判定该声音的声纹信息与预存的声纹信息一致；若该相似度小于预设值，则该电子设备判定该声音的声纹信息与预存的声纹信息不一致。

S903、音乐识别服务器根据该第一音频文件从音频资源库中查找第二音频文件，以及确定第二音频文件的起始播放位置。

在一种实施例中，音乐识别服务器根据该第一音频文件从音频资源库中查找出第二音频文件的方式可以为：对第一音频文件进行特征提取，并利用提取出的特征（例如，基频序列）进行检索，从预存的音频资源库（或者称为特征数据库）中选取和第一音频文件最相似的第二音频文件。也即，该第二音频文件的特征与该第一音频文件的特征的相似度，高于第三音频文件的特征与该声音的特征的相似度，该第三音频文件为上述音频资源库中除该第二音频文件的音频文件。

可选的，音乐识别服务器可以利用自动语音识别（automatic speech recognition, ASR）技术将该第一音频文件转化为文本信息，从而确定该第一音频文件所对应的歌词信息。进一步的，音乐识别服务器可以根据识别出的文字信息确定用户哼唱音乐的进度，进而确定第二

音频文件的起始播放位置。第二音频文件的起始播放位置与该第一音频文件的结束位置相对应，因此，电子设备从该起始播放位置播放第二音频文件，可以达到跟随用户的哼唱进度播放音频的效果。

S904、音乐识别服务器向该电子设备发送该第二音频文件以及第一指示信息，该第一指示信息指示该第二音频文件的起始播放位置。

S905、在接收了音乐识别服务器发送的该第二音频文件以及第一指示信息之后，该电子设备通过音频输出模块从该起始播放位置播放该第二音频文件。

以下将对上述方法中各个步骤的具体实施方式作进一步的补充说明。

在一种实施例中，在电子设备通过音频输入模块采集外部环境中的声音之前，电子设备需要判定自身的哼唱识别功能是否开启。电子设备可以通过如上述实施例中的用户界面 41，用户界面 42，用户界面 51，用户界面 52，用户界面 62，用户界面 72 示例性所示的用户界面接收用户对哼唱识别功能的设置。若电子设备判定自身的哼唱识别功能开启，则电子设备执行通过音频输入模块采集外部环境中的声音的步骤；若电子设备判定自身的哼唱识别功能未开启，则电子设备不执行通过音频输入模块采集外部环境中的声音的步骤。

在又一种可能的实现方式中，当检测到该电子设备处于锁定状态时，该电子设备停止通过音频输入模块采集外部环境中的声音。可以理解的是，当检测到电子设备解除锁定之后，该电子设备可以通过音频输入模块采集外部环境中的声音。这种实现方式，可以参照以上实施例中对用户界面 52 中的开关控件 556 的介绍，该开关控件 556 可用于设定哼唱识别功能的可使用状态。通过这种方式，可以在电子设备处于锁定状态时，停止对环境声音的采集，能够降低功耗，节省电子设备的电量。

在又一种可能的实现方式中，当检测到该电子设备处于预设地点时，该电子设备停止通过音频输入模块采集外部环境中的声音。可以理解的是，若检测到该电子设备不再位于预设地点时，该电子设备可通过音频输入模块采集外部环境中的声音。其中，该预设地点可以为用户设置的地点（例如，用户设置的公司所在地等等），该预设地点还可以为电子设备中预存的地点（例如，学校、医院、影院，等等）。其中，该电子设备可以通过全球定位系统（global positioning system, GPS），蓝牙（bluetooth, BT）或者无线局域网（wireless local area networks, WLAN）确定自身所在位置。这种可能的实现方式，可以参照以上实施例中，对用户界面 52 中的开关控件 557 的介绍。具体的，当该“环境勿扰”的开关控件（开关控件 557）处于开启状态时，电子设备实时检测（或者按照预设周期检测）自身是否位于预设地点，若检测到该电子设备处于预设地点时，该电子设备停止通过音频输入模块采集外部环境中的声音。需要说明的是，该预设地点是不适于播放音频文件的地点，通过这种方式，可以避免在不适宜的地方播放第二音频文件的问题，并节省电子设备的电量。

在又一种可能的实现方式中，当检测到环境光亮度小于预设值的持续时间，大于预设时间时，该电子设备停止通过音频输入模块采集外部环境中的声音。可以理解的是，当检测到环境光亮度大于或等于预设值的持续时间，大于预设时间时，该电子设备可通过音频输入模块采集外部环境中的声音。可选的，电子设备可以通过环境光传感器感知环境光亮度。需要说明的是，电子设备的环境光亮度小于预设值的持续时间，大于预设时间的情况，可能代表了该电子设备位于用户口袋，或者当前时间为夜晚的情况，在这种情况下，电子设备不适于播放音频文件，通过这种方式，可以避免在不适宜的地方播放第二音频文件的问题，并节省

电子设备的电量。

在又一种可能的实现方式中，该电子设备在第一时间段内停止通过音频输入模块采集外部环境中的声音。其中，该第一时间段可以为预设的时间段（例如晚上 11 点到早上 9 点），该第一时间段还可以为根据用户输入的时间信息确定的时间段。其中，该第一时间段为根据用户输入的时间信息确定的时间段的这种情况可以对应于，上述实施例中，对用户界面 52 中的开关控件 536 的介绍。用户可以输入哼唱识别功能的开启时间与结束时间，该第一时间段为从结束时间到开启时间的这一个时间段。

在一种实施例中，在步骤 S902 中，在电子设备判定采集的声音的声纹信息与预存的声纹信息是否一致之前，电子设备可以判断该声音信号是否为人声。若该电子设备判断该声音为人声，则该电子设备再判断采集的声音的声纹信息与预存的声纹信息是否一致；若该电子设备判定该声音不為人声，则该电子设备继续通过音频输入模块采集外部环境中的声音。其中，电子设备判断声音是否为人声的方式可以为：电子设备计算该声音的频率，若该频率位于预设频率范围内，则电子设备判定该声音为人声；若该频率不位于预设频率范围内，则电子设备判定该声音不為人声。该预设频率范围可以根据需求进行设定，举例而言，由于男声的基准音区为 64Hz ~ 523Hz，女声的基准音区为 160Hz ~ 1200Hz，该预设频率范围可以为 64Hz ~ 1200Hz。

在一种实施例中，电子设备还可以通过摄像头获取用户的口型信息。举例而言，电子设备可以通过如上述实施例中的用户界面 52 以及用户界面 53 示例性所示的用户界面接收用户对摄像头的访问权限的设置。可选的，电子设备的摄像头的开启状态可以与哼唱识别功能的开启状态保持一致。可选的，在电子设备通过摄像头获取用户的口型信息之前，该电子设备判断该声音是否为人声。若判定该声音为人声，电子设备可以通过摄像头获取用户的口型信息。判断该声音是否为人声的方式，可参照上述的描述，此处不再赘述。这种方式，可以降低电子设备的功耗，节约电子设备的电量。

在这种情况下，音乐识别服务器还可以接收哼唱识别服务器发送的口型信息，音乐识别服务器能够根据口型信息确定文字信息，结合口型确定的文字信息与第一音频文件共同确定最终的识别结果。也即，该第二音频文件对应的文本信息与该口型信息对应的文本信息的相似度，高于该第三音频文件对应的文本信息与该口型信息对应的文本信息的相似度。通过这种方式能够进一步提升识别第二音频文件的准确性。

在一种实施例中，在该音乐识别服务器根据该第一音频文件从音频资源库中查找出第二音频文件之前，该音乐识别服务器判断该第一音频文件是否为音乐片段。可选的，该音乐识别服务器音乐片段可以根据第一音频文件对应的文字信息，以及音频文件中连续的文字之间的多个间隔时间判断该第一音频文件是否为音乐片段。需要说明的是，音乐识别服务器中预存有个音频文件对应的文字信息（可以理解为歌词），以及音频文件中连续的文字之间的多个间隔时间。若该第一音频文件对应的文字信息与预存的一个或多个音频文件对应的文字信息的相似度不小于预设值，并且该第一音频文件中连续的文字之间的多个间隔时间与该一个或多个音频文件中连续的文字之间的多个间隔时间的相似度不小于预设值，则判定该第一音频文件为音乐片段。具体的，当该音乐识别服务器判定该声音信号为音乐片段时，音乐识别服务器根据该第一音频文件从音频资源库中查找出第二音频文件。可选的，该第二音频文件包

含于该一个或多个音频文件中。当该音乐识别服务器判断该声音信号不为音乐片段时，音乐识别服务器反馈给电子设备该声音信号不为音乐片段的的结果。

在一种可能的实施方式中，该第二音频文件的标签包含于第一用户的用户标签。其中，该第一用户为电子设备中登录的用户，或者为使用该电子设备的用户，该音乐识别服务器中预存有该第一用户的用户标签。通过这种方式，可以使得第二音频文件更加符合用户的喜好，提升用户体验。

在一种实施例，在该电子设备通过音频输出模块从该起始播放位置播放该第二音频文件之前，需要确定该电子设备的位置信息与预设地点是否一致。具体的，若该电子设备判定该电子设备的位置与预设地点不一致，则该电子设备通过该音频输出模块从该起始播放位置播放该第二音频文件。可选的，若电子设备确定自身所在的位置与预设地点一致，电子设备可以仅显示哼唱识别结果，但不对该音频文件进行播放，显示哼唱识别结果的用户界面可参照上述实施例介绍的用户界面 21、用户界面 31、用户界面 32、用户界面 33、用户界面 34、用户界面 81 以及用户界面 82，此处不再赘述。该预设地点的含义以及预设地点的确定方式，可参照前述介绍的内容，此处不再赘述。这种可能的实现方式，可以参照以上实施例，对用户界面 52 中的开关控件 557 的介绍。具体的，当该“环境勿扰”的开关控件（开关控件 557）处于开启状态时，电子设备在播放音频文件之前，需要确定自身所在的位置不为预设地点。通过这种方式，可以避免在不适宜的地方播放第二音频文件的问题，并节省电子设备的电量。

在又一种可能的实现方式中，在电子设备播放第二音频文件之前，确定电子设备所在环境的环境音量，该电子设备根据该环境音量确定播放该第二音频文件的音量。具体的，该环境音量越大，该电子设备播放第二音频文件的音量越大，该环境音量越小，该电子设备播放第二音频文件的音量越小。

在一种可能的实现方式中，在该电子设备通过音频输出模块从该起始播放位置播放该第二音频文件之后，该方法还包括：该电子设备显示第二音频文件的标识信息，以及播放控件；其中，该播放控件的显示状态为第一状态，该第一状态表示该第二音频文件正在被播放；若该电子设备检测到作用于处于该第一状态的该播放控件的第一用户操作，响应于该第一用户操作，该电子设备暂停播放该第二音频文件，并将该播放控件的显示状态设为第二状态，该第二状态表示该第二音频文件暂停播放。可选的，若该电子设备检测到作用于处于该第二状态的该播放控件的第二用户操作，响应于该第二用户操作，该电子设备继续播放该第二音频文件，并将该播放控件的显示状态设为第一状态。另外，该电子设备显示第二音频文件的标识信息，以及播放控件的用户界面，可参照上述实施例介绍的用户界面 21、用户界面 31、用户界面 32、用户界面 33、用户界面 34、用户界面 81 以及用户界面 82，此处不再赘述。

在一种可能的实现方式中，电子设备从开始播放第二音频文件的时刻到预设时刻（例如，第 5 秒，第 6 秒等时间值）的时间段内，将使播放第二音频文件的音量由低到高逐渐增大。例如，从音量的最小值逐渐增大到用户设定的音量值，或者，从用户设定的音量值的 30% 逐渐增大到用户设定的音量值的 100%，还可以存在其他的音量增大方式，本申请实施例不作限制。

在一种可能的实现方式中，在该电子设备通过音频输出模块从该起始播放位置播放该第

二音频文件之后，该电子设备还可以检测该第二音频文件是否存储在预存的音乐文件夹中，若是，该电子设备可以在播放完该第二音频文件之后，播放该音乐文件夹中的其他音频文件。

可以理解的，关于图 9 所述方法的各个步骤的具体实现方式，可参考前述图 1A-图 8B 所述的实施例，这里不赘述。

在一种实施例中，本申请提供的哼唱识别方法还可以应用在开放平台中。具体的，该开放平台获取第一音频文件，该第一音频文件中包括外部环境中的声音；若该开放平台判定该第一音频文件的声纹信息与预存的声纹信息一致，则该开放平台根据该第一音频文件从音频资源库中查找第二音频文件，以及确定该第二音频文件的起始播放位置；其中，该第二音频文件的特征与该第一音频文件的特征的相似度，高于第三音频文件的特征与该声音的特征的相似度，该第三音频文件为上述音频资源库中除该第二音频文件的音频文件，该第二音频文件的起始播放位置与该第一音频文件的结束位置相对应；该开放平台从该起始播放位置播放该第二音频文件，或该开发平台控制电子设备的其他应用程序从该起始播放位置播放该第二音频文件。

