



(12) 发明专利申请

(10) 申请公布号 CN 104123937 A

(43) 申请公布日 2014. 10. 29

(21) 申请号 201310157869. 2

(22) 申请日 2013. 04. 28

(71) 申请人 腾讯科技(深圳)有限公司

地址 518000 广东省深圳市福田区振兴路赛格科技园 2 栋东 403 室

(72) 发明人 卢鲤 饶丰 刘松 唐宗尧 张翔 岳帅 陈波

(74) 专利代理机构 北京三高永信知识产权代理有限公司 11138

代理人 罗振安

(51) Int. Cl.

G10L 15/26(2006. 01)

G10L 15/06(2013. 01)

G10L 15/30(2013. 01)

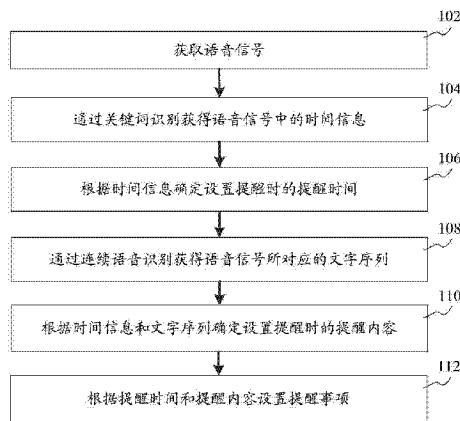
权利要求书4页 说明书19页 附图8页

(54) 发明名称

提醒设置方法、装置和系统

(57) 摘要

本发明公开了一种提醒设置方法、装置和系统,属于语音识别领域。所述方法包括:获取语音信号,通过关键词识别获得所述语音信号中的时间信息;根据所述时间信息确定设置提醒时的提醒时间;通过连续语音识别获得所述语音信号所对应的文字序列,根据所述时间信息和文字序列确定设置提醒时的提醒内容;根据所述提醒时间和所述提醒内容设置提醒事项。本发明通过关键词识别来获得语音信号中的时间信息,可以保证时间信息提取的正确性;达到了即便语音识别的全文识别准确率较差,导致识别出的文字序列是错误的情况下,仍然可以通过关键词识别获得正确的时间信息来设置提醒事项的效果。



1. 一种提醒设置方法,其特征在于,所述方法包括:
  - 获取语音信号;
  - 通过关键词识别获得所述语音信号中的时间信息;
  - 根据所述时间信息确定设置提醒时的提醒时间;
  - 通过连续语音识别获得所述语音信号所对应的文字序列,根据所述时间信息和所述文字序列确定设置提醒时的提醒内容;
  - 根据所述提醒时间和所述提醒内容设置提醒事项。
2. 根据权利要求 1 所述的方法,其特征在于,所述根据所述时间信息确定设置提醒时的提醒时间,包括:
  - 从预设的至少一个时间信息模板中匹配出与所述时间信息对应的时间信息模板;
  - 根据与所述时间信息对应的时间信息模板将所述时间信息转换为标准格式的时间;
  - 将所述标准格式的时间作为设置提醒时的提醒时间。
3. 根据权利要求 2 所述的方法,其特征在于,所述将所述标准格式的时间作为设置提醒时的提醒时间点,包括:
  - 检测与所述标准格式的时间匹配的未来时刻是否为两个或者两个以上;
  - 若检测结果为与所述标准格式的时间匹配的未来时刻是两个或者两个以上,则从所述两个或者两个以上的未来时刻中选择符合第一预设条件的未来时刻作为设置提醒时的提醒时间。
4. 根据权利要求 1 所述的方法,其特征在于,所述根据所述时间信息和所述文字序列确定设置提醒时的提醒内容,包括:
  - 根据所述时间信息和所述文字序列确定所述文字序列中的内容信息,所述内容信息是所述文字序列中与所述时间信息不对应的部分的子集;
  - 将所述内容信息和 / 或所述内容信息在所述语音信号中所对应的语音片段作为设置提醒时的提醒内容。
5. 根据权利要求 4 所述的方法,其特征在于,所述时间信息包括时间关键词和所述时间关键词在所述语音信号中的起止时刻,所述文字序列包括至少一个文字和每个文字在所述语音信号中的起止时刻,所述根据所述时间信息和所述文字序列确定所述文字序列中的内容信息,包括:
  - 根据所述时间关键词在所述语音信号中的起止时刻和所述各个文字在所述语音信号中的起止时刻,将所述文字序列区分为与所述时间信息对应的部分和与所述时间信息不对应的部分;
  - 若所述与所述时间信息不对应的部分为一段,则将所述与所述时间信息不对应的部分作为所述内容信息;
  - 若所述与所述时间信息不对应的部分为两段或者两段以上,则将所述与所述时间信息不对应的部分中符合第二预设条件的一段作为所述内容信息。
6. 根据权利要求 1 所述的方法,其特征在于,所述获取语音信号,包括:
  - 接收终端发送的所述语音信号;
  - 所述根据所述提醒时间点和所述提醒内容设置提醒事项,包括:
    - 将所述提醒时间点和所述提醒内容发送给所述终端,以便所述终端根据所述提醒时间

点和所述提醒内容设置提醒事项。

7. 根据权利要求1至6任一项所述的方法,其特征在于,所述时间信息包括时间关键词和所述时间关键词在所述语音信号中的起止时刻,所述通过关键词识别获得语音信号中的时间信息,包括:

将所述语音信号中的各个片段输入关键词检索网络,所述关键词检索网络包括有至少一个前景模型和至少一个背景模型;

所述前景模型为基于单音素或者三音素的隐马尔可夫模型;

所述背景模型为基于音素混淆度对所述前景模型中的音素进行聚类后训练得到的模型;或者,所述背景模型为通过与所述前景模型所采用的语种不同的其它语种训练得到的模型;

将所述前景模型和所述背景模型分别与所述各个片段进行匹配,得到所述时间关键词和所述时间关键词在所述语音信号中的起止时刻,将所述时间关键词和所述时间关键词在所述语音信号中的起止时刻作为所述语音信号所对应的的时间信息。

8. 根据权利要求1至6任一项所述的方法,其特征在于,所述文字序列包括至少一个文字和每个文字在所述语音信号中的起止时刻,所述通过连续语音识别获得所述语音信号所对应的文字序列,包括:

加载搜索空间;

提取所述语音信号中的音频特征;

通过维特比解码计算所述音频特征在所述搜索空间中的最优文字序列,所述最优文字序列包括至少一个文字和每个文字在所述语音信号中的起止时刻,将所述最优文字序列作为所述语音信号所对应的文字序列。

9. 一种提醒设置装置,其特征在于,所述装置包括:

语音获取模块,用于获取语音信号;

第一识别模块,用于通过关键词识别获得所述语音获取模块获取到的语音信号中的时间信息;

时间确定模块,用于根据所述第一识别模块识别出的所述时间信息确定设置提醒时的提醒时间;

第二识别模块,用于通过连续语音识别获得所述语音获取模块获取到的语音信号所对应的文字序列;

内容确定模块,用于通过所述第一识别模块获得的时间信息和通过所述第二识别模块获得的文字序列确定设置提醒时的提醒内容;

提醒设置模块,用于通过所述时间确定模块确定的提醒时间和所述内容确定模块确定的提醒内容设置提醒事项。

10. 根据权利要求9所述的装置,其特征在于,所述时间确定模块,包括:匹配单元、转换单元和时间确定单元;

所述匹配单元,用于从预设的至少一个时间信息模板中匹配出与所述时间信息对应的的时间信息模板;

所述转换单元,用于根据所述匹配单元匹配出的与所述时间信息对应的的时间信息模板将所述时间信息转换为标准格式的时间;

所述时间确定单元,用于将所述转换单元转换到的标准格式的时间作为设置提醒时的提醒时间。

11. 根据权利要求 10 所述的装置,其特征在于,所述时间确定单元,包括:

检测子单元和确定子单元;

所述检测子单元,用于检测与所述标准格式的时间匹配的未来时刻是否为两个或者两个以上;

所述确定子单元,用于若所述检测子单元的检测结果为与所述标准格式的时间匹配的未来时刻是两个或者两个以上,则从所述两个或者两个以上的未来时刻中选择符合第一预设条件的未来时刻作为设置提醒时的提醒时间。

12. 根据权利要求 9 所述的装置,其特征在于,所述内容确定模块,包括:内容信息确定单元和提醒内容确定单元;

所述内容信息确定单元,用于根据所述时间信息和所述文字序列确定所述文字序列中的内容信息,所述内容信息是所述文字序列中与所述时间信息不对应的部分的子集;

所述提醒内容确定单元,用于将所述内容信息确定单元区分出的内容信息和 / 或所述内容信息在所述语音信号中所对应的语音片段作为设置提醒时的提醒内容。

13. 根据权利要求 12 所述的装置,其特征在于,所述第一识别模块识别出的时间信息包括时间关键词和所述时间关键词在所述语音信号中的起止时刻,所述第二识别模块识别出的文字序列包括至少一个文字和每个文字在所述语音信号中的起止时刻,所述内容信息确定单元,包括:区分子单元、第一确定子单元和第二确定子单元;

所述区分子单元,用于根据所述第一识别模块识别出的时间关键词在所述语音信号中的起止时刻和所述第二识别模块识别出的各个文字在所述语音信号中的起止时刻,将所述文字序列区分为与所述时间信息对应的部分和与所述时间信息不对应的部分;

所述第一确定子单元,用于若所述区分子单元区分出的与所述时间信息不对应的部分为一段,则将所述与所述时间信息不对应的部分作为所述内容信息;

所述第二确定子单元,用于若所述区分子单元区分出的与所述时间信息不对应的部分为两段或者两段以上,则将所述与所述时间信息不对应的部分中符合第二预设条件的一段作为所述内容信息。

14. 根据权利要求 9 的装置,其特征在于:

所述语音获取模块,用于接收终端发送的所述语音信号;

所述提醒设置模块,用于将所述时间确定模块确定的提醒时间和所述内容确定模块确定的提醒内容发送给所述终端,以便所述终端根据所述提醒时间和所述提醒内容设置提醒事项。

15. 根据权利要求 9 至 14 任一所述的装置,其特征在于,所述时间信息包括时间关键词和所述时间关键词在所述语音信号中的起止时刻,所述第一识别模块,包括:输入单元和识别单元;

所述输入单元,用于将所述语音信号中的各个片段输入关键词检索网络,所述关键词检索网络包括有至少一个前景模型和至少一个背景模型;

所述前景模型为基于单音素或者三音素的隐马尔可夫模型;

所述背景模型为基于音素混淆度对所述前景模型中的音素进行聚类后训练得到的模

型；或者，所述背景模型为通过与所述前景模型所采用的语种不同的其它语种训练得到的模型；

所述识别单元，用于将所述前景模型和所述背景模型分别与所述各个片段进行匹配，得到所述时间关键词和所述时间关键词在所述语音信号中的起止时刻，将所述时间关键词和所述时间关键词在所述语音信号中的起止时刻作为所述语音信号所对应的时间信息。

16. 根据权利要求9至14任一所述的装置，其特征在于，所述文字序列包括至少一个文字和每个文字在所述语音信号中的起止时刻，所述第二识别模块，包括：

加载单元、提取单元和解码单元；

所述加载单元，用于加载搜索空间；

所述提取单元，用于提取所述语音信号中的音频特征；

所述解码单元，用于通过维特比解码计算所述提取单元提取到的音频特征在所述加载单元加载的搜索空间中的最优文字序列，所述最优文字序列包括至少一个文字和每个文字在所述语音信号中的起止时刻，将所述最优文字序列作为所述语音信号所对应的文字序列。

17. 一种提醒设置系统，其特征在于，所述系统包括：终端和服务器；

所述终端通过无线网络或者有线网络与所述服务器相连；

所述服务器中包括如权利要求9至16任一项所述的提醒设置装置。

## 提醒设置方法、装置和系统

### 技术领域

[0001] 本发明涉及语音识别领域,特别涉及一种提醒设置方法、装置和系统。

### 背景技术

[0002] 随着语音识别技术和移动互联网的快速发展,在诸如智能手机、平板电脑、智能电视之类的电子设备上已经越来越多地提供基于语音输入的应用程序,这些基于语音输入的应用程序可以根据用户输入的语音信号来提供服务。

[0003] 根据一定范围内的用户调研结果表明,用户使用基于语音输入的应用程序时,使用频率最高的三项功能分别为:通过语音输入设置提醒事项、通过语音输入查询天气和通过语音输入定位自身的地理位置。

[0004] 以用户在智能手机上通过语音输入设置提醒事项为例,目前的提醒设置方法包括:首先,智能手机采集用户输入的语音信号,该语音信号用于设置对应于某一个提醒时间的提醒事项,比如,该语音信号可以是“明天上午 8 点叫我起床”;智能手机将该语音信号转发给服务器后,服务器通过“连续语音识别”和“语义分析”两个阶段来处理该语音信号,也即服务器首先通过连续语音识别将该语音信号全部识别为对应的文字序列,然后通过语义分析技术从文字序列中提取时间信息“明天上午 8 点”和提醒内容“叫我起床”,并将提取结果反馈给智能手机;智能手机根据接收到的时间信息“明天上午 8 点”和提醒内容“叫我起床”设置相应的提醒事项。

[0005] 在实现本发明的过程中,发明人发现背景技术至少存在以下问题:

