



(12) 发明专利申请

(10) 申请公布号 CN 105210146 A

(43) 申请公布日 2015. 12. 30

(21) 申请号 201480025286. 7

(51) Int. Cl.

(22) 申请日 2014. 04. 24

G10L 15/22(2006. 01)

G06F 1/32(2006. 01)

(30) 优先权数据

G10L 15/08(2006. 01)

61/820, 526 2013. 05. 07 US

G10L 25/78(2006. 01)

61/864, 490 2013. 08. 09 US

H04W 52/02(2006. 01)

14/092, 527 2013. 11. 27 US

(85) PCT国际申请进入国家阶段日

2015. 11. 04

(86) PCT国际申请的申请数据

PCT/US2014/035241 2014. 04. 24

(87) PCT国际申请的公布数据

W02014/182458 EN 2014. 11. 13

(71) 申请人 高通股份有限公司

地址 美国加利福尼亚州

(72) 发明人 尹盛莱 真珉豪 金泰殊 黄奎雄

(74) 专利代理机构 北京律盟知识产权代理有限  
责任公司 11287

代理人 宋献涛

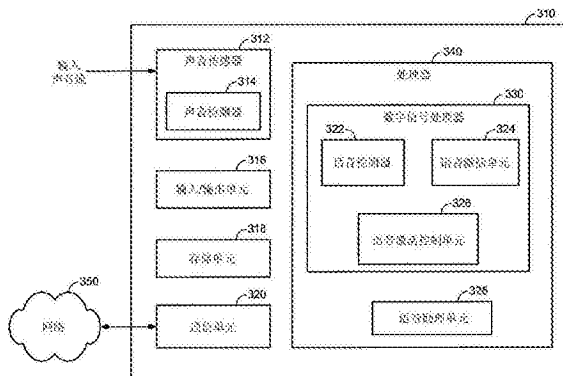
权利要求书3页 说明书15页 附图15页

(54) 发明名称

用于控制语音激活的方法和设备

(57) 摘要

本发明揭示一种用于在移动装置中通过目标关键字来控制语音激活的方法。所述方法包含接收输入声音流。当所述输入声音流指示话音时,激活所述语音激活单元以检测所述目标关键字,并且从所述输入声音流提取至少一个声音特征。此外,所述方法包含在所述至少一个声音特征指示非目标关键字时去激活所述语音激活单元。



1. 一种用于在移动装置中通过目标关键字控制语音激活的方法,所述方法包括:  
接收输入声音流;  
当所述输入声音流指示话音时,激活语音激活功能以检测所述目标关键字;  
从所述输入声音流提取至少一个声音特征;以及  
当所述至少一个声音特征指示非目标关键字时,去激活所述语音激活功能。
2. 根据权利要求 1 所述的方法,其进一步包括:  
确定与所述至少一个声音特征的非目标关键字状态相对应的第一度量值;以及  
将所述第一度量值与基于所述输入声音流的第二度量值比较,其中基于所述比较的结果确定所述至少一个声音特征指示所述非目标关键字。
3. 根据权利要求 2 所述的方法,其中所述第二度量值与所述至少一个声音特征的目标关键字状态相对应。
4. 根据权利要求 2 所述的方法,其中所述第二度量值与在第一时间段期间的平均非目标关键字状态相关联。
5. 根据权利要求 1 所述的方法,其进一步包括从所述至少一个声音特征确定关键字分数和非关键字分数。
6. 根据权利要求 5 所述的方法,其中所述关键字分数和所述非关键字分数通过所述语音激活功能确定。
7. 根据权利要求 5 所述的方法,其中去激活所述语音激活功能包含在所述非关键字分数超出所述关键字分数第一分数阈值以上时去激活所述语音激活功能。
8. 根据权利要求 5 所述的方法,其中去激活所述语音激活功能包含基于在第一时间段上的所述非关键字分数的第一平均值和在第二时间段上的所述非关键字分数的第二平均值来去激活所述语音激活功能。
9. 根据权利要求 8 所述的方法,其中去激活所述语音激活功能包含在所述第一平均值比所述第二平均值大第二分数阈值以上时,去激活所述语音激活功能。
10. 根据权利要求 9 所述的方法,其进一步包括基于所述移动装置的背景信息调整所述第二分数阈值。
11. 根据权利要求 8 所述的方法,其中所述第一时间段比所述第二时间段更短。
12. 根据权利要求 1 所述的方法,其进一步包括:  
当在去激活所述语音激活功能之后所接收的随后的输入声音流指示话音时,再激活所述语音激活功能;以及  
当从所述随后的输入声音流提取的至少一个声音特征指示非目标关键字时,去激活所述语音激活功能。
13. 根据权利要求 12 所述的方法,其进一步包括对所述语音激活功能去激活的次数计数,  
其中再激活所述语音激活功能包含在所述语音激活功能去激活的所述次数小于特定数目时再激活所述语音激活功能。
14. 根据权利要求 13 所述的方法,其进一步包括:  
如果所述语音激活功能去激活的所述次数大于或等于所述特定数目,那么确定在所述所接收的输入声音流中的沉默的持续时间;以及