其中，该开放平台是提供开放应用程序编程接口(application programming interface, API)或函数(function)的平台。也即，该开放平台可具备提供有 API 的应用程序的功能，或者函数的功能。可选的，该开放平台可以通过调用 API (或者函数) 以实现上述图 9 中电子设备和音乐识别服务器所执行的方法。举例而言，开放平台可以是语音助手平台，可以仅包括电子设备侧的语音助手，也可以包括电子设备侧和服务端侧与语音助手直接关联的平台，还可以仅是服务端侧与语音助手关联的平台，本发明实施例不做具体限定。开放平台获取第一音频的方式可以为，该开放平台通过自身所在装置的音频输入模块获取第一音频文件，或者，该开放平台接收与它自身具有连接关系的电子设备发送的第一音频文件。可选的，电子设备可主动向该开放平台发送第一音频文件，或者，开放平台主动从电子设备中获取第一音频文件。之后，该开放平台调用具有声纹识别功能的 API (或者函数) 判断该第一音频文件的声纹信息与预存的声纹信息是否一致，若判定该第一音频文件的声纹信息与预存的声纹信息一致，该开放文件调用具有哼唱识别功能的 API (或者函数) 以实现根据第一音频文件从音频资源库中查找第二音频文件。之后，该开放平台通过自身所在装置的音频输出模块从该起始播放位置播放该第二音频文件，或者，该开发平台控制电子设备的其他应用程序从该起始播放位置播放该第二音频文件。可选的，该开放平台可以向电子设备发送第二音频文件以及第一指示信息，该第一指示信息包括该起始播放位置，该第一指示信息用于指示该电子设备从该起始播放位置播放该第二音频文件。

需要说明的是，该开放平台执行本申请实施例提供的哼唱识别的方式，均可参照图 9 中所述方法的各个步骤的具体实现方式，此处不再赘述。

本申请的各实施方式可以任意进行组合，以实现不同的技术效果。

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。当使用软件实现时，可以全部或部分地以计算机程序产品的形式实现。所述计算机程序产品包括一个或多个计算机指令。在计算机上加载和执行所述计算机程序指令时，全部或部分地产生按照本申请所述的流程或功能。所述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。所述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，所述计算机指令可以从一个网站

站点、计算机、服务器或数据中心通过有线（例如同轴电缆、光纤、数字用户线）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。所述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用介质集成的服务器、数据中心等数据存储设备。所述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）等。

总之，以上所述仅为本申请技术方案的实施例而已，并非用于限定本申请的保护范围。凡根据本申请的揭露，所作的任何修改、等同替换、改进等，均应包含在本申请的保护范围之内。

## 权利要求书

1、一种哼唱识别方法，其特征在于，包括：

电子设备通过音频输入模块采集外部环境中的声音；

若所述电子设备判定所述声音的声纹信息与预存的声纹信息一致，则所述电子设备向音乐识别服务器发送第一音频文件，所述第一音频文件中包含所述声音，所述音乐识别服务器用于根据所述第一音频文件从音频资源库中查找出第二音频文件，以及确定所述第二音频文件的起始播放位置；其中，所述第二音频文件的特征与所述第一音频文件的特征的相似度，高于第三音频文件的特征与所述声音的特征的相似度，所述第三音频文件为上述音频资源库中除所述第二音频文件的音频文件，所述第二音频文件的起始播放位置与所述第一音频文件的结束位置相对应；

所述电子设备接收所述音乐识别服务器发送的所述第二音频文件以及第一指示信息，所述第一指示信息指示所述第二音频文件的起始播放位置；

所述电子设备通过音频输出模块从所述起始播放位置播放所述第二音频文件。

2、根据权利要求1所述的方法，其特征在于，所述方法还包括：

所述电子设备通过摄像头获取用户的口型信息；

若所述声音的声纹信息与预存的声纹信息一致，则所述电子设备向音乐识别服务器发送所述口型信息；

其中，所述音乐识别服务器还用于将所述口型信息转化为文本信息，所述根据所述第一音频文件从音频资源库中查找出第二音频文件，包括：

根据所述第一音频文件和所述口型信息对应的文本信息从音频资源库中查找出第二音频文件，其中，所述第二音频文件对应的文本信息与所述口型信息对应的文本信息的相似度，高于所述第三音频文件对应的文本信息与所述口型信息对应的文本信息的相似度。

3、根据权利要求2所述的方法，其特征在于，所述电子设备通过摄像头获取用户的口型信息，包括：

若所述电子设备判定所述声音为人声，则通过摄像头获取用户的口型信息。

4、根据权利要求1-3任一项所述的方法，其特征在于，所述电子设备通过音频输入模块采集外部环境中的声音，包括：

若所述电子设备判定所述音频输入模块和/或所述音频输出模块未被占用，则所述电子设备通过所述音频输入模块采集外部环境中的声音。

5、根据权利要求1-4任一项所述的方法，其特征在于，所述第二音频文件的标签包含于第一用户的用户标签。

6、根据权利要求1-5中任一项所述的方法，其特征在于，在所述电子设备通过音频输出模块从所述起始播放位置播放所述第二音频文件之后，所述方法还包括：

所述电子设备显示第二音频文件的标识信息，以及播放控件；

其中，所述播放控件的显示状态为第一状态，所述第一状态表示所述第二音频文件正在被播放；

若所述电子设备检测到作用于处于所述第一状态的所述播放控件的第一用户操作，响应于所述第一用户操作，所述电子设备暂停播放所述第二音频文件，并将所述播放控件的显示状态设为第二状态，所述第二状态表示所述第二音频文件暂停播放。

7、根据权利要求 1-6 任一项所述的方法，其特征在于，所述方法还包括：

当检测到所述电子设备处于锁定状态时，所述电子设备停止通过所述音频输入模块采集外部环境中的声音。

8、根据权利要求 1-7 任一项所述的方法，其特征在于，所述方法还包括：

当检测到所述电子设备处于预设地点时，所述电子设备停止通过所述音频输入模块采集外部环境中的声音。

9、根据权利要求 1-7 任一项所述的方法，其特征在于，所述电子设备通过音频输出模块从所述起始播放位置播放所述第二音频文件，包括：

若所述电子设备判定所述电子设备的位置与预设地点不一致，则所述电子设备通过所述音频输出模块从所述起始播放位置播放所述第二音频文件。

10、根据权利要求 1-9 任一项所述的方法，其特征在于，所述方法还包括：

所述电子设备在第一时间段内停止通过所述音频输入模块采集外部环境中的声音。

11、一种电子设备，其特征在于，包括音频输入模块，音频输出模块，处理器，存储器，其中：

所述存储器用于存储程序指令；

所述处理器用于根据所述程序指令执行以下操作：

通过音频输入模块采集外部环境中的声音；

若判定所述声音的声纹信息与预存的声纹信息一致，则向音乐识别服务器发送第一音频文件，所述第一音频文件中包含所述声音，所述音乐识别服务器用于根据所述第一音频文件从音频资源库中查找出第二音频文件，以及确定第二音频文件的起始播放位置；其中，所述第二音频文件的特征与所述第一音频文件的特征的相似度，高于第三音频文件的特征与所述声音的特征的相似度，所述第三音频文件为上述音频资源库中除所述第二音频文件的音频文件，所述第二音频文件的起始播放位置与所述第一音频文件的结束位置相对应；

接收所述音乐识别服务器发送的所述第二音频文件以及第一指示信息，所述第一指示信息指示所述第二音频文件的起始播放位置；

通过音频输出模块从所述起始播放位置播放所述第二音频文件。

12、根据权利要求 11 所述的电子设备，其特征在于，所述电子设备还包括摄像头，所述处理器还用于根据所述程序指令执行以下操作：

通过摄像头获取用户的口型信息；

若所述声音的声纹信息与预存的声纹信息一致,则向音乐识别服务器发送所述口型信息;  
其中,所述音乐识别服务器还用于将所述口型信息转化为文本信息;

所述音乐识别服务器还具体用于:根据所述第一音频文件和所述口型信息对应的文本信息从音频资源库中查找出第二音频文件,其中,所述第二音频文件对应的文本信息与所述口型信息对应的文本信息的相似度,高于所述第三音频文件对应的文本信息与所述口型信息对应的文本信息的相似度。

13、根据权利要求 12 所述的电子设备,其特征在于,所述处理器具体用于根据所述程序指令执行以下操作:

若判定所述声音为人声,则通过摄像头获取用户的口型信息。

14、根据权利要求 11-13 任一项所述的电子设备,其特征在于,所述处理器具体用于根据所述程序指令执行以下操作:

若判定所述音频输入模块和/或音频输出模块未被占用,则通过音频输入模块采集外部环境中的声音。

15、根据权利要求 11-14 任一项所述的电子设备,其特征在于,所述第二音频文件的标签包含于第一用户的用户标签。

16、根据权利要求 11-15 中任一项所述的电子设备,其特征在于,所述电子设备还包括显示屏,所述处理器还用于根据所述程序指令执行以下操作:

通过显示屏显示第二音频文件的标识信息,以及播放控件;

其中,所述播放控件的显示状态为第一状态,所述第一状态表示所述第二音频文件正在被播放;

若检测到作用于处于所述第一状态的所述播放控件的第一用户操作,响应于所述第一用户操作,暂停播放所述第二音频文件,并将所述播放控件的显示状态设为第二状态,所述第二状态表示所述第二音频文件暂停播放。

17、根据权利要求 11-16 任一项所述的电子设备,其特征在于,所述处理器还用于根据所述程序指令执行以下操作:

当检测到所述电子设备处于锁定状态时,停止通过音频输入模块采集外部环境中的声音。

18、根据权利要求 11-17 任一项所述的电子设备,其特征在于,所述处理器还用于根据所述程序指令执行以下操作:

当检测到所述电子设备处于预设地点时,停止通过音频输入模块采集外部环境中的声音。

19、根据权利要求 11-17 任一项所述的电子设备,其特征在于,所述处理器具体用于根据所述程序指令执行以下操作:

若判定所述电子设备的位置与预设地点不一致,则通过音频输出模块从所述起始播放位置播放所述第二音频文件。

20、根据权利要求 11-19 任一项所述的电子设备，其特征在于，所述处理器还用于根据所述程序指令执行以下操作：

在第一时间段内停止通过音频输入模块采集外部环境中的声音。

21、一种哼唱识别方法，其特征在于，包括：

开放平台获取第一音频文件，所述第一音频文件中包括外部环境中的声音；

若所述开放平台判定所述第一音频文件的声纹信息与预存的声纹信息一致，则所述开放平台根据所述第一音频文件从音频资源库中查找第二音频文件，以及确定所述第二音频文件的起始播放位置；其中，所述第二音频文件的特征与所述第一音频文件的特征的相似度，高于第三音频文件的特征与所述声音的特征的相似度，所述第三音频文件为上述音频资源库中除所述第二音频文件的音频文件，所述第二音频文件的起始播放位置与所述第一音频文件的结束位置相对应；

所述开放平台从所述起始播放位置播放所述第二音频文件，或所述开发平台控制电子设备的其他应用程序从所述起始播放位置播放所述第二音频文件。

22、根据权利要求 21 所述的方法，其特征在于，所述方法还包括：

所述开放平台通过所述电子设备获取用户的口型信息；

若所述开放平台判定所述第一音频文件的声纹信息与预存的声纹信息一致，所述开放平台将所述口型信息转化为文本信息；

所述根据所述第一音频文件从音频资源库中查找出第二音频文件，包括：

根据所述第一音频文件和所述口型信息对应的文本信息从音频资源库中查找出第二音频文件，其中，所述第二音频文件对应的文本信息与所述口型信息对应的文本信息的相似度，高于所述第三音频文件对应的文本信息与所述口型信息对应的文本信息的相似度。

23、根据权利要求 22 所述的方法，其特征在于，所述开放平台通过所述电子设备获取用户的口型信息，包括：

若所述开放平台判定所述第一音频文件中包括的声音为人声，则通过所述电子设备获取用户的口型信息。

24、根据权利要求 21-23 任一项所述的方法，其特征在于，开放平台获取第一音频文件，包括：

若所述音频输入模块和/或音频输出模块未被其他应用占用，则所述开放平台获取第一音频文件。

25、根据权利要求 21-24 任一项所述的方法，其特征在于，所述第二音频文件的标签包含于第一用户的用户标签。

26、根据权利要求 21-25 中任一项所述的方法，其特征在于，在所述开放平台从所述起始播放位置播放所述第二音频文件之后，所述方法还包括：

所述开放平台通过电子设备显示第二音频文件的标识信息，以及播放控件；

其中，所述播放控件的显示状态为第一状态，所述第一状态表示所述第二音频文件正在被播放；

若所述开放平台检测到作用于处于所述第一状态的所述播放控件的第一用户操作，响应于所述第一用户操作，所述开放平台暂停播放所述第二音频文件，或控制电子设备的其他应用程序暂停播放所述第二音频文件，并将所述播放控件的显示状态设为第二状态，所述第二状态表示所述第二音频文件暂停播放。

27、根据权利要求 21-26 任一项所述的方法，其特征在于，所述方法还包括：

当检测到所述电子设备处于锁定状态时，所述开放平台停止获取第一音频文件。

28、根据权利要求 21-27 任一项所述的方法，其特征在于，所述方法还包括：

当检测到所述电子设备处于预设地点时，所述开放平台停止获取第一音频文件。

29、根据权利要求 21-27 任一项所述的方法，其特征在于，所述开放平台从所述起始播放位置播放所述第二音频文件，或所述开发平台控制电子设备的其他应用程序从所述起始播放位置播放所述第二音频文件，包括：

若所述开放平台判定所述电子设备的位置与预设地点不一致，则所述开放平台从所述起始播放位置播放所述第二音频文件，或所述开发平台控制电子设备的其他应用程序从所述起始播放位置播放所述第二音频文件。