[0006] 第一,服务器在将语音信号全部识别为文字序列的过程中,全文识别准确率并不稳定,比如:在环境噪音较大的情况下,全文识别准确率会明显下降;又比如:由于连续语音识别的基本解码原理是寻求全局最优解,所以若语音信号的初始部分识别错误,则后续部分被识别错误的概率会非常大;

[0007] 第二,服务器从识别出的文字序列中提取时间信息和提醒内容的过程中,通常基于模板来匹配文字序列,按照模板匹配结果中的时间区域提取时间信息,按照模板匹配结果中的事件区域提取提醒内容。具体实现时,需要预先收集各种可能的文字序列模板,受限于收集工作的局限,最终的文字序列模板库无法覆盖所有可能的文字序列形式,使得即便语音信号被准确无误地识别为对应的文字序列,也可能因文字序列模板无法完全匹配而提取出不正确的时间信息,导致提醒事项的设置失败或者设置错误。

### 发明内容

[0008] 为了解决服务器在时间信息提取不正确时,导致提醒事项的设置失败或者设置错误的问题,本发明实施例提供了一种提醒设置方法、装置和系统。所述技术方案如下:

[0009] 第一方面,提供了一种提醒设置方法,所述方法包括:

[0010] 获取语音信号;

[0011] 通过关键词识别获得语音信号中的时间信息;

- [0012] 根据所述时间信息确定设置提醒时的提醒时间；
- [0013] 通过连续语音识别获得所述语音信号所对应的文字序列；
- [0014] 根据所述时间信息和所述文字序列确定设置提醒时的提醒内容；
- [0015] 根据所述提醒时间和所述提醒内容设置提醒事项。
- [0016] 第二方面,提供了一种提醒设置装置,所述装置包括：
- [0017] 语音获取模块,用于获取语音信号；
- [0018] 第一识别模块,用于通过关键词识别获得所述语音获取模块获取到的语音信号中的时间信息；
- [0019] 时间确定模块,用于根据所述第一识别模块识别出的所述时间信息确定设置提醒时的提醒时间；
- [0020] 第二识别模块,用于通过连续语音识别获得所述语音获取模块获取到的语音信号所对应的文字序列；
- [0021] 内容确定模块,用于通过所述第一识别模块获得的时间信息和通过所述第二识别模块获得的文字序列确定设置提醒时的提醒内容；
- [0022] 提醒设置模块,用于通过所述时间确定模块确定的提醒时间和所述内容确定模块确定的提醒内容设置提醒事项。
- [0023] 第三方面,提供了一种提醒设置系统,所述系统包括：终端和服务器；
- [0024] 所述终端通过无线网络或者有线网络与所述服务器相连；
- [0025] 所述服务器包括如第二方面所述的提醒设置装置。
- [0026] 本发明实施例提供的技术方案带来的有益效果是：
- [0027] 通过关键词识别来获得语音信号中的时间信息,可以保证时间信息提取的正确性；减少了在时间信息提取不正确时,导致提醒事项的设置失败或者设置错误的问题；达到了即便语音识别的全文识别准确率较差,导致识别出的文字序列是错误的情况下,仍然可以通过关键词识别获得正确的时间信息来设置提醒事项的效果。

#### 附图说明

[0028] 为了更清楚地说明本发明实施例中的技术方案,下面将对实施例描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅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。

- [0029] 图 1 是本发明一个实施例提供的提醒设置方法的方法流程图；
- [0030] 图 2A 是本发明另一实施例提供的提醒设置方法的方法流程图；
- [0031] 图 2B 是本发明实施例所涉及的关键词检索网络的结构示意图；
- [0032] 图 2C 是本发明实施例所涉及的前景模型的训练方法的方法流程图；
- [0033] 图 2D 是本发明实施例所涉及的背景模型的训练方法的方法流程图；
- [0034] 图 2E 是本发明实施例所涉及的连续语音识别过程的方法流程图；
- [0035] 图 2F 是图 2A 所对应实施例中两路语音识别的结果对齐示意图；
- [0036] 图 3A 是本发明再一个实施例提供的提醒设置方法的方法流程图；
- [0037] 图 3B 是图 3A 所对应实施例中终端采集语音信号时的界面示意图；

- [0038] 图 3C 是图 3A 所对应实施例中两路语音识别的结果对齐示意图；
- [0039] 图 3D 是图 3A 所对应实施例中终端成功设置提醒事项后的界面示意图；
- [0040] 图 4 是本发明一个实施例提供的提醒设置装置的结构方框图；
- [0041] 图 5 是本发明另一实施例提供的提醒设置装置的结构方框图；
- [0042] 图 6 是本发明再一实施例提供的提醒设置装置的结构方框图；
- [0043] 图 7 是本发明一个实施例提供的提醒设置系统的结构方框图；
- [0044] 图 8 是本发明一个实施例提供的电子设备的结构方框图；
- [0045] 图 9 是本发明另一实施例提供的服务器的结构方框图。

### 具体实施方式

[0046] 为使本发明的目的、技术方案和优点更加清楚，下面将结合附图对本发明实施方式作进一步地详细描述。

[0047] 发明人研究发现，通过语音输入设置提醒事项时，时间信息的重要性要大于提醒内容的重要性。比如，语音信号为“明天下午 3 点提醒我测体温”，如果时间信息识别错误，导致识别出的文字序列是“明天下午闪电提醒我测体温”，则提醒事项会设置失败；但如果仅仅是提醒内容识别错误，导致识别出的文字序列是“明天下午 3 点提醒我去提问”，则提醒事项仍然会设置成功。即便在明天下午 3 点提醒用户“去提问”，用户也会根据该错误的谐音文字获知正确的提醒内容是“测体温”。所以，在语音输入设置提醒事项的过程中，需要尽量保证时间信息的识别准确性。为此，本文实施例中同时提供两路并行的语音识别，一路语音识别为专用于识别时间信息的关键词识别，另一路语音识别为用于将整个语音信号识别为文字序列的连续语音识别。具体描述如下：

[0048] 请参考图 1，其示出了本发明一个实施例提供的提醒设置方法的方法流程图。该提醒设置方法可以用于提供基于语音输入设置提醒事项功能的独立电子设备中，该电子设备可以是智能手机、智能电视、平板电脑或者电子书阅读器之类的终端；该提醒设置方法也可以用于服务器中，该服务器属于提供基于语音输入设置提醒事项功能的电子系统，该电子系统不仅包括服务器，还包括至少一个终端。该提醒设置方法，包括：

[0049] 步骤 102，获取语音信号；

[0050] 该语音信号是用户通过语音输入的用于设置提醒事项的信号，可以由电子设备通过麦克风采集；也可以由终端通过麦克风采集后，发送给服务器。

[0051] 步骤 104，通过关键词识别获得语音信号中的时间信息；

[0052] 关键词识别是从连续语音流中识别出关键词的语音识别技术。本实施例中的关键词均为有关时间的关键词，比如：年、月、日、时、分、秒、x 点、x 号、傍晚、早上、上午、下午和明天等等。

[0053] 通过关键词识别获得步骤 102 中获取的语音信号中的时间信息，该时间信息是根据语音信号中有关时间的语音片段识别出的、以文字形式来表示的信息。时间信息包括时间关键词和时间关键词在语音信号中的起止时刻。

[0054] 步骤 106，根据时间信息确定设置提醒时的提醒时间；

[0055] 提醒时间是提醒事项的触发时间。换句话说，在当前时刻是提醒时间时，提醒事项会被触发执行。



- [0056] 步骤 108,通过连续语音识别获得语音信号所对应的文字序列;
- [0057] 连续语音识别是从连续语音流中识别出连续文字序列的语音识别技术。
- [0058] 通过连续语音识别可以获得步骤 102 中获取的语音信号所对应的文字序列。文字序列是根据语音信号中所有语音片段识别出的、包括至少一个文字的序列。文字序列包括至少一个文字和每个文字在语音信号中的起止时刻。
- [0059] 步骤 110,根据时间信息和文字序列确定设置提醒时的提醒内容;
- [0060] 提醒内容包括内容信息和 / 或内容信息在语音信号中对应的语音片段,内容信息是文字序列中与时间信息不对应的部分的子集。
- [0061] 需要说明的是,步骤 108 可以在步骤 104 之前、同时或者之后进行。
- [0062] 步骤 112,根据提醒时间和提醒内容设置提醒事项。
- [0063] 综上所述,本实施例提供的提醒设置方法,通过关键词识别来获得语音信号中的时间信息,可以保证时间信息提取的正确性;减少了在时间信息提取不正确时,导致提醒事项的设置失败或者设置错误的问题;达到了即便语音识别的全文识别准确率较差,导致识别出的文字序列是错误的情况下,仍然可以通过关键词识别获得正确的时间信息来设置提醒事项的效果。
- [0064] 请参考图 2A,其示出了本发明另一实施例提供的提醒设置方法的方法流程图。本实施例以该提醒设置方法应用于电子设备中实现,并以该电子设备为平板电脑来举例说明。该提醒设置方法,包括:
- [0065] 步骤 201,获取语音信号;
- [0066] 平板电脑运行有用于提供基于语音输入设置提醒事项功能的应用程序,该应用程序在检测到预定触发信号后,就可以触发平板电脑通过麦克风采集语音信号。该预定触发信号包括预定按键信号、预定声控信号、预定传感器信号和预定滑动轨迹信号中的至少一种。
- [0067] 比如,该预定触发信号是“摇一摇”信号,则该应用程序通过平板电脑内部的传感器检测到“摇一摇”信号后,平板电脑通过麦克风获取语音信号。此时,平板电脑可以通过内置或者外置麦克风采集用户输入的语音信号。
- [0068] 假设平板电脑采集到的语音信号的内容为“明天下午 3 点去会议室开会”。
- [0069] 步骤 202,通过关键词识别获得语音信号中的时间信息;
- [0070] 平板电脑通过关键词识别获得语音信号中的时间信息,时间信息具体包括时间关键词和时间关键词在语音信号中的起止时刻。具体来讲,平板电脑首先将语音信号中的各个片段输入关键词检索网络,该关键词检索网络包括有至少一个前景模型和至少一个背景模型;将前景模型和背景模型分别与各个片段进行匹配,得到时间关键词和时间关键词在语音信号中的起止时刻,将时间关键词和时间关键词在语音信号中的起止时刻作为语音信号所对应的时间信息。
- [0071] 参考图 2B,其示出了关键词检索网络的结构示意图。该关键词检索网络可以包括 M 个前景模型 KW (keyword,关键词)和 N 个背景模型 Filler。其中,前景模型用于识别语音信号中的时间关键词,背景模型用于识别语音信号中的非时间关键词,并与前景模型识别的时间关键词进行区分。换句话说,每个前景模型 KW 可以对应一个时间关键词,比如:x 年、x 月、x 日、x 时、x 分、x 秒、x 点、x 号、傍晚、早上、上午、下午和明天等等;而背景模型

Filler 则对应于语音信号中的非时间关键词。

[0072] 在语音信号的各个片段输入关键词检索网络后,对应于时间关键词的片段会被前景模型 KW 所识别,对应于非时间关键词的片段会被背景模型 Filler 所识别,背景模型识别出的非时间关键词可以对前景模型识别出的时间关键词形成更好的区分,通过对前景模型 KW 所识别出的关键词进行关键词确认,可以获得语音信号中的时间关键词,和时间关键词在语音信号中的起止时刻。起止时刻包括起始时刻和终止时刻。

[0073] 比如,平板电脑可以将语音信号“明天下午 3 点去会议室开会”输入关键词检索网络来识别时间关键词“明天下午 3 点”,和时间关键词“明天下午 3 点”在语音信号中的起始时刻和终止时刻。

[0074] 该关键词检索网络中的前景模型可以是基于单音素或者三音素的 HMM (Hidden Markov Model, 隐马尔可夫模型)。音素是语音的最小发音单元,三音素是根据一个音素的上下文的不同来决定的建模单元。前景模型的训练方式可以采用单音素或者三音素的 HMM 模型训练流程,大致包括如下几个步骤,如图 2C 所示:

[0075] 202a,为每个音素训练初始的种子 HMM 模型;

[0076] 音素是语音的最小发音单元,每个语种可以由若干个音素构成音素集合,该音素集合可以通过预先准备获得。HMM 模型的拓扑结构是声学模型常见的一种结构,不再赘述。

[0077] 202b,通过大量的标注数据对每一个音素的种子 HMM 模型进行模型参数更新,获得更新后的单音素模型;

[0078] 其中,标注数据是预先准备的音频数据,通过采集某一语种的音频数据,可以得到音频数据的字一级的标注数据,同时还需要提取标注数据的音频特征。本实施例中,音频特征采用 PLP (Perceptual Linear Predictive, 感知线性预测) 特征。在其它实施例中,还可能采用其它形式的音频特征,并不具体限定为 PLP 特征。

[0079] 202c,通过上下文扩展将每个音素与上下文的相邻音素相关联,使得每个单音素模型变成上下文链接的三音素模型;

[0080] 202d,通过决策树聚类对三音素模型进行聚类,适当减少三音素模型的模型参数;

[0081] 202e,基于聚类后的三音素模型通过高斯分裂训练重新进行模型参数更新;

[0082] 202f,对更新后的三音素模型进行 MPE (Minimum Phone Error, 最小音素错误) 鉴别性训练,获得具有更好鲁棒性的模型参数。至此,可以获得具有较为精准的前景模型。