如果所述沉默的持续时间超出特定时间段,那么再激活所述语音激活功能。

15. 根据权利要求 13 所述的方法,其进一步包括:

如果所述语音激活功能去激活的所述次数大于或等于所述特定数目,那么检测在所述所接收的输入声音流的背景中的改变;以及

如果检测到在所述所接收的输入声音流的所述背景中的所述改变,那么再激活所述语音激活功能。

16. 根据权利要求 13 所述的方法,其进一步包括如果所述语音激活功能去激活的所述次数大于或等于所述特定数目,那么在特定时间段上去激活接收所述输入声音流。

17. 根据权利要求 13 所述的方法,其中接收所述输入声音流包含确定所述输入声音流是否指示话音。

18. 根据权利要求 17 所述的方法,其进一步包括如果所述语音激活功能去激活的所述次数大于或等于所述特定数目时,那么在特定时间段上去激活确定所述输入声音流是否指示话音。

19. 根据权利要求 13 所述的方法,其进一步包括复位所述语音激活功能去激活的所述次数。

20. 根据权利要求 19 所述的方法,其中复位所述语音激活功能去激活的所述次数包含在所述语音激活功能去激活的所述次数在特定时间上保持不变时,复位所述语音激活功能去激活的所述次数。

21. 根据权利要求 1 所述的方法,其中所述语音激活功能由经配置以激活在所述移动装置中的至少一个应用的语音激活单元执行。

22. 一种用于通过目标关键字控制语音激活的移动装置,所述移动装置包括:

接收器,其经配置以接收输入声音流;

话音检测器,其经配置以确定所述输入声音流是否指示话音;

语音激活单元,其经配置以在所述输入声音流指示话音时,在所述话音检测器激活所述语音激活单元以检测所述目标关键字之后从所述输入声音流提取至少一个声音特征;以及

语音激活控制单元,其经配置以在所述至少一个声音特征指示非目标关键字时去激活所述语音激活单元。

23. 根据权利要求 22 所述的移动装置,其中所述语音激活单元进一步经配置以从所述至少一个声音特征确定关键字分数和非关键字分数。

24. 根据权利要求 23 所述的移动装置,其中所述语音激活控制单元经配置以在所述非关键字分数超出所述关键字分数第一分数阈值以上时去激活所述语音激活单元。

25. 根据权利要求 23 所述的移动装置,其中所述语音激活控制单元经配置以基于在第一时间段上的所述非关键字分数的第一平均值和在第二时间段上的所述非关键字分数的第二平均值来去激活所述语音激活单元。

26. 一种非暂时性计算机可读存储媒体,其存储用于在移动装置中通过目标关键字控制语音激活的指令,所述指令使得处理器执行操作,所述操作包括:

接收输入声音流;

当所述输入声音流指示话音时,激活语音激活功能以检测所述目标关键字;

从所述输入声音流提取至少一个声音特征 ;以及  
当所述至少一个声音特征指示非目标关键字时,去激活所述语音激活功能。

27. 根据权利要求 26 所述的非暂时性计算机可读存储媒体,其中所述操作进一步包括 :

确定与所述至少一个声音特征的非目标关键字状态相对应的第一度量值 ;以及  
将所述第一度量值与基于所述输入声音流的第二度量值比较,其中基于所述比较的结果确定所述至少一个声音特征指示所述非目标关键字。

28. 根据权利要求 27 所述的非暂时性计算机可读存储媒体,其中所述第二度量值与所述至少一个声音特征的目标关键字状态相对应。

29. 一种用于通过目标关键字控制语音激活的移动装置,所述移动装置包括 :

用于接收输入声音流的装置 ;

用于确定所述输入声音流是否指示语音的装置 ;