30、根据权利要求 21-29 任一项所述的方法，其特征在于，所述方法还包括：

所述开放平台在第一时间段内停止获取第一音频文件。

31、一种包含指令的计算机程序产品，其特征在于，当所述计算机程序产品在电子设备上运行时，使得所述电子设备执行如权利要求 1-10，21-30 中任一项所述的方法。

32、一种计算机可读存储介质，包括指令，其特征在于，当所述指令在电子设备上运行时，使得所述电子设备执行如权利要求 1-10，21-30 中任一项所述的方法。

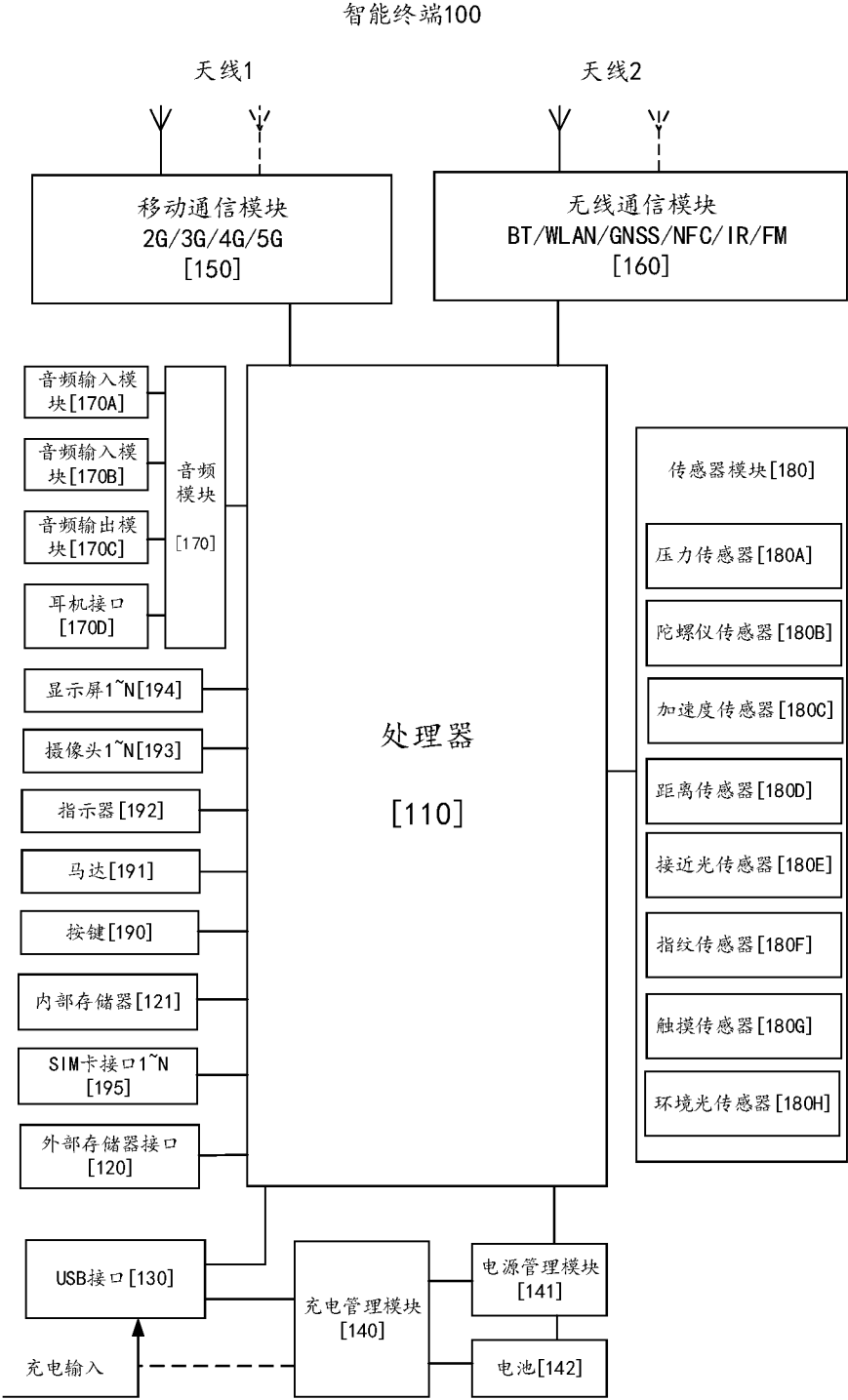


图 1A

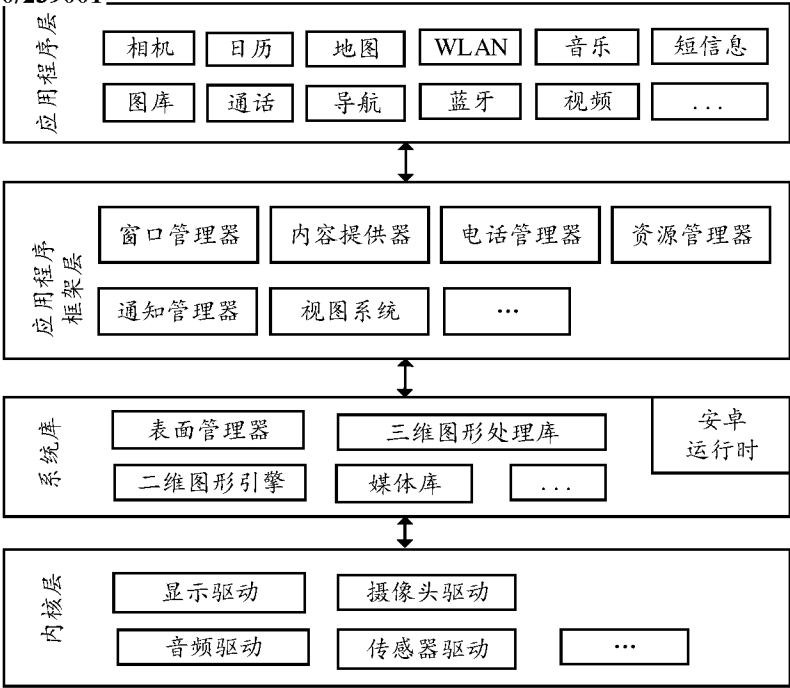


图 1B

智能家居设备110

