

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 12 月 10 日 (10.12.2020)



(10) 国际公布号
WO 2020/244355 A1

- (51) 国际专利分类号:
G10L 25/78 (2013.01) *G10L 25/51* (2013.01)
G10L 15/22 (2006.01)
- (21) 国际申请号: PCT/CN2020/089551
- (22) 国际申请日: 2020 年 5 月 11 日 (11.05.2020)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201910475949.X 2019年6月3日 (03.06.2019) CN
- (71) 申请人: 清华大学(TSINGHUA UNIVERSITY) [CN/
CN]; 中国北京市海淀区清华园 1 号,
Beijing 100084 (CN)。
- (72) 发明人: 史元春(SHI, Yuanchun); 中国北京市
海淀区清华园 1 号, Beijing 100084 (CN)。 喻

纯(YU, Chun); 中国北京市海淀区清华园
1 号, Beijing 100084 (CN)。

(74) 代理人: 北京集佳知识产权代理有限公司
(UNITALEN ATTORNEYS AT LAW); 中国北京
市朝阳区建国门外大街 22 号赛特广场
7 层, Beijing 100004 (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: MICROPHONE SIGNAL-BASED VOICE INTERACTION WAKE-UP ELECTRONIC DEVICE, METHOD, AND MEDIUM

(54) 发明名称: 基于麦克风信号的语音交互唤醒电子设备、方法和介质

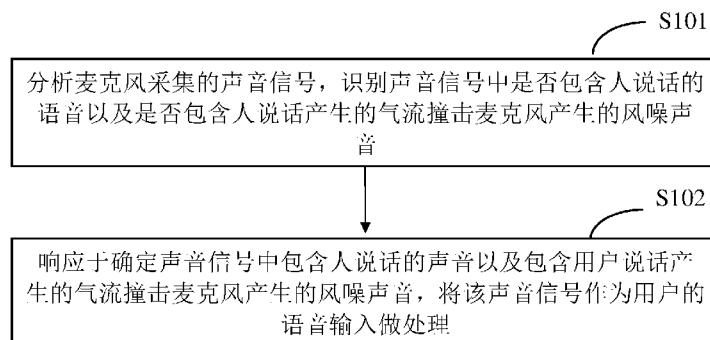


图 1

- S101 Analyze a sound signal collected by a microphone, and identify whether the sound signal contains a voice of a person speaking and whether the sound signal contains wind noise generated by the airflow generated by the person speaking hitting the microphone
- S102 In response to determining that the sound signal contains the voice of the person speaking and contains wind noise generated by the airflow generated by the user speaking hitting the microphone, use the sound signal as a voice input of a user and process same

(57) Abstract: A smart electronic device having a built-in microphone; the smart electronic portable device performs voice input-based interaction with a user by means of the following operations: processing a sound signal captured by a microphone to determine whether a voice signal is present in the sound signal; in response to confirming that a voice signal is present in the sound signal, further determining on the basis of the sound signal collected by the microphone whether the distance between the smart electronic device and the mouth of the user is less than a predetermined threshold; and in response to determining that the distance between the electronic device and the mouth of the user is less than the predetermined threshold, using the sound signal collected by the microphone as a voice input and processing same. The described interaction method is suitable for a user performing voice input while carrying a smart electronic



WO 2020/244355 A1

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))。

device, operation being natural and simple; in addition, the step of voice input is simplified, and the burden and difficulty of interaction are reduced so that interaction is more natural.

(57) 摘要: 一种内置有麦克风的智能电子设备, 所述智能电子便携设备如下操作与用户进行基于语音输入的交互: 处理麦克风捕获的声音信号判断声音信号中是否存在语音信号; 响应于确认声音信号中存在语音信号, 基于麦克风采集的声音信号进一步判断智能电子设备与用户的嘴部距离是否小于预定阈值; 响应于确定电子设备与用户嘴部距离小于预定阈值, 将麦克风采集的声音信号作为语音输入处理。该交互方法适用于用户在携带智能电子设备时进行语音输入, 操作自然且简单, 简化了语音输入的步骤, 降低交互负担和难度, 使得交互更加自然。

基于麦克风信号的语音交互唤醒电子设备、方法和介质

本申请要求于 2019 年 06 月 03 日提交中国专利局、申请号为 201910475949.X 发明名称为“基于麦克风信号的语音交互唤醒电子设备、方法和介质”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

本申请涉及语音输入领域，且更为具体地，涉及智能电子设备、语音输入触发方法。

10 背景技术

随着计算机技术的发展，语音识别算法日益成熟，语音输入因其在交互方式上的高自然性与有效性而正变得越来越重要。用户可以通过语音与移动设备（手机、手表等）进行交互，完成指令输入、信息查询、语音聊天等多种任务。

15 而在何时触发语音输入这一点上，现有的解决方案都有一些缺陷：

1. 物理按键触发

按下（或按住）移动设备的某个（或某些）物理按键后，激活语音输入。

20 该方案的缺点是：需要物理按键；容易误触发；需要用户按键。

2. 界面元素触发

点击（或按住）移动设备的屏幕上的界面元素（如图标），激活语音输入。

25 该方案的缺点是：需要设备具备屏幕；触发元素占用屏幕内容；受限于软件 UI 限制，可能导致触发方式繁琐；容易误触发。

3. 唤醒词（语音）检测

-2-

以某个特定词语（如产品昵称）为唤醒词，设备检测到对应的唤醒词后激活语音输入。

该方案的缺点是：隐私性和社会性较差；交互效率较低。

5 发明内容

鉴于上述情况，提出了本申请：

根据本申请的一个方面，一种内置有麦克风的智能电子设备，所述智能电子便携设备如下操作与用户进行基于语音输入的交互：处理麦克风捕获的声音信号判断声音信号中是否存在语音信号；响应于确认声音信号中存在语音信号，基于麦克风采集的声音信号进一步判断智能电子设备与用户的嘴部距离是否小于预定阈值；响应于确定电子设备与用户嘴部距离小于预定阈值，将麦克风采集的声音信号作为语音输入处理。

优选的，预定阈值为 3 厘米。

优选的，预定阈值为 1 厘米。

15 优选的，电子设备的麦克风处还有接近光传感器，通过接近光传感器判断是否有物体接近电子设备。

优选的，电子设备的麦克风处还有距离传感器，通过距离传感器直接测量电子设备与用户嘴部的距离。

20 优选的，通过麦克风收集的声音信号特征来判断智能电子设备与用户的嘴部距离是否小于预定阈值。

优选的，所述语音信号包括下面各项之一或者组合：用户以正常音量说话发出的声音；用户低声说话发出的声音；用户声带不发声说话产生的声音。

25 优选的，电子设备还可操作来：响应于确定用户正在近距离对着电子设备说话；判断用户在以如下方式中的一种在发声，包括：用户以正常音量说话的声音；用户以小音量说话的声音；用户以声带不发声方式说话发出的声音；以及根据判断的结果不同，对声音信号做不同的处理。

优选的，所述不同的处理为激活不同的应用程序处理语音输入。

优选的，判断中使用的特征包括音量、频谱特征，能量分布等。

—3—

优选的，判断智能电子设备与用户的嘴部距离是否小于预定阈值时使用的特征包括声音信号中的时域特征和频域特征，包括音量、频谱能量。

5 优选的，所述判断智能电子设备与用户的嘴部距离是否小于预定阈值包括：从麦克风采集到的声音信号通过滤波器提取语音信号；判断所述语音信号的能量是否超过一定阈值；响应于语音信号强度超过一定阈值，判断电子设备与用户嘴部距离小于预定阈值。

优选的，所述判断智能电子设备与用户的嘴部距离是否小于预定阈值包括：利用深度神经网络模型处理麦克风采集的数据，判断智能电子设备与用户的嘴部距离是否小于预定阈值。

10 优选的，所述判断智能电子设备与用户的嘴部距离是否小于预定阈值包括：记录用户在未做语音输入时的语音信号；将麦克风当前采集的语音信号与未做语音输入时的语音信号作比较；如果判断麦克风当前采集的语音信号音量超过未做语音输入时的语音信号的音量一定阈值，判断智能电子设备与用户的嘴部距离小于预定阈值。