用于在所述输入声音流指示语音时在所述用于确定的装置激活用于提取的装置以检测所述目标关键字之后从所述输入声音流提取至少一个声音特征的装置 ;以及

用于在所述至少一个声音特征指示非目标关键字时去激活所述用于提取的装置的装置。

30. 根据权利要求 29 所述的移动装置,其进一步包括用于从所述至少一个声音特征确定关键字分数和非关键字分数的装置。

## 用于控制语音激活的方法和设备

### [0001] 优先权主张

[0002] 本申请案主张共同拥有的 2013 年 5 月 7 日提交的第 61/820,526 号美国临时专利申请案、2013 年 8 月 9 日提交的第 61/864,490 号美国临时专利申请案,以及 2013 年 11 月 27 日提交的第 14/092,527 号美国非临时申请案的优先权,所述申请案中的每一者的内容以全文引用的方式明确地并入本文中。

### 技术领域

[0003] 本发明大体上涉及语音激活,且更具体来说,涉及在移动装置中控制语音激活。

### 背景技术

[0004] 近年来,移动装置(例如智能电话、个人数字助理(PDA)、平板计算机以及膝上型计算机)的使用已经变得分布广泛。这些装置允许用户执行多种功能,例如浏览因特网、拍摄相片或视频、打电话或进行视频通话等。另外,此类装置通常提供具有各种功能性的应用。

[0005] 此外,许多此类移动装置可以包含用于捕获音频信号的一或多个麦克风。这些装置中的一些可以允许捕获且处理音频信号以用于语音激活。使用此类语音激活特征,可以通过在移动装置中输入语音关键字来激活应用。

[0006] 为了充分利用此类语音激活特征,移动装置通常需要经配置以针对可能的关键字连续监视且处理环境声音。然而,语音激活特征的连续操作可能耗尽移动装置的有限功率资源。

### 发明内容

[0007] 本发明提供用于在移动装置中控制语音激活的方法、系统和装置。在这些方法、系统和装置中,一旦接收输入声音流,就可以在输入声音流指示话音时激活语音激活单元以检测目标关键字。此外,可以从输入声音流提取至少一个声音特征,且当所述至少一个声音特征指示非目标关键字时可以去激活语音激活单元。

[0008] 根据本发明的一个方面,揭示一种用于在移动装置中通过目标关键字控制语音激活的方法。在所述方法中,接收输入声音流,且当所述输入声音流指示话音时激活语音激活单元以检测目标关键字。从输入声音流提取至少一个声音特征,且当所述至少一个声音特征指示非目标关键字时去激活语音激活单元。本发明还描述了与此方法相关的装置、系统、装置的组合以及计算机可读媒体。

[0009] 根据本揭示内容的另一方面,揭示一种用于通过目标关键字控制语音激活的移动装置。所述移动装置包含接收器、话音检测器、语音激活单元以及语音激活控制单元。所述接收器经配置以接收输入声音流。此外,所述话音检测器经配置以在所述输入声音流指示话音时激活语音激活单元以检测目标关键字。语音激活单元经配置以从输入声音流提取至少一个声音特征,且语音激活控制单元经配置以在所述至少一个声音特征指示非目标关键

字时去激活所述语音激活单元。

## 附图说明

[0010] 图 1 图示根据本发明的一个实施例的移动装置,其经配置以捕获输入声音且在于所述输入声音中检测到目标关键字时执行语音激活功能。

[0011] 图 2 图示根据本发明的一个实施例的在会议中的用户,其中所述用户的移动装置经配置以在所述会议期间的对话不包含目标关键字时去激活语音激活单元。

[0012] 图 3 图示根据本发明的一个实施例的移动装置的框图,所述移动装置经配置以检测在输入声音流中的目标关键字以用于激活在所述移动装置中的功能。

[0013] 图 4A 图示根据本发明的一个实施例的 DSP 的框图,所述 DSP 经配置以从声音传感器接收激活信号和输入声音流且产生用于激活语音助理单元的激活信号。

[0014] 图 4B 图示根据本发明的另一实施例的 DSP 的框图,所述 DSP 经配置以从声音传感器接收激活信号和输入声音流且产生用于激活语音助理单元的激活信号。

[0015] 图 5 图示根据本发明的一个实施例的语音激活单元的更详细框图,所述语音激活单元经配置以接收输入声音流且在所述输入声音流中检测目标关键字。

[0016] 图 6 图示根据本发明的一个实施例的输入声音流的多个置信度分数、多个当前非关键字分数、多个当前关键字分数以及多个总平均非关键字分数的示范性曲线图。

[0017] 图 7 图示根据本发明的一个实施例的一种方法的流程图,所述方法用于基于由语音激活单元计算的关键字分数和非关键字分数去激活移动装置的语音激活单元。

[0018] 图 8 图示根据本发明的一个实施例的一种方法的流程图,所述方法用于通过语音激活控制单元去激活语音激活单元以终止接收输入声音流和处理正被接收的输入声音流。