[0083] 该关键词检索网络中的背景模型并非是常见的通过非关键词训练的模型,在本实施例中,背景模型为基于音素混淆度对前景模型中的音素进行聚类后训练得到的模型;或者,背景模型为通过与前景模型所采用的语种不同的其它语种训练得到的模型,比如前景模型是采用中文普通话训练的模型,则背景模型是采用粤语或者越南语等不同于中文普通话,但是有一定相似性的语种来训练的模型。这两种背景模型可以模糊掉一些单音素的发音个性,从而和前景模型中的关键词音素形成更为鲜明的对比,可以尽量避免背景模型对语音信号中的关键词的吸收,提高关键词的检出召回率。检出召回率包括检出率和召回率,检出率和召回率都用于表征关键词检索网络的检出性能,检出率是指正确识别的关键词数与参考关键词总数的比值;召回率则是指对于某一个查询项,检测出的正确识别的关键词数与参考关键词总数的比值。

[0084] 在背景模型为基于音素混淆度对前景模型中的音素进行聚类后训练得到的模型时,该背景模型的训练方式如图 2D 所示:

[0085] 202A,获得前景模型中的各个音素模型;

[0086] 由于前景模型在训练时,基于单音素或者三音素来训练音素模型,可以获得各个音素对应的单音素模型或者三音素模型。

[0087] 202B,通过计算各个音素模型之间的 GMM (Gaussian mixture model,混合高斯模型) 距离,获得相似音素之间的音素混淆矩阵;

[0088] 202C,通过音素混淆矩阵将相似程度达到阈值的 n 个音素聚类后,用一个音素代替,获得粗粒度的背景模型。

[0089] 比如音素“b”和“d”比较相似,可以统一用一个音素“b”代替;又比如,音素“z”和“zh”比较相似,可以统一用一个音素“z”代替。

[0090] 步骤 203,根据时间信息确定设置提醒时的提醒时间;

[0091] 平板电脑可以根据关键词识别获得的时间信息确定设置提醒时的提醒时间。由于识别出的时间信息可能存在不同的格式,所以本步骤还可以包括如下几个子步骤:

[0092] 第一,从预设的至少一个时间信息模板中匹配出与时间信息对应的时间信息模板;

[0093] 平板电脑可以将识别出的时间信息与预设的至少一个时间信息模板进行匹配,从该至少一个时间信息模板中找出时间格式及语义与所述时间信息相同或相似的一个时间信息模板,作为与该时间信息相匹配的时间信息模板。本实施例中,所述时间信息模板是预先通过包含大量时间信息的语料库来搜集的。这个过程中,首先将搜集时间信息的所有可能的语料作为样本库;然后分析和总结这些语料的序列特点;最后根据分析结果获得时间信息模板。比如一种常见的的时间信息模板为“日期+时钟”,与之符合的时间信息可以是“明天 8 点”;另一种常见的的时间信息模板为“月+日期”,与之符合的时间信息可以是“4 月 8 号”。

[0094] 需要说明的是,虽然用户在语音输入时对时间信息的口语说法可能非常多,但是相较于传统的语义分析技术需要对所有可能的语句(不局限于时间)搜集模板来讲,仅搜集有关时间信息的模板的工作量要小很多,而且搜集到的模板可以基本覆盖用户对时间信息的所有口语说法。

[0095] 第二,根据与所述时间信息对应的的时间信息模板将所述时间信息转换为标准格式的时间;

[0096] 其中,该标准格式的时间通常为某一种标准时间格式的时间,比如,当所述与时间信息对应的的时间信息模板为“星期+时钟”,标准时间格式为“年/月/日/时/分/秒”时,假设当前时刻为“2013/4/18/10:24:10”,则平板电脑可以将时间信息“下周三 8 点”转化为标准格式的时间“2013/4/24/8:00:00”。

[0097] 第三,将标准格式的时间作为设置提醒时的提醒时间。

[0098] 然后平板电脑将标准格式的时间作为设置提醒时的提醒时间,比如将“2013/4/24/8:00:00”作为设置提醒时的提醒时间。

[0099] 由于识别出的时间信息有时并不是一个完整的时间信息,就会导致存在多个未来时刻与之对应。比如语音信号为“8 点提醒我开会”,则识别出的时间信息在转换成标准格

式的时间后为“xxxx/xx/xx/8:00”，存在多个未来时刻与之对应。为此，本步骤还可以具体包括：

[0100] 1) 平板电脑检测与标准格式的时间匹配的未来时刻是否为两个或者两个以上；

[0101] 平板电脑可以检测与标准格式的时间匹配的未来时刻是否为两个或者两个以上。比如，标准格式的时间为“xxxx/xx/xx/8:00”，当前时刻为“2013/04/19/07:01:19”，则存在“2013/04/19/AM08:00:00”、“2013/04/19/PM08:00:00”、“2013/04/20/AM08:00:00”、“2013/04/20/PM08:00:00”等等未来时刻与标准格式的时间匹配。

[0102] 2) 若检测结果为与标准格式的时间匹配的未来时刻是两个或者两个以上，则从两个或者两个以上的未来时刻中选择符合第一预设条件的未来时刻作为设置提醒时的提醒时间。

[0103] 该第一预设条件可以是“距离当前时刻最近的”，比如，平板电脑可以从两个或者两个以上的未来时刻中选择距离当前时刻最近的一个“2013/04/19/AM08:00:00”作为设置提醒时的提醒时间。在其它实施例中，第一预设条件还可以是与历史设置规律匹配的、与用户的再次语音输入所对应的等等其他条件。还需要说明的是，符合第一预设条件的未来时刻通常为一个，但不排除两个或者两个以上的可能，比如设置每周提醒或者每年提醒等提醒事项时，符合第一预设条件的未来时刻可以为两个或者两个以上。

[0104] 通过上述子步骤，可以实现时间信息不完整时，自动对时间信息进行“智能调整”。当然，作为优选，平板电脑在调整时间信息的过程中还可以与用户进行交互，来让用户选择与时间信息匹配的所有可能的未来时刻中更为准确的一个。

[0105] 步骤 204，通过连续语音识别获得语音信号所对应的文字序列；

[0106] 平板电脑还通过连续语音识别获得语音信号所对应的文字序列。比如，平板电脑通过连续语音识别获得语音信号所对应的文字序列为“明天下午闪电去会议室开会”。

[0107] 具体来讲，通过连续语音识别获得语音信号所对应的文字序列，可以包括如下几个子步骤，如图 2E 所示：

[0108] 204a，加载搜索空间；

[0109] 该搜索空间包括声学模型、词典和语言模型，声学模型为基于单音素或者三音素的隐马尔可夫模型，词典包括词语与音素之间的对应关系，语言模型包括词语与词语之间组合成文字序列时的比例关系。

[0110] 本步骤中的声学模型是与上述前景模型相同或者相似的声学模型。

[0111] 204b，提取语音信号中的音频特征；

[0112] 此处的音频特征仍然采用 PLP (Perceptual Linear Predictive, 感知线性预测) 特征。一段语音信号可以对应一段音频特征序列。

[0113] 204c，通过维特比解码计算音频特征在搜索空间中的最优文字序列，最优文字序列包括至少一个文字和每个文字在语音信号中的起止时刻，将最优文字序列作为语音信号所对应的文字序列。

[0114] 维特比 (Viterbi) 解码基于下式：

$$[0115] \quad \hat{W} = \arg \max_w P(W|O) = \arg \max_w \frac{P(O|W)P(W)}{P(O)}$$

[0116] 其中，O 代表音频特征序列，P(O|W) 代表了来自声学模型的似然值，其表征了语

音信号本身的声学特征与文本序列  $W$  的相似程度,这个概率可以通过在训练声学模型时得到,  $P(W)$  表征了该文本序列在文字搭配上出现的概率,这个概率通常由上下文相关的语言模型确定。根据贝叶斯决策理论,本子步骤的任务就是要找到一个最优文本序列  $\hat{W}$ ,使得上式输出的后验概率最大化。

[0117] 基于以上子步骤,语音信号在经过连续语音识别处理之后,将获得对应的文字序列。同时从上式可以看出,在获得相应的文本序列  $W$  的同时,也可以得到文本序列  $W$  和音频特征序列  $O$  的对应关系,通过该对应关系可以获得文字序列中各个文字在语音信号中的起止时刻。

[0118] 步骤 205,根据时间信息和文字序列确定文字序列中的内容信息,内容信息是文字序列中与时间信息不对应的部分的子集;

[0119] 具体来讲,时间信息包括时间关键词和时间关键词在语音信号中的起止时刻,文字序列包括至少一个文字和每个文字在语音信号中的起止时刻。

[0120] 平板电脑根据时间关键词在语音信号中的起止时刻和各个文字各自在语音信号中的起止时刻,将文字序列区分为与时间信息对应的部分和与时间信息不对应的部分。

[0121] 若与时间信息不对应的部分为一段,则将与时间信息不对应的部分作为内容信息。比如,与时间信息对应的部分位于文字序列中的开始部分,则与时间信息不对应的部分为一段,并且将该与时间信息不对应的部分作为内容信息。

[0122] 若与时间信息不对应的部分为两段或者两段以上,则将与时间信息不对应的部分中符合第二预设条件的一段作为内容信息。比如,与时间信息对应的部分位于文字序列中的中间部分,则与时间信息不对应的部分为两段,并且将这两段中符合第二预设条件的一段作为内容信息。

[0123] 第二预设条件可以是“长度最长”,也即,与时间信息不对应的部分中符合第二预设条件的一段可以是两段或者两段以上与时间信息不对应的部分中长度最长的一段。

[0124] 补充说明的是,此处所述的“对应”和“不对应”是指在时间维度上的对应关系。

[0125] 比如,平板电脑在获得关键词识别输出的时间信息:时间关键词、时间关键词在语音信号中的起止时刻;以及连续语音识别输出的文字序列:语音信号所对应的至少一个文字、各个文字在语音信号中的起止时刻之后,可以将两组结果进行对齐,对齐结果可以参考图 2F 所示。

[0126] 其中,关键词识别由于没有语言模型带来的干扰,从而准确地识别出了时间信息“明天下午 3 点”;而连续语音识别则错误地将文字序列中的两个文字“3 点”识别为了“闪电”。但是根据对齐结果,可以将文字序列区分为与时间信息对应的部分“明天下午 3 点”71 和与时间信息不对应的部分“去会议室开会”72。

[0127] 在图 2F 所示中,与时间信息不对应的部分只有一段,所以将与时间信息不对应的部分“去会议室开会”72 作为内容信息。

[0128] 步骤 206,将内容信息和 / 或内容信息在语音信号中所对应的语音片段作为设置提醒时的提醒内容;

[0129] 平板电脑将内容信息和 / 或内容信息在语音信号中所对应的语音片段作为设置提醒时的提醒内容。也即,分为三种情况:

[0130] 平板电脑将内容信息作为设置提醒时的提醒内容;

[0131] 平板电脑将内容信息在语音信号中所对应的语音片段作为设置提醒时的提醒内容；

[0132] 平板电脑将内容信息和 / 或内容信息在语音信号中所对应的语音片段作为设置提醒时的提醒内容。

[0133] 步骤 207, 根据提醒时间和提醒内容设置提醒事项。

[0134] 平板电脑根据提醒时间和提醒内容设置提醒事项。提醒内容可以是内容信息和内容信息在语音信号中所对应的语音片段两者中的任意一种或者全部两种, 以同时将两者作为设置提醒时的提醒内容为例, 在提醒时间到来时, 平板电脑既可以在自身的屏幕上以文字形式显示内容信息“去会议室开会”, 也可以在自身的扬声器中播放内容信息在语音信号中所对应的语音片段“去会议室开会”。当然, 平板电脑也可以根据用户的选择仅展示内容信息, 或者, 仅展示内容信息在语音信号中所对应的语音片段。

[0135] 与上一实施例相比, 本发明实施例还通过基于音素混淆度对前景模型中的音素进行聚类后训练得到的背景模型, 或者通过与前景模型所采用的语种不同的其它语种训练得到的背景模型来进行关键词识别, 可以尽量避免背景模型对语音信号中的关键词的吸收, 提高关键词的检出召回率。

[0136] 另外, 通过对两路语音识别的结果进行比对, 将文字序列中不与时间信息对应的内容信息作为设置提醒时的提醒内容, 使得即便连续语音识别获得的文字序列是错误的, 也不影响提醒事项的正确设置和运行。换句话说, 即便此时提醒内容中包含的内容信息是错误的, 用户也不会错过这个提醒事项; 而且提醒内容包括内容信息在语音信号中所对应的语音片段时, 用户也可以通过自己的原始语音信号来获知正确的提醒内容, 使得通过语音输入设置提醒事项这项功能的容错率更高, 有效减少用户重复输入语音信号才能正确设置提醒事项的现象发生。

[0137] 请参考图 3A, 其示出了本发明再一个实施例提供的提醒设置方法的方法流程图。与上一实施例不同的是, 本实施例以该提醒设置方法应用于包含有至少一个终端和服务器的电子系统中实现来举例说明。该提醒设置方法, 包括:

[0138] 步骤 301, 终端采集语音信号;