15 优选的，将声音信号作为用户的语音输入所做的处理包括以下一种或多种：将声音信号存储到电子设备上的可存储介质；将声音信号通过互联网发送出去；将声音信号中的语音信号识别为文字，存储到电子设备上的可存储介质；将声音信号中的语音信号识别为文字，通过互联网发送出去；将声音信号中的语音信号识别为文字，理解用户的语音指令，执行相应操作。

20 优选的，电子设备还通过声纹分析识别特定用户，只对包含特定用户语音的声音信号做处理。

优选的，电子设备为智能手机、智能手表、智能戒指等。

25 此处的移动设备包括但不限于手机、头戴式显示器、手表，以及智能戒指、腕表等更小型的智能穿戴设备。

根据本申请的另一方面，提供了一种由配置有麦克风的智能电子设备执行的语音输入触发方法，包括如下操作：用于智能电子设备与用户进行基于语音输入的交互：处理麦克风捕获的声音信号判断声音信号中是否存在语音信号；响应于确认声音信号中存在语音信号，基于麦克风采集的声音

—4—

信号进一步判断智能电子设备与用户的嘴部距离是否小于预定阈值；响应于确定电子设备与用户嘴部距离小于预定阈值，将麦克风采集的声音信号作为语音输入处理。

根据本申请的另一方面，提供了一种计算机可读介质，其上存储有计算机可执行指令，计算机可执行指令被计算机执行时能够执行语音交互唤醒方法，所述语音交互唤醒方法包括：处理麦克风捕获的声音信号判断声音信号中是否存在语音信号；响应于确认声音信号中存在语音信号，基于麦克风采集的声音信号进一步判断智能电子设备与用户的嘴部距离是否小于预定阈值；响应于确定电子设备与用户嘴部距离小于预定阈值，将麦克风采集的声音信号作为语音输入处理。

根据本申请的一个方面，提供了一种配置有麦克风的电子设备，电子设备具有存储器和中央处理器，存储器上存储有计算机可执行指令，计算机可执行指令被中央处理器执行时能够执行如下操作：分析麦克风采集的声音信号，识别声音信号中是否包含人说话的语音以及是否包含人说话产生的气流撞击麦克风产生的风噪声音，响应于确定声音信号中包含人说话的声音以及包含用户说话产生的气流撞击麦克风产生的风噪声音，将该声音信号作为用户的语音输入做处理。

优选的，用户说话的语音包括：用户以正常音量说话的声音，用户以小音量说话的声音，用户以声带不发声方式说话发出的声音。

优选的，电子设备还可操作来：响应于确定用户正在近距离对着电子设备说话，判断用户在以如下方式中的一种在发声，包括：用户以正常音量说话的声音，用户以小音量说话的声音，用户以声带不发声方式说话发出的声音；根据判断的结果不同，对声音信号做不同的处理。

优选的，所述不同的处理为激活不同的应用程序处理语音输入。

优选的，判断使用的特征包括音量、频谱特征，能量分布等。

优选的，将声音信号作为用户的语音输入所做的处理包括以下一种或多种：将声音信号存储到电子设备上的可存储介质；将声音信号通过互联网发送出去；将声音信号中的语音信号识别为文字，存储到电子设备上的可存储介质；将声音信号中的语音信号识别为文字，通过互联网发送出去；

—5—

将声音信号中的语音信号识别为文字，理解用户的语音指令，执行相应操作。

优选的，电子设备还可操作来通过声纹分析识别特定用户，只对包含特定用户语音的声音信号做处理。

5 优选的，电子设备为智能手机、智能手表、智能戒指中的一种。

优选的，电子设备还可操作来：使用神经网络模型判断声音信号中是否包含用户说话的语音以及说话产生的气流撞击麦克风产生的风噪声。

10 优选的，电子设备还可操作来识别声音信号中是否包含人说话的语音以及是否包含人说话产生的气流撞击麦克风产生的风噪声。包括识别声音信号中是否包含用户说话的语音；响应于确定声音信号中包含用户说话的语音，识别语音中的音素，将语音信号表示为音素序列；针对音素序列中的每个音素，确定该音素是否为吐气音素，即：用户发声该音素时有气流从嘴中出来；将声音信号按照固定窗口长度切分为声音片段序列；利用频率特征，识别每个声音片段是否包含风噪声；将语音音素序列中的吐气音素和声音片段序列中识别为风噪声的片段做比较，同时将音素序列中的非吐气音素和声音片段序列中识别为风噪声的片段作比较，当吐气音素与风噪声片段重合度高于一定阈值，且非吐气音素与非风噪声片段重合度低于一定阈值时，判断该声音信号中包含用户说话产生的气流撞击麦克风产生的风噪声。

20 优选的，识别声音信号中是否包含人说话的语音以及是否包含人说话产生的气流撞击麦克风产生的风噪声。包括：识别声音信号中包含风噪声的声音特征；响应于确定声音信号中包含风噪声，识别声音信号包含语音信号；响应于确定声音信号中包含语音信号，识别语音信号对应的音素序列；针对声音信号中的风噪声特征，计算每一时刻的风噪声特征强度；针对音素序列中的每个音素，根据预先定义的数据模型获得该音素吐气的强度；通过
25 基于高斯混合贝叶斯模型分析风噪声特征与音素序列的一致性，重合度高于一定阈值时，判断该声音信号中包含用户说话产生的气流撞击麦克风产生的风噪声。

根据本申请的另一方面，提供了一种内置有多个麦克风的电子设备，电子设备具有存储器和中央处理器，存储器上存储有计算机可执行指令，

—6—

计算机可执行指令被中央处理器执行时能够执行如下操作：分析多个麦克风采集的声音信号；判断用户是否正在近距离对着电子设备说话；响应于确定用户正在近距离对着电子设备说话，将麦克风采集的声音信号作为用户的语音输入处理。

5 优选的，多个麦克风构成麦克风阵列系统。

优选的，所述判断用户是否正在近距离对着电子设备说话包括：利用到达阵列上各传声器的声音信号之间的时间差计算用户嘴部相对于麦克风阵列的位置；当用户嘴部距离电子设备的距离小于一定阈值时，确定用户正在近距离对着电子设备说话。

10 优选的，所述距离阈值为 10 厘米。

优选的，所述将该声音信号作为用户的语音输入做处理包括：根据说话人嘴部和电子设备之间距离的不同，对用户的语音输入做不同处理。

15 优选的，所述判断用户是否正在近距离对着电子设备说话包括：判断是否至少有一个麦克风采集的声音信号中包含用户说话的语音信号；响应于确定至少有一个麦克风采集的声音信号中包含用户说话的语音信号，从麦克风采集的声音信号中提取语音信号；判断从不同麦克风采集的声音信号中提取的语音信号的幅度差异是否超过预定阈值时；响应于确定幅度差值超过预定阈值，确认用户正在近距离对着电子设备说话。

20 优选的，电子设备还可操作来：定义多个麦克风中，语音信号幅度最大的麦克风为响应麦克风；根据响应麦克风的不同，对用户的语音输入做不同的处理。

优选的，所述判断用户是否正在近距离对着电子设备说话包括：利用提前训练的机器学习模型，处理多个麦克风的的声音信号，判断用户是否正在近距离对着电子设备说话。

25 优选的，用户说话的语音包括：用户以正常音量说话的声音；用户以小音量说话的声音；用户以声带不发声方式说话发出的声音。

优选的，用户说话的语音包括：用户以正常音量说话的声音，用户以小音量说话的声音，用户以声带不发声方式说话发出的声音。

优选的，电子设备还可操作来：响应于确定用户正在近距离对着电子

-7-

设备说话；判断用户在以如下方式中的一种在发声，包括：用户以正常音量说话的声音；用户以小音量说话的声音；用户以声带不发声方式说话发出的声音；以及根据判断的结果不同，对声音信号做不同的处理。