[0019] 图 9 图示根据本发明的一个实施例的一种方法的流程图,所述方法用于在去激活语音激活单元以终止接收输入声音流和处理正被接收的输入声音流之后再激活所述语音激活单元。

[0020] 图 10 图示根据本发明的一个实施例的一种方法的详细流程图,所述方法用于检测输入声音流的状态改变以在去激活的次数等于预定值时再激活语音激活单元。

[0021] 图 11 图示根据本发明的另一实施例的一种方法的详细流程图,所述方法用于检测输入声音的状态改变以在去激活的次数等于预定值时再激活语音激活单元。

[0022] 图 12A 图示示出根据本发明的一个实施例的来自用于确定平均非关键字分数阈值的模拟的多条线的曲线图。

[0023] 图 12B 图示示出根据本发明的另一实施例的来自用于确定平均非关键字分数阈值的模拟的多条线的曲线图。

[0024] 图 13 是根据本发明的一个实施例的具有无线通信能力的示范性移动装置的框图。

## 具体实施方式

[0025] 现在将详细参考各种实施例,在附图中图示所述实施例的实例。在以下详细描述中,阐述众多具体细节以便提供对本发明的透彻理解。然而,对于所属领域的技术人员将显而易见的是,可在没有这些具体细节的情况下实践本发明。在其它情况下,未详细描述众所

周知的方法、程序、系统和组件,以便不会不必要地混淆各种实施例的各方面。

[0026] 图 1 图示根据本发明的一些实施例的移动装置 110,其经配置以捕获输入声音流且在于所述输入声音流中检测到目标关键字时执行语音激活功能。在所说明的实施例中,用户 120 可以在躺在床上 130 上时说出目标关键字且移动装置 110 接收所说出的声音。响应于所说出的声音,移动装置 110 激活语音激活单元以在所接收的输入声音流中检测目标关键字。语音激活单元从输入声音流提取至少一个声音特征。当至少一个声音特征指示非目标关键字时,移动装置 110 去激活语音激活单元。

[0027] 移动装置 110 可以时任何合适的装置,例如蜂窝式电话、智能电话、配备有声音捕获能力(例如,允许检测目标关键字以用于激活语音激活功能的麦克风)的膝上型计算机或平板计算机。例如,在入睡之前,用户 120 可以口头地在移动装置 110 中设置闹钟。在此情况下,用户 120 可以向移动装置 110 说出目标关键字“设置闹钟”,所述移动装置激活闹钟应用。移动装置 110 随后接收所述声音且确定所接收的声音(即,输入声音流)是否超出预定阈值声音强度。此外,当所接收的声音超出预定阈值声音强度时,移动装置 110 通过使用声音分类方法(例如,支持向量机技术)确定所接收的声音是否是话音。

[0028] 如果移动装置 110 确定所接收的声音是超出预定阈值声音强度的话音,那么移动装置 110 激活语音激活单元以确定所接收的语音声音是否包含目标关键字,所述目标关键字可以事先存储在移动装置 110 的关键字数据库中。在确定所接收的声音包含目标关键字后,移动装置 110 激活与目标关键字相关联的闹钟应用且允许用户 120 将闹钟设置成所需时间。如本文中所使用,术语“声音流”是指一或多个声音信号或声音数据的序列。此外,术语“关键字”是指可以用于激活在移动装置中的功能或应用的一或多个单词或声音的任何数字或模拟表示,且可以包含呈人的话音或对话的口头形式或经记录形式的任何数目的单词。如本文中所使用,术语“语音激活”意指通过在所捕获的声音中检测预定目标关键字来激活移动装置的至少一个功能或应用。

[0029] 为了允许语音激活,移动装置 110 可以经配置以连续监视输入声音流以用于确定所捕获的声音是否包含目标关键字,直到移动装置 110 从所捕获的声音检测到目标关键字。以此方式操作移动装置 110 通常导致增加的功率的使用,由此缩短电池寿命。因此,为了节约功率,可以通过控制语音激活来减少移动装置 110 的功率消耗,如下文在一些实施例中所述。

[0030] 在图 1 中,一旦激活移动装置 110 的语音激活单元,所述语音激活单元就可以活动模式继续操作,直到所述语音激活单元检测到目标关键字。在此情况下,因为用户 120 在睡觉,所以到移动装置 110 的输入声音流将不具有包含用户 120 的目标关键字的任何话音。因此,移动装置 110 经配置以自动地去激活移动装置 110 的语音激活单元,如下文更详细地描述。通过监视输入声音流且按需要自动地去激活语音激活单元,移动装置 110 可以节约功率且提高电池寿命。当移动装置 110 确定随后接收的声音是话音时,可以再激活去激活语音激活单元。