[0139] 终端运行有用于提供基于语音输入设置提醒事项功能的应用程序, 结合参考图 3B, 终端可以提供用于启用设置提醒事项的第一界面 32, 在用户点击“语音提醒”按钮 34 后, 跳转至第二界面 36, 并且在用户点击录音按钮 38 后, 终端通过内置麦克风或者外接麦克风开始采集用户输入的语音信号, 假设终端采集到的语音信号的内容为“提醒我下周三 8 点和杰克去郊游”。

[0140] 步骤 302, 服务器接收终端发送的语音信号;

[0141] 服务器可以通过有线网络或者无线网络接收终端发送的语音信号。

[0142] 步骤 303, 服务器通过关键词识别获得语音信号中的时间信息;

[0143] 服务器通过关键词识别获得语音信号中的时间信息, 时间信息具体包括时间关键词和时间关键词在语音信号中的起止时刻。具体来讲, 平板电脑首先将语音信号中的各个片段输入关键词检索网络, 该关键词检索网络包括有至少一个前景模型和至少一个背景模型; 将前景模型和背景模型分别与各个片段进行匹配, 得到时间关键词和时间关键词在语音信号中的起止时刻, 将时间关键词和时间关键词在语音信号中的起止时刻作为语音信号

所对应的的时间信息。

[0144] 参考图 2B 可知,该关键词检索网络可以包括 M 个前景模型 KW (keyword,关键词)和 N 个背景模型 Filler。其中,前景模型用于识别语音信号中的时间关键词,背景模型用于识别语音信号中的非时间关键词,并与前景模型识别的时间关键词进行区分。换句话说,每个前景模型 KW 可以对应一个时间关键词,比如:年、月、日、时、分、秒、x 点、x 号、傍晚、早上、上午、下午和明天等等;而背景模型 Filler 则对应于语音信号中的非时间关键词。

[0145] 在语音信号中的各个片段输入关键词检索网络后,对应于时间关键词的片段会被前景模型 KW 所识别,对应于非时间关键词的片段会被背景模型 Filler 所识别,背景模型识别出的非时间关键词可以对前景模型识别出的时间关键词形成更好的区分,通过对前景模型 KW 所识别出的关键词进行关键词确认,可以获得语音信号中的时间关键词,和时间关键词在语音信号中的起始时刻和终止时刻。起止时刻包括起始时刻和终止时刻。

[0146] 该关键词检索网络中的前景模型可以是基于单音素或者三音素的 HMM (Hidden Markov Model,隐马尔可夫模型),音素是语音的最小发音单元,三音素是根据一个音素的上下文的不同来决定不同的建模单元。前景模型的训练方式可以采用单音素或者三音素的 HMM 模型训练流程,训练过程可以参考图 2C 及上一实施例所示,本实施例中不再赘述。

[0147] 该关键词检索网络中的背景模型并非是常见的通过非关键词训练的模型,在本实施例中,背景模型为基于音素混淆度对前景模型中的音素进行聚类后训练得到的模型;或者,背景模型为通过与前景模型所采用的语种不同的其它语种训练得到的模型,比如前景模型是采用中文普通话训练的模型,则背景模型是采用粤语或者越南语等不同于中文普通话,但是有一定相似性的语种来训练的模型。这两种背景模型可以模糊掉一些单音素的发音个性,从而和前景模型中的关键词音素形成更为鲜明的对比,可以尽量避免背景模型对语音信号中的关键词的吸收,提高关键词的检出召回率。

[0148] 在背景模型为基于音素混淆度对前景模型中的音素进行聚类后训练得到的模型时,该背景模型的训练方式如图 2D 及上一实施例所示,本实施例中不再赘述。

[0149] 比如,服务器可以将语音信号“提醒我下周三 8 点和杰克去郊游”输入关键词检索网络来检出时间关键词“下周三 8 点”,和时间关键词“下周三 8 点”在语音信号中的起始时刻和终止时刻。

[0150] 步骤 304,服务器根据时间信息确定设置提醒时的提醒时间;

[0151] 服务器可以根据关键词识别出的时间信息确定设置提醒时的提醒时间。

[0152] 由于识别出的时间信息可能存在不同的格式,所以本步骤还可以包括如下几个子步骤:

[0153] 第一,从预设的至少一个时间信息模板中匹配出与时间信息对应的的时间信息模板;

[0154] 服务器可以将识别出的时间信息与预设的至少一个时间信息模板进行匹配,从该至少一个时间信息模板中找出时间格式及语义与所述时间信息相同或相似的一个时间信息模板,作为与该时间信息相匹配的时间信息模板。本实施例中,所述时间信息模板是预先通过包含大量时间信息的语料库来搜集的。这个过程中,首先将搜集时间信息的所有可能的语料作为样本库;然后分析和总结这些语料的序列特点;最后根据分析结果获得时间信息模板。比如一种常见的的时间信息模板为“日期+时钟”,与之符合的时间信息可以是“明天

8 点”；另一种常见的信息模板为“月 + 日期”，与之符合的时间信息可以是“4 月 8 号”。

[0155] 需要说明的是，虽然用户在语音输入时对时间信息的口语说法可能非常多，但是相较于传统的语义分析技术需要对所有可能的语句（不局限于时间）搜集模板来讲，仅搜集有关时间信息的模板的工作量要小很多，而且搜集到的模板可以基本覆盖用户对时间信息的所有口语说法。

[0156] 第二，根据与上述时间信息对应的信息模板将上述时间信息转换为标准格式的时间；

[0157] 服务器根据与时间信息对应的信息模板将时间信息转换为标准格式的时间。其中，该标准格式的时间通常为某一种标准时间格式的时间，比如，当上述与时间信息对应的信息模板为“星期 + 时钟”，标准时间格式为“年 / 月 / 日 / 时 / 分 / 秒”时，假设当前时刻为“2013/4/18/10:24:10”，则服务器可以将时间信息“下周三 8 点”转化为标准格式的时间“2013/4/24/8 :00 :00”。

[0158] 第三，将标准格式的时间作为设置提醒时的提醒时间。

[0159] 然后服务器将标准格式的时间作为设置提醒时的提醒时间，比如将“2013/4/24/8 :00 :00”作为设置提醒时的提醒时间。

[0160] 由于识别出的时间信息有时并不是一个完整的时间信息，就会导致存在多个未来时刻与之对应。比如语音信号为“8 点提醒我开会”，则识别出的时间信息在转换成标准格式的时间后为“xxxx/xx/xx/8 :00”，存在多个未来时刻与之对应。为此，本步骤还可以具体包括：

[0161] 1)，服务器检测与标准格式的时间匹配的未来时刻是否为两个或者两个以上；

[0162] 服务器可以检测与标准格式的时间匹配的未来时刻是否为两个或者两个以上。比如，标准格式的时间为“xxxx/xx/xx/8 :00”，当前时刻为“2013/04/19/07 :01 :19”，则存在“2013/04/19/AM08 :00 :00”、“2013/04/19/PM08 :00 :00”、“2013/04/20/AM08 :00 :00”、“2013/04/20/PM08 :00 :00”等等未来时刻与标准格式的时间匹配。

[0163] 2) 若检测结果为与标准格式的时间匹配的未来时刻是两个或者两个以上，则从两个或者两个以上的未来时刻中选择符合第一预设条件的未来时刻作为设置提醒时的提醒时间。

[0164] 该第一预设条件可以是“距离当前时刻最近的”，比如，服务器可以从两个或者两个以上的未来时刻中选择距离当前时刻最近的一个“2013/04/19/AM08 :00 :00”作为设置提醒时的提醒时间。在其它实施例中，第一预设条件还可以是与历史设置规律更匹配的、与用户的再次语音输入所对应的等等其他条件。还需要说明的是，符合第一预设条件的未来时刻通常为一个，但不排除两个或者两个以上的可能，比如设置每周提醒或者每年提醒等提醒事项时，符合第一预设条件的未来时刻可以为两个或者两个以上。

[0165] 通过上述子步骤，可以实现时间信息不完整时，自动对时间信息进行“智能调整”。当然，作为优选，服务器在调整时间信息的过程中还可以通过终端与用户进行交互，来让用户选择与时间信息匹配的所有可能的未来时刻中更为准确的一个。

[0166] 步骤 305，服务器通过连续语音识别获得语音信号所对应的文字序列；

[0167] 服务器还通过连续语音识别获得语音信号所对应的文字序列，比如，服务器通过连续语音识别获得语音信号所对应的文字序列为“提醒我下周三 8 点和杰克去郊游”。



[0168] 具体来讲,服务器首先加载搜索空间,然后提取语音信号中的音频特征,最后通过维特比解码计算音频特征在搜索空间中的最优文字序列,最优文字序列包括至少一个文字和每个文字在语音信号中的起止时刻,将最优文字序列作为语音信号所对应的文字序列。具体如图 2E 及上一实施例所示。

[0169] 步骤 306,服务器根据时间信息和文字序列确定文字序列中的内容信息,内容信息是文字序列中与时间信息不对应的部分的子集;

[0170] 具体来讲,时间信息包括时间关键词和所述时间关键词在所述语音信号中的起止时刻,文字序列包括至少一个文字和每个文字在所述语音信号中的起止时刻,服务器根据时间关键词在语音信号中的起止时刻和文字序列中各个文字各自在语音信号中的起止时刻,将文字序列区分为与时间信息对应的部分和与时间信息不对应的部分。

[0171] 若与时间信息不对应的部分为一段,则将与时间信息不对应的部分作为内容信息。比如,与时间信息对应的部分位于文字序列中的开始部分,则与时间信息不对应的部分为一段,并且将该段作为内容信息。

[0172] 若与时间信息不对应的部分为两段或者两段以上,则将与时间信息不对应的部分中符合第二预设条件的一段作为内容信息。比如,与时间信息对应的部分位于文字序列中的中间部分,则与时间信息不对应的部分为两段,并且将这两段中符合第二预设条件的一段作为内容信息。

[0173] 第二预设条件可以是“长度最长”,也即,与时间信息不对应的部分中符合第二预设条件的一段可以是两段或者两段以上与时间信息不对应的部分中长度最长的一段。

[0174] 补充说明的是,此处所述的“对应”和“不对应”是指在时间维度上的对应关系。

[0175] 比如,服务器在获得关键词识别输出的结果:时间信息、时间信息在语音信号中的起止时刻;以及连续语音识别输出的结果:语音信号所对应的文字序列、文字序列中各个文字在语音信号中的起止时刻之后,可以将两组结果进行对齐,对齐结果可以参考图 3C 所示。

[0176] 其中,关键词识别识别出的时间信息为“下周三 8 点”;而连续语音识别识别出的文字序列为“提醒我下周三 8 点和杰克去郊游”。根据对齐结果,可以将文字序列分为与时间信息对应的部分“下周三 8 点”81、与时间信息不对应的部分“提醒我”82 和与时间信息不对应的部分“和杰克去郊游”83。

[0177] 由于图 3C 所示中,与时间信息不对应的部分包含两段,所以将该与时间信息不对应的部分中长度最长的一段“和杰克去郊游”83 作为内容信息。

[0178] 步骤 307,服务器将内容信息和 / 或内容信息在语音信号中所对应的语音片段作为设置提醒时的提醒内容;

[0179] 服务器还可以将内容信息和 / 或内容信息在语音信号中所对应的语音片段作为设置提醒时的提醒内容。也即,分为三种情况:

[0180] 服务器将内容信息作为设置提醒时的提醒内容;

[0181] 服务器将内容信息在语音信号中所对应的语音片段作为设置提醒时的提醒内容;

[0182] 服务器将内容信息和 / 或内容信息在语音信号中所对应的语音片段作为设置提醒时的提醒内容。

[0183] 步骤 308,服务器将提醒时间和提醒内容发送给终端;

[0184] 服务器可以将提醒时间和提醒内容按照与终端约定的格式发送给终端。

[0185] 步骤 309,终端根据提醒时间和提醒内容设置提醒事项。

[0186] 终端在接收到提醒时间和提醒内容后,根据提醒时间和提醒内容设置提醒事项。同时可以将设置完毕的提醒事项展示给用户,如图 3D 所示。

[0187] 由于提醒内容可以是内容信息和内容信息在语音信号中所对应的语音片段两者中的任意一种或者全部两种,以同时将两者作为设置提醒时的提醒内容为例,在提醒时间到来时,终端既可以在自身的屏幕上以文字形式显示内容信息“提醒我开会”,也可以在自身的扬声器中播放内容信息在语音信号中所对应的语音片段“提醒我开会”,该非时间信息部分是用户设置提醒时输入的语音信号中的一部分音频信号。当然,终端也可以根据用户的选择仅展示内容信息,或者,仅展示内容信息在语音信号中所对应的语音片段。

[0188] 以下为本发明的装置实施例,在装置实施例中未详尽描述的细节,可以参考上述对应的方法实施例。

[0189] 请参考图 4,其示出了本发明一个实施例中提供的提醒设置装置的结构方框图。该提醒设置装置可以通过软件、硬件或者两者的结合实现成为电子设备的全部或者一部分,也可以实现成为服务器的全部或者一部分。该提醒设置装置,包括:语音获取模块 410、第一识别模块 420、时间确定模块 430、第二识别模块 440、内容确定模块 450 和提醒设置模块 460。

[0190] 语音获取模块 410,用于获取语音信号;

[0191] 第一识别模块 420,用于通过关键词识别获得所述语音获取模块 410 获取到的语音信号中的时间信息;