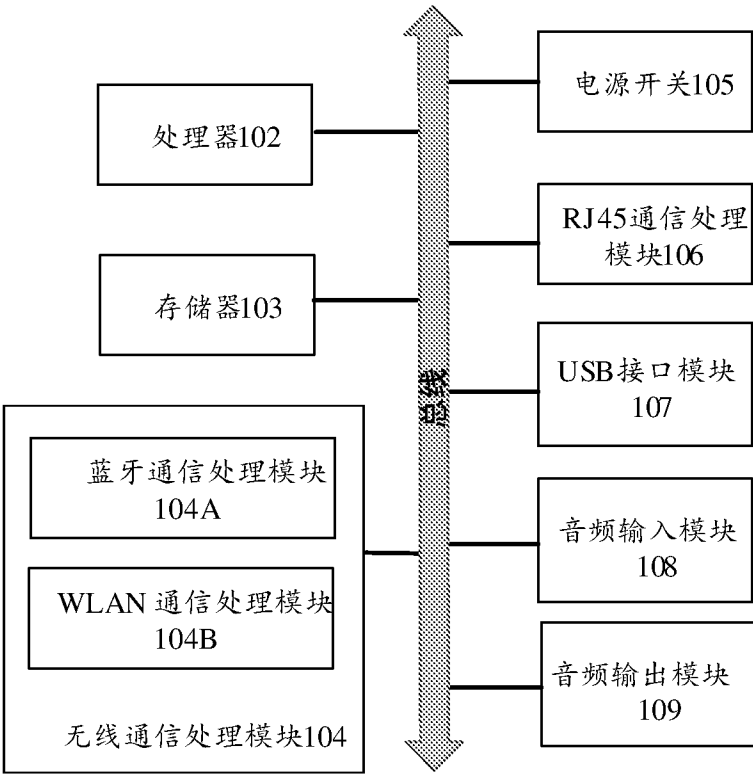


图 1C

车载设备120

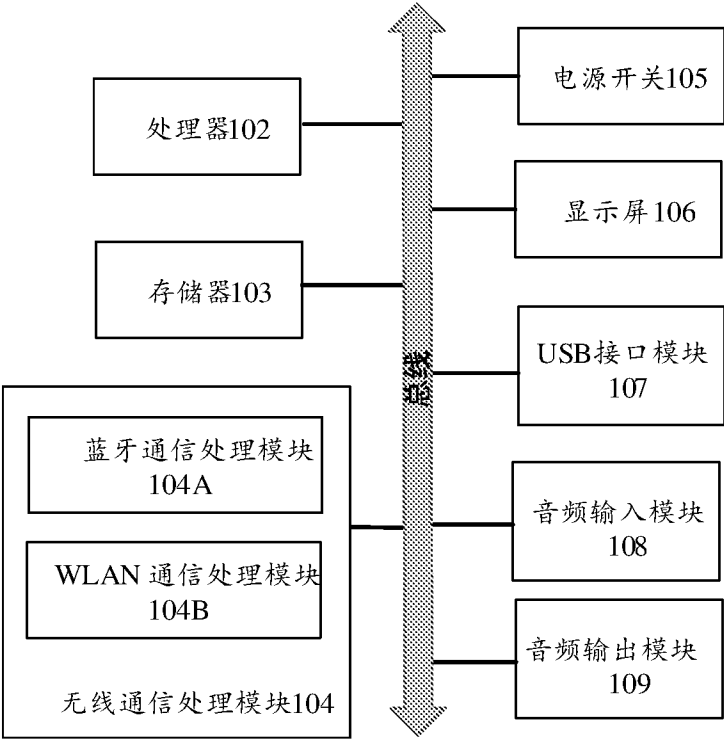


图 1D

用户界面21

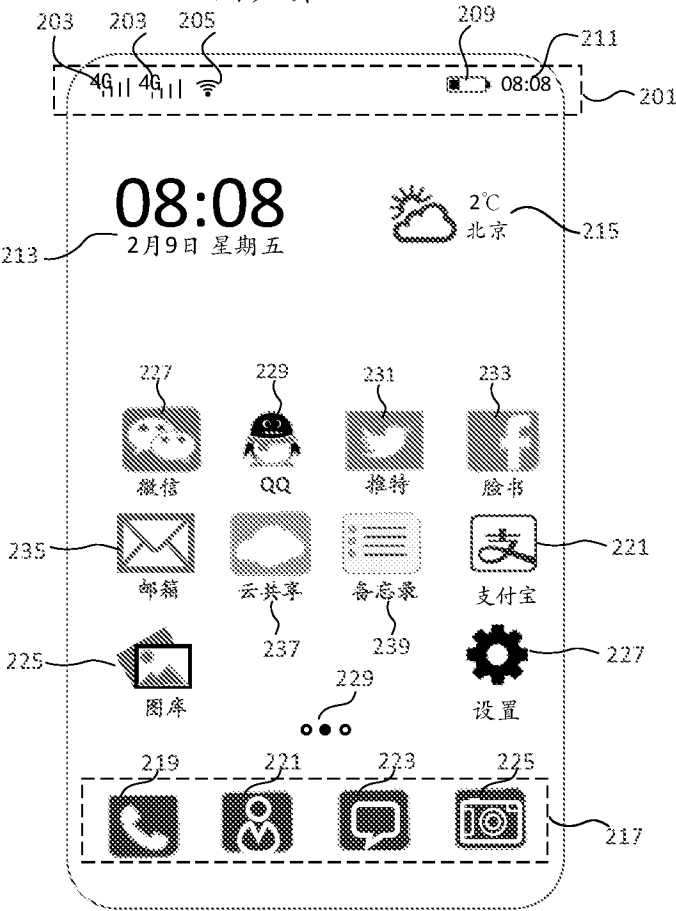


图 2

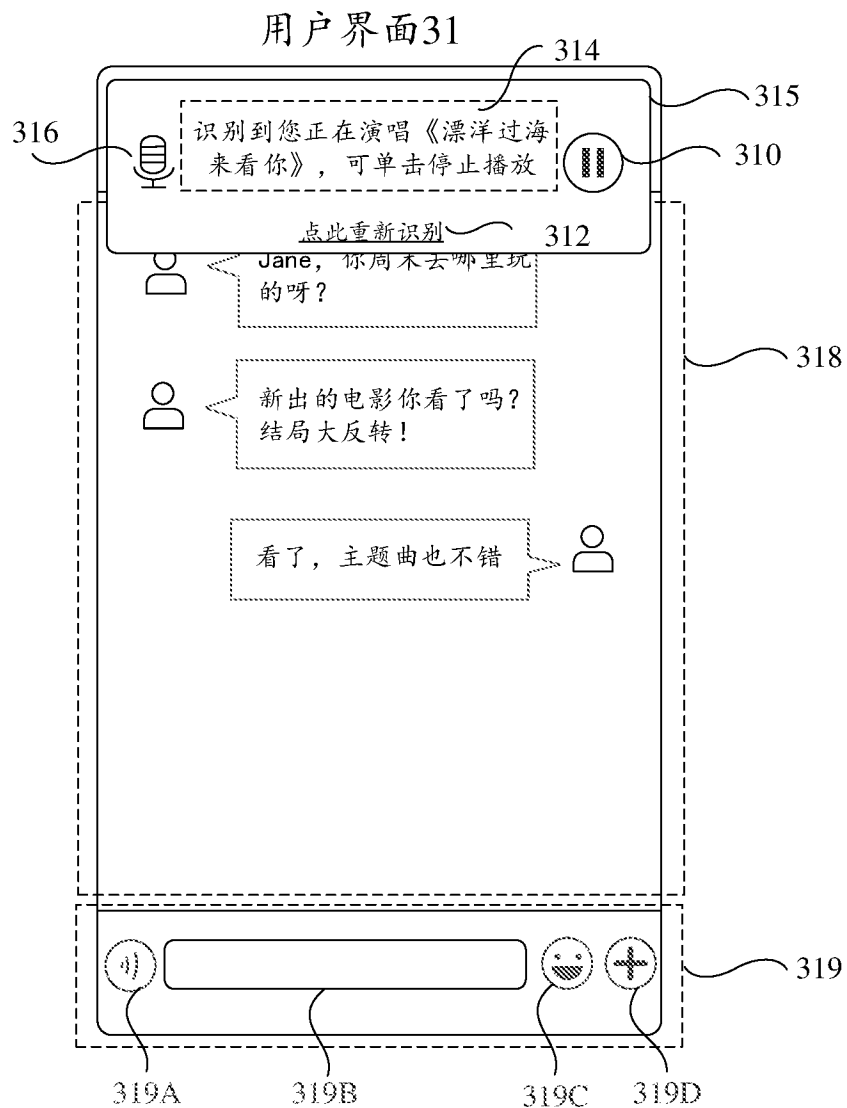


图 3A

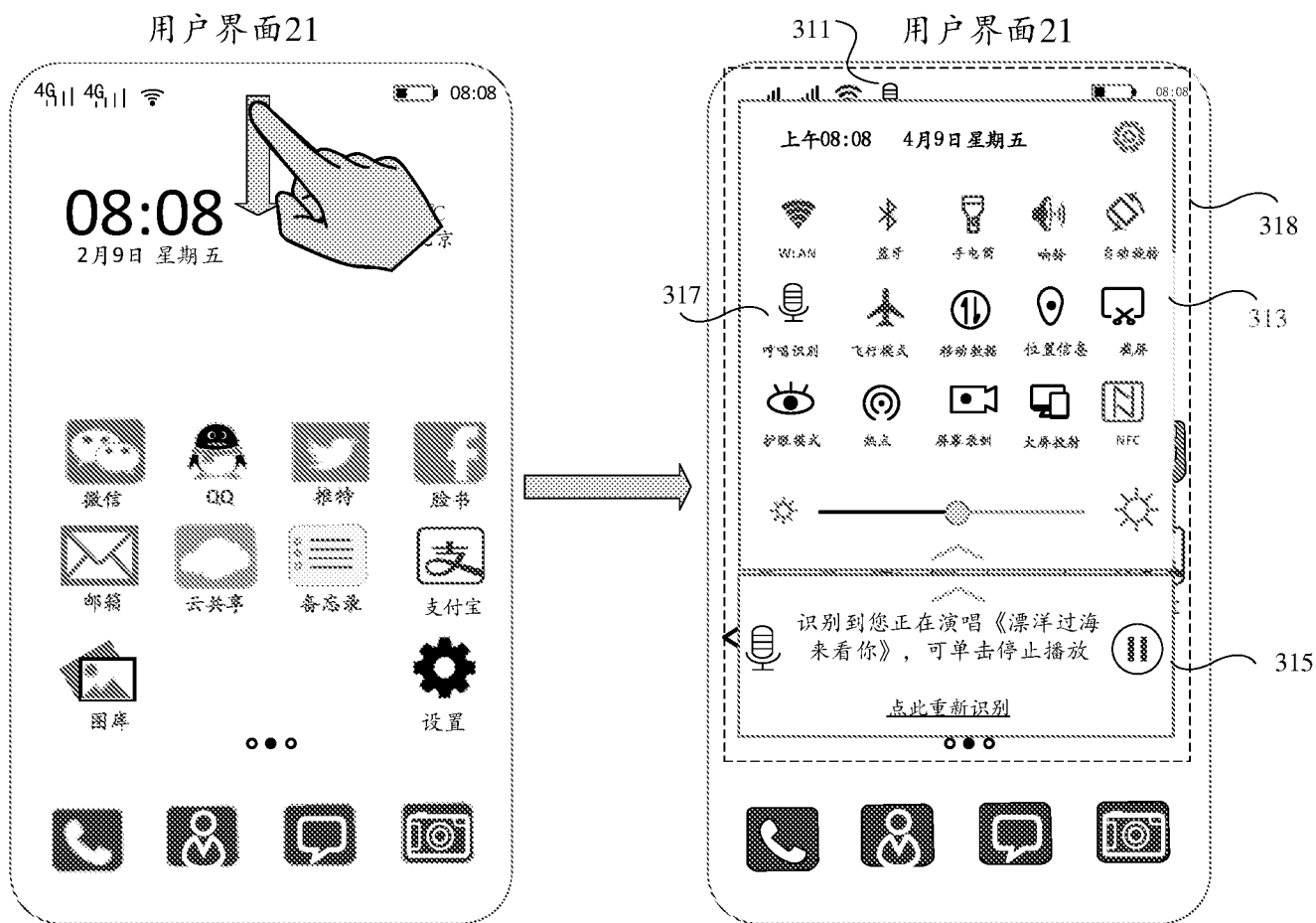


图 3B

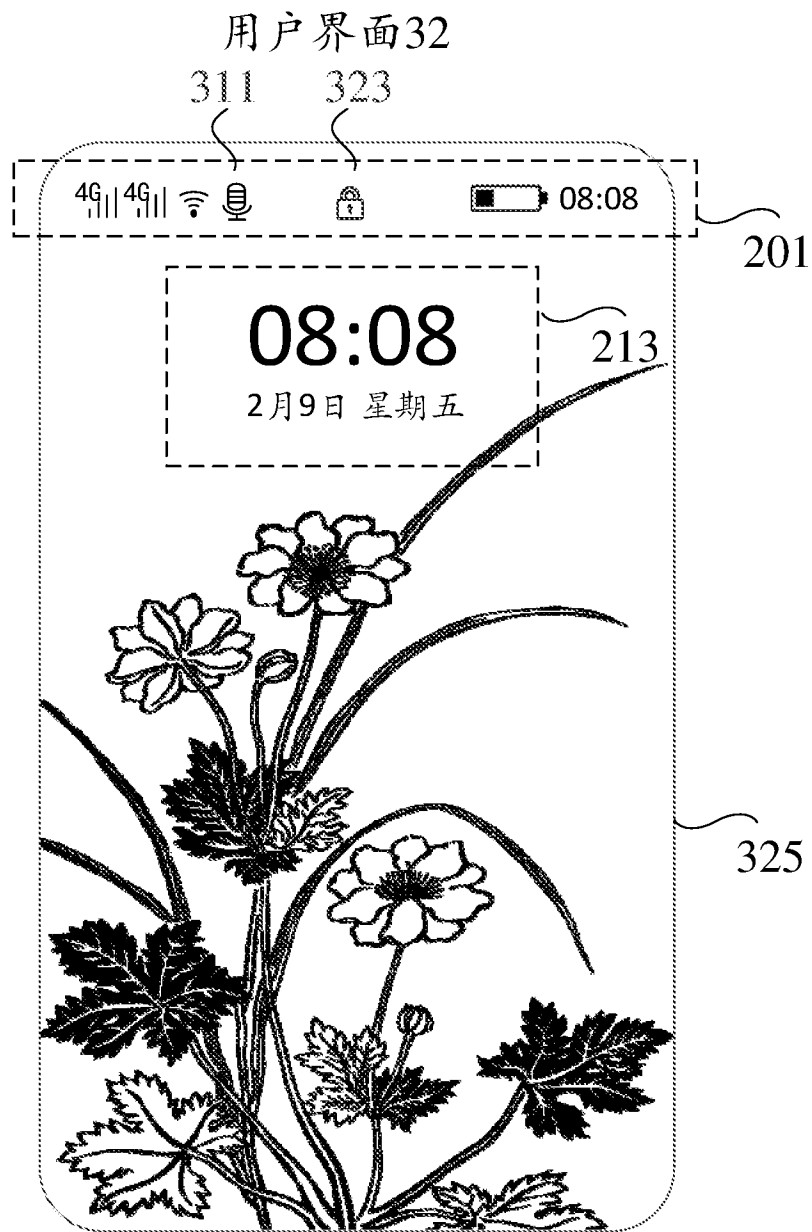


图 3C

用户界面32

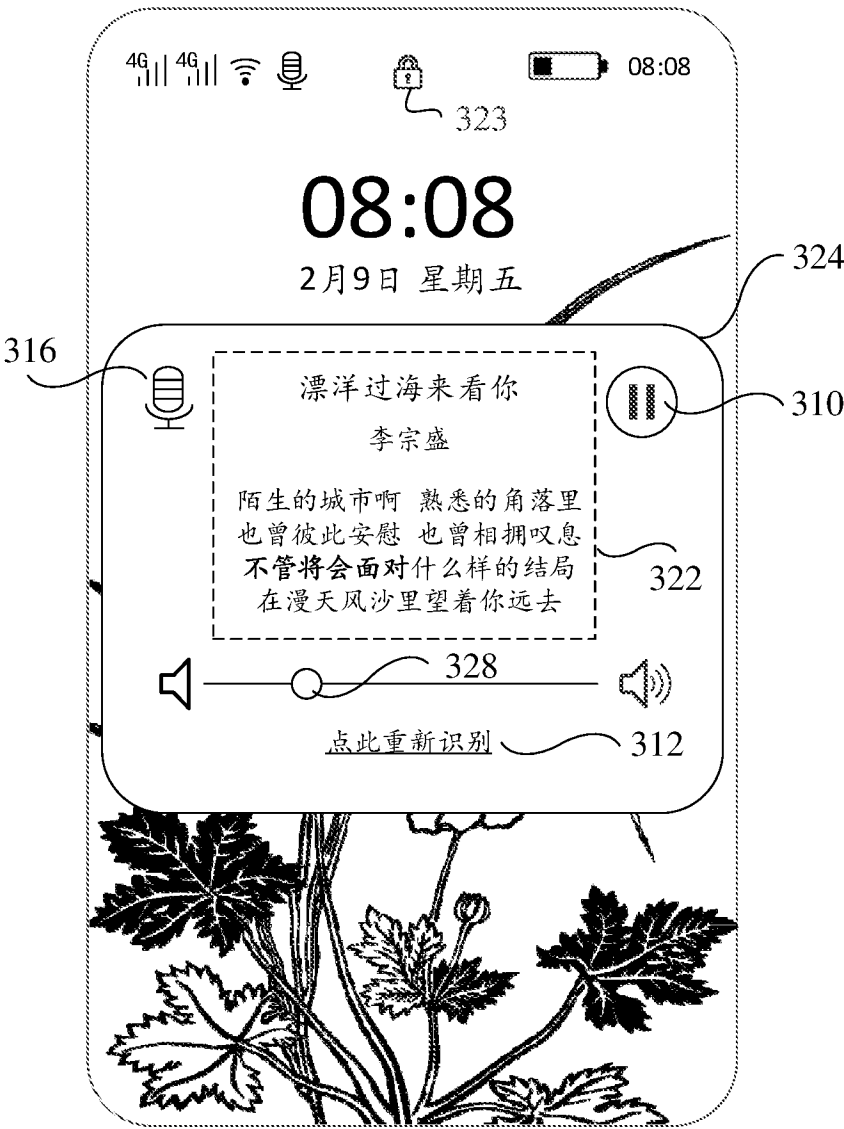