[0031] 如上文所述,当移动装置 110 确定所接收的声音是超出预定阈值声音强度的话音时,可以激活移动装置 110 的语音激活单元。在一些情况下,移动装置 110 的语音激活单元可能在没有任何语音输入的情况下由于噪声或故障而被激活。在此事件中,移动装置 110 经配置以检测激活且去激活语音激活单元以减少功率消耗。例如,如果在用户 120 如图 1

中所图示在睡觉时移动装置 110 的语音激活单元由于噪声或故障而被意外地激活,那么移动装置 110 可以基于随后接收的声音去激活语音激活单元。如本文中所使用,术语“激活”是指从断开状态或非活动状态(例如,空闲态)启用或开启装置或单元以允许所述装置或所述单元执行一或多个相关联功能。此外,术语“去激活”是指从激活状态(例如,“开启”状态)停用或断开装置或单元。术语“再激活”是指在装置或单元已经去激活之后启用或开启所述装置或所述单元。

[0032] 图 2 图示根据本发明的一个实施例的在会议中的用户 220,其中所述用户 220 的移动装置 210 经配置以在所述会议期间的对话不包含目标关键字时去激活语音激活单元。在所图示的情境中,移动装置 210 经配置以捕获输入声音流,且在于所述输入声音流中检测到目标关键字时执行语音激活功能。在一个实施例中,移动装置 210 经配置以通过声音传感器(例如麦克风)接收输入声音流,且确定所接收的声音是否是话音。如果移动装置 210 确定所接收的声音是话音,那么移动装置 210 激活语音激活单元以在所接收的输入声音流中检测目标关键字。

[0033] 在图 2 的会议情境中,移动装置 210 可以接收会议的会话声音且确定所述对话是话音。基于所检测到的话音,随后激活移动装置 210 的语音激活单元以用于确定所接收的声音是否包含预定目标关键字。如果所接收的对话声音在预定时间段上不包含目标关键字,那么可以假设,随后的对话声音将不太可能包含目标关键字。因此,当移动装置 210 确定所接收的声音在预定时间段上不包含目标关键字时,所述移动装置可以去激活语音激活单元以终止处理包含具有相似背景的随后的声音的会话声音。

[0034] 如果会议结束且不再接收会话声音,那么移动装置 210 可以再激活语音激活单元以检测目标关键字。例如,移动装置 210 可以通过监视被接收的输入声音流来检测对话结束的时间。在一个实施例中,当检测到随后的会话声音且在先前接收的会话声音和随后的会话声音之间存在预定的沉默的持续时间时,移动装置 210 可以再激活语音激活单元。如本文中所使用,术语“沉默”意指其中接收到在预定阈值声音强度之下的声音的状态。在另一实施例中,当检测到随后的声音是话音且在先前的声音和随后的声音之间检测到背景的改变时,移动装置 210 可以再激活语音激活单元。

[0035] 图 3 描绘根据本发明的一个实施例的移动装置 310 的框图,所述移动装置经配置以检测在输入声音流中的目标关键字以用于激活在移动装置 310 中的功能。移动装置 310 包含声音传感器 312、输入/输出(I/O)单元 316、存储单元 318、通信单元 320 以及处理器 340。类似于移动装置 110 和 210,移动装置 310 可以是配备有声音捕获和处理能力的任何合适的装置,例如蜂窝式电话、智能电话、个人计算机、膝上型计算机、平板计算机、智能电视、游戏装置、多媒体播放器等。

[0036] 处理器 340 包含数字信号处理器(DSP)330 和语音助理单元 328,并且可以是用于管理和操作移动装置 310 的应用程序处理器或中央处理单元(CPU)。DSP 330 包含话音检测器 322、语音激活单元 324 以及语音激活控制单元 326。在一个实施例中,DSP 330 是用于减少在处理声音流时的功率消耗的低功率处理器。在此配置中,DSP 330 中的语音激活单元 324 经配置以在于输入声音流中检测到目标关键字时激活语音助理单元 328。尽管在所说明的实施例中语音激活单元 324 经配置以激活语音助理单元 328,但所述语音激活单元也可以激活可能与目标关键字相关联的任何功能或应用。

