



(12)发明专利申请

(10)申请公布号 CN 109903751 A

(43)申请公布日 2019.06.18

(21)申请号 201711294885.0

(22)申请日 2017.12.08

(71)申请人 阿里巴巴集团控股有限公司

地址 英属开曼群岛大开曼资本大厦一座四
层847号邮箱

(72)发明人 刘勇 姚海涛

(74)专利代理机构 北京润泽恒知识产权代理有
限公司 11319

代理人 冀晓恺

(51) Int. Cl.

G10L 15/08(2006.01)

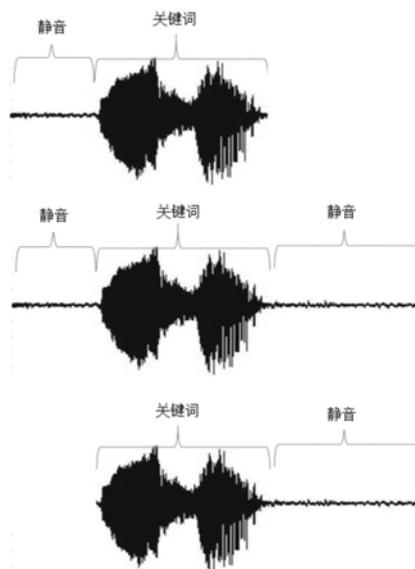
权利要求书3页 说明书16页 附图8页

(54)发明名称

关键词确认方法和装置

(57)摘要

本申请公开了一种关键词确认方法和装置。关键词确认方法包括：获取第一音频数据，所述第一音频数据被识别为关键词；获取所述第一音频数据和第二音频数据至少一者的片段对应的相似发音单元的发音相似概率；确定与所述第一音频数据时间上连续的第二音频数据中存在多个连续的静音片段；利用所述静音概率，以及所述第一音频数据的片段对应的发音相似概率和/或所述第二音频数据的片段对应的发音相似概率，判断所述第二音频数据是否为静音；确认所述第一音频数据为有效关键词。



1. 一种关键词确认方法,其特征在于,包括:
获取第一音频数据,所述第一音频数据被识别为关键词;
确定与所述第一音频数据时间上连续的第二音频数据中存在多个连续的静音片段;
确认所述第一音频数据为有效关键词。
2. 根据权利要求1所述的方法,其特征在于,所述确定与所述第一音频数据时间上连续的第二音频数据中存在多个连续的静音片段的步骤包括:
确定所述片段的发音相似概率,所述发音相似概率为所述片段与多个发音单元之间的最大相似概率;
确定所述片段的静音概率,所述静音概率为所述片段与静音单元的相似概率;
当所述发音相似概率与所述静音概率的关系满足预设条件时,确定所述片段为静音片段;
根据所确定的静音片段,确定所述第二音频数据中存在多个连续的所述静音片段。
3. 根据权利要求1所述的方法,其特征在于,所述确定所述第二音频数据的至少一个片段是否为静音片段的步骤包括:
确定所述片段的发音相似概率,所述发音相似概率为所述片段与多个发音单元之间的最大相似概率;
确定所述片段的静音概率,所述静音概率为所述最大相似概率对应的发音单元与静音单元的相似概率;
当所述发音相似概率与所述静音概率的关系满足预设条件时,确定所述片段为静音片段;
根据所确定的静音片段,确定所述第二音频数据中存在多个连续的所述静音片段。
4. 根据权利要求2或3所述的方法,其特征在于,所述确定所述第二音频数据中存在多个连续的所述静音片段的步骤包括:
确定所述第二音频数据中存在三个以上连续的静音片段。
5. 根据权利要求2或3所述的方法,其特征在于,所述预设条件包括:
所述片段的发音相似概率与静音概率的差值的绝对值小于第一阈值。
6. 根据权利要求1所述的方法,其特征在于,所述获取音频数据的步骤之前,所述方法还包括:
检测所采集到的音频数据中是否包括关键词。
7. 根据权利要求6所述的方法,其特征在于,所述关键词具有属性信息,所述确认所述第一音频数据为有效关键词的步骤包括:
当所述关键词的属性信息为主关键词时,且所述关键词之前的第二音频数据为静音时,确认所述关键词为有效主关键词。
8. 根据权利要求6所述的方法,其特征在于,所述关键词具有属性信息,所述确认所述第一音频数据为有效关键词的步骤包括:
当所述关键词的属性信息为副关键词时,且所述关键词之前和之后的第二音频数据均为静音时,确认所述关键词为有效副关键词。
9. 一种关键词确认方法,其特征在于,包括:
获取第一音频数据,所述第一音频数据被识别为关键词;

确定与所述第一音频数据时间上连续的第二音频数据的多个片段的累积静音概率；

确定所述第一音频数据的多个片段的累积关键词概率；

当所述累积静音概率与累积关键词概率的关系满足第二预设条件时，确认所述第一音频数据为有效关键词。

10. 根据权利要求9所述的方法，其特征在于，所述第二预设条件包括：

所述累积静音概率与累积关键词概率比值的绝对值大于第二阈值。

11. 根据权利要求9所述的方法，其特征在于，所述获取音频数据的步骤之前，所述方法还包括：

检测所采集到的音频数据中是否包括关键词。

12. 根据权利要求11所述的方法，其特征在于，所述关键词具有属性信息，所述当所述累积静音概率与累积关键词概率的关系满足第二预设条件时，确认所述第一音频数据为有效关键词的步骤包括：

当所述关键词的属性信息为主关键词时，且所述关键词之前的第二音频数据为静音时，确认所述关键词为有效主关键词。

13. 根据权利要求11所述的方法，其特征在于，所述关键词具有属性信息，所述当所述累积静音概率与累积关键词概率的关系满足第二预设条件时，确认所述第一音频数据为有效关键词的步骤包括：

当所述关键词的属性信息为副关键词时，且所述关键词之前和之后的第二音频数据均为静音时，确认所述关键词为有效副关键词。

14. 一种车载终端的关键词确认方法，其特征在于，包括：

通过车载音频采集装置获取第一音频数据，所述第一音频数据被识别为关键词；

确定与所述第一音频数据时间上连续的第二音频数据中存在多个连续的静音片段；

确认所述第一音频数据为有效关键词，其中所述有效关键词用于唤醒车载终端执行所述关键词对应的指令。

15. 一种车载终端的关键词确认方法，其特征在于，包括：

通过车载音频采集装置获取第一音频数据，所述第一音频数据被识别为关键词；

确定与所述第一音频数据时间上连续的第二音频数据的多个片段的累积静音概率；

确定所述第一音频数据的多个片段的累积关键词概率；

当所述累积静音概率与累积关键词概率的关系满足第二预设条件时，确认所述第一音频数据为有效关键词，其中所述有效关键词用于唤醒车载终端执行所述关键词对应的指令。

16. 一种关键词确认装置，其特征在于，包括：

音频数据获取模块，用于获取第一音频数据，所述第一音频数据被识别为关键词；

静音片段确定模块，用于确定与所述第一音频数据时间上连续的第二音频数据中存在多个连续的静音片段；

有效关键词确定模块，用于确认所述第一音频数据为有效关键词。

17. 一种关键词确认装置，其特征在于，包括：

音频数据获取模块，用于获取第一音频数据，所述第一音频数据被识别为关键词；

累积静音片段确定模块，用于确定与所述第一音频数据时间上连续的第二音频数据的

多个片段的累积静音概率；

累积关键词概率确定模块,用于确定所述第一音频数据的多个片段的累积关键词概率；

有效关键词确定模块,用于当所述累积静音概率与累积关键词概率的关系满足第二预设条件时,确认所述第一音频数据为有效关键词。

18. 一种终端设备,其特征在于,包括:

一个或多个处理器;和

其上存储有指令的一个或多个机器可读介质,当由所述一个或多个处理器执行时,使得所述终端设备执行如权利要求1-15中一个或多个所述的方法。

19. 根据权利要求18所述的终端设备,其特征在于,所述终端设备包括车载终端、移动终端、服务器、物联网设备或智能家居硬件。

20. 一个或多个机器可读介质,其上存储有指令,当由一个或多个处理器执行时,使得终端设备执行如权利要求1-15中一个或多个所述的方法。

关键词确认方法和装置

技术领域

[0001] 本申请涉及计算机领域,特别是涉及一种关键词确认方法和装置。

背景技术

[0002] 在计算机人机交互领域中,利用关键词对例如车载终端等电子装置进行唤醒,以开启人机交互功能或者执行某项功能,已被多种电子装置使用。

[0003] 例如,当使用者说出如启动系统、放大地图等特定关键词时,电子装置检测到使用者的上述关键词后执行相应的启动系统、放大地图的功能。通过这种语音控制的方式极大地增加了使用者的便利。

[0004] 但是,如何确定使用者发出的关键词,而不会造成误判,例如将不是关键词的正常对话识别为关键词,或者将关键词错误地识别为非关键词而拒绝启动,成为本领域需要解决的问题。

[0005] 为了解决这一问题,现有技术提出的一种解决方案是,通过将使用者发出的语音与关键词库中的关键词对比,如果能够匹配,则认为使用者发出的语音指令为关键词,电子装置对应地执行相应指令;如果未能匹配,则不是关键词,电子装置不执行指令。

[0006] 但是,这一方案容易将使用者会话中不是关键词的内容识别为关键词,例如使用者说出的“现在还不需要启动系统”或者“我就是想知道放大地图的功能能不能用”,这类的正常对话中包含“启动系统”或“放大地图”的关键词,会被电子装置识别为关键词,从而错误地执行指令。

发明内容

[0007] 鉴于上述问题,本发明一实施例提出一种关键词确认方法和装置,以解决现有技术存在的问题。

[0008] 为了解决上述问题,本申请一实施例公开一种关键词确认方法,包括:

[0009] 获取第一音频数据,所述第一音频数据被识别为关键词;