优选的，所述不同的处理为激活不同的应用程序处理语音输入。

5 优选的，判断的特征包括音量、频谱特征，能量分布等。

优选的，将声音信号作为用户的语音输入所做的处理包括以下一种或多种：将声音信号存储到电子设备上的可存储介质；将声音信号通过互联网发送出去；将声音信号中的语音信号识别为文字，存储到电子设备上的可存储介质；将声音信号中的语音信号识别为文字，通过互联网发送出去；

10 将声音信号中的语音信号识别为文字，理解用户的语音指令，执行相应操作。

优选的，电子设备还可操作来通过声纹分析识别特定用户，只对包含特定用户语音的声音信号做处理。

15 优选的，电子设备为智能手机、智能手表、智能戒指、平板电脑中的一种。

根据本申请的另一方面，一种内置有麦克风的电子设备，电子设备具有存储器和中央处理器，存储器上存储有计算机可执行指令，计算机可执行指令被中央处理器执行时能够执行如下操作：判断麦克风采集的声音信号中是否包含语音信号；响应于确认麦克风采集的声音信号中包含语音信号，判断用户是否在做低声说话，即以低于正常音量的方式说话；响应于确定用户正在做低声说话，无需任何唤醒操作地将声音信号作为语音输入处理。

优选的，所述低声说话包括声带不发声的低声说话和声带发声的低声说话两种方式。

25 优选的，电子设备还操作来：响应于确定用户在做低声说话；判断用户在做声带不发声的低声说话还是在做声带发声的低声说话；根据判断的结果不同，对声音信号做不同的处理。

优选的，不同的处理为激活不同的应用程序来响应语音输入。

优选的，判断用户是否在做低声说话时使用的信号特征包括音量、频

谱特征，能量分布。

优选的，判断用户在做声带不发声的低声说话或者在做声带发声的低声说话时使用的信号特征包括音量、频谱特征，能量分布。

- 5 优选的，所述判断用户是否在做低声说话包括：利用机器学习模型，处理麦克风采集的声音信号，判断用户是否在低声说话。

优选的，所述机器学习模型为卷积神经网络模型或者循环神经网络模型。

- 10 优选的，所述判断用户在做声带不发声的低声说话还是在做声带发声的低声说话包括：利用机器学习模型，处理麦克风采集的声音信号，判断用户在做声带不发声的低声说话或者在做声带发声的低声说话。

优选的，所述机器学习模型为卷积神经网络模型或者循环神经网络模型。

- 15 优选的，将声音信号作为用户的语音输入所做的处理包括以下一种或多种：将声音信号存储到电子设备上的可存储介质；将声音信号通过互联网发送出去；将声音信号中的语音信号识别为文字，存储到电子设备上的可存储介质；将声音信号中的语音信号识别为文字，通过互联网发送出去；将声音信号中的语音信号识别为文字，理解用户的语音指令，执行相应操作。

- 20 优选的，通过声纹分析识别特定用户，只对包含特定用户语音的声音信号做处理。

优选的，电子设备为智能手机、智能手表、智能戒指等。

本方案优势：

- 25 1. 交互更加自然。将设备放在嘴前即触发语音输入，符合用户习惯与认知。

2. 使用效率更高。单手即可使用。无需在不同的用户界面/应用之间切换，也不需按住某个按键，直接抬起手到嘴边就能使用。

3. 收音质量高。设备的录音机在用户嘴边，收取的语音输入信号清晰，受环境音的影响较小。

4. 高隐私性与社会性。设备在嘴前，用户只需发出相对较小的声音即可完成高质量的语音输入，对他人的干扰较小，同时用户姿势可以包括捂嘴等，具有较好的隐私保护。

5 附图说明

图 1 是根据本申请实施例的语音输入交互方法的示意性流程图；

图 2 示出了根据本申请的另一实施例的配置有多个麦克风的电子设备使用多个麦克风接收的声音信号的差别的语音输入触发方法的总体流程图；

10 图 3 示出了根据本申请实施例的内置有麦克风的电子设备基于低声说话方式识别的语音输入触发方法的总体流程图；

图 4 描述基于麦克风的语音信号的距离判断的语音输入触发方法的总体流程图；

15 图 5 是根据本申请实施例的触发姿势中的将手机上端麦克风贴近嘴部的正面示意图；

图 6 是根据本申请实施例的触发姿势中的将手机上端麦克风贴近嘴部的侧面示意图；

图 7 是根据本申请实施例的触发姿势中的将手机下端麦克风贴近嘴部的示意图；

20 图 8 是根据本申请实施例的触发姿势中的将智能手表麦克风贴近嘴部的示意图。

具体实施方式

25 为了使本领域技术人员更好地理解本申请，下面结合附图和具体实施方式对本申请作进一步详细说明。

本公开针对智能电子设备的语音输入触发，基于配置的麦克风捕捉的声音内在特征，来确定是否触发语音输入应用，其中无需传统的物理按键

-10-

触发、界面元素触发、唤醒词检测，交互更加自然。将设备放在嘴前即触发语音输入，符合用户习惯与认知。

下面将从以下几个方面来继续本公开：1、基于人类说话时风噪声特征的语音输入触发，具体地，通过识别人说话时候的语音和风噪声来直接启动语音输入并将接收的声音信号作为语音输入处理；2、基于多个麦克风接收的声音信号的差别的语音输入触发；3、基于低声说话方式识别的语音输入触发；4、基于麦克风的距离判断的语音输入触发。。

一、基于人类说话时风噪声特征的语音输入触发

10 当用户近距离对着麦克风说话时，即使声音很小或者不触发声带发声，麦克风采集到的声音信号中包含两种声音成分，一是人声带后者口腔震动发出的声音，二是人说话产生的气流撞击麦克风产生的风噪声。可以基于这个特性来触发电子设备的语音输入应用。

15 图 1 示出了根据本申请实施例的语音输入交互方法 100 的示意性流程图。

在步骤 S101 中，分析麦克风采集的声音信号，识别声音信号中是否包含人说话的语音以及是否包含人说话产生的气流撞击麦克风产生的风噪声，

20 在步骤 S102 中，响应于确定声音信号中包含人说话的声音以及包含用户说话产生的气流撞击麦克风产生的风噪声，将该声音信号作为用户的语音输入做处理。

本申请实施例的语音输入交互方法特别适合于在隐私要求比较高的情况下，不用声带发声地进行语音输入。

25 这里用户说话的语音可以包括：用户以正常音量说话的声音、用户以小音量说话的声音、用户以声带不发声方式说话发出的声音。

在一个示例中，可以识别上述不同的说话方式，根据识别结果产生不同的反馈，比如正常说话就是控制手机的语音助理，低声说话就是控制微信，声带不发声说话就是做语音转录笔记。

作为示例，将声音信号作为用户的语音输入所做的处理包括以下一种

或多种:

将声音信号存储到电子设备上的可存储介质;

将声音信号通过互联网发送出去;

将声音信号中的语音信号识别为文字, 存储到电子设备上的可存储介

5 质;

将声音信号中的语音信号识别为文字, 通过互联网发送出去;

将声音信号中的语音信号识别为文字, 理解用户的语音指令, 执行相应操作。

10 在一个示例中, 还包括通过声纹分析识别特定用户, 只对包含特定用户语音的声音信号做处理。

在一个示例中, 电子设备为智能手机、智能手表、智能戒指中的一种。

在一个示例中, 使用神经网络模型判断声音信号中是否包含用户说话的语音以及说话产生的气流撞击麦克风产生的风噪声音。此仅为示例, 可以使用其他机器学习算法。

15 在一个示例中, 识别声音信号中是否包含人说话的语音以及是否包含人说话产生的气流撞击麦克风产生的风噪声音包括:

识别声音信号中是否包含用户说话的语音;

响应于确定声音信号中包含用户说话的语音, 识别语音中的音素, 将语音信号表示为音素序列;

20 针对音素序列中的每个音素, 确定该音素是否为吐气音素, 即: 用户发声该音素时有气流从嘴中出来;

将声音信号按照固定窗口长度切分为声音片段序列;

利用频率特征, 识别每个声音片段是否包含风噪声;