图 3D

用户界面33

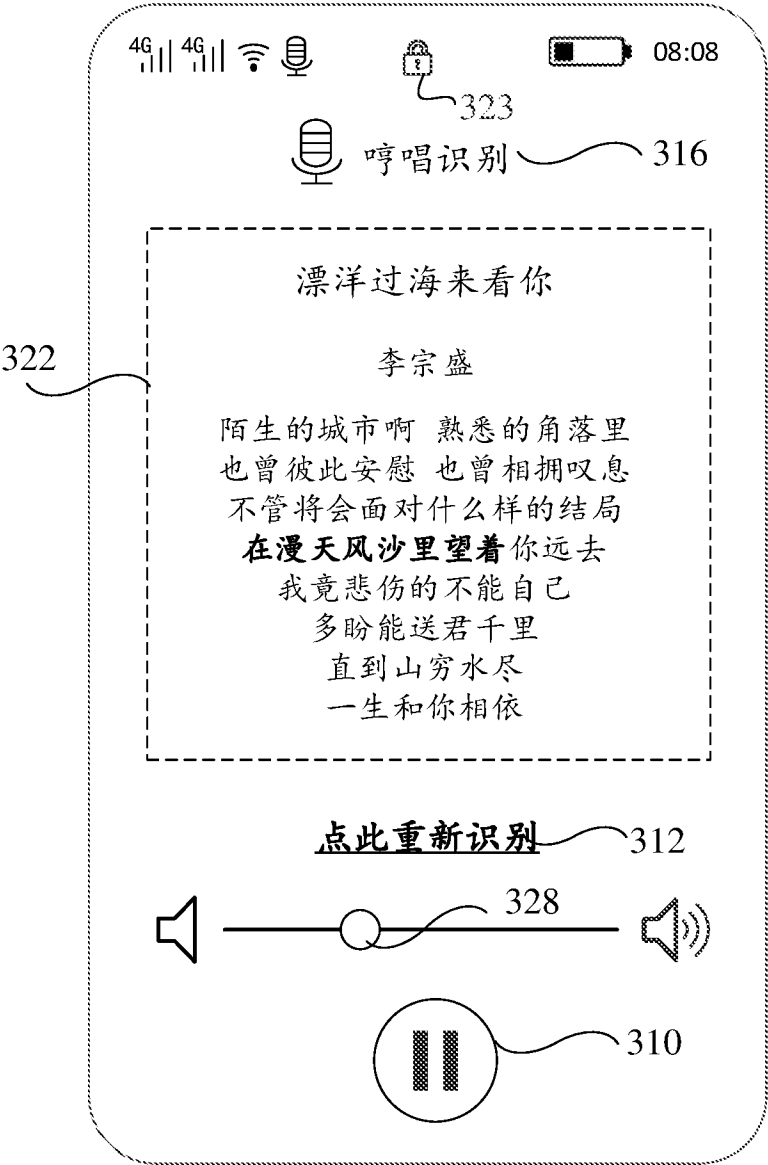


图 3E

用户界面34

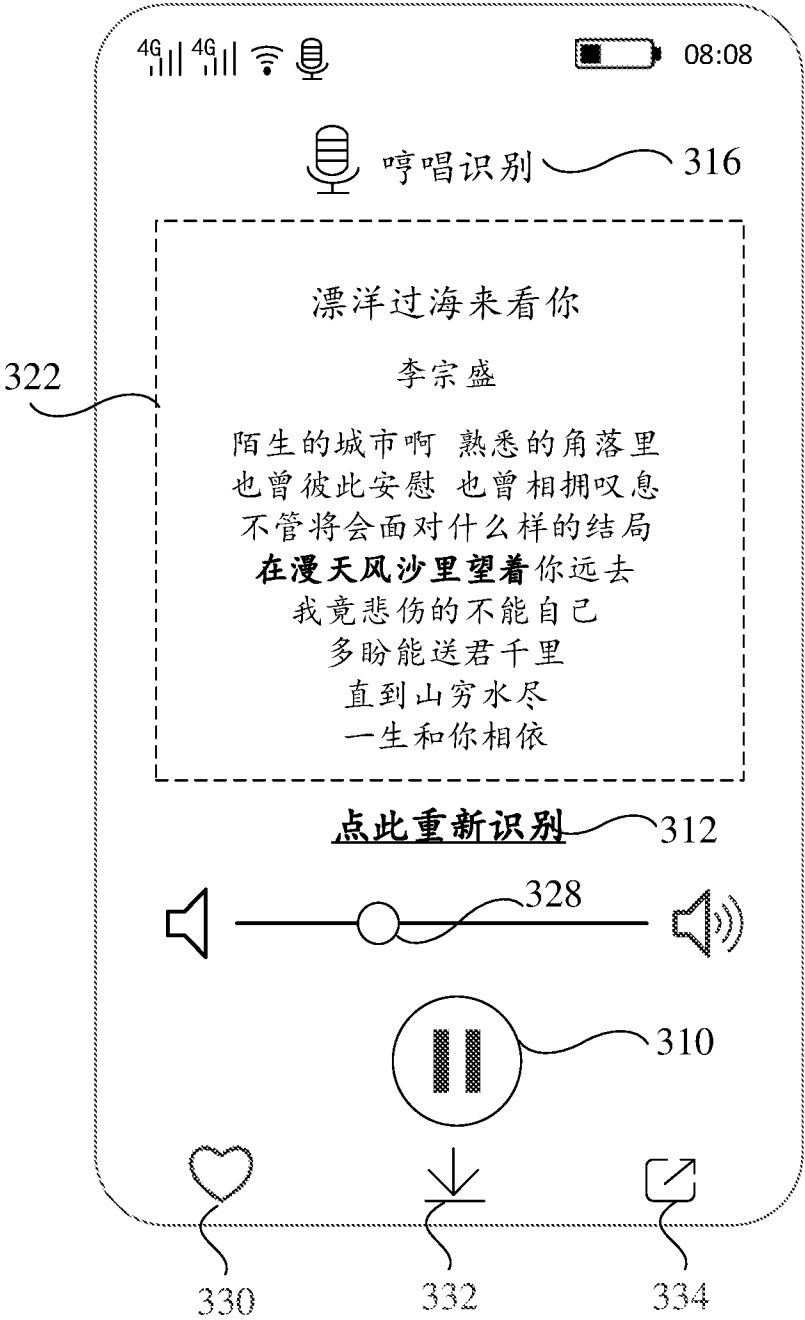


图 3F

## 用户界面35

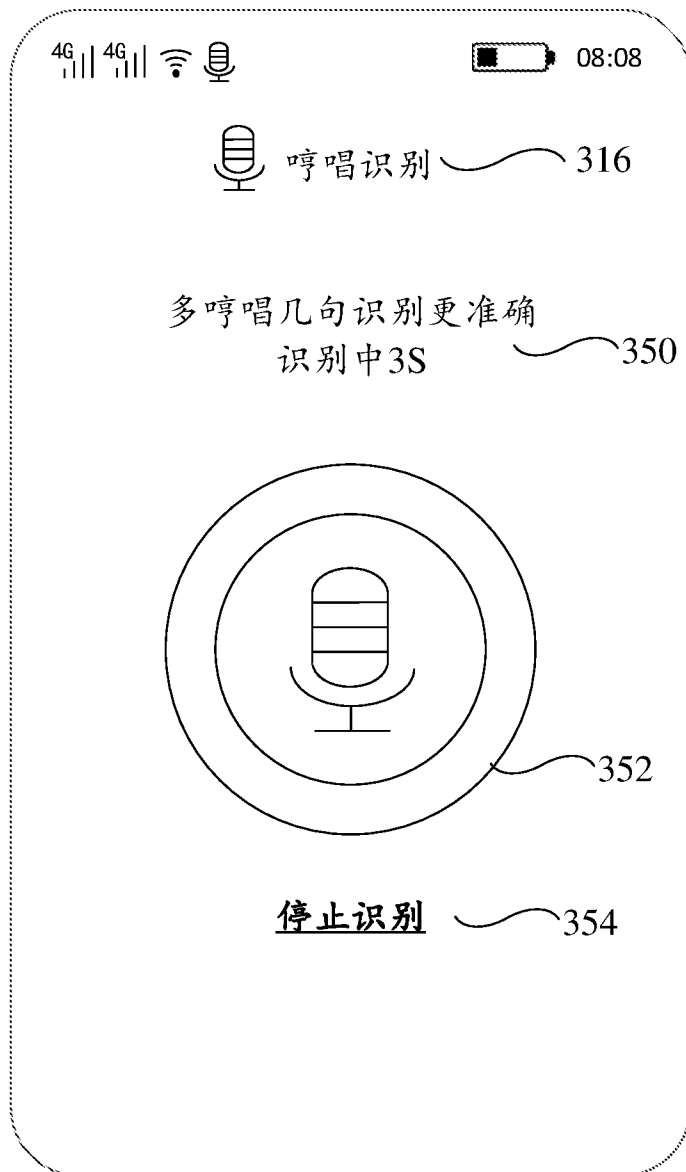


图 3G

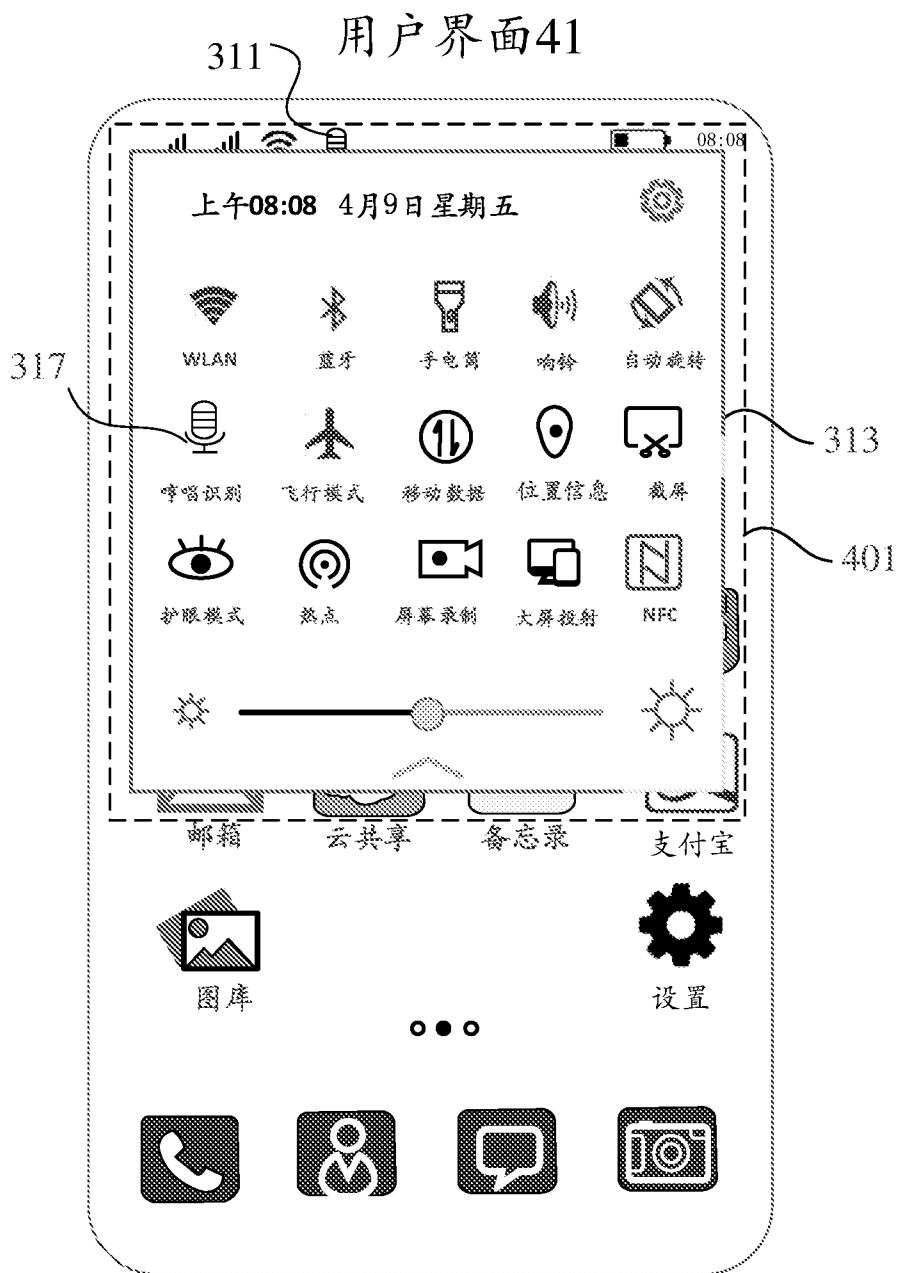


图 4A

用户界面42

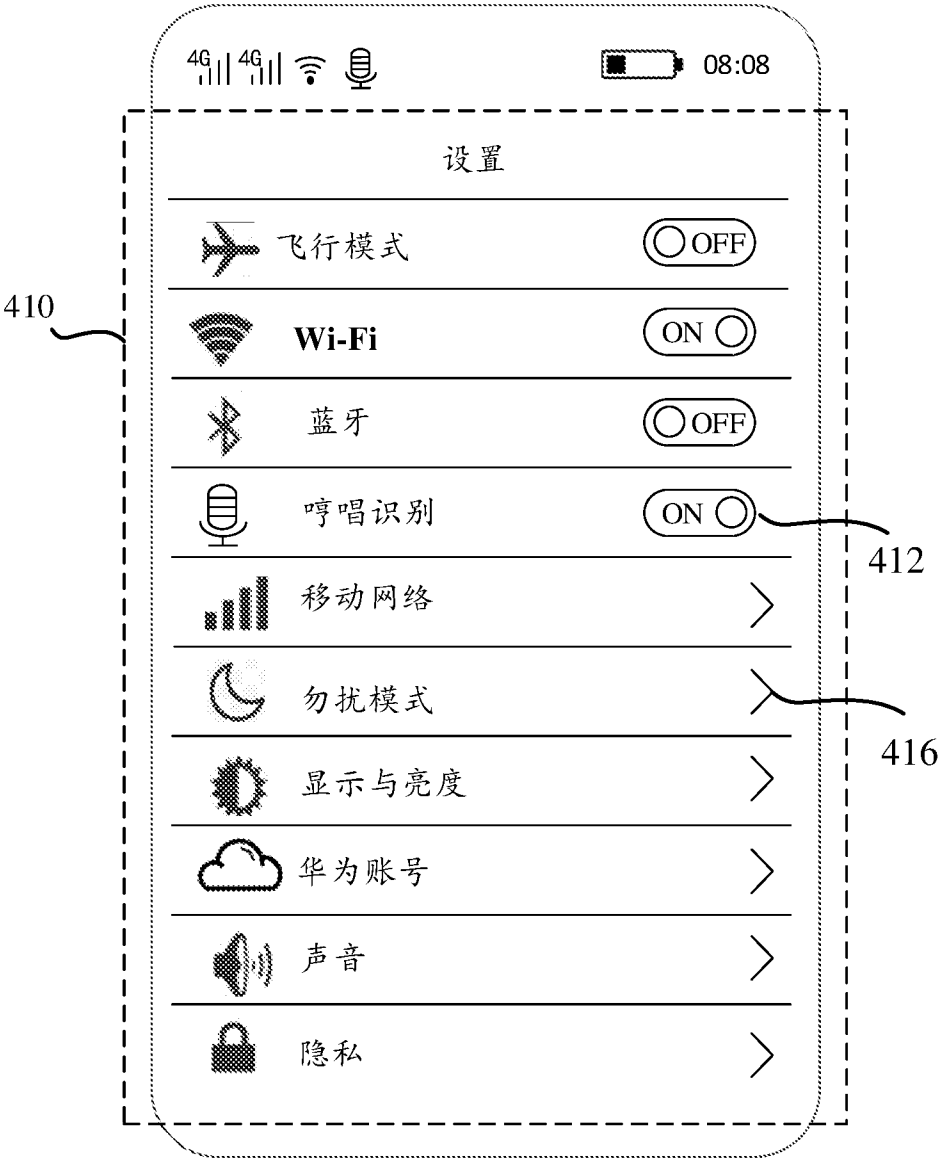
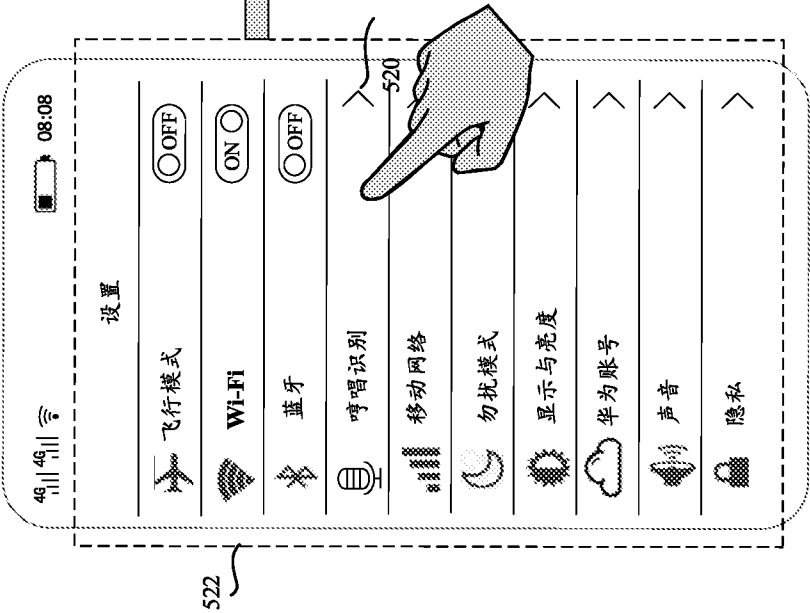