[0010] 确定与所述第一音频数据时间上连续的第二音频数据中存在多个连续的静音片段;

[0011] 确认所述第一音频数据为有效关键词。

[0012] 为了解决上述问题,本申请第二实施例公开一种关键词确认方法,包括:

[0013] 获取第一音频数据,所述第一音频数据被识别为关键词;

[0014] 确定与所述第一音频数据时间上连续的第二音频数据的多个片段的累积静音概率;

[0015] 确定所述第一音频数据的多个片段的累积关键词概率;

[0016] 当所述累积静音概率与累积关键词概率的关系满足第二预设条件时,确认所述第一音频数据为有效关键词。

[0017] 为了解决上述问题,本申请一实施例公开一种车载终端的关键词确认方法,包括:

- [0018] 通过车载音频采集装置获取第一音频数据,所述第一音频数据被识别为关键词;
- [0019] 确定与所述第一音频数据时间上连续的第二音频数据中存在多个连续的静音片段;
- [0020] 确认所述第一音频数据为有效关键词,其中所述有效关键词用于唤醒车载终端执行所述关键词对应的指令。
- [0021] 为了解决上述问题,本申请一实施例公开一种车载终端的关键词确认方法,包括:
- [0022] 通过车载音频采集装置获取第一音频数据,所述第一音频数据被识别为关键词;
- [0023] 确定与所述第一音频数据时间上连续的第二音频数据的多个片段的累积静音概率;
- [0024] 确定所述第一音频数据的多个片段的累积关键词概率;
- [0025] 当所述累积静音概率与累积关键词概率的关系满足第二预设条件时,确认所述第一音频数据为有效关键词,其中所述有效关键词用于唤醒车载终端执行所述关键词对应的指令。
- [0026] 为了解决上述问题,本申请一实施例公开一种关键词确认装置,包括:
- [0027] 音频数据获取模块,用于获取第一音频数据,所述第一音频数据被识别为关键词;
- [0028] 静音片段确定模块,用于确定与所述第一音频数据时间上连续的第二音频数据中存在多个连续的静音片段;
- [0029] 有效关键词确定模块,用于确认所述第一音频数据为有效关键词。
- [0030] 为了解决上述问题,本申请一实施例公开一种关键词确认装置,包括:
- [0031] 音频数据获取模块,用于获取第一音频数据,所述第一音频数据被识别为关键词;
- [0032] 累计静音片段确定模块,用于确定与所述第一音频数据时间上连续的第二音频数据的多个片段的累积静音概率;
- [0033] 累积关键词概率确定模块,用于确定所述第一音频数据的多个片段的累积关键词概率;
- [0034] 有效关键词确定模块,用于当所述累积静音概率与累积关键词概率的关系满足第二预设条件时,确认所述第一音频数据为有效关键词。
- [0035] 本申请一实施例还公开一种终端设备,包括:
- [0036] 一个或多个处理器;和
- [0037] 其上存储有指令的一个或多个机器可读介质,当由所述一个或多个处理器执行时,使得所述终端设备执行上述的方法。
- [0038] 本申请一实施例还公开一个或多个机器可读介质,其上存储有指令,当由一个或多个处理器执行时,使得终端设备执行上述的方法。
- [0039] 由上述可知,本申请实施例提出的关键词确认方法,至少包括以下优点:
- [0040] 本发明实施例提出的关键词确认方法中,利用了使用者的一般习惯,即在发出关键词的之前或之后会有停顿而产生静音,通过检测关键词前后是否存在静音来检测该关键词是否为有效关键词,提高了检测准确率,避免了识别错误;同时在针对静音的检测中,利用了音频数据的静音片段的检测,提高了判断是否为静音的准确性,进一步避免了将关键词误判为非关键词。

附图说明

[0041] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作一简单地介绍,显而易见地,下面描述中的附图是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0042] 图1所示为本发明一实施例的正常的关键词及前后音频数据的示意图

[0043] 图2是本申请第一实施例的关键词确认方法的流程图。

[0044] 图3是本申请第二实施例的关键词确认方法的流程图。

[0045] 图4是图3中步骤的子步骤流程图。

[0046] 图5是本申请第三实施例的关键词确认方法的流程图。

[0047] 图6是包括车辆环境的车载终端的示意图。

[0048] 图7是本申请第五实施例的车载终端的关键词确认方法的流程图。

[0049] 图8是本申请第六实施例的车载终端的关键词确认方法的流程图。

[0050] 图9是本申请第七实施例的关键词确认装置的方框图。

[0051] 图10是本申请第八实施例的关键词确认装置的方框图。

[0052] 图11示意性地示出了用于执行根据本发明的方法的终端设备的框图。

[0053] 图12示意性地示出了用于保持或者携带实现根据本发明的方法的程序代码的存储单元。

具体实施方式

[0054] 下面将结合本申请实施例中的附图,对本申请实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本申请一部分实施例,而不是全部的实施例。基于本申请中的实施例,本领域普通技术人员所获得的所有其他实施例,都属于本申请保护的范围。

[0055] 本申请的核心思想之一在于,提出一种关键词确认方法,利用关键词前后的静音来判断关键词是否确实为有效关键词,同时在针对静音的检测中,利用了音频数据的连续静音片段或者多个片段的累积静音概率来进行判断,提高了判断的准确性。

[0056] 第一实施例

[0057] 本发明第一实施例提出一种关键词确认方法。图1所示为本发明一实施例的正常的关键词及前后音频数据的示意图,如图1所示,按照使用者一般的习惯,在发出关键词之前或者之后会有停顿导致静音,所以正常的关键词前后可以采到静音。利用这一方式可以判断使用者发出的语音是否为关键词。

[0058] 图1中是正常的关键词的三种可能情形,分别包括:在关键词前有静音、在关键词后有静音、在关键词前后均有静音。以下将关键词对应的音频数据作为第一音频数据,将关键词前后的静音部分对应的音频数据作为第二音频数据。

[0059] 图2所示为本发明第一实施例的关键词确认方法的步骤流程图。如图2所示,本发明实施例的关键词确认方法例如可以包括如下步骤:

[0060] S101,获取第一音频数据,所述第一音频数据被识别为关键词;

[0061] 在这一步骤中,执行主体,例如是车载终端、手机、平板电脑等电子装置,可以获取音频数据,其至少包括第一音频数据和前/后的第二音频数据。第二音频数据与第一音频数据在时间上连续。此时检测到的第一音频数据已被识别为关键词,即,此时检测到的音频数据已确认与预存的关键词匹配。

[0062] 如图1所示,电子装置可以获取并检测如图1所示中关键词对应的第一音频数据,以及关键词之前、之后或之前和之后的第二音频数据,在实际使用中,电子装置的声音采集装置例如麦克风可以持续采集音频,音频数据例如是按照“帧”为单位获取的,一帧例如为10ms,则在检测到第一音频数据为关键词之后,获取该第一音频数据前/后若干帧,例如10帧的第二音频数据,进行后续分析。

[0063] 在一实施例中,需要进一步判断第一音频数据是否为“有效关键词”,只有在后续确认为有效关键词时才可以根据有效关键词执行相应的指令。

[0064] S102,确定与所述第一音频数据时间上连续的第二音频数据中存在多个连续的静音片段;

[0065] 在这一步骤中,将第二音频数据的片段输入电子装置的声音单元匹配模型后,可以获知其与声音单元库中的静音单元的相似度,作为该片段的静音概率。例如,针对第二音频数据的片段,将其输入声音单元匹配模型后,计算出其与静音单元的相似度为90%,则将90%作为该片段对应的静音概率,当这一静音概率满足一定的要求时,则认为第二音频数据的该片段为静音片段。

[0066] 在一实施例中,可以将多个片段输入电子装置的声音单元匹配模型,分别获取该片段对应的静音概率,并利用该静音概率确定该片段是否为静音片段。

[0067] 在判断该片段为静音片段之后,可以判断该第二音频数据中是否包括多个连续的静音片段。例如,针对第二音频数据的多个片段,获知了每一个片段是否为静音片段后可以检测这些静音片段中是否为连续的静音片段,例如,每一个片段具有一个是否为静音片段的标识 f ,当检测到时间上连续的三个片段均具有静音标识 f 时,则认为第二音频数据中存在多个连续的静音片段。

[0068] S103,确认所述第一音频数据为有效关键词。

[0069] 在这一步骤中,当第二音频数据中存在多个连续的静音片段时,则判断第二音频数据为静音,由此可以确认该关键词为有效关键词,后续可以根据该有效关键词执行对应的指令。

[0070] 例如,当第二音频数据中包括多个(例如3个以上)连续的静音片段时,认为第二音频数据为静音,继而判断出第一音频数据为有效关键词。

[0071] 值得注意的是,前述和后述的关键词可以包括多种内容。例如,用于唤醒电子装置中的操作系统的唤醒词、使用者的语音指令、指令中的关键参数等。例如,使用者针对电子装置的语音操作过程中,输入的“开启系统”、“调频到87.6”、“87.6”等等,均属于本发明实施例提出的“关键词”的范畴,本发明并不特别限制。

[0072] 由上述可知,本发明第一实施例提出的关键词确认方法至少具有如下技术效果:

[0073] 本发明实施例提出的关键词确认方法中,利用了使用者的一般习惯,即在发出关键词的之前或之后会有停顿而产生静音,通过检测关键词前后是否存在静音来检测该关键词是否为有效关键词,提高了检测准确率,避免了识别错误;同时在针对静音的检测中,利

用了音频数据的静音片段的检测,提高了判断是否为静音的准确性,进一步避免了将关键词误判为非关键词。

[0074] 第二实施例

[0075] 本发明第二实施例提出一种关键词确认方法。图3所示为本发明第二实施例的关键词确认方法的步骤流程图。如图3所示,本发明实施例的关键词确认方法如下步骤:

[0076] S201,获取第一音频数据,所述第一音频数据被识别为关键词;