[0192] 时间确定模块 430,用于根据所述第一识别模块 420 识别出的所述时间信息确定设置提醒时的提醒时间;

[0193] 第二识别模块 440,用于通过连续语音识别获得所述语音获取模块 410 获取到的语音信号所对应的文字序列;

[0194] 内容确定模块 450,用于通过所述第一识别模块 420 获得的时间信息和通过所述第二识别模块获得的文字序列确定设置提醒时的提醒内容;

[0195] 提醒设置模块 460,用于所述时间确定模块 430 确定的提醒时间和所述内容确定模块 450 确定的提醒内容设置提醒事项。

[0196] 综上所述,本实施例提供的提醒设置装置,通过关键词识别来获得语音信号中的时间信息,可以保证时间信息提取的正确性;减少了服务器在时间信息提取不正确时,导致提醒事项的设置失败或者设置错误的问题;达到了即便语音识别的全文识别准确率较差,导致识别出的文字序列是错误的情况下,仍然可以通过关键词识别获得正确的时间信息来设置提醒事项的效果。

[0197] 请参考图 5,其示出了本发明一个实施例中提供的提醒设置装置的结构方框图。该提醒设置装置可以通过软件、硬件或者两者的结合实现成为电子设备的全部或者一部分。该提醒设置装置,包括:语音获取模块 410、第一识别模块 420、时间确定模块 430、第二识别模块 440、内容获取模块 450 和提醒设置模块 460。

[0198] 语音采集模块 410 用于采集用户输入的语音信号;

[0199] 第一识别模块 420 用于通过关键词识别获得所述语音采集模块 410 采集到的语音信号中的时间信息；

[0200] 时间确定模块 430 用于根据所述第一识别模块 420 识别出的所述时间信息确定设置提醒时的提醒时间；

[0201] 第二识别模块 440 用于通过连续语音识别获得所述语音采集模块 410 采集到语音信号所对应的文字序列；

[0202] 内容确定模块 450, 用于通过所述第一识别模块 420 获得的时间信息和通过所述第二识别模块获得的文字序列确定设置提醒时的提醒内容。

[0203] 具体来讲, 所述内容确定模块 450, 包括: 内容信息确定单元 452 和提醒内容确定单元 454；

[0204] 所述内容信息确定单元 452, 用于根据所述时间信息和所述文字序列确定所述文字序列中的内容信息, 所述内容信息是所述文字序列中与所述时间信息不对应的部分的子集；

[0205] 所述提醒内容确定单元 454, 用于将所述内容信息确定单元 452 区分出的内容信息和 / 或所述内容信息在所述语音信号中所对应的语音片段作为设置提醒时的提醒内容。

[0206] 所述第一识别模块 420 识别出的时间信息包括时间关键词和所述时间关键词在所述语音信号中的起止时刻, 所述第二识别模块 440 识别出的文字序列包括至少一个文字和每个文字在所述语音信号中的起止时刻。

[0207] 所述内容信息确定单元 452 包括：

[0208] 区分子单元、第一确定子单元和第二确定子单元；

[0209] 所述区分子单元, 用于根据所述第一识别模块 420 识别出的时间关键词在所述语音信号中的起止时刻和所述第二识别模块 440 识别出的各个文字在所述语音信号中的起止时刻, 将所述文字序列区分为与所述时间信息对应的部分和与所述时间信息不对应的部分；

[0210] 所述第一确定子单元, 用于若所述区分子单元区分出的与所述时间信息不对应的部分为一段, 则将所述与所述时间信息不对应的部分作为所述内容信息；

[0211] 所述第二确定子单元, 用于若所述区分子单元区分出的与所述时间信息不对应的部分为两段或者两段以上, 则将所述与所述时间信息不对应的部分中符合第二预设条件的一段作为所述内容信息。

[0212] 提醒设置模块 460, 用于根据所述时间确定模块 430 确定的提醒时间和所述内容确定模块 450 确定的提醒内容设置提醒事项。

[0213] 与上一实施例相比, 本发明实施例还通过基于音素混淆度对前景模型中的音素进行聚类后训练得到的背景模型, 或者通过与前景模型所采用的语种不同的其它语种训练得到的背景模型来进行关键词识别, 可以尽量避免背景模型对语音信号中的关键词的吸收, 提高关键词的检出召回率。

[0214] 另外, 通过对两路语音识别的结果进行比对, 将文字序列中的内容信息作为设置提醒时的提醒内容, 使得即便连续语音识别获得的文字序列是错误的, 也不影响提醒事项的正确设置和运行。换句话说, 即便此时提醒内容中包含的内容信息是错误的, 用户也不会错过这个提醒事项, 而且提醒内容包括内容信息在语音信号中所对应的语音片段时, 用户

也可以通过自己的原始语音信号来获知正确的提醒内容,使得通过语音输入设置提醒事项这项功能的容错率更高,有效减少用户重复输入语音信号才能正确设置提醒事项的现象发生。

[0215] 请参考图6,其示出了本发明一个实施例中提供的提醒设置装置的结构方框图。该提醒设置装置可以通过软件、硬件或者两者的结合实现成为服务器的全部或者一部分。该提醒设置装置,包括:语音获取模块410、第一识别模块420、时间确定模块430、第二识别模块440、内容获取模块450和提醒设置模块460。

[0216] 语音获取模块410,用于接收终端发送的语音信号;

[0217] 第一识别模块420用于通过关键词识别获得所述语音获取模块410接收到的语音信号中的时间信息。具体来讲,所述时间信息包括时间关键词和所述时间关键词在所述语音信号中的起止时刻,所述第一识别模块420具体包括:输入单元和识别单元;

[0218] 所述输入单元,用于将所述语音信号中的各个片段输入关键词检索网络,所述关键词检索网络包括有至少一个前景模型和至少一个背景模型;

[0219] 所述前景模型为基于单音素或者三音素的隐马尔可夫模型;

[0220] 所述背景模型为基于音素混淆度对所述前景模型中的音素进行聚类后训练得到的模型;或者,所述背景模型为通过与所述前景模型所采用的语种不同的其它语种训练得到的模型;

[0221] 所述识别单元,用于通过所述前景模型和所述背景模型分别与所述各个片段进行匹配,得到所述时间关键词和所述时间关键词在所述语音信号中的起止时刻,将所述时间关键词和所述时间关键词在所述语音信号中的起止时刻作为所述语音信号所对应的时间信息。

[0222] 时间确定模块430用于根据所述第一识别模块420识别出的所述时间信息确定设置提醒时的提醒时间。所述时间确定模块430包括:匹配单元、转换单元和时间确定单元;

[0223] 所述匹配单元,用于从预设的至少一个时间信息模板中匹配出与所述时间信息对应的时间信息模板;

[0224] 所述转换单元,用于根据所述匹配单元匹配出的与所述时间信息对应的时间信息模板将所述时间信息转换为标准格式的时间;

[0225] 所述时间确定单元,用于将所述转换单元转换到的标准格式的时间作为设置提醒时的提醒时间。所述时间确定单元,包括:检测子单元和确定子单元;

[0226] 所述检测子单元,用于检测与所述标准格式的时间匹配的未来时刻是否为两个或者两个以上;

[0227] 所述确定子单元,用于若所述检测子单元的检测结果为与所述标准格式的时间匹配的未来时刻是两个或者两个以上,则从所述两个或者两个以上的未来时刻中选择符合第一预设条件的未来时刻作为设置提醒时的提醒时间。

[0228] 第二识别模块440用于通过连续语音识别获得所述信号接收模块414接收到的语音信号所对应的文字序列。具体来讲,所述文字序列包括至少一个文字和每个文字在所述语音信号中的起止时刻,所述第二识别模块440,包括:加载单元、提取单元和解码单元;

[0229] 所述加载单元,用于加载搜索空间;

[0230] 所述提取单元,用于提取所述语音信号中的音频特征;

[0231] 所述解码单元,用于通过维特比解码计算所述提取单元提取到的音频特征在所述加载单元加载的搜索空间中的最优文字序列,所述最优文字序列包括至少一个文字和每个文字在所述语音信号中的起止时刻,将所述最优文字序列作为所述语音信号所对应的文字序列。

[0232] 内容确定模块 450,用于通过所述第一识别模块 420 获得的时间信息和通过所述第二识别模块 440 获得的文字序列确定设置提醒时的提醒内容。

[0233] 具体来讲,所述内容确定模块 450,包括:内容信息确定单元 452 和提醒内容确定单元 454;

[0234] 所述内容信息确定单元 452,用于根据所述时间信息和所述文字序列确定所述语音信号中的内容信息,所述内容信息是所述文字序列中与所述时间信息不对应的部分的子集;

[0235] 所述提醒内容确定单元 454,用于将所述内容信息确定单元 452 区分出的内容信息和/或所述内容信息在所述语音信号中所对应的语音片段作为设置提醒时的提醒内容。

[0236] 所述内容信息确定单元 452 包括:

[0237] 区分子单元、第一确定子单元和第二确定子单元;

[0238] 所述区分子单元,用于根据所述第一识别模块 420 识别出的时间信息在所述语音信号中的起止时刻和所述第二识别模块 440 识别出的各个文字各自在所述语音信号中的起止时刻,将所述文字序列区分为与所述时间信息对应的部分和与所述时间信息不对应的部分;

[0239] 所述第一确定子单元,用于若所述区分子单元区分出的与所述时间信息不对应的部分为一段,则将所述与所述时间信息不对应的部分作为所述内容信息;

[0240] 所述第二确定子单元,用于若所述区分子单元区分出的与所述时间信息不对应的部分为两段或者两段以上,则将所述与所述时间信息不对应的部分中符合第二预设条件的一段作为所述内容信息。

[0241] 所述提醒设置模块 460,用于将所述时间确定模块 430 确定的提醒时间和所述内容确定模块 450 确定的提醒内容发送给所述终端,以便所述终端根据所述提醒时间和所述提醒内容设置提醒事项。

[0242] 请参考图 7,其示出了本发明一个实施例提供的提醒设置系统的结构示意图。该提醒设置系统包括终端 720 和服务器 760。

[0243] 该终端 720 可以是智能手机、平板电脑、电子书阅读器、MP3 播放器(Moving Picture Experts Group Audio Layer III,动态影像专家压缩标准音频层面 3)、MP4 (Moving Picture Experts Group Audio Layer IV,动态影像专家压缩标准音频层面 3)播放器、膝上型便携计算机和台式计算机等等。

[0244] 终端 720 和服务器 760 之间可以通过无线网络或者有线网络相连。

[0245] 该服务器 760 包括图 6 所对应实施例提供的提醒设置装置。

[0246] 请参考图 8,其示出了本发明一个实施例提供的电子设备的结构示意图。该电子设备用于实施上述实施例中提供的提醒设置方法,具体来讲:

[0247] 电子设备 800 可以包括 RF (Radio Frequency,射频)电路 810、包括有一个或一个以上计算机可读存储介质的存储器 820、输入单元 830、显示单元 840、传感器 850、音频电路

860、短距离无线传输模块 870、包括有一个或者一个以上处理核心的处理器 880、以及电源 890 等部件。本领域技术人员可以理解,图 8 中示出的电子设备结构并不构成对电子设备的限定,可以包括比图示更多或更少的部件,或者组合某些部件,或者不同的部件布置。其中:

[0248] RF 电路 810 可用于收发信息或通话过程中,信号的接收和发送,特别地,将基站的下行信息接收后,交由一个或者一个以上处理器 880 处理;另外,将涉及上行的数据发送给基站。通常,RF 电路 810 包括但不限于天线、至少一个放大器、调谐器、一个或多个振荡器、用户身份模块(SIM)卡、收发信机、耦合器、LNA(Low Noise Amplifier,低噪声放大器)、双工器等。此外,RF 电路 810 还可以通过无线通信与网络和其他设备通信。所述无线通信可以使用任一通信标准或协议,包括但不限于 GSM(Global System of Mobile communication,全球移动通讯系统)、GPRS(General Packet Radio Service,通用分组无线服务)、CDMA(Code Division Multiple Access,码分多址)、WCDMA(Wideband Code Division Multiple Access,宽带码分多址)、LTE(Long Term Evolution,长期演进)、电子邮件、SMS(Short Messaging Service,短消息服务)等。

[0249] 存储器 820 可用于存储软件程序以及模块,比如,存储器 820 可以用于存储采集语音信号的软件程序、实现关键词识别的软件程序、实现连续语音识别的软件程序以及实现设置提醒事项的软件程序等等。处理器 880 通过运行存储在存储器 820 的软件程序以及模块,从而执行各种功能应用以及数据处理,比如本发明实施例中“获取语音信号”的功能、“通过关键词识别获得语音信号中的时间信息”的数据处理功能、“通过连续语音识别获得语音信号所对应的文字序列”的数据处理功能以及“设置提醒事项”的功能等等。存储器 820 可主要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、至少一个功能所需的应用程序(比如声音播放功能、图像播放功能等)等;存储数据区可存储根据电子设备 800 的使用所创建的数据(比如音频数据、电话本等)等。此外,存储器 820 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、或其他易失性固态存储器件。相应地,存储器 820 还可以包括存储器控制器,以提供处理器 880 和输入单元 830 对存储器 820 的访问。