[0037] 声音传感器 312 可以经配置以接收输入声音流并且将它提供给 DSP 330 中的话音检测器 322。声音传感器 312 可以包含一或多个麦克风或可以用于接收、捕获、感测和 / 或检测到移动装置 310 的声音输入流的任何其它类型的声音传感器。另外,声音传感器 312 可以采用任何合适的软件和 / 或硬件以用于执行此类功能。

[0038] 在一个实施例中,声音传感器 312 可以经配置以根据占空比周期性地接收输入声音流。声音传感器 312 可以进一步包含声音检测器 314,其用于确定输入声音流的所接收的部分是否超出预定阈值声音强度。当输入声音流的所接收的部分超出阈值声音强度时,声音传感器 312 的声音检测器 314 激活话音检测器 322 且将所接收的部分提供到 DSP 330 中的话音检测器 322。替代地,在不确定所接收的部分是否超过阈值声音强度的情况下,声音传感器 312 可以周期性地接收输入声音流的一部分且激活话音检测器 322 以将所接收的部分提供给话音检测器 322。可以了解,声音检测器 314 可以独立地存在于声音传感器 312 的外部或可以包含于话音检测器 322 等其它单元中。

[0039] 为了供检测目标关键字使用,存储单元 318 存储目标关键字和与所述目标关键字的多个部分相关联的信息。在一个实施例中,目标关键字可被划分成多个基本声音单元,例如,音素、音位或其子单元,并且表示目标关键字的多个部分可基于基本声音单元产生。在一些实施例中,目标关键字的检测可以使用时间模式识别方法来执行,所述方法包含马尔可夫链模型,例如隐马尔可夫模型 (HMM)、半马尔可夫模型 (SMM) 或其组合。在此情况下,目标关键字的每个部分随后可以与在 HMM 或 SMM 下的状态相关联。状态信息可以包含在目标关键字的状态当中的输入状态和从所述状态中的每一者转换到包含自身的下一状态的转移信息。存储单元 318 可以使用任何合适的存储装置或存储器装置实施,例如, RAM (随机存取存储器)、ROM (只读存储器)、EEPROM (电可擦除可编程只读存储器)、快闪存储器、固态驱动器 (SSD) 或类似者。

[0040] 当激活时, DSP 330 中的话音检测器 322 从声音传感器 312 接收输入声音流的部分。在一个实施例中,话音检测器 322 从所接收的部分提取多个声音特征且通过使用任何合适的声音分类方法确定提取到的声音特征是否指例如话音等所关注的声音,所述声音分类方法例如,基于高斯混合模型 (GMM) 的分类器、HMM、神经网络、图解模型以及支持向量机 (SVM)。如本文中所使用,术语“声音特征”是指描述或表示指定的持续时间的声音的任何信息或数据,且可以包含从所述声音提取的频率或功率分量、表示声音的类的声音类别识别符和 / 或从所述声音提取的任何其它信息或数据。

[0041] 如果所接收的部分被确定为是所关注的声音,那么话音检测器 322 激活语音激活单元 324 并且输入声音流的所接收的部分和其余部分被提供给语音激活单元 324。在一些其它实施例中,可以在 DSP 330 中省略话音检测器 322。在此情况下,当所接收的部分超出阈值声音强度时,声音传感器 312 激活语音激活单元 324 并且将输入声音流的所接收的部分和其余部分直接提供给语音激活单元 324。

[0042] 当激活时,语音激活单元 324 经配置以连续地接收输入声音流并且从输入声音流检测目标关键字。当接收到输入声音流时,语音激活单元 324 可从输入声音流依次提取多个声音特征。在将 HMM 用于目标关键字的检测的情况下,语音激活单元 324 可以从存储单元 318 获得目标关键字的状态信息,包含多个状态、输入状态以及转移信息。对于每个声音特征,可以通过使用例如 GMM、神经网络以及 SVM 等任何合适的概率模型针对状态中的每一

者确定观察分数。依据转移信息,语音激活单元 324 可以获得从可能用于目标关键字的多个状态序列中的状态中的每一者到下一个状态的转移分数。

[0043] 在确定观察分数并且获得转移分数之后,语音激活单元 324 确定用于可能的状态序列的关键字分数。在一个实施例中,如果所确定的关键字分数当中的最大关键字分数超出预定分数阈值,那么语音激活单元 324 将输入声音流检测为目标关键字。在检测目标关键字之后,语音激活单元 324 产生且发射激活信号以激活与目标关键字相关联的语音助理单元 328。