[0077] S202,确定与所述第一音频数据时间上连续的第二音频数据中存在多个连续的静音片段;

[0078] S203,确认所述第一音频数据为有效关键词。

[0079] 上述步骤S201至步骤S203与上一实施例的步骤S101至步骤S103相同或相似,在此不再赘述。本实施例重点说明与上一实施例的不同之处。

[0080] 在本发明关键词确认方法的一实施例中,所述步骤S202,即确定与所述第一音频数据时间上连续的第二音频数据中存在多个连续的静音片段的步骤可以包括如下子步骤:

[0081] S2021,确定所述片段的发音相似概率,所述发音相似概率为所述片段与多个发音单元之间的最大相似概率;

[0082] 在这一步骤中,第二音频数据的片段例如可以为上述的音频帧,也可以为其他单位的片段,在此并不限制,只要是按照特定的规律,例如时间、存储方式等对音频数据划分获得的片段,均属于本发明保护的范围。例如,该片段可以为10ms或者20ms一帧的音频帧,也可以为1s的音频段落等。

[0083] 发音单元可以是根据使用者的发音划分获得的音素、音节、字、词等单元,例如,当使用者发出的是“斑马”,发音单元是以音素为单位划分的,音素是比音节更小的发音单元,从现有的音素集中可知,“斑马”对应的音素“b a_h nn_h m a_l a_l”;当发音单元是以音节为单位划分的,“斑马”对应的音节是“b an m a”;当发音单元是以字的发音为单位划分的,“斑马”对应的划分方式是“ban ma”;当发音单元是以词为单位划分的,“斑马”对应的划分方式是“banma”。

[0084] 针对每一种划分方式,可以构建对应的声音单元库。词库中除了包含上述的发音单元,还可以包括静音单元等。当第二音频数据的一个片段与声音单元库中预存的其中一个发音单元的相似概率最高,则认为该片段与该发音单元匹配,则将该发音单元作为相似发音单元,同时将该片段与该相似发音单元的相似率作为发音相似概率。

[0085] 针对第二音频数据中的片段,将该片段输入电子装置的声音单元匹配模型进行判断,如果该片段与声音单元库中的第五发音单元的相似度最高,为80%,与第六发音单元的相似度其次,为70%,则可以记录相似度80%的第五发音单元,作为片段对应的相似发音单元,并记录发音相似概率80%,用于后续处理。

[0086] S2022,确定所述片段的静音概率,所述静音概率为所述片段与静音单元的相似概率;

[0087] 在这一步骤中,将第二音频数据的片段输入电子装置的声音单元匹配模型后,可以获知其与声音单元库中的静音单元的相似度,作为该片段的静音概率。例如,针对第二音频数据的片段,将其输入声音单元匹配模型后,计算出其与静音单元的相似度为90%,则将

90%作为该片段对应的静音概率

[0088] 值得注意的是,上述的静音单元可以是预存在声音单元库中的,可以通过大量数据迭代训练模型获得,例如综合考虑声音的能量、环境噪音(包括风声、音乐声、汽车鸣笛等)等获得静音单元,并不限于绝对的无声。静音单元的长度、属性等可以与发明单元对应。例如当发音单元是按照音素来划分的,则静音单元可以为静音音素;当发音单元是按照音节来划分的,则静音单元可以为静音音节,在此并不限制。

[0089] S2023,当所述发音相似概率与所述静音概率的关系满足预设条件时,确定所述片段为静音片段。

[0090] 所述预设条件例如包括:

[0091] 所述片段的发音相似概率与静音概率的差值的绝对值小于第一阈值。

[0092] 在这一步骤中,可以利用之前获得的第二音频数据的片段对应的发音相似概率和对应的静音概率,判断该第二音频数据的片段是否为静音。

[0093] 如上述可知,本发明实施例提出的方案中,静音概率的判断中并不是将音频数据的片段与绝对静音进行对比,而是将发音相似概率和对应的静音概率做对比,综合考虑了环境噪声等因素,因此本发明提供的方案可以避免因为静音判断不准确而拒绝正确的关键词。

[0094] 利用发音相似概率和所述静音概率判断音频数据的片段是否为静音的方式有多种,在此举例进行介绍。

[0095] 例如,片段满足“发音相似概率 $p_{\max}(\text{indexframe})$ 和所述静音概率 $p_{\text{sil}}(\text{indexframe})$ 的差值的绝对值小于15%”这一要求,即:

[0096] $|p_{\max}(\text{indexframe}) - p_{\text{sil}}(\text{indexframe})| < 15\%$

[0097] 则认定该片段为静音片段。

[0098] 在本发明关键词确认方法的一实施例中,上述子步骤S2022中,静音概率的判断也可以是利用最大相似概率对应的发音单元与静音单元的相似概率对应的发音单元与静音单元的相似概率来判定。即,子步骤S2022可以替换为如下子步骤:

[0099] S2024:确定所述片段的静音概率,所述静音概率为所述最大相似概率对应的发音单元与静音单元的相似概率。

[0100] 在子步骤S2021中已确定了该片段的发音相似概率,例如前述举例中,第二音频数据中的片段通过电子装置的声音单元匹配模型进行判断获知,该片段与声音单元库中的第五发音单元的相似度最高,为80%,则此处最大相似概率80%对应的第五发音单元为相似发音单元。在这一子步骤中,可以计算第五发音单元与静音单元的相似概率,作为片段的静音概率。

[0101] 依据以上列举的方式和本领域技术人员的技术能力,本领域技术人员可以利用这些发音相似概率和静音概率设置诸多判断方式,以判断第二音频数据的片段是否为静音,本发明并不特别限制。

[0102] 在步骤S2023或步骤S2024之后,即确定所述片段为静音片段之后,所述步骤S202,即确定与所述第一音频数据时间上连续的第二音频数据中存在多个连续的静音片段的步骤还可以包括如下子步骤:

[0103] S2025,根据所确定的静音片段,确定所述第二音频数据中存在多个连续的所述静

音片段。

[0104] 在这一步骤中,可以判断该第二音频数据中是否包括多个连续的静音片段。例如,针对第二音频数据的多个片段,例如,在子步骤S2023或S2024为每一个片段设置一个为静音片段的标识f,当检测到时间上连续的三个片段均具有静音标识f时,则认为第二音频数据中存在多个连续的静音片段。

[0105] 在本发明关键词确认方法的一实施例中,上述子步骤S2025,即确定所述第二音频数据中存在多个连续的所述静音片段的步骤中,上述“多个”可以为三个以上,即,子步骤S2025可以为:

[0106] 确定所述第二音频数据中存在三个以上连续的静音片段。

[0107] 在本发明关键词确认方法的一实施例中,所述步骤S201,即获取音频数据的步骤之前,所述方法还可以包括:

[0108] S200,检测所采集到的音频数据中是否包括关键词。

[0109] 在这一步骤中,电子装置的关键词库中可以预存多个关键词,比如“你好斑马”、“开启系统”、“放大地图”,“缩小地图”,“退出导航”等。第一音频数据中的关键词可能是这里的任何一个,可以利用关键词词库,计算输入的第一音频数据与这些关键词相似的概率,选出概率最高且高于设定阈值的词,作为检测到的关键词。

[0110] 具体地,例如可以利用本发明的声音单元匹配的方式,将音频数据切分为多个片段,当一个片段与声音单元库中预存的其中一个发音单元的相似概率最高,则认为该片段与该发音单元匹配,则将该发音单元作为相似发音单元,同时将该片段与该相似发音单元的相似率作为发音相似概率。

[0111] 在针对一段语音,例如第一音频数据时,将多个片段的发音相似概率进行处理,例如经过相乘,获得概率最大的路径,将该路径对应的词作为匹配的关键词。

[0112] 在本发明关键词确认方法的一实施例中,所述关键词具有属性信息,所述步骤S203,确认所述第一音频数据为有效关键词的步骤可以包括:

[0113] 当所述关键词的属性信息为主关键词时,且所述关键词之前的第二音频数据为静音时,确认所述关键词为有效主关键词。

[0114] 在本发明实施例中,每一个关键词可以对应一个属性信息,该信息记载了关键词为主关键词还是副关键词。电子装置的关键词库中预存的多个关键词例如可以分为主关键词和副关键词,比如“你好斑马”、“开启系统”等设置为主关键词,“放大地图”,“缩小地图”,“退出导航”等设置为副关键词。

[0115] 对于主关键词,如果考虑主关键词后面可以没有内容,也可能会直接识别的语音,例如“你好斑马请帮我查一下到中关村的路”,则可以设置关键词之前的音频数据为静音,且关键词的属性信息为主关键词,则确认这一关键词为主关键词;而不去检测关键词之后是否为静音。

[0116] 在本发明关键词确认方法的一实施例中,所述关键词具有属性信息,所述关键词具有属性信息,所述步骤S203,确认所述第一音频数据为有效关键词的步骤可以包括:

[0117] 当所述关键词的属性信息为副关键词时,且所述关键词之前和之后的第二音频数据均为静音时,确认所述关键词为有效副关键词。

[0118] 在这一步骤中,副关键词可能是使用者要求电子装置直接执行的命令,例如“放大

地图”。可以设置关键词之前和之后的内容均为静音,并且关键词的属性为副关键词,才确认该关键词为副关键词。当使用者说出“我就是想试试放大地图能不能用”或“不知道能不能放大地图”或“放大地图就可以了”这类语音时,虽然可以检测到关键词,但是不满足前后静音的条件,仍不会将其判断为有效关键词。

[0119] 综上所述,本实施例提出的关键词确认方法至少具有如下优点:

[0120] 本发明实施例提出的应用于车载终端的关键词确认方法中,利用了使用者的一般习惯,即在发出关键词的之前或之后会有停顿而产生静音,通过检测关键词前后是否存在静音来检测该关键词是否为有效关键词,提高了检测准确率,避免了识别错误;同时在针对静音的检测中,利用了音频数据的静音片段的检测,提高了判断是否为静音的准确性,进一步避免了将关键词误判为非关键词。