25 将语音音素序列中的吐气音素和声音片段序列中识别为风噪声的片段做比较, 同时将音素序列中的非吐气音素和风噪声片段作比较, 当吐气音素与风噪声片段重合度高于一定阈值, 且非吐气音素与非风噪声片段重合度低于一定阈值时, 判断该声音信号中包含用户说话产生的气流撞击麦克风产生的风噪声音。

在一个示例中, 识别声音信号中是否包含人说话的语音以及是否包含

-12-

人说话产生的气流撞击麦克风产生的风噪声音包括:

识别声音信号中包含风噪的声音特征;

响应于确定声音信号中包含风噪声, 识别声音信号包含语音信号;

响应于确定声音信号中包含语音信号, 识别语音信号对应的音素序列;

5 针对声音信号中的风噪特征, 计算每一时刻的风噪特征强度;

针对音素序列中的每个音素, 根据预先定义的数据模型获得该音素吐气的强度;

通过基于高斯混合贝叶斯模型分析风噪特征与音素序列的一致性, 重合度高于一定阈值时, 判断该声音信号中包含用户说话产生的气流撞击麦克风产生的风噪声音。
10

二、基于多个麦克风接收的声音信号的差别的语音输入触发

图 2 示出了根据本申请的另一实施例的配置有多个麦克风的电子设备使用多个麦克风接收的声音信号的差别的语音输入触发方法的总体流程图。
15

电子设备例如手机内置有多个麦克风的电子设备, 电子设备具有存储器和中央处理器, 存储器上存储有计算机可执行指令, 计算机可执行指令被中央处理器执行时能够执行本实施例的语音输入触发方法。

20 如图 2 所示, 在步骤 S201 中, 分析多个麦克风采集的声音信号。

在一个示例中, 多个麦克风包括至少 3 个麦克风, 构成麦克风阵列系统, 通过声音信号达到各个麦克风的时间差可以估计声源相对于智能设备的空间位置。

这里的声音信号包括例如声音信号的幅度、频率等等。

25 在步骤 S202 中, 基于多个麦克风采集的声音信号, 判断用户是否正在近距离对着电子设备说话。

在一个示例中, 判断用户是否正在近距离对着电子设备说话包括:

利用到达阵列上各传声器的声音信号之间的时间差计算用户嘴部相对于麦克风阵列的位置,

-13-

当用户嘴部距离电子设备的距离小于一定阈值时，确定用户正在近距离对着电子设备说话。

在一个示例中，所述距离阈值为 10 厘米。

在步骤 S203 中，响应于确定用户正在近距离对着电子设备说话，将麦克风采集的声音信号作为用户的语音输入处理。

在一个示例中，将该声音信号作为用户的语音输入做处理包括：

根据说话人嘴部和电子设备之间距离的不同，对用户的语音输入做不同处理。例如，当距离为 0-3cm 时，激活语音助手响应用户的语音输入；当距离为 3-10cm 时，激活微信应用程序响应用户的语音输入，将语音信息发送给好友；

在一个示例中，判断用户是否正在近距离对着电子设备说话包括：

判断是否至少有一个麦克风采集的声音信号中包含用户说话的语音信号，

响应于确定至少有一个麦克风采集的声音信号中包含用户说话的语音信号，从麦克风采集的声音信号中提取语音信号，

判断从不同麦克风采集的声音信号中提取的语音信号的幅度差异是否超过预定阈值时，

响应于确定幅度差值超过预定阈值，确认用户正在近距离对着电子设备说话。

在上面的例子中，还可以包括：

定义多个麦克风中语音信号幅度最大的麦克风为响应麦克风，

根据响应麦克风的的不同，对用户的语音输入做不同的处理。例如，当响应麦克风是智能手机底部的麦克风时，激活智能手机上的语音助手；当响应麦克风是智能手机顶部的麦克风时，激活录音机功能将用户的语音记录到存储设备；

在一个示例中，判断用户是否正在近距离对着电子设备说话包括：利用提前训练的机器学习模型，处理多个麦克风的的声音信号，判断用户是否正在近距离对着电子设备说话。一般地，准备训练样本数据，然后利用训练样本数据来训练选定的机器学习模型，在实际应用时（有时也叫测试），

-14-

将多个麦克风捕获的声音信号（作为测试样本）输入机器学习模型，得到的输出表示用户是否正在近距离对着电子设备说话。作为示例，机器学习模型例如为深度学习神经网络、支持向量机、决策树等。

- 5 在一个示例中，用户说话的语音包括：用户以正常音量说话的声音，用户以小音量说话的声音，用户以声带不发声方式说话发出的声音。

- 10 在一个示例中，将声音信号作为用户的语音输入所做的处理包括以下一种或多种：将声音信号存储到电子设备上的可存储介质；将声音信号通过互联网发送出去；将声音信号中的语音信号识别为文字，存储到电子设备上的可存储介质；将声音信号中的语音信号识别为文字，通过互联网发送出去；将声音信号中的语音信号识别为文字，理解用户的语音指令，执行相应操作。

在一个示例，还包括通过声纹分析识别特定用户，只对包含特定用户语音的声音信号做处理。

- 15 作为示例，电子设备为智能手机、智能手表、智能戒指、平板电脑等。本实施例利用内置的不同麦克风之间声音信号的差别来识别用户是否近距离对着电子设备说话，进而决定是否启动语音输入，具有识别可靠，计算方法简单等优点。

三、基于低声说话方式识别的语音输入触发

- 20 低声说话是指说话音量小于正常说话（比如与他人正常交谈）音量的方式。低声说话包括两种方式。一种是声带不震动的低声说话（俗称悄悄话），另一种是声带发生震动的低声说话。在声带不震动的低声说话方式下，产生的声音主要包含空气通过喉部、嘴部发出的声音以及嘴内舌头牙齿发出的声音。在声带震动的低声说话方式下，发出的声音除了包含声带不震动的低声说话方式下产生的声音，还包括声带震动产生的声音。但相比于正常音量的说话方式，声带震动的低声说话过程中，声带震动程度较小，产生的声带震动声音较小。声带不震动低声说话产生的声音和声带震动产生的声音的频率范围不同，可以区分。声带震动低声说话和声带震动的正常音量说话可以通过音量阈值来区分，具体的阈值可以提前设定，也可以
- 25

由用户来设定。

示例方法：对麦克风采集的语音信号做滤波处理，提取两部分信号，分别为声带震动产生的声音成分 V1 和空气通过喉部、嘴部发出的以及嘴内舌头牙齿发出的声音 V2。当 V1 和 V2 的能量比值小于一定阈值时，判定用户在做低声说话。

一般情况下，低声说话只有当用户距离麦克风比较近的时候才能检测，比如距离小于 30 厘米时。而定义近距离情况下的低声说话作为语音输入，对用户而言是一种易于学习和理解和方便操作的交互方式，可以免除显式的唤醒操作，比如按压特定的唤醒按键或者是通过语音唤醒词。且这种方式在绝大多数的实际使用情况下，不会被误触发。

图 3 示出了根据本申请实施例的配备有麦克风的电子设备基于低声说话方式识别的语音输入触发方法的总体流程图。配备有麦克风的电子设备具有存储器和中央处理器，存储器上存储有计算机可执行指令，计算机可执行指令被中央处理器执行时能够执行根据本申请实施例的语音输入触发方法。

如图 3 所示，在步骤 S301 中，判断麦克风采集的声音信号中是否包含语音信号。

在步骤 S302 中，响应于确认麦克风采集的声音信号中包含语音信号，判断用户是否在做低声说话，即以低于正常音量的方式说话。

在步骤 S303 中，响应于确定用户正在做低声说话，无需任何唤醒操作地将声音信号作为语音输入处理。

低声说话可以包括声带不发声的低声说话和声带发声的低声说话两种方式。

在一个示例中，语音输入触发方法还可以包括：响应于确定用户在做低声说话，判断用户在做声带不发声的低声说话还是在做声带发声的低声说话，根据判断的结果不同，对声音信号做不同的处理。