图 4B

用户界面51



用户界面52

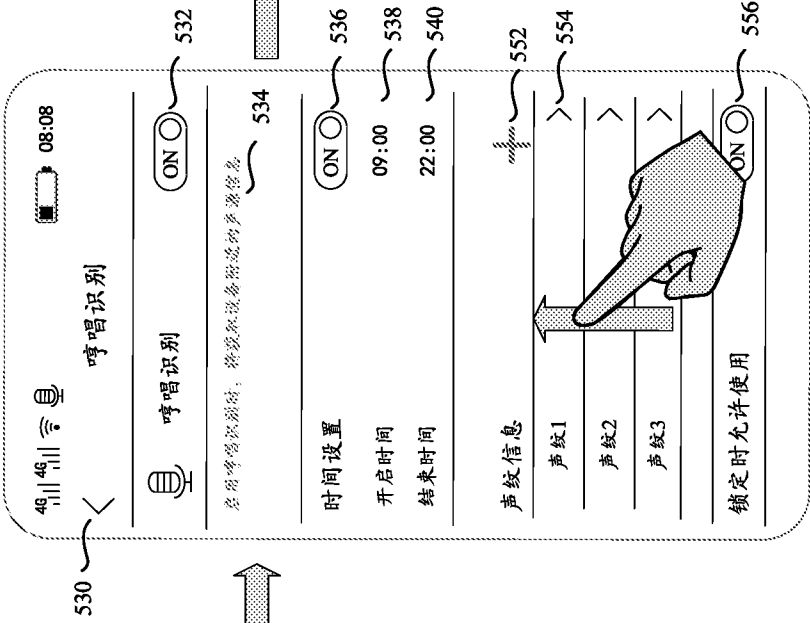


图5A

用户界面52

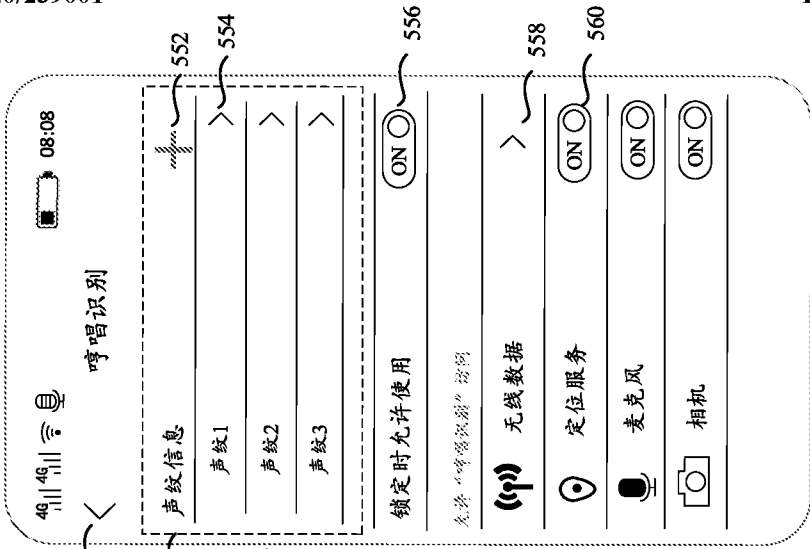


图5C

用户界面51

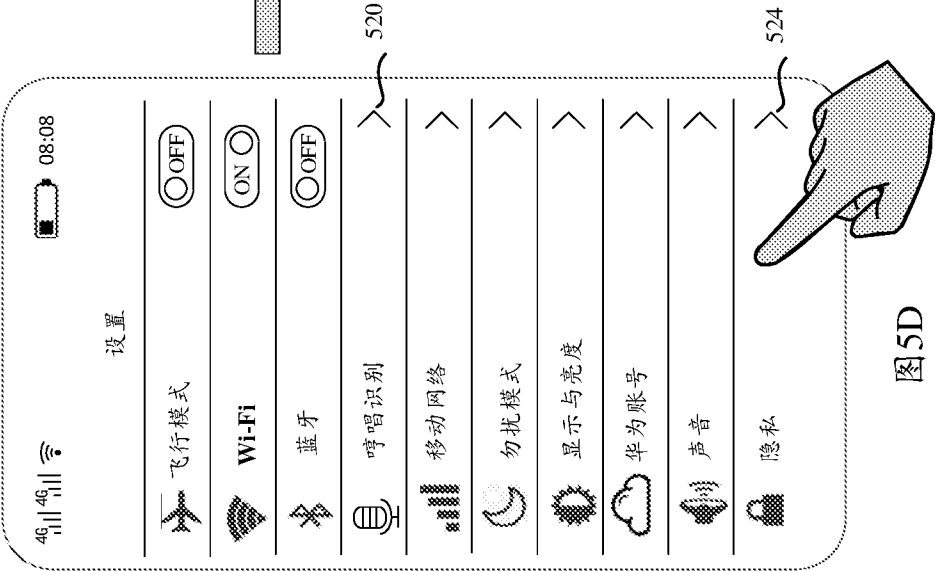


图5D

用户界面53

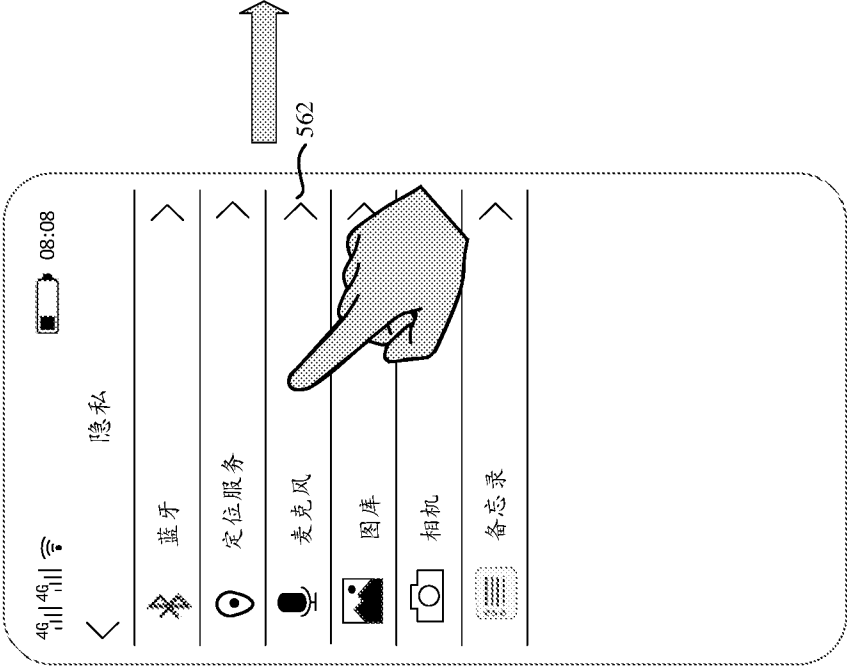


图5E

用户界面54

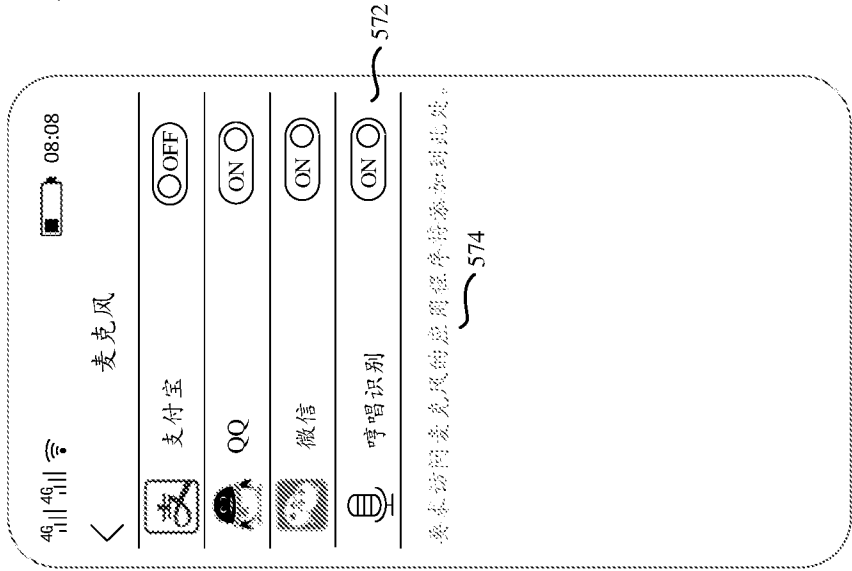


图5F