[0121] 除此之外,本实施例提出的关键词确认方法至少还包括如下优点:

[0122] 本发明实施例提出的关键词确认方法的一可选实施例中,提出了优选的判断方法,利用第二音频数据的累积静音概率和第一音频数据的累积关键词概率的比值和/或第二音频数据的发音相似概率和静音概率的差值判断第二音频数据是否为静音,使判断结果更加准确;此外针对关键词的不同类型——主关键词和副关键词,设置了不同的进一步确认方式,使判断结果更加可靠。

[0123] 第三实施例

[0124] 本发明第三实施例提出一种关键词确认方法。图5所示为本发明第三实施例的关键词确认方法的步骤流程图。如图5所示,本发明实施例的关键词确认方法如下步骤:

[0125] S301,获取第一音频数据,所述第一音频数据被识别为关键词;

[0126] 在这一步骤中,执行主体,例如是车载终端、手机、平板电脑等电子装置,可以获取音频数据,其至少包括第一音频数据和前/后的第二音频数据。第二音频数据与第一音频数据在时间上连续。此时检测到的第一音频数据已被识别为关键词,即,此时检测到的音频数据已确认与预存的关键词匹配。

[0127] 电子装置可以获取并检测关键词对应的第一音频数据,以及关键词之前、之后或之前和之后的第二音频数据,在实际使用中,电子装置的声音采集装置例如麦克风可以持续采集音频,音频数据例如是按照“帧”为单位获取的,一帧例如为10ms,则在检测到第一音频数据为关键词之后,获取该第一音频数据前/后若干帧,例如10帧的第二音频数据,进行后续分析。

[0128] S302,确定与所述第一音频数据时间上连续的第二音频数据的多个片段的累积静音概率;

[0129] 第二音频数据的多个片段的累积静音概率 $p(\text{sil})$ 可以利用第二音频数据的每一个片段的静音概率的乘积或者总和获得。每一个片段可以通过前述第一和第二实施例中提供的方式计算出静音概率,在步骤S302中,可以将这些静音概率通过相加或相乘的方式获得累积静音概率。

[0130] S303,确定所述第一音频数据的多个片段的累积关键词概率;

[0131] 在这一步骤中,第一音频数据的累积关键词概率 $p(\text{kws})$ 可以是第一音频数据的多个片段对应的发音相似概率的乘积;例如,针对第一音频数据的第一片段和第二片段,将这两个片段输入电子装置的声音单元匹配模型进行判断,如果第一片段与声音单元库中的第

一发音单元的相似度最高,为90%,与第二发音单元相似度其次,为80%;则将第一发音单元作为第一片段的相似发音单元,90%作为第一片段的发音相似概率;第二片段的发音相似概率例如为70%,则第一音频数据的累积关键词概率为 $80\% \times 70\%$ 。

[0132] 值得注意的是,第一音频数据的累积关键词概率 $p(kws)$ 可以是从前序的处理中获得,并非限定为即时计算获取。本发明并不限定其获取方式。

[0133] S304,当所述累积静音概率与累积关键词概率的关系满足第二预设条件时,确认所述第一音频数据为有效关键词。

[0134] 在本发明关键词确认方法的一实施例中,所述第二预设条件包括:

[0135] 所述累积静音概率与累积关键词概率比值大于第二阈值

[0136] 在这一实施例中,举例来说, $p(sil)/p(kws)$ 的比值越大,则第二音频数据为静音的判断越准确。因此可以设置第二预设条件包括所述累积静音概率与累积关键词概率比值大于第二阈值,当 $p(sil)/p(kws)$ 大于第二阈值(例如1.5)时认为第二音频数据为静音。

[0137] 在本发明一可选实施例中,所述步骤S301,即获取音频数据的步骤之前,所述方法还包括:

[0138] S300,检测所采集到的音频数据中是否包括关键词。

[0139] 在这一步骤中,电子装置的关键词库中可以预存多个关键词,比如“你好斑马”、“开启系统”、“放大地图”,“缩小地图”,“退出导航”等。第一音频数据中的关键词可能是这里的任何一个,可以利用关键词词库,计算输入的第一音频数据与这些关键词相似的概率,选出概率最高且高于设定阈值的词,作为检测到的关键词。

[0140] 具体地,例如可以利用本发明的声音单元匹配的方式,将音频数据切分为多个片段,当一个片段与声音单元库中预存的其中一个发音单元的相似概率最高,则认为该片段与该发音单元匹配,则将该发音单元作为相似发音单元,同时将该片段与该相似发音单元的相似率作为发音相似概率。

[0141] 在针对一段语音,例如第一音频数据时,将多个片段的发音相似概率进行处理,例如经过相乘,获得概率最大的路径,将该路径对应的词作为匹配的关键词。

[0142] 在本发明一可选实施例中,所述关键词具有属性信息,所述当所述累积静音概率与累积关键词概率的关系满足第二预设条件时,确认所述第一音频数据为有效关键词的步骤包括:

[0143] 当所述关键词的属性信息为主关键词时,且所述关键词之前的第二音频数据为静音时,确认所述关键词为有效主关键词。

[0144] 在本发明实施例中,每一个关键词可以对应一个属性信息,该信息记载了关键词为主关键词还是副关键词。电子装置的关键词库中预存的多个关键词例如可以分类为主关键词和副关键词,比如“你好斑马”、“开启系统”等设置为主关键词,“放大地图”,“缩小地图”,“退出导航”等设置为副关键词。

[0145] 对于主关键词,如果考虑主关键词后面可以没有内容,也可能会直接识别的语音,例如“你好斑马请帮我查一下到中关村的路”,则可以设置关键词之前的音频数据为静音,且关键词的属性信息为主关键词,则确认这一关键词为主关键词;而不去检测关键词之后是否为静音。

[0146] 在本发明一可选实施例中,所述关键词具有属性信息,所述当所述累积静音概率

与累积关键词概率的关系满足第二预设条件时,确认所述第一音频数据为有效关键词的步骤包括:

[0147] 当所述关键词的属性信息为副关键词时,且所述关键词之前和之后的第二音频数据均为静音时,确认所述关键词为有效副关键词。

[0148] 在这一步骤中,副关键词可能是使用者要求电子装置直接执行的命令,例如“放大地图”。可以设置关键词之前和之后的内容均为静音,并且关键词的属性为副关键词,才确认该关键词为副关键词。当使用者说出“我就是想试试放大地图能不能用”或“不知道能不能放大地图”或“放大地图就可以了”这类语音时,虽然可以检测到关键词,但是不满足前后静音的条件,仍不会将其判断为有效关键词。

[0149] 综上所述,本实施例提出的关键词确认方法至少具有如下优点:

[0150] 本发明实施例提出的关键词确认方法中,利用了使用者的一般习惯,即在发出关键词的之前或之后会有停顿而产生静音,通过检测关键词前后是否存在静音来检测该关键词是否为有效关键词,提高了检测准确率,避免了识别错误;同时在针对静音的检测中,利用了音频数据的静音片段的检测,提高了判断是否为静音的准确性,进一步避免了将关键词误判为非关键词。

[0151] 除此之外,本实施例提出的关键词确认方法至少还包括如下优点:

[0152] 本发明实施例提出的关键词确认方法的一可选实施例中,提出了优选的判断方法,利用第二音频数据的累积静音概率和第一音频数据的累积关键词概率的比值和/或第二音频数据的发音相似概率和静音概率的差值判断第二音频数据是否为静音,使判断结果更加准确;此外针对关键词的不同类型——主关键词和副关键词,设置了不同的进一步确认方式,使判断结果更加可靠。

[0153] 第四实施例

[0154] 本发明第四实施例提出一种车载终端的关键词确认方法。图6是包括车辆环境的车载终端的示意图。图7是本申请第五实施例的车载终端的关键词确认方法的流程图。如图6所示,车辆包括设置在车内的车载终端200,车载终端200包括扬声器400和麦克风700,还可以包括屏幕、按键等(图未示)。扬声器400除了可以集成于车载终端,还可以设置在车辆内部的其他位置,供乘坐者600收听信息。车载终端200具有计算处理功能,其可以安装操作系统和应用程序,还可以通过互联网500与服务器300远程联网进行数据交互。

[0155] 如图7所示,本发明实施例的车载终端的关键词确认方法如下步骤:

[0156] S401,通过车载音频采集装置获取第一音频数据,所述第一音频数据被识别为关键词;

[0157] 在这一步骤中,车载终端200可以获取音频数据,其至少包括第一音频数据和前/后的第二音频数据。第二音频数据与第一音频数据在时间上连续。此时检测到的第一音频数据已被识别为关键词,即,此时麦克风700检测到的音频数据已确认与预存的关键词匹配。

[0158] 如图1所示,车载终端200可以获取并检测如图1所示中关键词对应的第一音频数据,以及关键词之前、之后或之前和之后的第二音频数据,在实际使用中,车载终端的声音采集装置例如麦克风可以持续采集音频,音频数据例如是按照“帧”为单位获取的,一帧例如为10ms,则在检测到第一音频数据为关键词之后,获取该第一音频数据前/后若干帧,例

如10帧的第二音频数据,进行后续分析。

[0159] S402,确定与所述第一音频数据时间上连续的第二音频数据中存在多个连续的静音片段;

[0160] 在这一步骤中,将第二音频数据的片段输入车载终端的声音单元匹配模型后,可以获知其与声音单元库中的静音单元的相似度,作为该片段的静音概率。例如,针对第二音频数据的片段,将其输入声音单元匹配模型后,计算出其与静音单元的相似度为90%,则将90%作为该片段对应的静音概率,当这一静音概率满足一定的要求时,则认为第二音频数据的该片段为静音片段。