作为示例，不同的处理为将语音输入交给不同的应用程序来处理。比如正常说话就是控制手机的语音助理，低声说话就是控制微信，声带不发声说话就是做语音转录笔记。

作为示例，判断用户是否在做低声说话时使用的信号特征可以包括音量、频谱特征，能量分布等。

作为示例，判断用户在做声带不发声的低声说话或者在做声带发声的低声说话时使用的信号特征包括音量、频谱特征，能量分布等。

- 5 作为示例，判断用户是否在做低声说话可以包括：利用机器学习模型，处理麦克风采集的声音信号，判断用户是否在低声说话。

作为示例，机器学习模型可以为卷积神经网络模型或者循环神经网络模型。

- 10 作为示例，判断用户在做声带不发声的低声说话还是在做声带发声的低声说话包括：利用机器学习模型，处理麦克风采集的声音信号，判断用户在做声带不发声的低声说话或者在做声带发声的低声说话。

作为示例，将声音信号作为用户的语音输入所做的处理包括以下一种或多种：

- 15 将声音信号存储到电子设备上的可存储介质；
将声音信号通过互联网发送出去；
将声音信号中的语音信号识别为文字，存储到电子设备上的可存储介质；
将声音信号中的语音信号识别为文字，通过互联网发送出去；
将声音信号中的语音信号识别为文字，理解用户的语音指令，执行相应操作。
- 20

作为示例，语音输入触发方法还可以包括：通过声纹分析识别特定用户，只对包含特定用户语音的声音信号做处理。

作为示例，电子设备可以为智能手机、智能手表、智能戒指等。

- 25 有关低声说话模式以及检测方法，作为示例，可以参考下述参考文献：

Zhang, Chi, and John HL Hansen. "Analysis and classification of speech mode: whispered through shouted." Eighth Annual Conference of the International Speech Communication Association. 2007.

Meenakshi, G. Nisha, and Prasanta Kumar Ghosh. "Robust whisper activity detection using long-term log energy variation of sub-band signal." IEEE Signal Processing Letters 22.11 (2015): 1859-1863.

5 四、基于麦克风的语音信号的距离判断的语音输入触发

下面结合图 4 描述基于麦克风的语音信号的距离判断的语音输入触发方法的总体流程图。

如图 4 所示，在步骤 401 中，处理麦克风捕获的声音信号判断声音信号中是否存在语音信号。

10 在步骤 402 中，响应于确认声音信号中存在语音信号，基于麦克风采集的声音信号进一步判断智能电子设备与用户的嘴部距离是否小于预定阈值。

在步骤 403 中，响应于确定电子设备与用户嘴部距离小于预定阈值，将麦克风采集的声音信号作为语音输入处理。

15 在一个示例中，预定阈值为 10 厘米。

语音信号可以包括下面各项之一或者组合：用户以正常音量说话发出的声音；用户低声说话发出的声音；用户声带不发声说话产生的声音。

20 在一个示例中，判断智能电子设备与用户的嘴部距离是否小于预定阈值时使用的特征包括声音信号中的时域特征和频域特征，包括音量、频谱能量。

在一个示例中，所述判断智能电子设备与用户的嘴部距离是否小于预定阈值包括：利用深度神经网络模型处理麦克风采集的数据，判断智能电子设备与用户的嘴部距离是否小于预定阈值。

25 在一个示例中，所述判断智能电子设备与用户的嘴部距离是否小于预定阈值包括：记录用户在做语音输入时的语音信号，将麦克风当前采集的语音信号与未做语音输入时的语音信号作比较，如果判断麦克风当前采集的语音信号音量超过未做语音输入时的语音信号的音量一定阈值，判断智能电子设备与用户的嘴部距离小于预定阈值。

在一个示例中，将声音信号作为用户的语音输入所做的处理包括以

下一种或多种：将声音信号存储到电子设备上的可存储介质；将声音信号通过互联网发送出去；将声音信号中的语音信号识别为文字，存储到电子设备上的可存储介质；将声音信号中的语音信号识别为文字，通过互联网发送出去；将声音信号中的语音信号识别为文字，理解用户的语音指令，
5 执行相应操作。

在一个示例中，语音输入触发还包括通过声纹分析识别特定用户，只对包含特定用户语音的声音信号做处理。

在一个示例中，电子设备为智能手机、智能手表、智能戒指等。

图 5 至图 8 显示了几例用户将智能电子便携设备的麦克风放置到嘴边
10 较近距离的位置，此时用户发出的语音将作为语音输入。其中，图 5 与图 6 分别是手机上端有麦克风的情况，在这种情况下，用户有语音交互意图时，可以将手机的麦克风移动到嘴边 0~10 厘米处，直接说话即可作为语音输入。图 7 是手机下端有麦克风的情况，与前述上端有麦克风相类似，两种姿势不是互斥的，如果手机上下端均有麦克风则任意一种姿势均可以实
15 施交互方案。图 8 是对应设备为智能手表时的情况，与上述设备为手机的情况类似。上述对触发姿势的说明是示例性的，并非穷尽性的，并且也不限于所披露的各种设备和麦克风情况。

作为使用单个麦克风来接收声音输入并触发语音输入的一个具体实施例子，可以首先通过分析单麦克风接收到的声音输入，判断是否为语音，
20 并通过分析语音的近距离特有的特征，如麦克风爆破音、近场风噪、吹气声、能量、频谱特征、时域特征等，判断电子设备自身与用户的嘴的距离是否小于给定阈值，以及通过声纹识别判断语音输入来源是否属于可服务用户，结合以上几点来判断是否将麦克风信号作为语音输入。

作为使用双麦克风来接收声音输入并触发语音输入的一个具体实施例子，通过分析双麦克风输入信号的特征差异，如能量特征、频谱特征，判断发声位置是否贴近其中一个麦克风，通过双麦克风的信号差异从而屏蔽环境噪音、分离语音到对应的单声道，然后通过上述单麦克风的特征分析方法，判断电子设备自身与用户的嘴的距离小于给定阈值，以及通过声纹识别判断语音输入来源是否属于可服务用户，结合以上几点来判断是否将
25

信号作为语音输入。

作为使用多麦克风阵列来接收声音输入并触发语音输入的一个具体实施例子，通过比较分析不同麦克风接收到的声音输入的信号的差异，通过从环境中分离近场语音信号，识别与检测声音信号是否包括语音，通过多
5 麦克风阵列的声源定位技术判断语音信号的用户嘴的位置与设备之间的距离是否小于预定阈值，以及通过声纹识别判断语音输入来源是否属于可服务用户，结合以上几点来判断是否将信号作为语音输入。

在一个示例中，在智能电子便携设备通过分析语音信号，检测到发音位置位于自身附近，也即移动设备位于用户嘴部较近位置，智能电子便携
10 设备便将声音信号作为语音输入，根据任务与上下文的不同，再结合自然语言处理技术理解用户的语音输入并完成相应的任务。

麦克风并不局限于前述示例，而是可以包括下面各项之一或者其组合：设备内置单麦克风；设备内置双麦克风；设备内置多麦克风阵列；外接无线麦克风；以及外接有线麦克风。

15 如前所述，智能电子便携设备可以为手机，装备有双耳蓝牙耳机、带有麦克风的有线耳机或者其他麦克风传感器。

智能电子便携设备可以为手表，以及智能戒指、腕表中的一种智能穿戴设备。

20 智能电子便携设备为头戴式智能显示设备，装备有麦克风或者多麦克风组。

在一个示例中，在电子设备激活语音输入应用后，可以做出反馈输出，反馈输出包括震动、语音、图像中的一种或者其组合。

本申请各个实施例的方案可以提供下述一种或几种优势：

25 1. 交互更加自然。将设备放在嘴前即触发语音输入，符合用户习惯与认知。

2. 使用效率更高。单手即可使用。无需在不同的用户界面/应用之间切换，也不需按住某个按键，直接抬起手到嘴边就能使用。

3. 收音质量高。设备的录音机在用户嘴边，收取的语音输入信号清晰，受环境音的影响较小。

—20—

4. 高隐私性与社会性。设备在嘴前，用户只需发出相对较小的声音即可完成高质量的语音输入，对他人的干扰较小，同时用户姿势可以包括捂嘴等，具有较好的隐私保护。

5 以上已经描述了本申请的各实施例，上述说明是示例性的，并非穷尽性的，并且也不限于所披露的各实施例。在不偏离所说明的各实施例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。因此，本申请的保护范围应该以权利要求的保护范围为准。