[0250] 输入单元 830 可用于接收输入的数字或字符信息,以及产生与用户设置以及功能控制有关的键盘、鼠标、操作杆、光学或者轨迹球信号输入。具体地,输入单元 830 可包括触敏表面 831 以及其他输入设备 832。触敏表面 831,也称为触摸显示屏或者触控板,可收集用户在其上或附近的触摸操作(比如用户使用手指、触笔等任何适合的物体或附件在触敏表面 831 上或在触敏表面 831 附近的操作),并根据预先设定的程式驱动相应的连接装置。可选的,触敏表面 831 可包括触摸检测装置和触摸控制器两个部分。其中,触摸检测装置检测用户的触摸方位,并检测触摸操作带来的信号,将信号传送给触摸控制器;触摸控制器从触摸检测装置上接收触摸信息,并将它转换成触点坐标,再送给处理器 880,并能接收处理器 880 发来的命令并加以执行。此外,可以采用电阻式、电容式、红外线以及表面声波等多种类型实现触敏表面 831。除了触敏表面 831,输入单元 830 还可以包括其他输入设备 832。具体地,其他输入设备 832 可以包括但不限于物理键盘、功能键(比如音量控制按键、开关按键等)、轨迹球、鼠标、操作杆等中的一种或多种。

[0251] 显示单元 840 可用于显示由用户输入的信息或提供给用户的信息以及电子设备

800 的各种图形用户接口,这些图形用户接口可以由图形、文本、图标、视频和其任意组合来构成。显示单元 840 可包括显示面板 841,可选的,可以采用 LCD(Liquid Crystal Display, 液晶显示器)、OLED(Organic Light-Emitting Diode, 有机发光二极管)等形式来配置显示面板 841。进一步的,触敏表面 831 可覆盖在显示面板 841 之上,当触敏表面 831 检测到在其上或附近的触摸操作后,传送给处理器 880 以确定触摸事件的类型,随后处理器 880 根据触摸事件的类型在显示面板 841 上提供相应的视觉输出。虽然在图 8 中,触敏表面 831 与显示面板 841 是作为两个独立的部件来实现输入和输入功能,但是在某些实施例中,可以将触敏表面 831 与显示面板 841 集成而实现输入和输出功能。

[0252] 电子设备 800 还可包括至少一种传感器 850,比如光传感器、运动传感器以及其他传感器。具体地,光传感器可包括环境光传感器及接近传感器,其中,环境光传感器可根据环境光线的明暗来调节显示面板 841 的亮度,接近传感器可在电子设备 800 移动到耳边时,关闭显示面板 841 和 / 或背光。作为运动传感器的一种,重力加速度传感器可检测各个方向上(一般为三轴)加速度的大小,静止时可检测出重力的大小及方向,可用于识别手机姿态的应用(比如横竖屏切换、相关游戏、磁力计姿态校准)、振动识别相关功能(比如计步器、敲击)等;至于电子设备 800 还可配置的陀螺仪、气压计、湿度计、温度计、红外线传感器等其他传感器,在此不再赘述。

[0253] 音频电路 860、扬声器 861,传声器 862 可提供用户与电子设备 800 之间的音频接口。音频电路 860 可将接收到的音频数据转换后的电信号,传输到扬声器 861,由扬声器 861 转换为声音信号输出;另一方面,传声器 862 将收集的声音信号转换为电信号,由音频电路 860 接收后转换为音频数据,再将音频数据输出处理器 880 处理后,经 RF 电路 810 以发送给另一终端,或者将音频数据输出至存储器 820 以便进一步处理。音频电路 860 还可能包括耳塞插孔,以提供外设耳机与电子设备 800 的通信。

[0254] 短距离无线传输模块 870 可以是 WIFI (wireless fidelity,无线保真)模块或者蓝牙模块等。电子设备 800 通过短距离无线传输模块 870 可以帮助用户收发电子邮件、浏览网页和访问流式媒体等,它为用户提供了无线的宽带互联网访问。虽然图 8 示出了短距离无线传输模块 870,但是可以理解的是,其并不属于电子设备 800 的必须构成,完全可以根据需要在不改变发明的本质的范围内而省略。

[0255] 处理器 880 是电子设备 800 的控制中心,利用各种接口和线路连接整个电子设备的各个部分,通过运行或执行存储在存储器 820 内的软件程序和 / 或模块,以及调用存储在存储器 820 内的数据,执行电子设备 800 的各种功能和处理数据,从而对电子设备进行整体监控。可选的,处理器 880 可包括一个或多个处理核心;优选的,处理器 880 可集成应用处理器和调制解调处理器,其中,应用处理器主要处理操作系统、用户界面和应用程序等,调制解调处理器主要处理无线通信。可以理解的是,上述调制解调处理器也可以不集成到处理器 880 中。

[0256] 电子设备 800 还包括给各个部件供电的电源 890(比如电池),优选的,电源可以通过电源管理系统与处理器 880 逻辑相连,从而通过电源管理系统实现管理充电、放电、以及功耗管理等功能。电源 890 还可以包括一个或一个以上的直流或交流电源、再充电系统、电源故障检测电路、电源转换器或者逆变器、电源状态指示器等任意组件。

[0257] 尽管未示出,电子设备 800 还可以包括摄像头、蓝牙模块等,在此不再赘述。具体

在本实施例中,电子设备 800 的显示单元是触摸屏显示器。

[0258] 电子设备 800 还包括有存储器,以及一个或者一个以上的程序,其中一个或者一个以上程序存储于存储器中,且经配置以由一个或者一个以上处理器执行。上述一个或者一个以上程序包含的指令用于执行一个提醒设置方法,该提醒设置方法可以是图 2A 及图 2A 所对应实施例所示出的提醒设置方法。

[0259] 请参考图 9,其示出了本发明另一实施例提供的服务器的结构示意图。该服务器可以是提醒设置系统中的服务器,该提醒设置系统中还包括通过无线网络或者有线网络与服务器相连的终端。该服务器包括有发射机 920、接收机 940、处理器 960 和存储器 980,以及一个或者一个以上的程序,其中一个或者一个以上程序存储于存储器 980 中,且经配置以由一个或者一个以上处理器 960 执行。上述一个或者一个以上程序包含的指令用于执行一个提醒设置方法,该提醒设置方法可以是图 3A 及图 3A 所对应实施例所示出的提醒设置方法。

[0260] 作为另一方面,本发明再一实施例还提供了一种计算机可读存储介质,该计算机可读存储介质可以是上述实施例中的存储器中所包含的计算机可读存储介质;也可以是单独存在,未装配入电子设备或者服务器中的计算机可读存储介质。所述计算机可读存储介质存储有一个或者一个以上程序,所述一个或者一个以上程序被一个或者一个以上的处理器用来执行一个提醒设置方法,上述一个或者一个以上程序包含的指令用于执行一个提醒设置方法,该提醒设置方法是图 2A 及图 2A 所对应实施例所示出的提醒设置方法,和/或,该提醒设置方法是图 3A 及图 3A 所对应实施例所示出的提醒设置方法。

[0261] 应当理解的是,本文中虽然使用术语第一、第二等描述各个元件,但是这些元件应该不受这些术语的限制。这些术语仅被用于彼此区分元件。还应当理解的是,在本文中使用的,除非上下文清楚地支持例外情况,单数形式“一个”(“a”、“an”、和“the”)旨在也包括复数形式。还应该理解的是,在本文中使用的“和/或”是指包括一个或者一个以上相关联地列出的项目的任意和所有可能组合。

[0262] 上述本发明实施例序号仅仅为了描述,不代表实施例的优劣。

[0263] 本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成,也可以通过程序来指令相关的硬件完成,所述的程序可以存储于一种计算机可读存储介质中,上述提到的存储介质可以是只读存储器,磁盘或光盘等。

[0264] 以上所述仅为本发明的较佳实施例,并不用以限制本发明,凡在本发明的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。