[0161] 在一实施例中,可以将多个片段输入车载终端的声音单元匹配模型,分别获取该片段对应的静音概率,并利用该静音概率确定该片段是否为静音片段。

[0162] S403,确认所述第一音频数据为有效关键词,其中所述有效关键词用于唤醒车载终端执行所述关键词对应的指令。

[0163] 在这一步骤中,当第二音频数据中存在多个连续的静音片段时,则判断第二音频数据为静音时,可以确认该关键词为有效关键词,后续可以根据该有效关键词执行对应的指令。

[0164] 例如,针对第二音频数据的多个片段,前述已获知了每一个片段是否为静音片段。在这一步骤中可以检测这些静音片段中是否为连续的静音片段,当包括多个(例如3个以上)连续的静音片段时,认为第二音频数据为静音,继而判断出第一音频数据为有效关键词,其中所述有效关键词用于唤醒车载终端执行所述关键词对应的指令。

[0165] 综上所述,本实施例提出的车载终端的关键词确认方法至少具有如下优点:

[0166] 本发明实施例提出的应用于车载终端的关键词确认方法中,利用了使用者的一般习惯,即在发出关键词的之前或之后会有停顿而产生静音,通过检测关键词前后是否存在静音来检测该关键词是否为有效关键词,提高了检测准确率,避免了识别错误;同时在针对静音的检测中,利用了音频数据的静音片段的检测,提高了判断是否为静音的准确性,进一步避免了将关键词误判为非关键词。

[0167] 第五实施例

[0168] 本发明第五实施例提出一种车载终端的关键词确认方法。图8是本申请第六实施例的车载终端的关键词确认方法的流程图。如图8所示,本发明实施例的车载终端的关键词确认方法如下步骤:

[0169] S501,通过车载音频采集装置获取第一音频数据,所述第一音频数据被识别为关键词;

[0170] 在这一步骤中,车载终端可以获取音频数据,其至少包括第一音频数据和前/后的第二音频数据。第二音频数据与第一音频数据在时间上连续。此时检测到的第一音频数据已被识别为关键词,即,此时检测到的音频数据已确认与预存的关键词匹配。

[0171] 车载终端可以获取并检测关键词对应的第一音频数据,以及关键词之前、之后或之前和之后的第二音频数据,在实际使用中,车载终端的声音采集装置例如麦克风可以持续采集音频,音频数据例如是按照“帧”为单位获取的,一帧例如为10ms,则在检测到第一音频数据为关键词之后,获取该第一音频数据前/后若干帧,例如10帧的第二音频数据,进行后续分析。

[0172] S502,确定与所述第一音频数据时间上连续的第二音频数据的多个片段的累积静音概率;

[0173] 第二音频数据的多个片段的累积静音概率 $p(\text{sil})$ 可以利用第二音频数据的每一个片段的静音概率的乘积或者总和获得。每一个片段可以通过前述第一和第二实施例中提供的方式计算出静音概率,在步骤S502中,可以将这些静音概率通过相加或相乘的方式获得累积静音概率。

[0174] S503,确定所述第一音频数据的多个片段的累积关键词概率;

[0175] 在这一步骤中,第一音频数据的累积关键词概率 $p(\text{kws})$ 可以是第一音频数据的多个片段对应的发音相似概率的乘积;例如,针对第一音频数据的第一片段和第二片段,将这两个片段输入车载终端的声音单元匹配模型进行判断,如果第一片段与声音单元库中的第一发音单元的相似度最高,为90%,与第二发音单元相似度其次,为80%;则将第一发音单元作为第一片段的相似发音单元,90%作为第一片段的发音相似概率;第二片段的发音相似概率例如为70%,则第一音频数据的累积关键词概率为 $80\% \times 70\%$ 。

[0176] S504,当所述累积静音概率与累积关键词概率的关系满足第二预设条件时,确认所述第一音频数据为有效关键词,其中所述有效关键词用于唤醒车载终端执行所述关键词对应的指令。

[0177] 在这一实施例中,举例来说, $p(\text{sil})/p(\text{kws})$ 的比值越大,则第二音频数据为静音的判断越准确。因此可以设置第二预设条件包括所述累积静音概率与累积关键词概率比值大于第二阈值,当 $p(\text{sil})/p(\text{kws})$ 大于第二阈值(例如1.5)时认为第二音频数据为静音。

[0178] 在本发明关键词确认方法的一实施例中,所述第二预设条件包括:

[0179] 所述累积静音概率与累积关键词概率比值大于第二阈值。

[0180] 综上所述,本实施例提出的车载终端的关键词确认方法至少具有如下优点:

[0181] 本发明实施例提出的应用于车载终端的关键词确认方法中,利用了使用者的一般习惯,即在发出关键词的之前或之后会有停顿而产生静音,通过检测关键词前后是否存在静音来检测该关键词是否为有效关键词,提高了检测准确率,避免了识别错误;同时在针对静音的检测中,利用了音频数据的静音片段的检测,提高了判断是否为静音的准确性,进一步避免了将关键词误判为非关键词。

[0182] 值得注意的是,尽管上述第四和第五实施例提出了车载终端的关键词确认方法,但是本领域技术人员可以明确的是,本发明提出的关键词确认方法可以不限于车载终端,还可以应用于其他各种智能设备中。例如手机、服务器、智能家居硬件等各种具有计算、处理功能的智能设备。智能家居硬件例如包括微波炉、烤箱、洗衣机、洗碗机、空调、路由器、智能音箱,电视,电冰箱,吸尘器等。

[0183] 在一个实施例中,当上述的关键词确认方法应用于智能音箱中时,上述关键词例如可以包括“播放音乐”或“下一曲”等。当智能音箱通过声音接收装置接收到“播放音乐”这一关键词(第一音频数据)时,该智能音箱即通过判断第二音频数据是否为静音,继而确认该关键词为有效关键词。在一实施例中,在确认为有效关键词之后,智能音箱可以开始执行有效关键词对应的播放音乐的指令。

[0184] 以上实施例以智能音箱为例进行了说明,但是本领域技术人员可以明确的是,本发明提供的关键词确认方法可以应用于各式智能设备,在此并不限制。

[0185] 第六实施例

[0186] 本发明第六实施例提出一种关键词确认装置。图9是本申请第七实施例的关键词确认装置的方框图。如图9所示,本发明实施例的关键词确认装置包括:

[0187] 音频数据获取模块601,用于获取第一音频数据,所述第一音频数据被识别为关键词;

[0188] 静音片段确定模块602,用于确定与所述第一音频数据时间上连续的第二音频数据中存在多个连续的静音片段;

[0189] 有效关键词确定模块603,用于确认所述第一音频数据为有效关键词。

[0190] 综上所述,本实施例提出的关键词确认装置至少具有如下优点:

[0191] 本发明实施例提出的关键词确认装置中,利用了使用者的一般习惯,即在发出关键词的之前或之后会有停顿而产生静音,通过检测关键词前后是否存在静音来检测该关键词是否为有效关键词,提高了检测准确率,避免了识别错误;同时在针对静音的检测中,利用了音频数据的静音片段的检测,提高了判断是否为静音的准确性,进一步避免了将关键词误判为非关键词。

[0192] 第七实施例

[0193] 本发明第七实施例提出一种关键词确认装置。图10是本申请第八实施例的关键词确认装置的方框图。如图10所示,本发明实施例的关键词确认装置包括:

[0194] 音频数据获取模块701,用于获取第一音频数据,所述第一音频数据被识别为关键词;

[0195] 累积静音片段确定模块702,用于确定与所述第一音频数据时间上连续的第二音频数据的多个片段的累积静音概率;

[0196] 累积关键词概率确定模块703,用于确定所述第一音频数据的多个片段的累积关键词概率;

[0197] 有效关键词确定模块704,用于当所述累积静音概率与累积关键词概率的关系满足第二预设条件时,确认所述第一音频数据为有效关键词。

[0198] 综上所述,本实施例提出的关键词确认装置至少具有如下优点:

[0199] 本发明实施例提出的关键词确认装置中,利用了使用者的一般习惯,即在发出关键词的之前或之后会有停顿而产生静音,通过检测关键词前后是否存在静音来检测该关键词是否为有效关键词,提高了检测准确率,避免了识别错误;同时在针对静音的检测中,利用了音频数据的静音片段的检测,提高了判断是否为静音的准确性,进一步避免了将关键词误判为非关键词。

[0200] 对于装置实施例而言,由于其与方法实施例基本相似,所以描述得比较简单,相关之处参见方法实施例的部分说明即可。

[0201] 图11为本申请一实施例提供的终端设备的硬件结构示意图。如图11所示,该终端设备可以包括输入设备90、处理器91、输出设备92、存储器93和至少一个通信总线94。通信总线94用于实现元件之间的通信连接。存储器93可能包含高速RAM存储器,也可能还包括非易失性存储NVM,例如至少一个磁盘存储器,存储器93中可以存储各种程序,用于完成各种处理功能以及实现本实施例的方法步骤。

[0202] 可选的,上述处理器91例如可以为中央处理器(Central Processing Unit,简称

CPU)、应用专用集成电路(ASIC)、数字信号处理器(DSP)、数字信号处理设备(DSPD)、可编程逻辑器件(PLD)、现场可编程门阵列(FPGA)、控制器、微控制器、微处理器或其他电子元件实现,该处理器91通过有线或无线连接耦合到上述输入设备90和输出设备92。

[0203] 可选的,上述输入设备90可以包括多种输入设备,例如可以包括面向用户的用户接口、面向设备的设备接口、软件的可编程接口、摄像头、传感器中至少一种。可选的,该面向设备的设备接口可以是用于设备与设备之间进行数据传输的有线接口、还可以是用于设备与设备之间进行数据传输的硬件插入接口(例如USB接口、串口等);可选的,该面向用户的用户接口例如可以是面向用户的控制按键、用于接收语音输入的语音输入设备以及用户接收用户触摸输入的触摸感知设备(例如具有触摸感应功能的触摸屏、触控板等);可选的,上述软件的可编程接口例如可以是供用户编辑或者修改程序的入口,例如芯片的输入引脚接口或者输入接口等;可选的,上述收发信机可以是具有通信功能的射频收发芯片、基带处理芯片以及收发天线等。麦克风等音频输入设备可以接收语音数据。输出设备92可以包括显示器、音响等输出设备。