-21-

权 利 要 求

1、一种配置有麦克风的智能电子设备，其特征在于，所述智能电子便携设备如下操作与用户进行基于语音输入的交互：

处理麦克风捕获的声音信号判断声音信号中是否存在语音信号；

5 响应于确认声音信号中存在语音信号，基于麦克风采集的声音信号进一步判断智能电子设备与用户的嘴部距离是否小于预定阈值；以及

响应于确定电子设备与用户嘴部距离小于预定阈值，将麦克风采集的声音信号作为语音输入处理。

2、根据权利要求1的智能电子设备，其特征在于，预定阈值为3厘米。

10 3、根据权利要求1的智能电子设备，其特征在于，预定阈值为1厘米。

4、根据权利要求1的智能电子设备，其特征在于，电子设备的麦克风处还有接近光传感器，通过接近光传感器判断是否有物体接近电子设备。

5、根据权利要求1的智能电子设备，其特征在于，电子设备的麦克风处还有距离传感器，通过距离传感器直接测量电子设备与用户嘴部的距离。

15 6、根据权利要求1的智能电子设备，其特征在于，通过麦克风收集的声音信号特征来判断智能电子设备与用户的嘴部距离是否小于预定阈值。

7、根据权利要求1的智能电子设备，其特征在于，所述语音信号包括下面各项之一或者组合：

用户以正常音量说话发出的声音；

20 用户低声说话发出的声音；

用户声带不发声说话产生的声音。

8、根据权利要求1的智能电子设备，其特征在于，还包括：

响应于确定用户正在近距离对着电子设备说话，

判断用户在以如下方式中的一种在发声，包括：

25 用户以正常音量说话的声音，

用户以小音量说话的声音，

用户以声带不发声方式说话发出的声音；以及

根据判断的结果不同，对声音信号做不同的处理。

9、根据权利要求8的智能电子设备，其特征在于，所述不同的处理为

激活不同的应用程序处理语音输入。

10、根据权利要求 8 的智能电子设备，其特征在于，判断中使用的特征包括音量、频谱特征，能量分布。

5 11、根据权利要求 1 的智能电子设备，其特征在于，判断智能电子设备与用户的嘴部距离是否小于预定阈值时使用的特征包括声音信号中的时域特征和频域特征，包括音量、频谱能量。

12、根据权利要求 11 的智能电子设备，其特征在于，所述判断智能电子设备与用户的嘴部距离是否小于预定阈值包括：

10 从麦克风采集到的声音信号通过滤波器提取语音信号，
判断所述语音信号的能量是否超过一定阈值，
响应于语音信号强度超过一定阈值时，
判断电子设备与用户嘴部距离小于预定阈值。

13、根据权利要求 1 的智能电子设备，其特征在于，所述判断智能电子设备与用户的嘴部距离是否小于预定阈值包括：

15 利用深度神经网络模型处理麦克风采集的数据，判断智能电子设备与用户的嘴部距离是否小于预定阈值。

14、根据权利要求 1 的智能电子设备，其特征在于，所述判断智能电子设备与用户的嘴部距离是否小于预定阈值包括：

20 记录用户在未做语音输入时的语音信号，
将麦克风当前采集的语音信号与未做语音输入时的语音信号作比较，
如果判断麦克风当前采集的语音信号音量超过未做语音输入时的语音信号的音量一定阈值，判断智能电子设备与用户的嘴部距离小于预定阈值。

15、根据权利要求 1 的智能电子设备，其特征在于，将声音信号作为用户的语音输入所做的处理包括以下一种或多种：

25 将声音信号存储到电子设备上的可存储介质；
将声音信号通过互联网发送出去；
将声音信号中的语音信号识别为文字，存储到电子设备上的可存储介质；
将声音信号中的语音信号识别为文字，通过互联网发送出去；

-23-

将声音信号中的语音信号识别为文字，理解用户的语音指令，执行相应操作。

16、根据权利要求 1 的智能电子设备，其特征在于，还包括通过声纹分析识别特定用户，只对包含特定用户语音的声音信号做处理。

5 17、根据权利要求 1 的智能电子设备，其特征在于，电子设备为智能手机、智能手表或智能戒指。

18、一种由配置有麦克风的智能电子设备执行的语音交互唤醒方法，其特征在于，包括如下操作用于智能电子设备与用户进行基于语音输入的交互：

10 处理麦克风捕获的声音信号判断声音信号中是否存在语音信号，
响应于确认声音信号中存在语音信号，基于麦克风采集的声音信号进一步判断智能电子设备与用户的嘴部距离是否小于预定阈值；