## 用户界面55

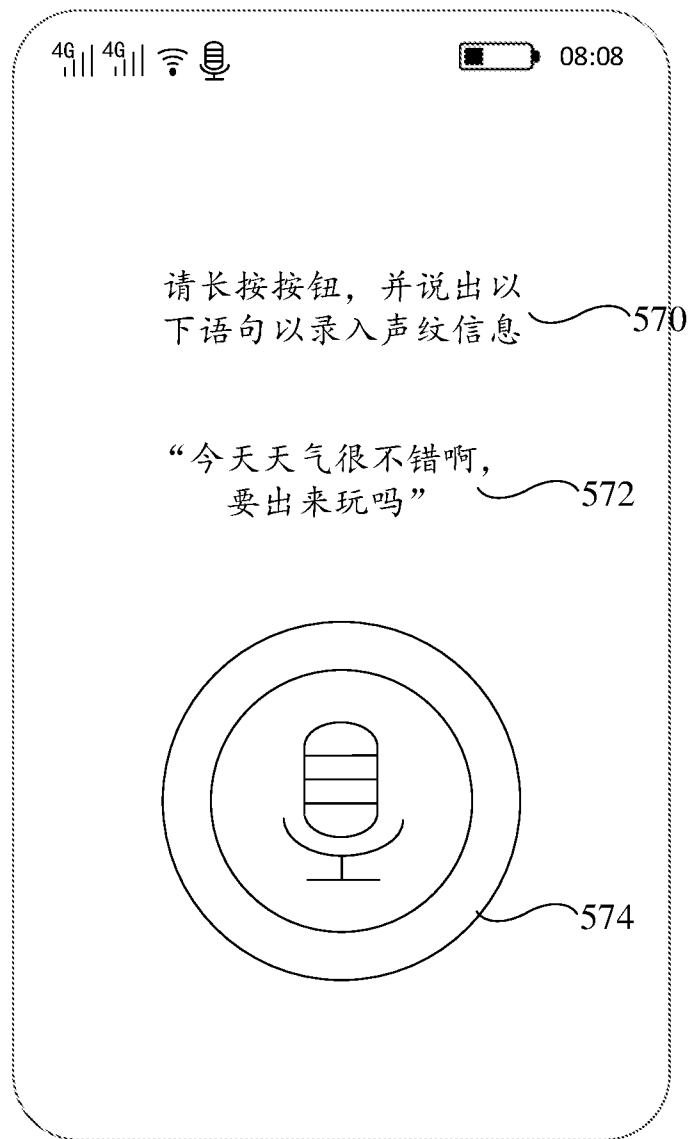


图 5G

用户界面62

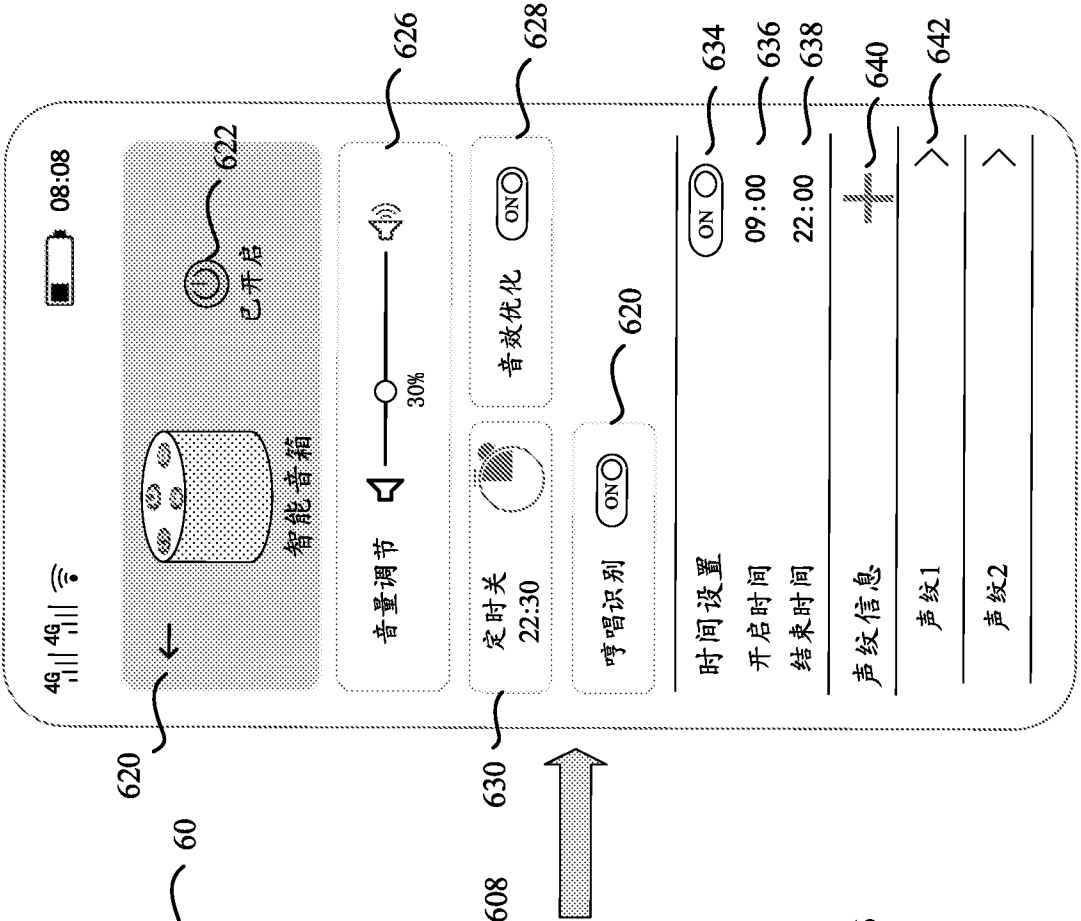


图6B

用户界面61

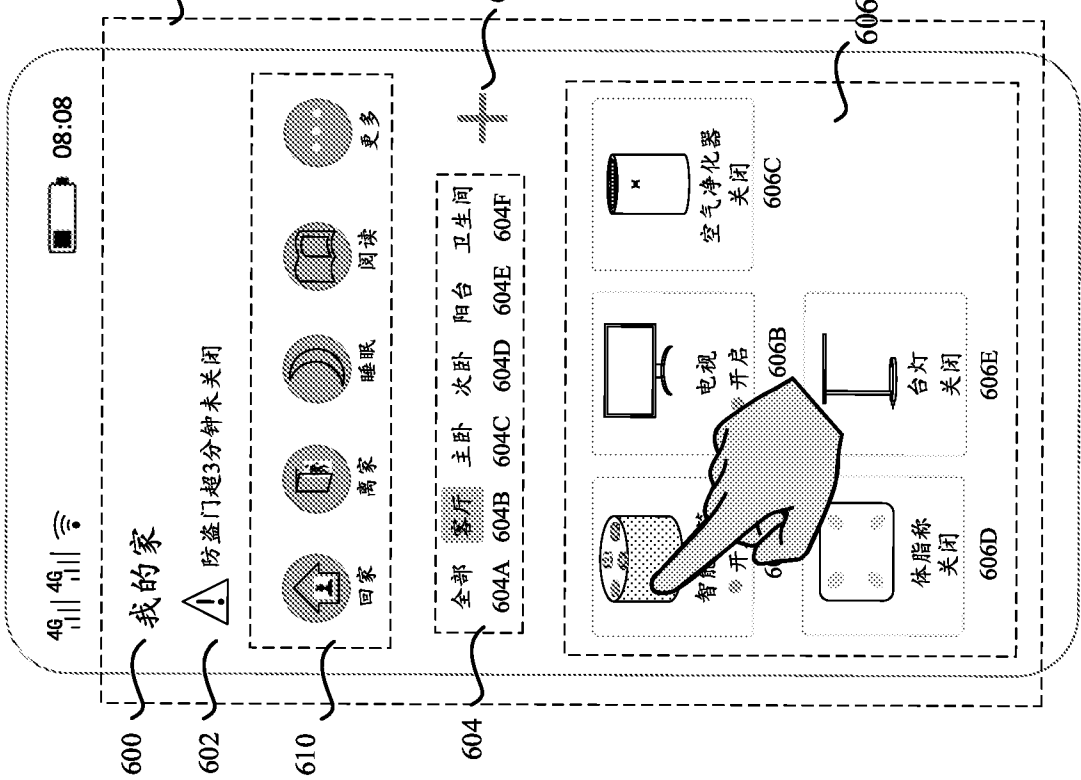


图6A

用户界面63

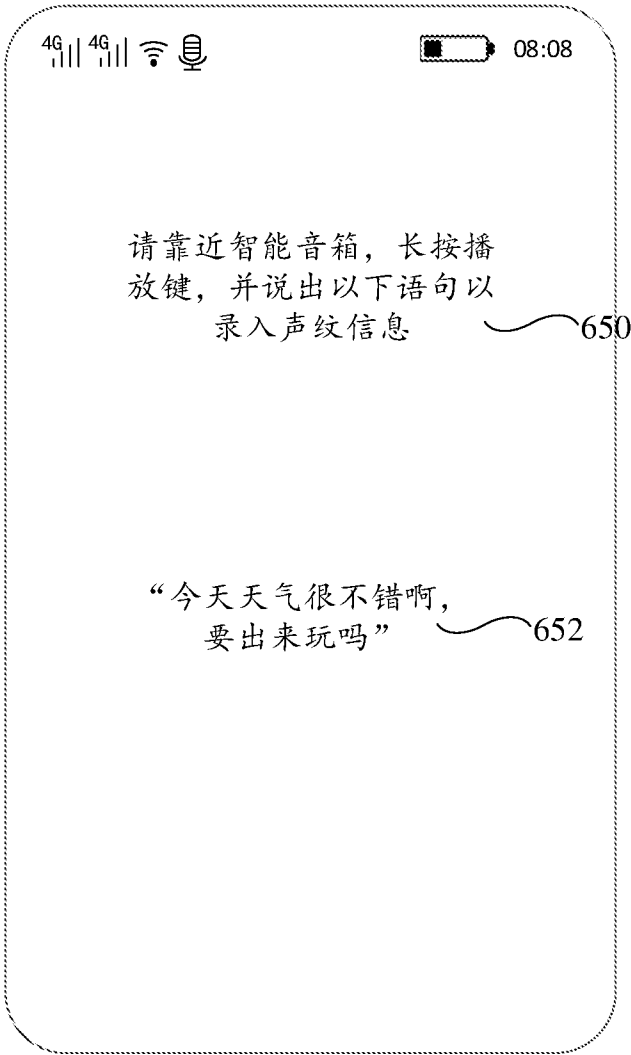


图 6C

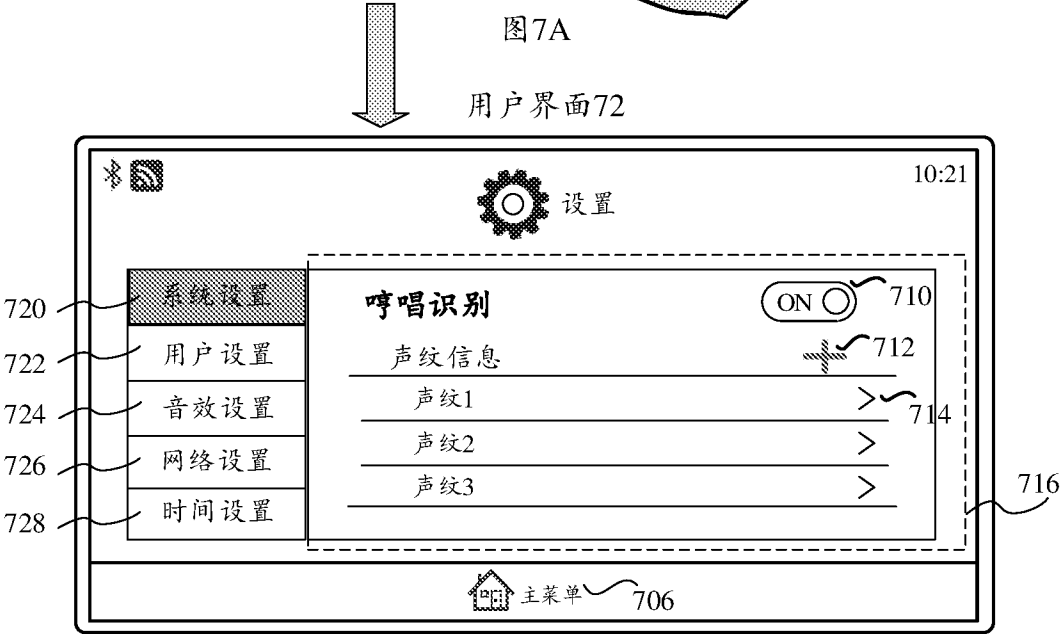
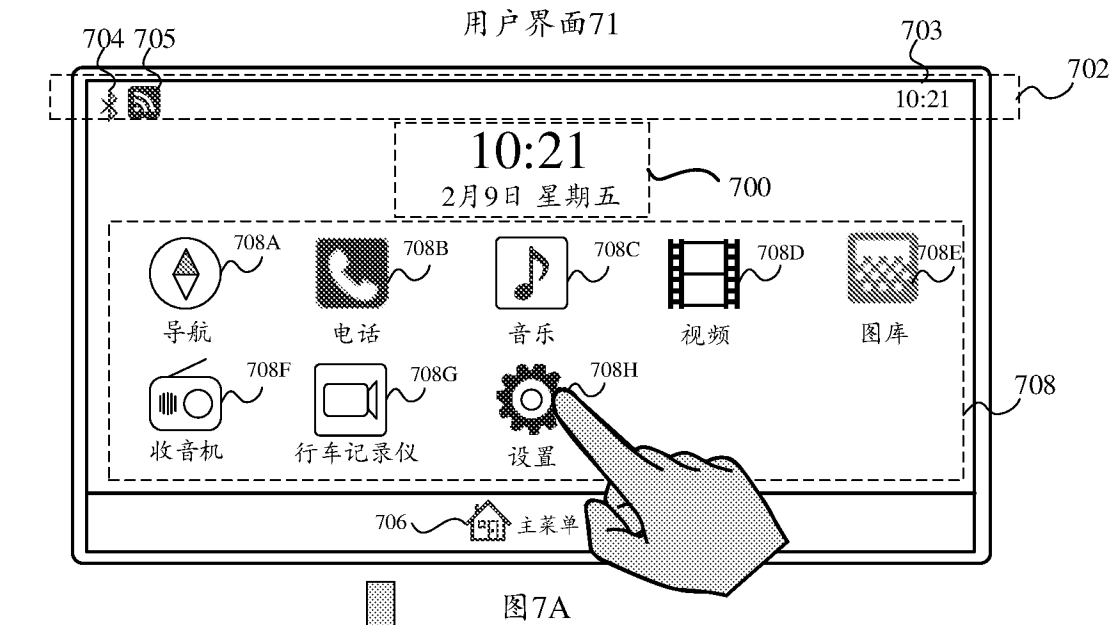