[0204] 在本实施例中,该终端设备的处理器包括用于执行各设备中数据处理装置各模块的功能,具体功能和技术效果参照上述实施例即可,此处不再赘述。

[0205] 图12为本申请另一实施例提供的终端设备的硬件结构示意图。图12是对图11在实现过程中的一个具体的实施例。如图12所示,本实施例的终端设备包括处理器101以及存储器102。

[0206] 处理器101执行存储器102所存放的计算机程序代码,实现上述实施例中图1至图7的方法。

[0207] 存储器102被配置为存储各种类型的数据以支持在终端设备的操作。这些数据的示例包括用于在终端设备上操作的任何应用程序或方法的指令,例如消息,图片,视频等。存储器102可能包含随机存取存储器(random access memory,简称RAM),也可能还包括非易失性存储器(non-volatile memory),例如至少一个磁盘存储器。

[0208] 可选地,处理器101设置在处理组件100中。该终端设备还可以包括:通信组件103,电源组件104,多媒体组件105,音频组件106,输入/输出接口107和/或传感器组件108。终端设备具体所包含的组件等依据实际需求设定,本实施例对此不作限定。

[0209] 处理组件100通常控制终端设备的整体操作。处理组件100可以包括一个或多个处理器101来执行指令,以完成上述图1至图7方法的全部或部分步骤。此外,处理组件100可以包括一个或多个模块,便于处理组件100和其他组件之间的交互。例如,处理组件100可以包括多媒体模块,以方便多媒体组件105和处理组件100之间的交互。

[0210] 电源组件104为终端设备的各种组件提供电力。电源组件104可以包括电源管理系统,一个或多个电源,及其他与为终端设备生成、管理和分配电力相关联的组件。

[0211] 多媒体组件105包括在终端设备和用户之间的提供一个输出接口的显示屏。在一些实施例中,显示屏可以包括液晶显示器(LCD)和触摸面板(TP)。如果显示屏包括触摸面板,显示屏可以被实现为触摸屏,以接收来自用户的输入信号。触摸面板包括一个或多个触摸传感器以感测触摸、滑动和触摸面板上的手势。所述触摸传感器可以不仅感测触摸或滑动动作的边界,而且还检测与所述触摸或滑动操作相关的持续时间和压力。

[0212] 音频组件106被配置为输出和/或输入音频信号。例如,音频组件106包括一个麦克

风(MIC),当终端设备处于操作模式,如语音识别模式时,麦克风被配置为接收外部音频信号。所接收的音频信号可以被进一步存储在存储器102或经由通信组件103发送。在一些实施例中,音频组件106还包括一个扬声器,用于输出音频信号。

[0213] 输入/输出接口107为处理组件100和外围接口模块之间提供接口,上述外围接口模块可以是点击轮,按钮等。这些按钮可包括但不限于:音量按钮、启动按钮和锁定按钮。

[0214] 传感器组件108包括一个或多个传感器,用于为终端设备提供各个方面的状态评估。例如,传感器组件108可以检测到终端设备的打开/关闭状态,组件的相对定位,用户与终端设备接触的存在或不存在。传感器组件108可以包括接近传感器,被配置用来在没有任何的物理接触时检测附近物体的存在,包括检测用户与终端设备间的距离。在一些实施例中,该传感器组件108还可以包括摄像头等。

[0215] 通信组件103被配置为便于终端设备和其他设备之间有线或无线方式的通信。终端设备可以接入基于通信标准的无线网络,如WiFi,2G或3G,或它们的组合。在一个实施例中,该终端设备中可以包括SIM卡插槽,该SIM卡插槽用于插入SIM卡,使得终端设备可以登录GPRS网络,通过互联网与服务端建立通信。

[0216] 由上可知,在图12实施例中所涉及的通信组件103、音频组件106以及输入/输出接口107、传感器组件108均可以作为图11实施例中的输入设备的实现方式。

[0217] 本申请实施例提供了一种终端设备,包括:一个或多个处理器;和其上存储有指令的一个或多个机器可读介质,当由所述一个或多个处理器执行时,使得所述终端设备执行如本申请实施例中一个或多个所述的视频摘要的生成方法。

[0218] 在一实施例中,上述终端设备可以包括车载终端、移动终端(例如手机、平板电脑、个人数字助理等)、服务器、物联网设备或智能家居硬件等各种具有计算、处理功能的智能终端设备。智能家居硬件例如包括微波炉、烤箱、洗衣机、洗碗机、空调、路由器、智能音箱,电视,电冰箱,吸尘器等。上述智能终端设备可以安装应用程序,提供人机交互的操作界面,执行前述各实施例的关键词确认方法。

[0219] 例如,这些智能终端设备可以通过自身或者外接的音频接收部件接收音频数据,在确认该第一音频数据前后的第二音频数据为静音后,确认第一音频数据为有效关键词。例如,针对手机,通过这一方式可以判断使用者发出的语音指令是否为指示手机中安装的应用程序执行对应操作的指令——例如开启音乐、导航等;针对物联网设备或智能家居硬件,通过这一方式可以判断使用者发出的语音指令是否为指示其中安装的软件或者系统执行对应的操作的指令——例如连接其他设备、调高空调温度、开启烤箱的高温烘烤模式等。在此并不特别限制。因此,通过上述举例说明可知,本发明可以应用于各类终端设备。

[0220] 本说明书中的各个实施例均采用递进的方式描述,每个实施例重点说明的都是与其他实施例的不同之处,各个实施例之间相同相似的部分互相参见即可。

[0221] 尽管已描述了本申请实施例的优选实施例,但本领域内的技术人员一旦得知了基本创造性概念,则可对这些实施例做出另外的变更和修改。所以,所附权利要求意欲解释为包括优选实施例以及落入本申请实施例范围的所有变更和修改。

[0222] 最后,还需要说明的是,在本文中,诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来,而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且,术语“包括”、“包含”或者其任何其他变体意

在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者终端设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者终端设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括所述要素的过程、方法、物品或者终端设备中还存在另外的相同要素。

[0223] 以上对本申请所提供的一种关键词确认方法和装置,进行了详细介绍,本文中应用了具体个例对本申请的原理及实施方式进行了阐述,以上实施例的说明只是用于帮助理解本申请的方法及其核心思想;同时,对于本领域的一般技术人员,依据本申请的思想,在具体实施方式及应用范围上均会有改变之处,综上所述,本说明书内容不应理解为对本申请的限制。