响应于确定电子设备与用户嘴部距离小于预定阈值，将麦克风采集的声音信号作为语音输入处理。

15 19、一种计算机可读介质，其特征在于，其上存储有计算机可执行指令，计算机可执行指令被计算机执行时能够执行语音交互唤醒方法，所述语音交互唤醒方法包括：

处理麦克风捕获的声音信号判断声音信号中是否存在语音信号，

20 响应于确认声音信号中存在语音信号，基于麦克风采集的声音信号进一步判断智能电子设备与用户的嘴部距离是否小于预定阈值；

响应于确定电子设备与用户嘴部距离小于预定阈值，将麦克风采集的声音信号作为语音输入处理。

— 1/7 —

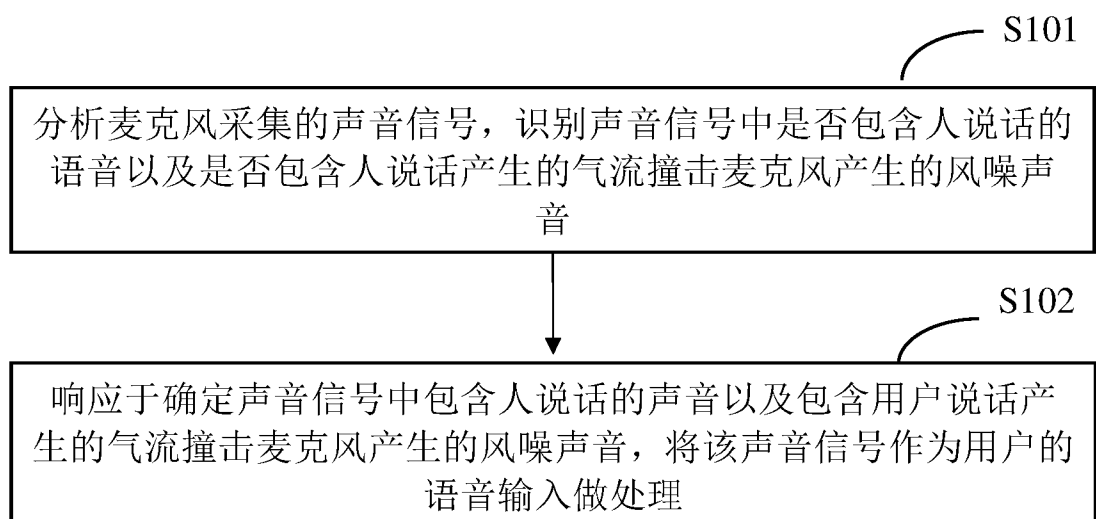


图 1

—2/7—

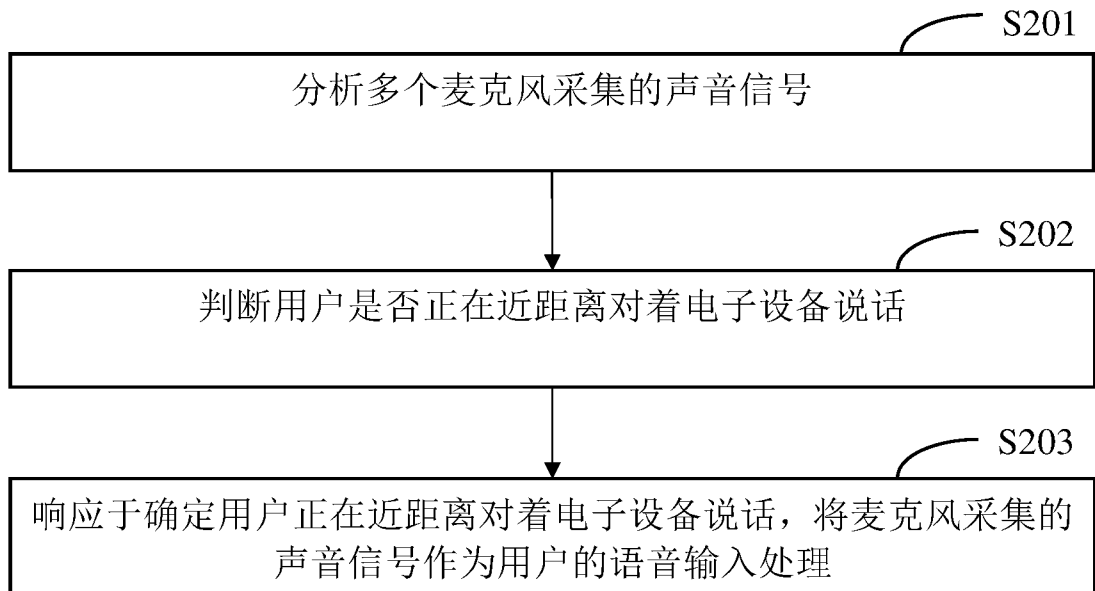


图 2

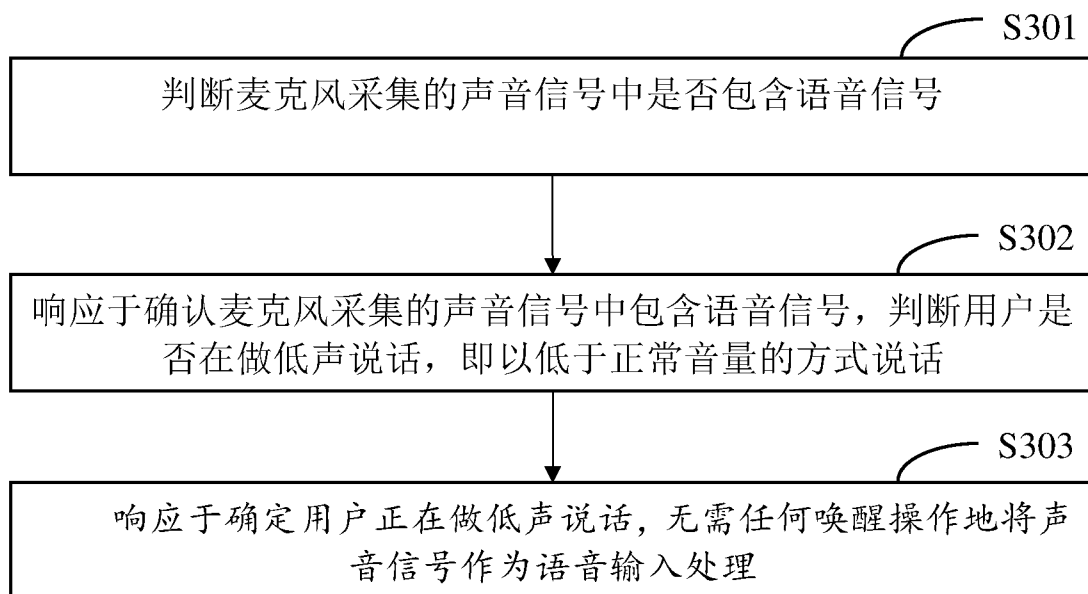


图 3

— 3/7 —

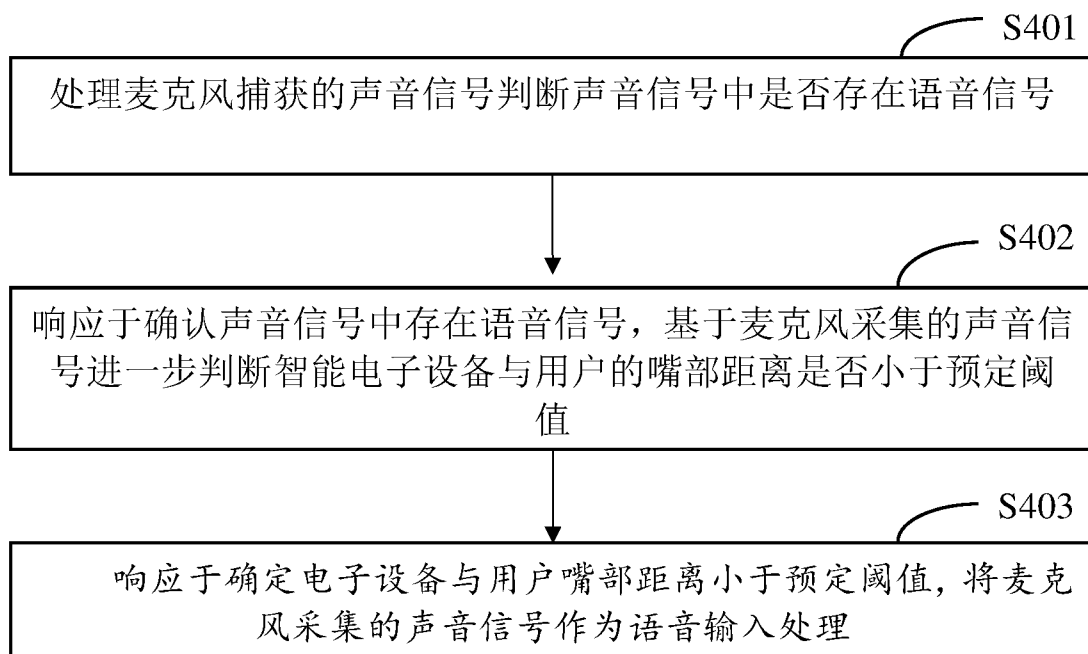


图 4



图 5



图 6



图 7

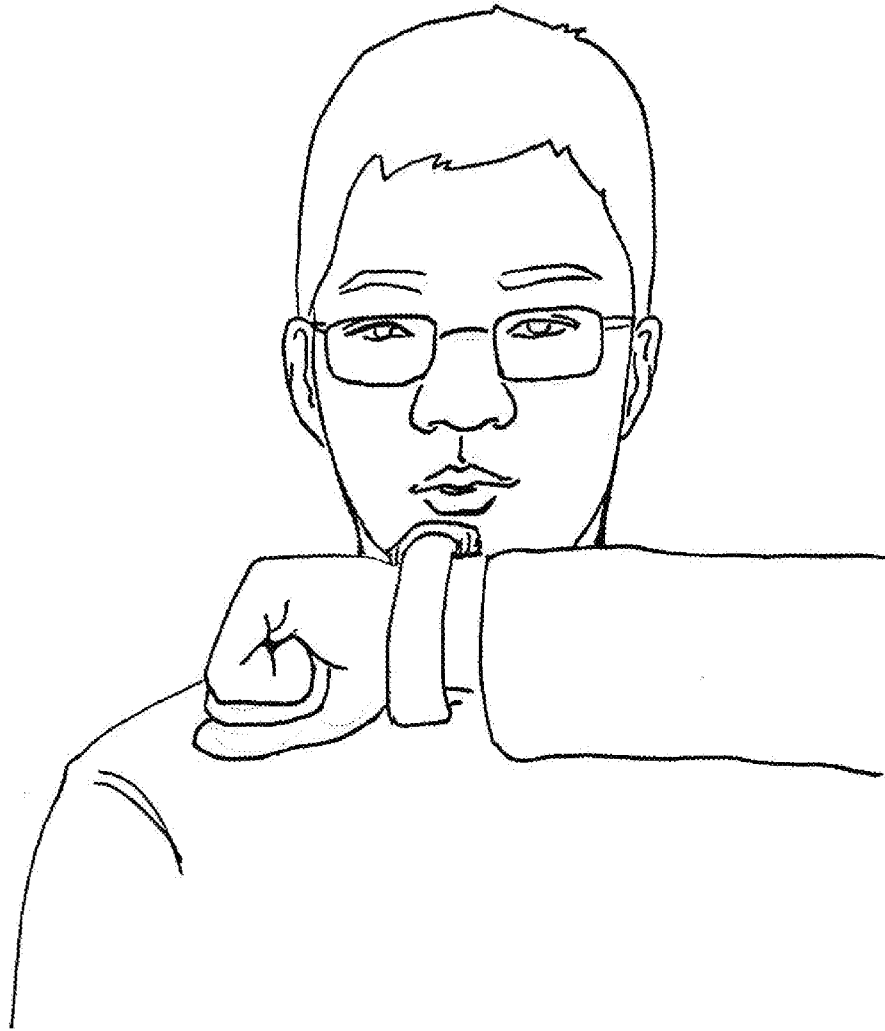


图 8

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/089551

A. CLASSIFICATION OF SUBJECT MATTER

G10L 25/78(2013.01)i; G10L 15/22(2006.01)i; G10L 25/51(2013.01)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)

G10L:H04M1/725

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)

CNKI, CNPAT, WPI, EPODOC: 声音, 语音, 音频, 低声, 轻声, 低语, 小声, 悄声, 悄悄话, 声带, 震动, 吹气声, 耳语, 近距离, 距离近, 贴近, 接近, 靠近, 唇, 嘴, 脸, 检测, 距离, 捂嘴, 接近传感器, whispered, whisper, silent, voice, nasal air, flow, airflow, ingressive, utterance, shouted, speech mode, activity detection, proximity sensor, low voice, near, lower, wake-up, awakening, mouth, lips, covering, action, vocal cords, vibrate

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 110097875 A (TSINGHUA UNIVERSITY) 06 August 2019 (2019-08-06) claims 1-10, description paragraphs [0013]-[0078]	1-19
PX	CN 110111776 A (TSINGHUA UNIVERSITY) 09 August 2019 (2019-08-09) description, paragraphs [0014]-[0204]	1-19
PX	CN 110223711 A (TSINGHUA UNIVERSITY) 10 September 2019 (2019-09-10) description, paragraphs [0014]-[0204]	1-19
PX	CN 110428806 A (TSINGHUA UNIVERSITY) 08 November 2019 (2019-11-08) description, paragraphs [0014]-[0202]	1-19
Y	WO 2018223388 A1 (MICROSOFT TECHNOLOGY LICENSING L.L.C. et al.) 13 December 2018 (2018-12-13) description, paragraphs [0016]-[0072]	1-19
Y	CN 105120059 A (HUIZHOU TCL MOBILE COMMUNICATION CO., LTD.) 02 December 2015 (2015-12-02) description, paragraphs [0025]-[0040]	1-19



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

12 July 2020

Date of mailing of the international search report

10 August 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

International application No.

PCT/CN2020/089551**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 109686378 A (HUAWEI TECHNOLOGIES CO., LTD.) 26 April 2019 (2019-04-26) description, paragraphs [0110]-[0170]	1-19
Y	CN 109313898 A (APPLE INC.) 05 February 2019 (2019-02-05) description, paragraphs [0247]-[0340]	1-19
Y	CN 105847584 A (GOERTEK INC.) 10 August 2016 (2016-08-10) description, paragraphs [0024]-[0047]	1-19

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2020/089551

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	110097875	A	06 August 2019	None			
CN	110111776	A	09 August 2019	None			
CN	110223711	A	10 September 2019	None			
CN	110428806	A	08 November 2019	None			
WO	2018223388	A1	13 December 2018	EP	3613206	A1	26 February 2020