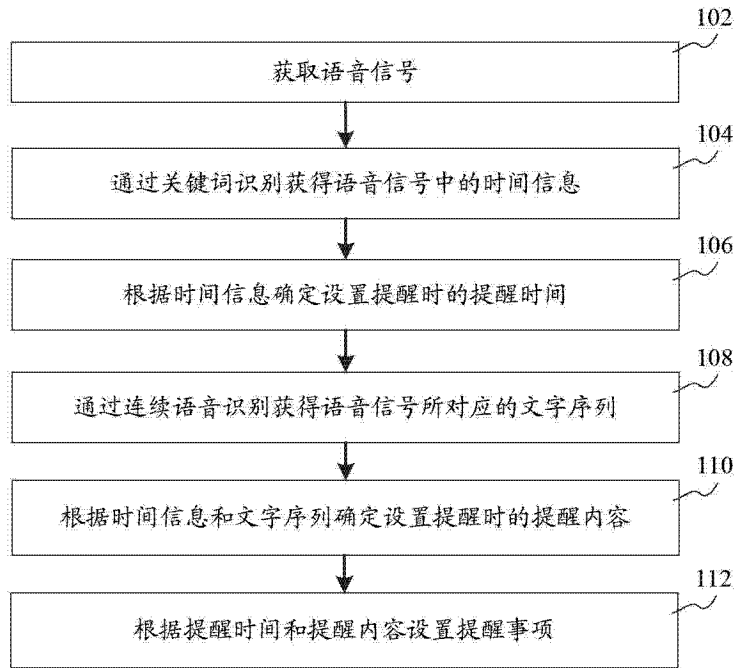


图 1

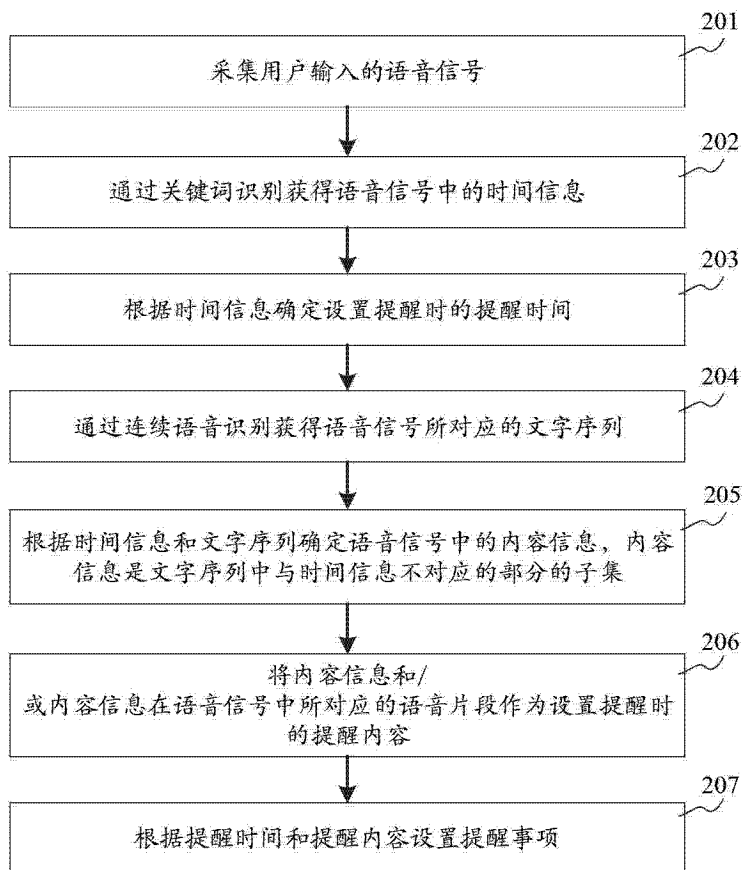


图 2A

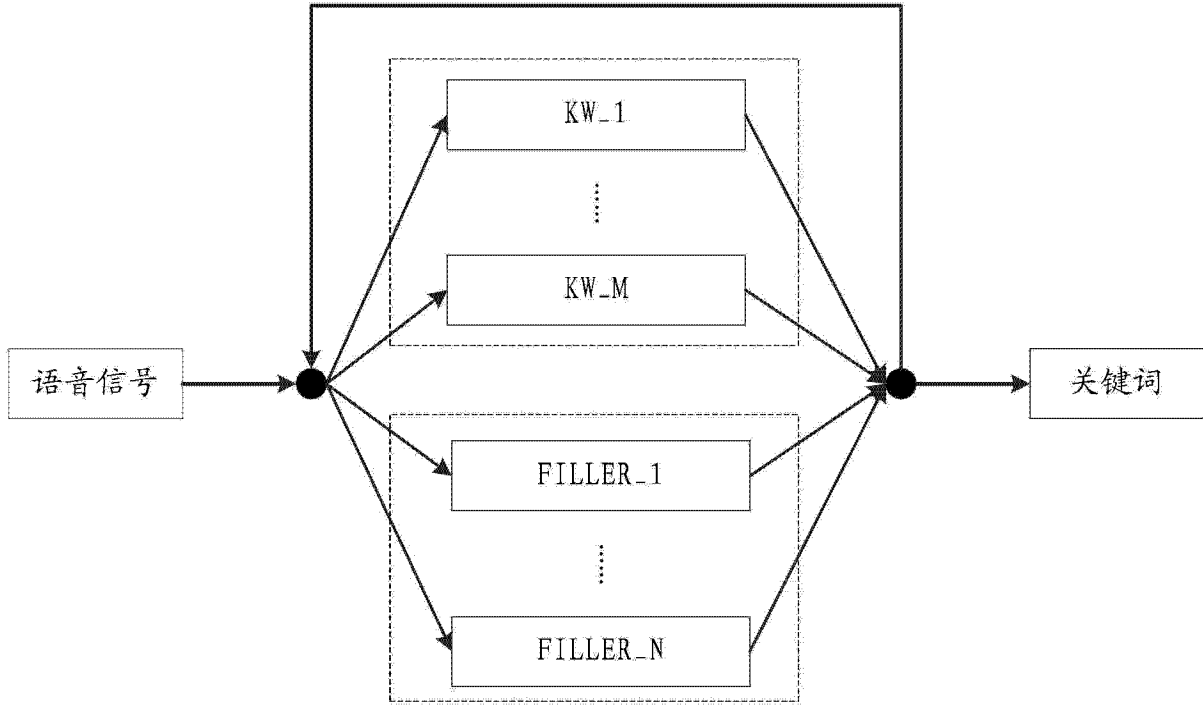


图 2B

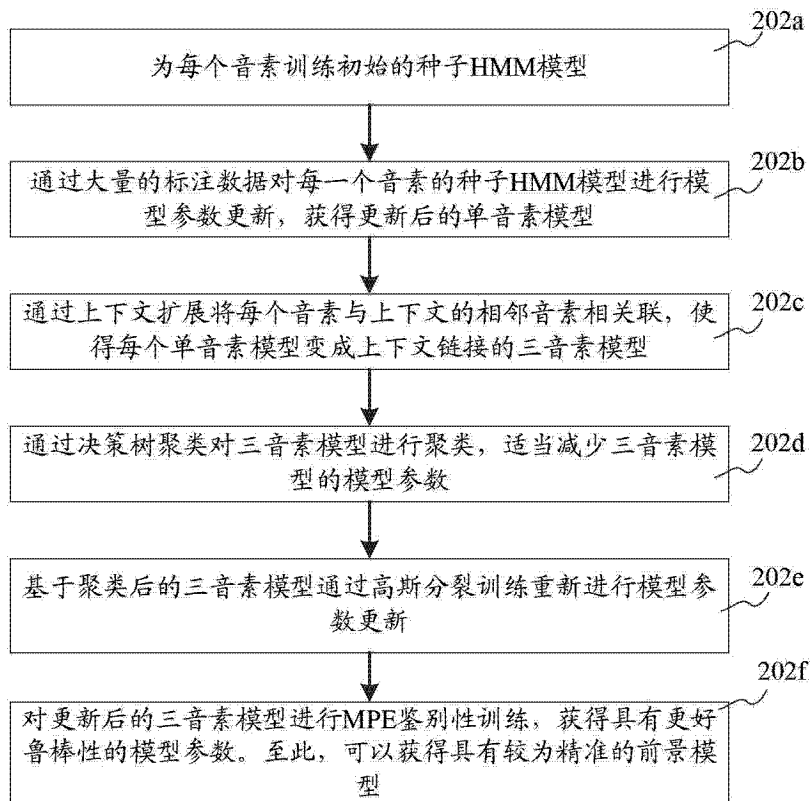


图 2C

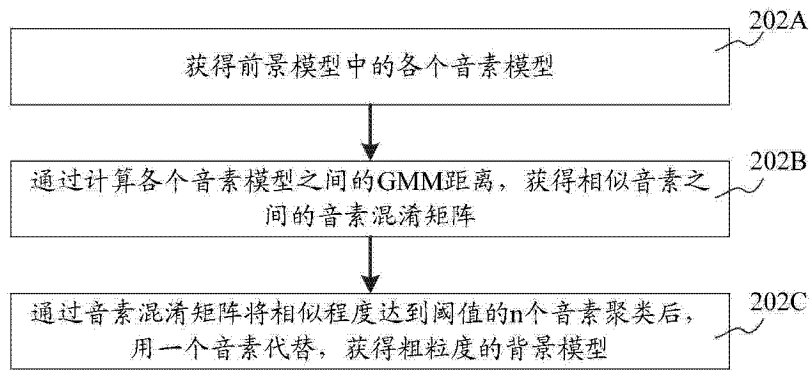


图 2D

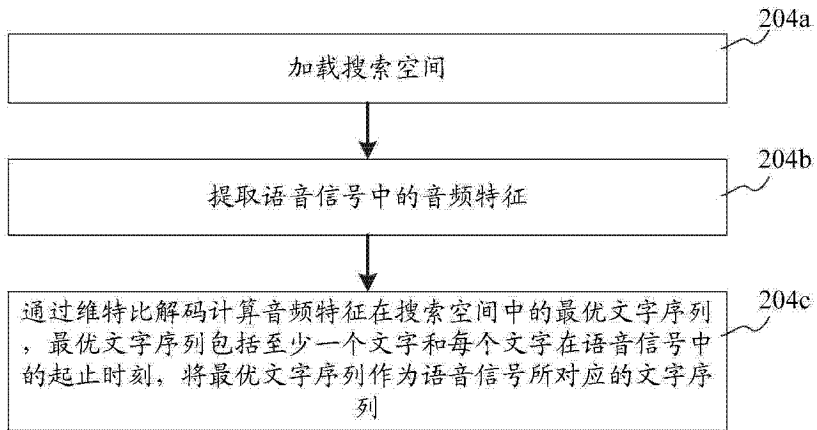


图 2E

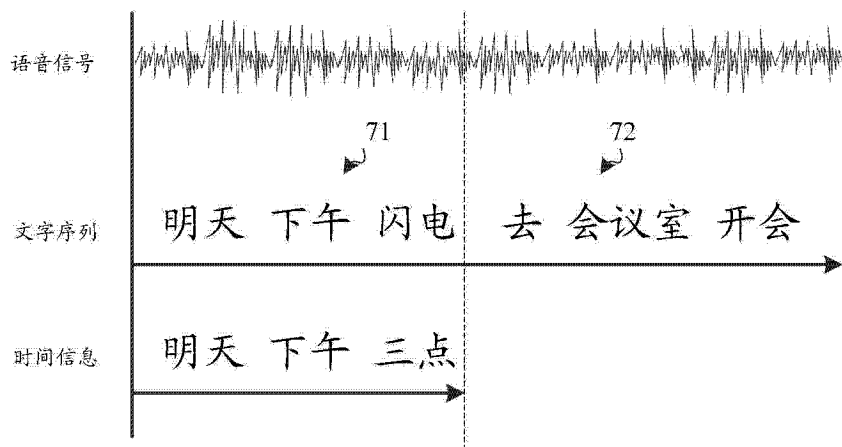


图 2F

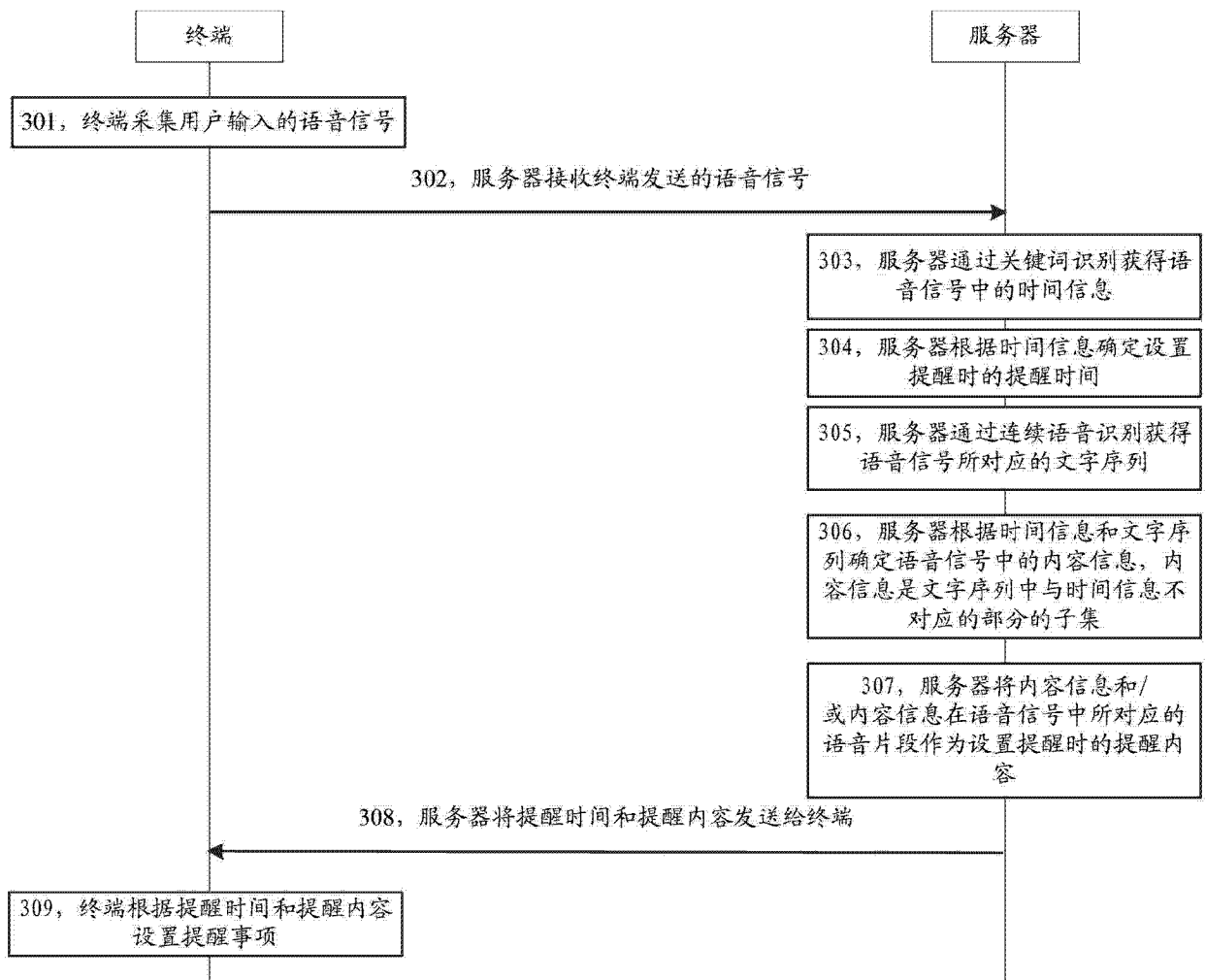


图 3A