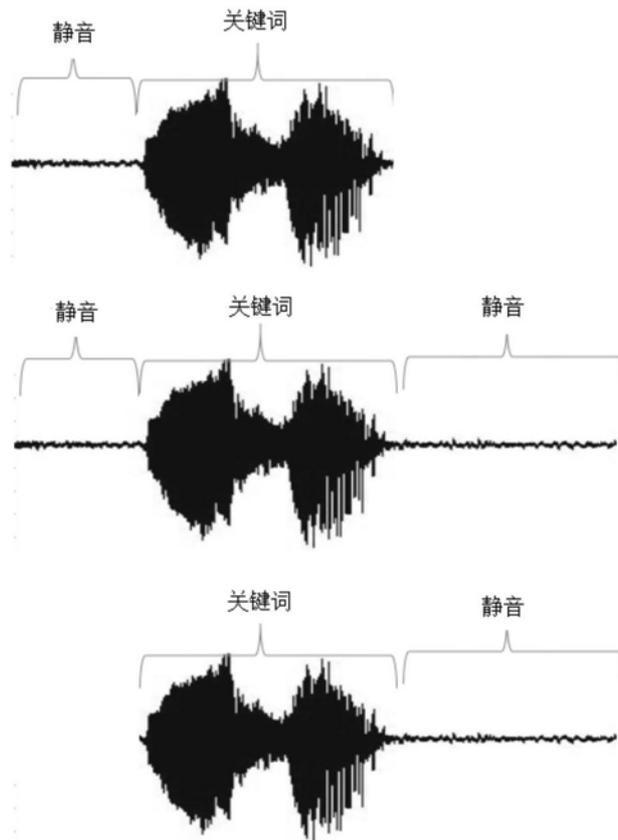


图1

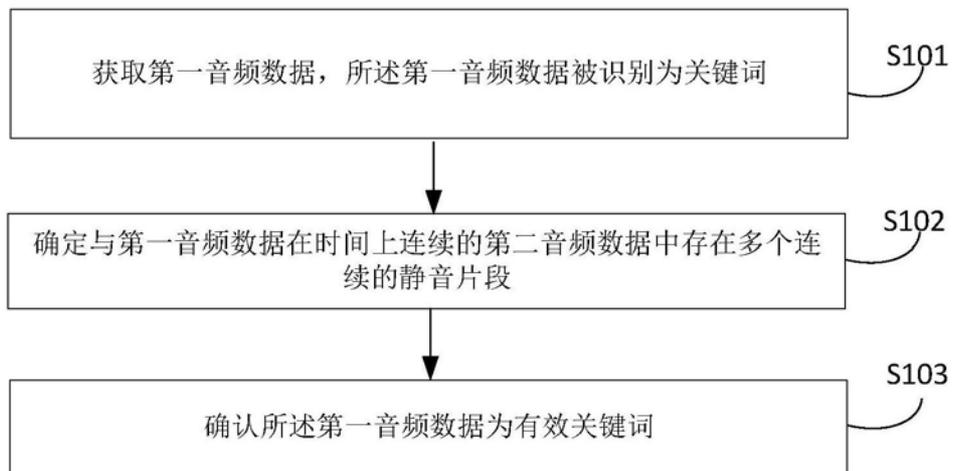


图2

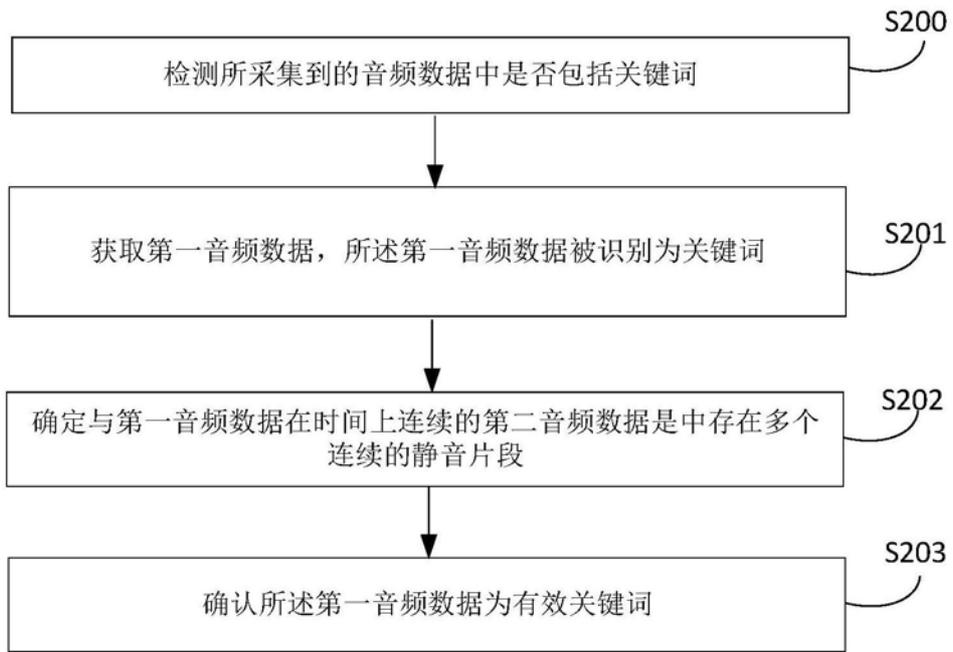


图3

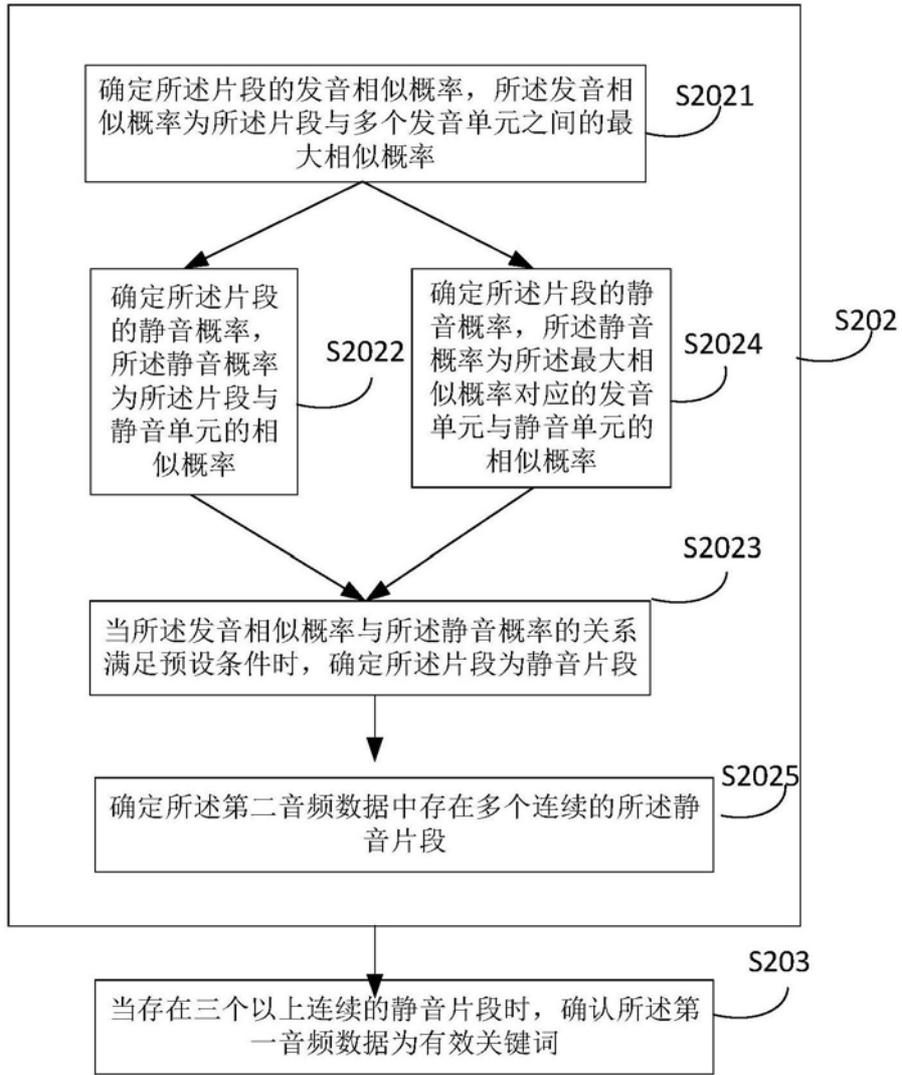


图4

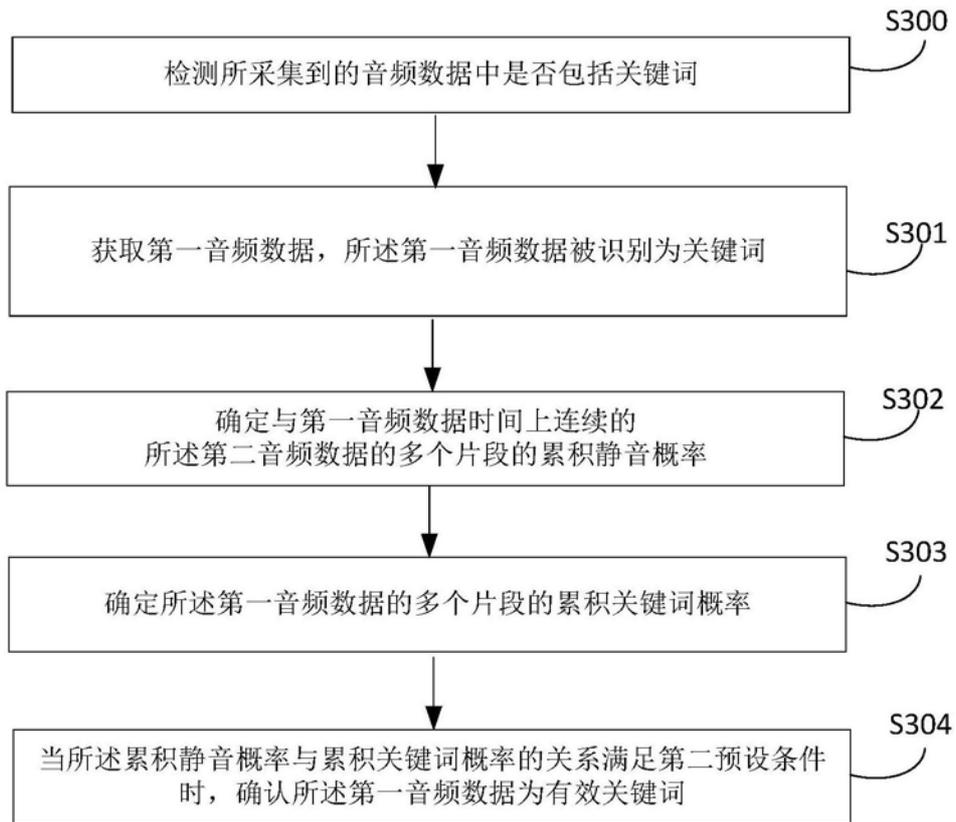


图5

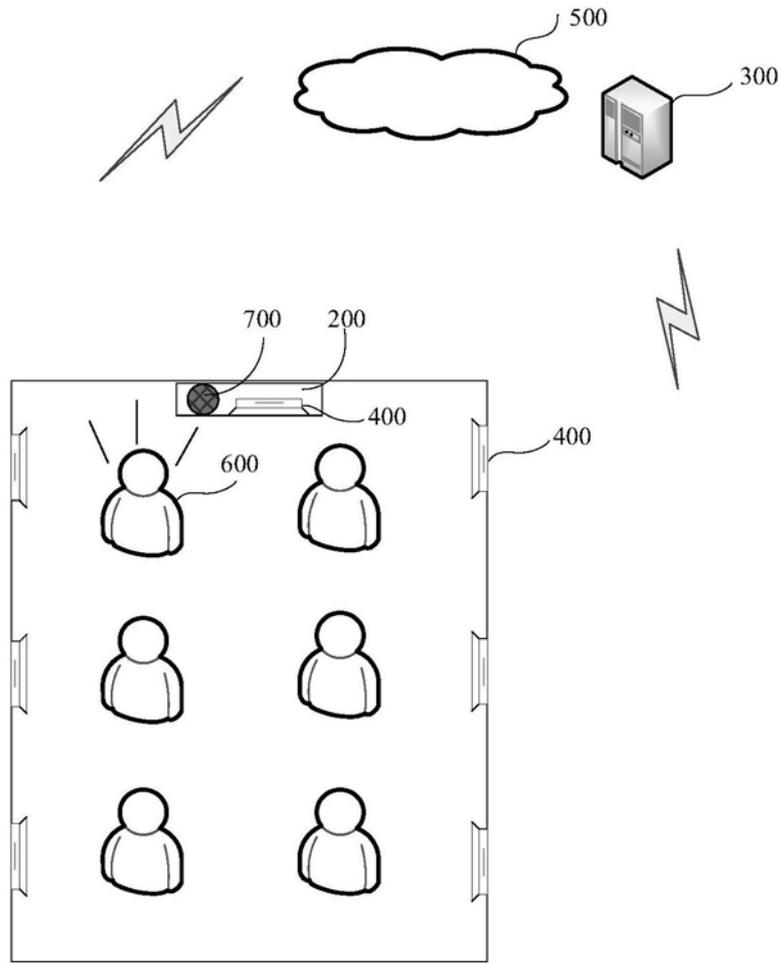


图6