[0044] 语音助理单元 328 响应于来自语音激活单元 324 的激活信号被激活。一旦被激活,语音助理单元 328 可以通过在触摸显示器单元上输出例如“我可以帮助你吗?”等消息和/或通过 I/O 单元 316 的扬声器执行语音助理功能。作为响应,用户可以说出语音命令以激活移动装置 310 的各种相关联的功能。例如,当接收到用于因特网搜索的语音命令时,语音助理单元 328 可以识别语音命令为搜索命令并且经由通信单元 320 通过网络 350 执行网络搜索。

[0045] 一旦语音激活单元 324 被激活,DSP 330 中的语音激活控制单元 326 就确定随后接收的输入声音流是否指示非目标关键字。如本文中所使用,术语“非目标关键字”是指除目标关键字外的所有声音且可以包含沉默、噪声以及不是目标关键字的任何单词或声音。当确定目标关键字并未包含于输入声音流中时,语音激活控制单元 326 可以去激活语音激活单元 324。这防止语音激活单元 324 连续处理输入声音流,由此减少移动装置 310 的功率消耗。

[0046] 图 4A 图示根据本发明的一个实施例的 DSP 330 的框图,所述 DSP 330 经配置以从声音传感器 312 接收激活信号和输入声音流且产生用于激活语音助理单元 328 的激活信号。如上文参考图 3 所描述,声音传感器 312 接收输入声音流,且当确定输入声音流是超出预定阈值声音强度的声音时,将激活信号(“开启”)输出到 DSP 330 中的语音检测器 322。连同激活信号,声音传感器 312 还可以将所接收的输入声音流提供到语音检测器 322。

[0047] 来自声音传感器 312 的激活信号激活语音检测器 322 以接收且处理输入声音流。在一个实施例中,语音检测器 322 确定所接收的输入声音流是否是语音。如果确定所接收的输入声音流是语音,那么语音检测器 322 产生激活信号(“开启”),所述激活信号连同输入声音流一起提供到语音激活单元 324。

[0048] 响应于来自语音检测器 322 的激活信号,开启语音激活单元 324 以接收且处理来自语音检测器 322 的输入声音流。语音激活单元 324 可以将输入声音流分段成多个帧且从经分段帧中的每一者提取声音特征。在特定实施例中,语音激活单元 324 可以确定与所提取的声音特征的非目标关键字状态相对应的第一度量值。例如,语音激活单元 324 可以确定与所提取的声音特征与非目标关键字相关联的似然性相对应的度量值。语音激活单元 324 可以将第一度量值与基于输入声音流的第二度量值比较。例如,第二度量值可以与所提取的声音特征的目标关键字状态(例如,所提取的声音特征与目标关键字相关联的似然性)相对应,或第二度量值可以与在特定时间段期间的平均非目标关键字状态相对应。语音激活单元 324 可以基于比较的结果确定所提取的声音特征是否指示非目标关键字。例如,如果第一度量值大于第二度量值,那么语音激活单元 324 可以确定所提取的声音特征指示非目标关键字。

[0049] 在另一个特定实施例中,基于所提取的声音特征,语音激活单元 324 确定指示所接收的输入声音流包含目标关键字的概率的关键字分数和指示所接收的输入声音流不包含目标关键字的概率的非关键字分数。在一个实施例中,如果关键字分数超出预定分数阈值,那么语音激活单元 324 将输入声音流检测为目标关键字。在检测到目标关键字后,语音激活单元 324 产生且发射激活信号以激活语音助理单元 328。

[0050] 一旦激活语音激活单元 324,所述语音激活单元就通过处理随后的输入声音流继续以活动状态操作。这可能不必要地耗尽移动装置 310 的功率资源,尤其是在随后的输入声音流不包含目标关键字时。在此类情况下,语音激活控制单元 326 用于断开语音激活单元 324。

[0051] 在一个实施例中,在确定关键字分数和非关键字分数之后,语音激活单元 324 将关键字分数和非关键字分数输出到语音激活控制单元 326。基于关键字分数和非关键字分数,语音激活控制单元 326 确定输入声音流是否不包含目标关键字。例如,语音激活控制单元 326 可以确定指示在关键字分数和非关键字分数之间的差值的置信度分数,且在置信度分数小于置信度分数阈值时检测输入声音流不包含目标关键字。语音激活控制单元 326 还可以确定当前非关键字分数和总非关键字平均分数,且在当前非关键字分数超出总非关键字平均分数超过平均非关键字分数阈值时,检测输入声音流不包含目标关键字。如果语音激活控制单元 326 确定输入声音流不包含目标关键字,那么语音激活控制单元 326 产生且发射去激活信号(“断开”)以断开语音激活单元 324。如上文所描述,语音激活控制单元 326 基于由语音激活单元 324 确定的分数控制语音激活单元 324。因此,语音激活控制单元 326 不需要使用另外的计算资源来计算分数且语音激活控制单元 326 可以有效地控制语音激活单元 324。