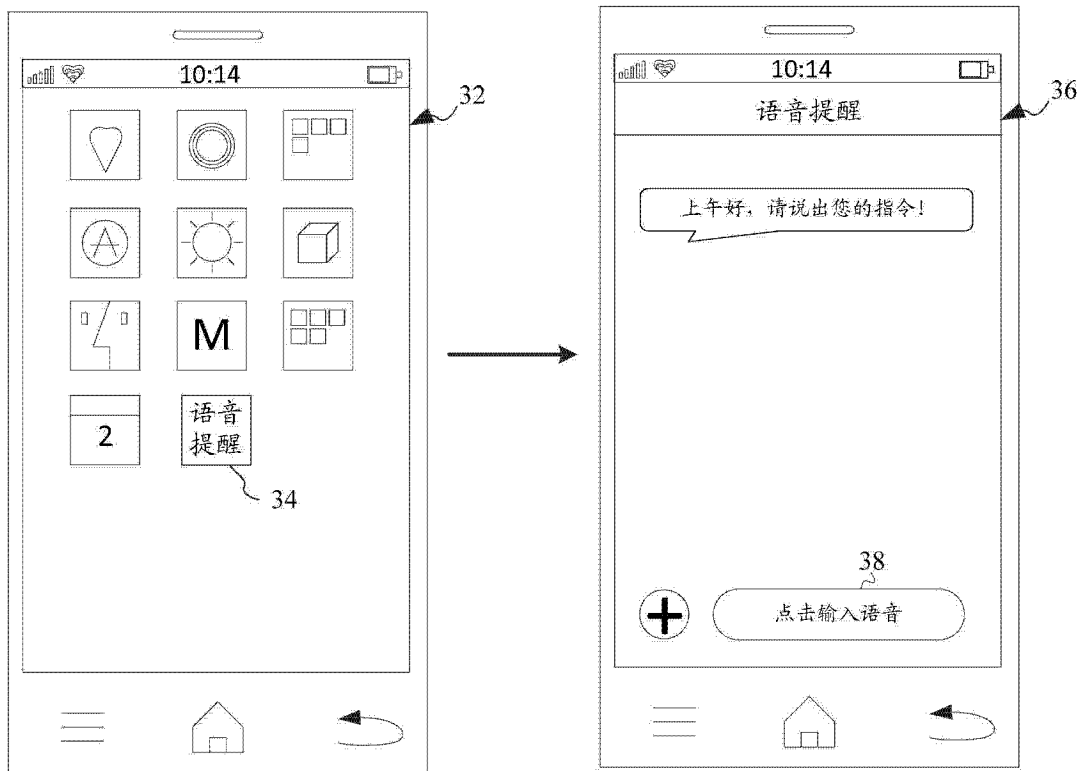


图 3B

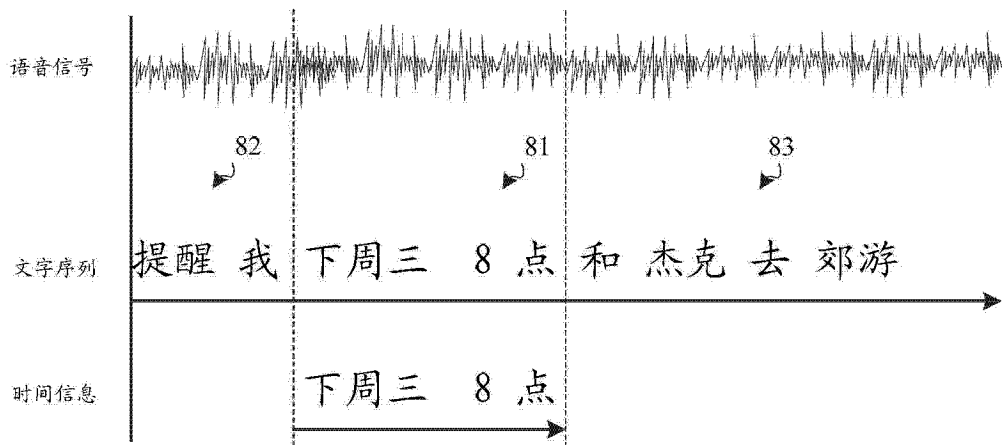


图 3C

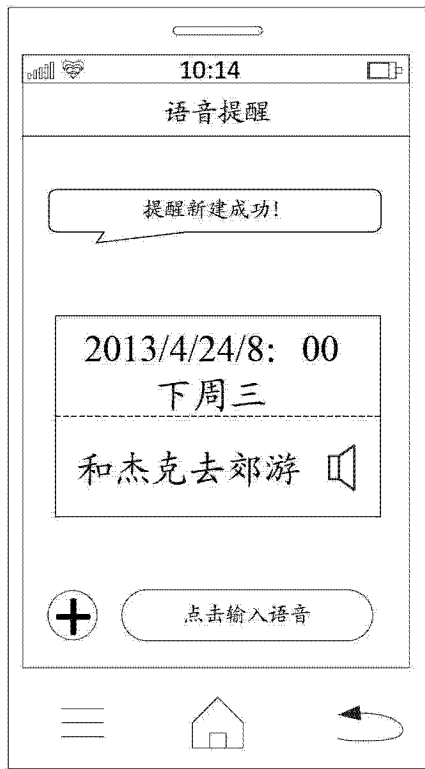


图 3D

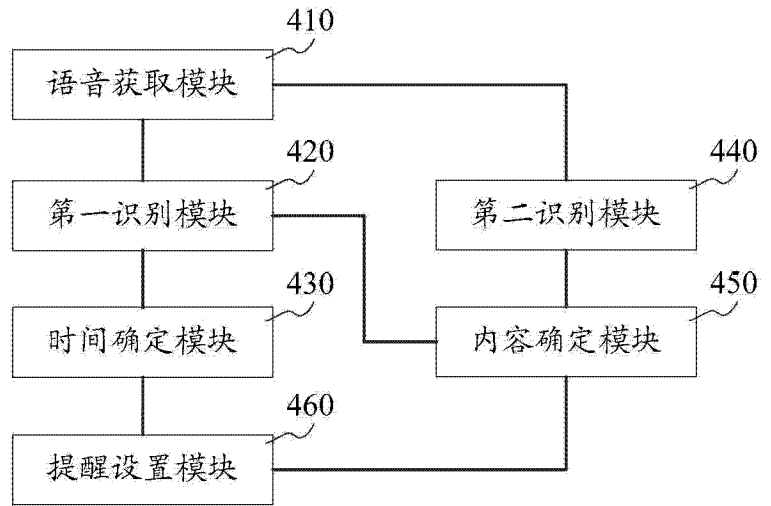


图 4

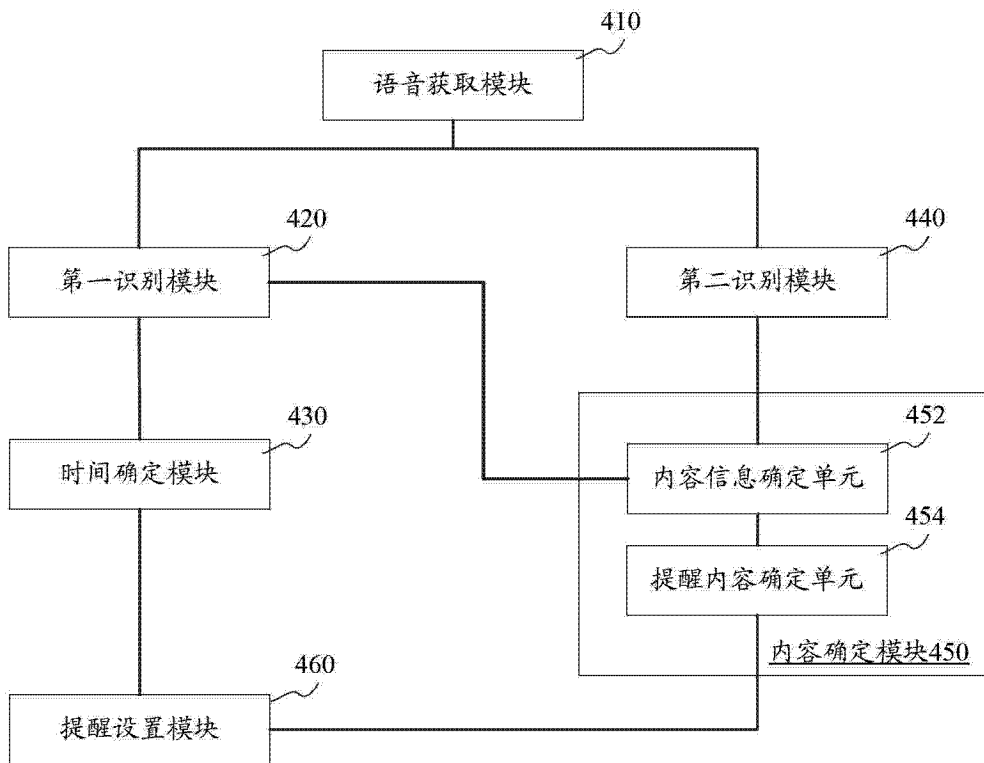


图 5

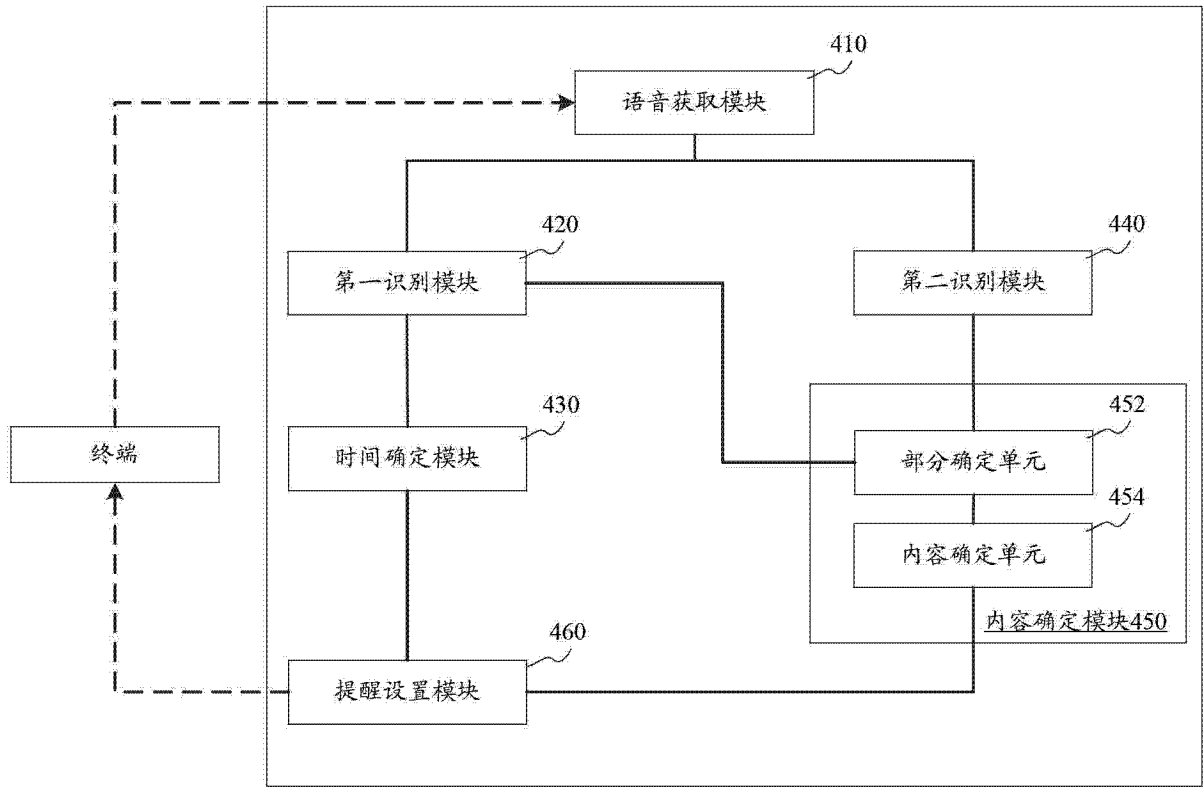


图 6

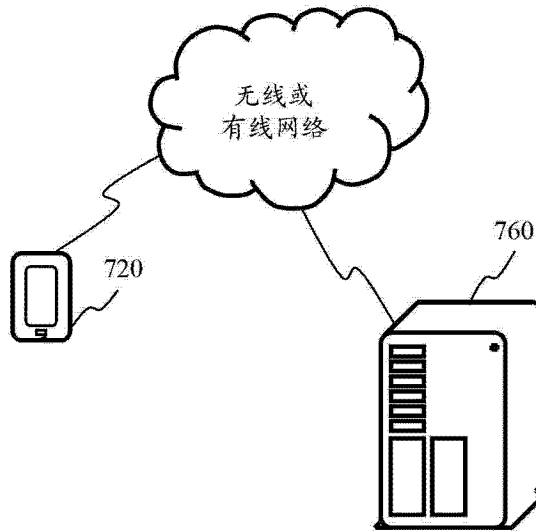


图 7

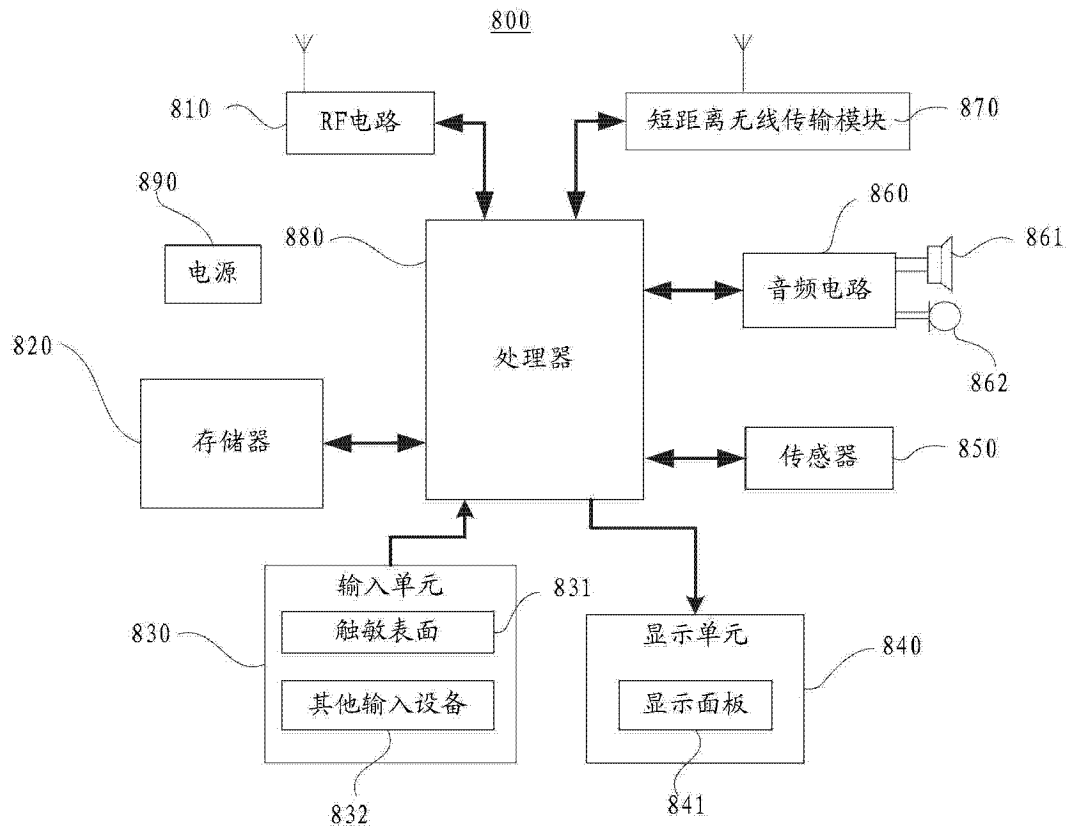


图 8

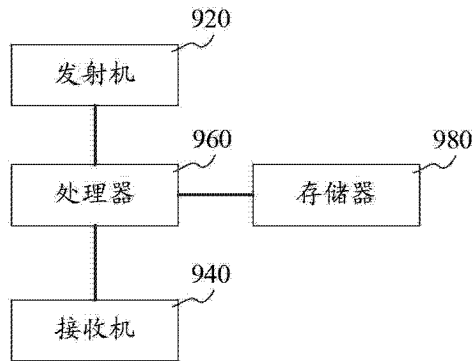


图 9