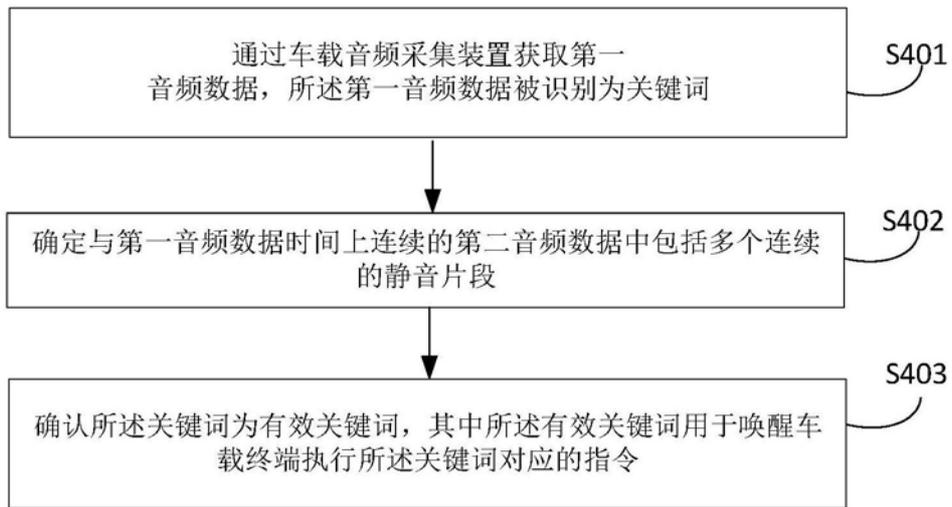


图7

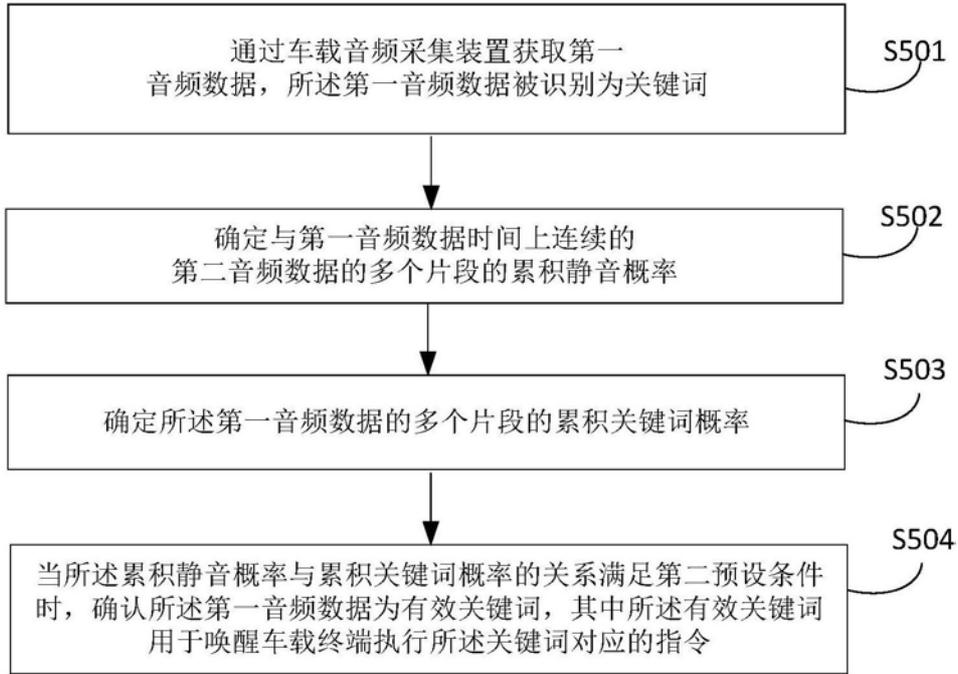


图8

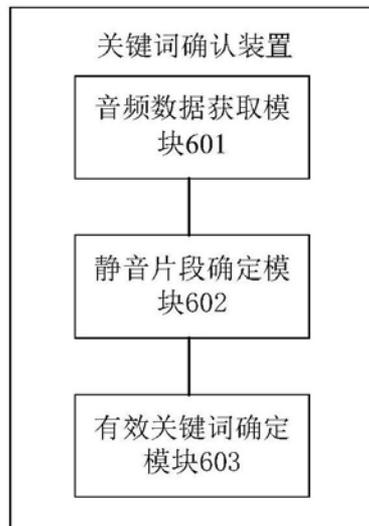


图9

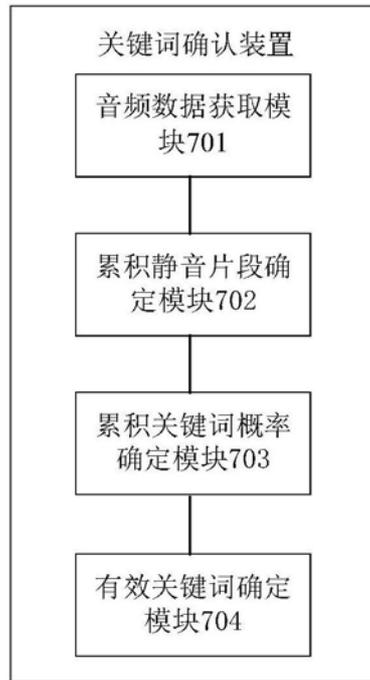


图10

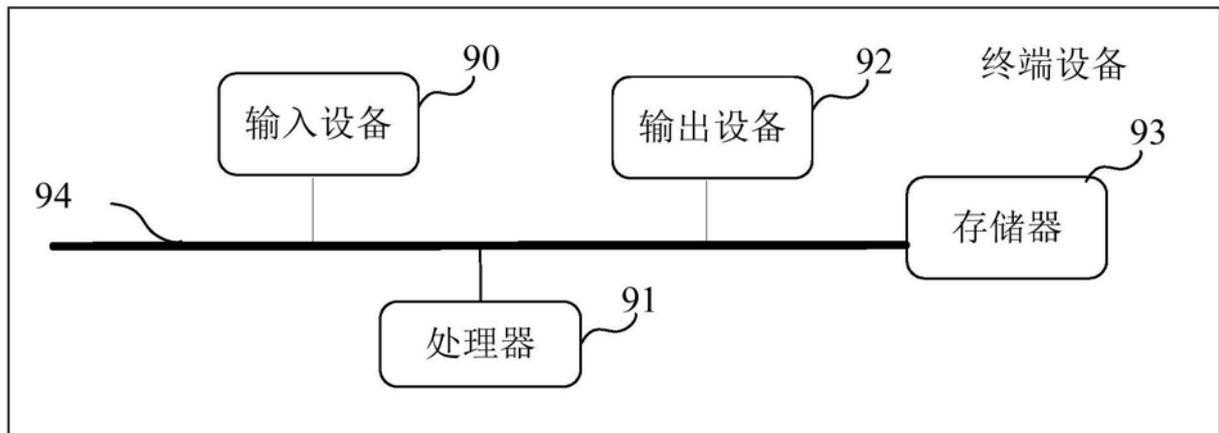


图11

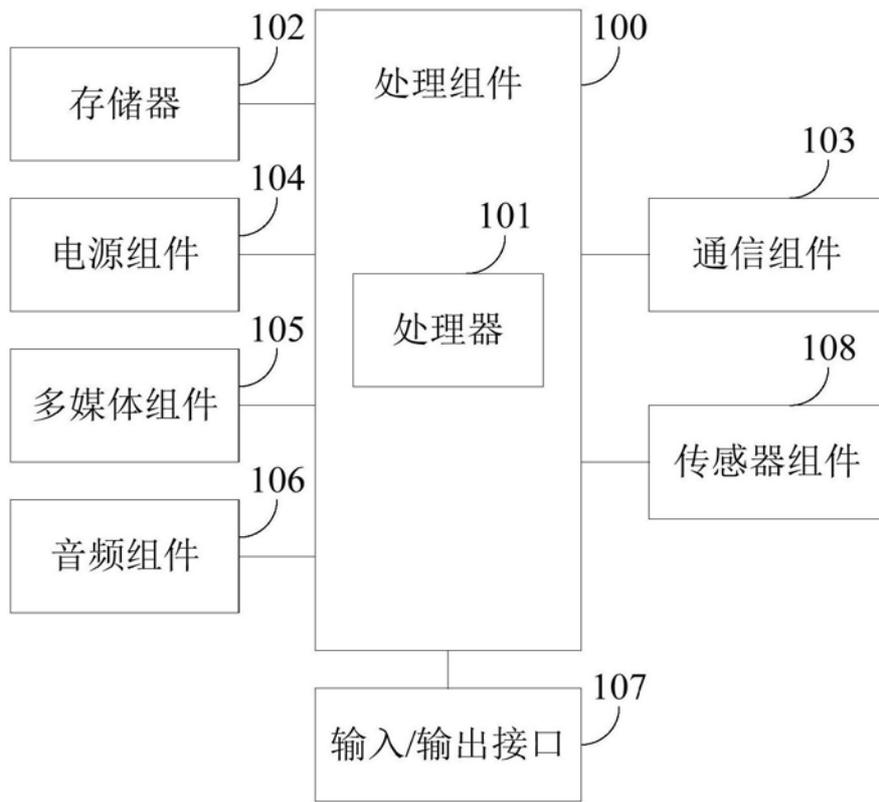


图12