				US	2020154187	A1	14 May 2020
CN	105120059	A	02 December 2015	CN	105120059	B	26 March 2019
CN	109686378	A	26 April 2019	EP	3471099	A2	17 April 2019
				US	2019115039	A1	18 April 2019
CN	109313898	A	05 February 2019	US	2017358301	A1	14 December 2017
				WO	2017213683	A1	14 December 2017
				EP	3455852	A1	20 March 2019
				US	2019122666	A1	25 April 2019
				KR	20190007450	A	22 January 2019
				US	10192552	B2	29 January 2019
CN	105847584	A	10 August 2016	CN	105847584	B	05 March 2019

国际检索报告

国际申请号

PCT/CN2020/089551

A. 主题的分类

G10L 25/78(2013.01)i; G10L 15/22(2006.01)i; G10L 25/51(2013.01)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)

G10L;H04M1/725

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))

CNKI, CNPAT, WPI, EP0D0C: 声音, 语音, 音频, 低声, 轻声, 低语, 小声, 悄声, 悄悄话, 声带, 震动, 吹气声, 耳语, 近距离, 距离近, 贴近, 接近, 靠近, 唇, 嘴, 脸, 检测, 距离, 捂嘴, 接近传感器, whispered, whisper, silent, voice, nasal air, flow, airflow, ingressive, utterance, shouted, speech mode, activity detection, proximity sensor, low voice, near, lower, wake-up, awakening, mouth, lips, covering, action, vocal cords, vibrate

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
PX	CN 110097875 A (清华大学) 2019年 8月 6日 (2019 - 08 - 06) 权利要求1-10, 说明书第[0013]-[0078]段	1-19
PX	CN 110111776 A (清华大学) 2019年 8月 9日 (2019 - 08 - 09) 说明书第[0014]-[0204]段	1-19
PX	CN 110223711 A (清华大学) 2019年 9月 10日 (2019 - 09 - 10) 说明书第[0014]-[0204]段	1-19
PX	CN 110428806 A (清华大学) 2019年 11月 8日 (2019 - 11 - 08) 说明书第[0014]-[0202]段	1-19
Y	WO 2018223388 A1 (MICROSOFT TECHNOLOGY LICENSING L.L.C. 等) 2018年 12月 13日 (2018 - 12 - 13) 说明书第[0016]-[0072]段	1-19
Y	CN 105120059 A (惠州TCL移动通信有限公司) 2015年 12月 2日 (2015 - 12 - 02) 说明书第[0025]-[0040]段	1-19

☒ 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

“&” 同族专利的文件

国际检索实际完成的日期

2020年 7月 12日

国际检索报告邮寄日期

2020年 8月 10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

王玥

电话号码 86-(10)- 53962545

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 109686378 A (华为技术有限公司) 2019年 4月 26日 (2019 - 04 - 26) 说明书第[0110]-[0170]段	1-19
Y	CN 109313898 A (苹果公司) 2019年 2月 5日 (2019 - 02 - 05) 说明书第[0247]-[0340]段	1-19
Y	CN 105847584 A (歌尔声学股份有限公司) 2016年 8月 10日 (2016 - 08 - 10) 说明书第[0024]-[0047]段	1-19

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2020/089551

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	110097875	A	2019年 8月 6日	无			
CN	110111776	A	2019年 8月 9日	无			
CN	110223711	A	2019年 9月 10日	无			
CN	110428806	A	2019年 11月 8日	无			
WO	2018223388	A1	2018年 12月 13日	EP	3613206	A1	2020年 2月 26日
				US	2020154187	A1	2020年 5月 14日
CN	105120059	A	2015年 12月 2日	CN	105120059	B	2019年 3月 26日
CN	109686378	A	2019年 4月 26日	EP	3471099	A2	2019年 4月 17日
				US	2019115039	A1	2019年 4月 18日
CN	109313898	A	2019年 2月 5日	US	2017358301	A1	2017年 12月 14日
				WO	2017213683	A1	2017年 12月 14日
				EP	3455852	A1	2019年 3月 20日
				US	2019122666	A1	2019年 4月 25日
				KR	20190007450	A	2019年 1月 22日
				US	10192552	B2	2019年 1月 29日
CN	105847584	A	2016年 8月 10日	CN	105847584	B	2019年 3月 5日