图7B

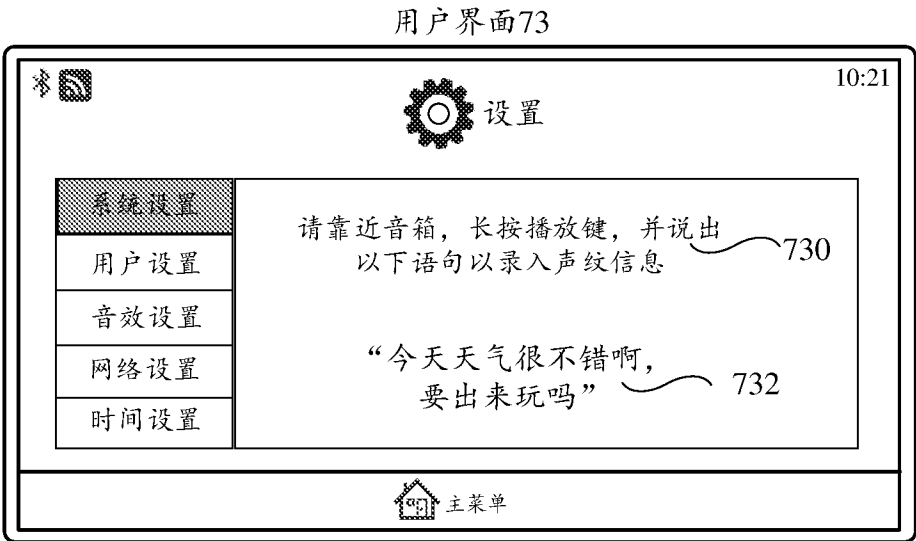


图 7C

用户界面81

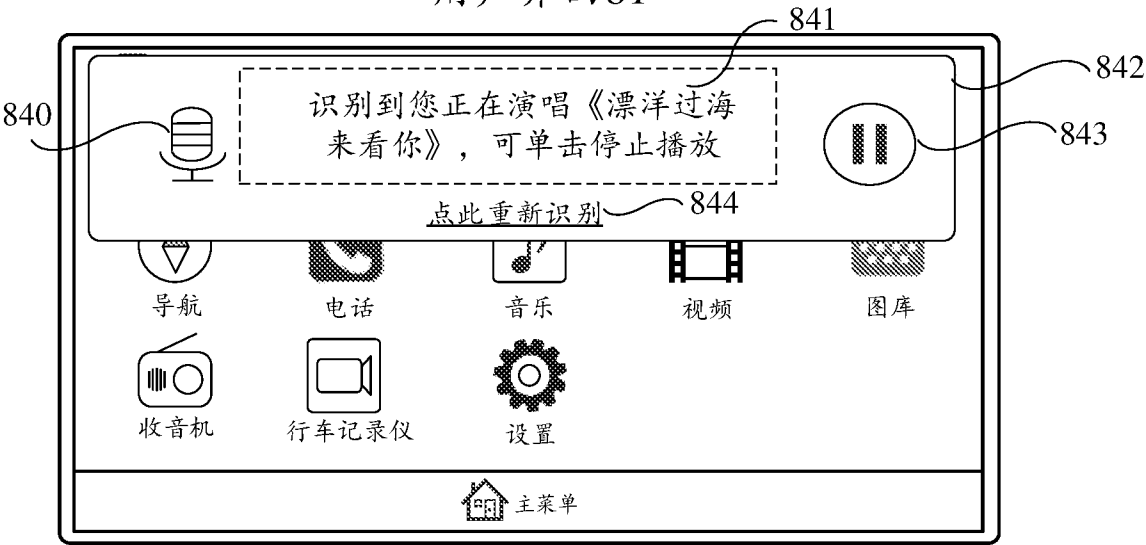


图 8A

用户界面82

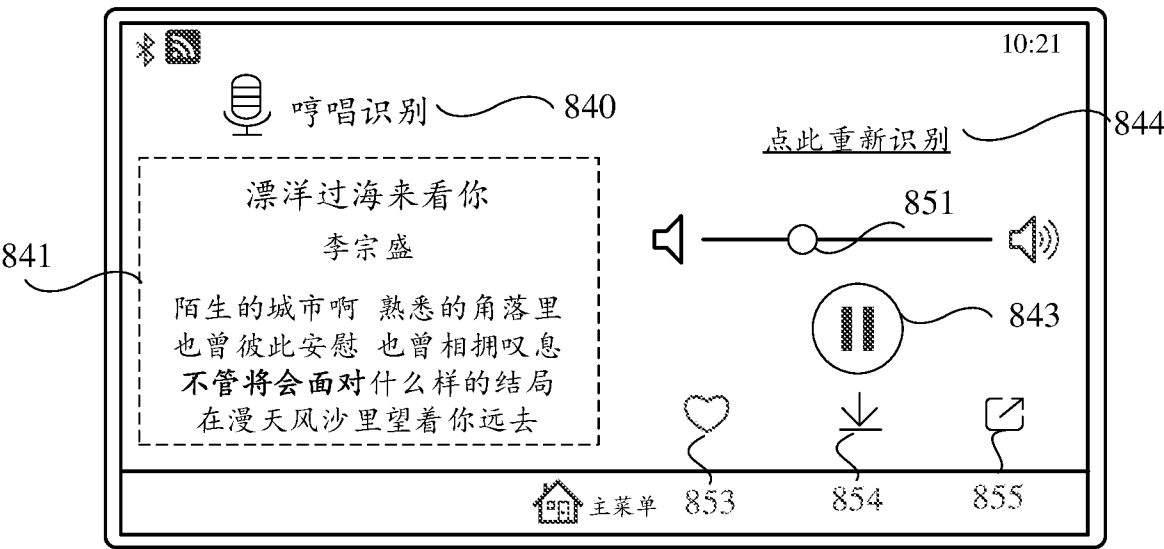


图 8B

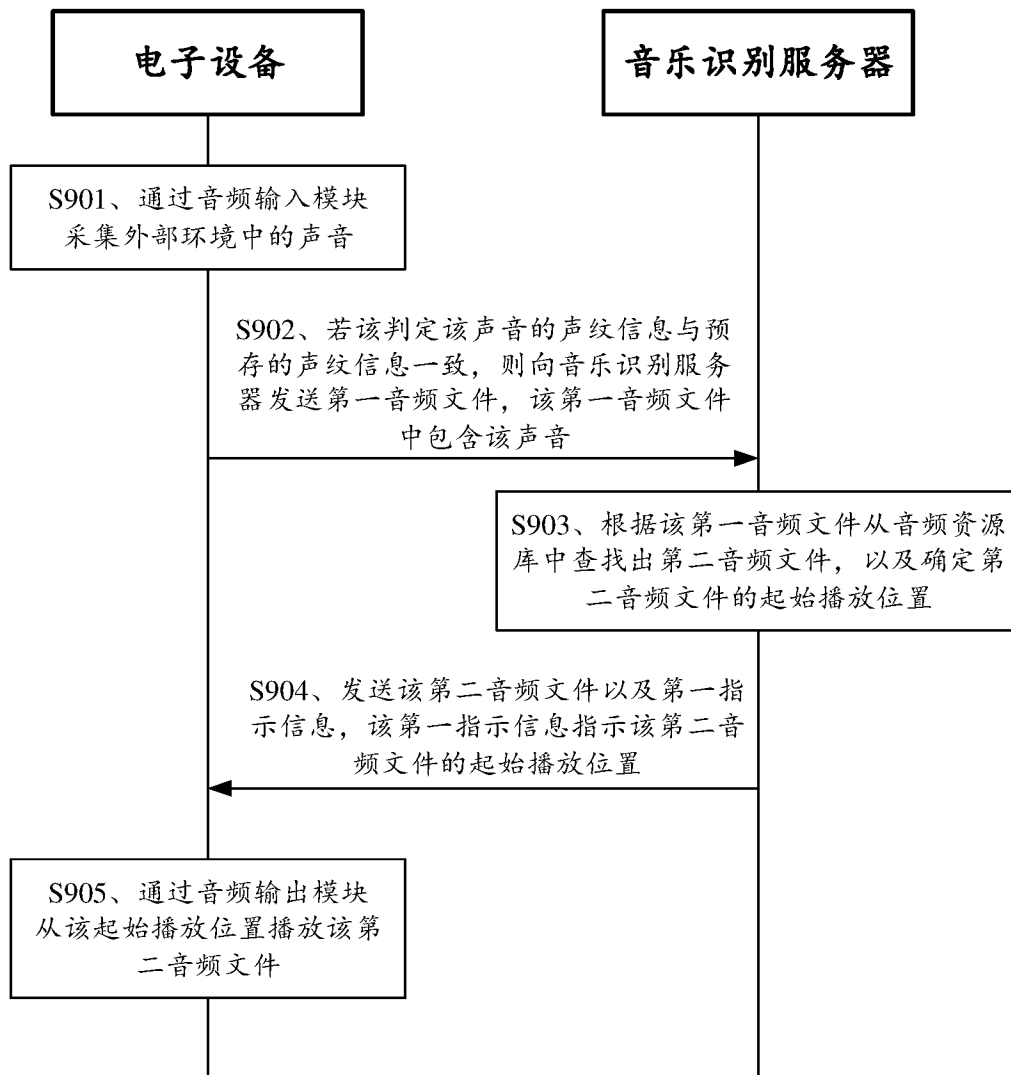


图 9

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/092802

**A. CLASSIFICATION OF SUBJECT MATTER**

G10L 15/02(2006.01)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)

G10L15/-; G06F17/-; H04L29/-

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)

CNABS; CNTXT; DWPI; SIPOABS; CNKI: 音频, 声音, 检索, 输入, 声纹, 口型, 相似度, audio, voice, searching, input, voiceprint, mouth, similarity

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                          | Relevant to claim No. |
|-----------|---------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| Y         | CN 103678308 A (XU, Feng) 26 March 2014 (2014-03-26)<br>claims 1-10, description paragraphs [0037]-[0044]                                   | 1-32                  |
| Y         | CN 104092654 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 08 October 2014<br>(2014-10-08)<br>claims 1-11, description paragraphs [0033]-[0122] | 1-32                  |
| Y         | CN 108320318 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 24 July 2018<br>(2018-07-24)<br>claims 1-20, description paragraphs [0067]-[0225]    | 2, 3, 12, 13, 22, 23  |



Further documents are listed in the continuation of Box C.



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

03 July 2020

Date of mailing of the international search report

20 July 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/  
CN)  
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing  
100088  
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

**INTERNATIONAL SEARCH REPORT**  
**Information on patent family members**

International application No.

**PCT/CN2020/092802**

| Patent document<br>cited in search report |           |   | Publication date<br>(day/month/year) | Patent family member(s) |            |    | Publication date<br>(day/month/year) |
|-------------------------------------------|-----------|---|--------------------------------------|-------------------------|------------|----|--------------------------------------|
| CN                                        | 103678308 | A | 26 March 2014                        | None                    |            |    |                                      |
| CN                                        | 104092654 | A | 08 October 2014                      | US                      | 2016286264 | A1 | 29 September 2016                    |
|                                           |           |   |                                      | US                      | 10097884   | B2 | 09 October 2018                      |
|                                           |           |   |                                      | WO                      | 2015110000 | A1 | 30 July 2015                         |
|                                           |           |   |                                      | CN                      | 104092654  | B  | 02 March 2016                        |
| CN                                        | 108320318 | A | 24 July 2018                         | None                    |            |    |                                      |

## 国际检索报告

国际申请号

PCT/CN2020/092802

## A. 主题的分类

G10L 15/02 (2006.01) i

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

## B. 检索领域

检索的最低限度文献 (标明分类系统和分类号)

G10L15/-; G06F17/-; H04L29/-

包含在检索领域中的除最低限度文献以外的检索文献

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))

CNABS; CNTXT; DWPI; SIPOABS; CNKI: 音频, 声音, 检索, 输入, 声纹, 口型, 相似度, audio, voice, searching, input, voice-print, mouth, similarity

## 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             |
|------|-------------------------------------------------------------------------------------------|---------------------|
| Y    | CN 103678308 A (许丰) 2014年 3月 26日 (2014 - 03 - 26)<br>权利要求1-10, 说明书第[0037]-[0044]段         | 1-32                |
| Y    | CN 104092654 A (腾讯科技深圳有限公司) 2014年 10月 8日 (2014 - 10 - 08)<br>权利要求1-11, 说明书第[0033]-[0122]段 | 1-32                |
| Y    | CN 108320318 A (腾讯科技深圳有限公司) 2018年 7月 24日 (2018 - 07 - 24)<br>权利要求1-20, 说明书第[0067]-[0225]段 | 2、3、12、<br>13、22、23 |

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

\* 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

“&amp;” 同族专利的文件

国际检索实际完成的日期

2020年 7月 3日

国际检索报告邮寄日期

2020年 7月 20日

ISA/CN的名称和邮寄地址

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

受权官员

邢爽

电话号码 010-62411289

国际检索报告  
关于同族专利的信息

国际申请号

PCT/CN2020/092802

| 检索报告引用的专利文件 |           |   | 公布日<br>(年/月/日) | 同族专利 |            |    | 公布日<br>(年/月/日) |
|-------------|-----------|---|----------------|------|------------|----|----------------|
| CN          | 103678308 | A | 2014年 3月 26日   | 无    |            |    |                |
| CN          | 104092654 | A | 2014年 10月 8日   | US   | 2016286264 | A1 | 2016年 9月 29日   |
|             |           |   |                | US   | 10097884   | B2 | 2018年 10月 9日   |
|             |           |   |                | WO   | 2015110000 | A1 | 2015年 7月 30日   |
|             |           |   |                | CN   | 104092654  | B  | 2016年 3月 2日    |
| CN          | 108320318 | A | 2018年 7月 24日   | 无    |            |    |                |