[0052] 另外,当随后的输入声音流指示话音时,语音激活控制单元 326 发射控制信号以控制话音检测器 322 开启语音激活单元 324。响应于控制信号,话音检测器 322 确定在去激活语音激活单元 324 之后接收的随后的输入声音流是否是话音。如果话音检测器 322 确定随后的输入声音流不是话音(例如,沉默),那么话音检测器 322 不激活语音激活单元 324。因此,语音激活单元 324 可以保持去激活而不消耗功率。另一方面,如果话音检测器 322 确定随后的输入声音流是话音,那么话音检测器 322 可以激活语音激活单元 324 以用于检测在随后的输入声音流中的目标关键字。

[0053] 在一个实施例中,在激活之后由语音激活单元 324 接收且处理的输入声音流可以是不包含目标关键字的连续对话(例如,一般的话音、一般的对话等)。在此情况下,语音激活控制单元 326 基于关键字分数和非关键字分数去激活语音激活单元 324,如上文所描述。一旦语音激活单元 324 已经去激活,声音传感器 312 就可以接收随后的输入声音流。当随后的输入声音流指示超出阈值声音强度的声音时,声音传感器 312 的声音检测器 314 将激活信号发射到话音检测器 322,如果随后的输入声音流包含话音,那么所述话音检测器可以转而将激活信号发射到语音激活单元 324。

[0054] 当再激活语音激活单元 324 时,语音激活单元 324 可以针对随后的输入声音流的每个声音特征计算关键字分数和非关键字分数。随后将关键字分数和非关键字分数发射到语音激活控制单元 326。基于关键字和非关键字分数,语音激活控制单元 326 可以确定随后的输入声音流不包含目标关键字且将去激活信号发射到语音激活单元 324。以此方式,语音

激活单元 324 可以反复地激活和去激活。

[0055] 如果语音激活控制单元 326 连续地确定输入声音流不包含目标关键字的次数超出预定计数,那么可以假设当前被接收的输入声音流不太可能包含目标关键字。根据一个实施例,语音激活控制单元 326 基于所产生且发射到语音激活单元 324 的去激活信号的数目对确定输入声音流不包含目标关键字的次数计数。如果去激活的次数大于或等于预定计数,那么语音激活控制单元 326 可以将控制信号发射到语音检测器 322 以防止语音激活单元 324 的激活,即使随后的输入声音流指示语音。因此,语音激活控制单元 326 可以去激活语音激活单元 324 以终止接收输入声音流且处理正被接收的输入声音流。

[0056] 在语音激活单元 324 已经基于连续的去激活的计数而去激活之后,所述语音激活单元可以被再激活以接收且处理随后的声音流。在一个实施例中,如果语音检测器 322 已经结束处理在输入声音流中的前一对话且接收新的对话,那么所述语音检测器可以再激活语音激活单元 324,因为存在最新接收的对话包含目标关键字的可能性。例如,当语音检测器 322 检测到在先前接收的对话和随后的对话之间的预定的沉默的持续时间时,所述语音检测器可以检测到前一对话结束。替代地,在声音传感器 312 中的声音检测器 314 可以经配置以检测沉默的持续时间以用于激活语音激活单元 324。

[0057] 在另一实施例中,在语音激活单元 324 基于连续的去激活的计数被去激活之后,当检测到在输入声音流的声音背景中的改变时,语音检测器 322 可以再激活所述语音激活单元。在背景中的改变是指在声音环境中的改变,例如从一个声音环境改变到另一个声音环境。例如,当移动装置 310 从汽车移动到室内时,从汽车和室内接收的环境声音通常显示可以用于识别每个环境的背景的不同声音特性。所接收的声音的背景信息可以通过采用任何合适的背景确定方法来确定。例如,可以将输入声音的声音特征与从具体背景训练出的数据库中的声音特征比较和匹配。之后,所接收的声音的背景信息可以从与数据库中的最佳匹配的声音特征相关联的背景信息确定。为了再激活语音激活单元 324,语音检测器 322 可以检测在前一输入声音和随后的输入声音流之间在声音背景中的改变。可以了解,任何其它单元(例如,语音激活控制单元 326)都可以替代语音检测器 322 检测在声音背景中的改变。

[0058] 图 4B 图示根据本发明的另一实施例的 DSP 330 的框图,所述 DSP 330 经配置以从声音传感器 312 接收激活信号和输入声音流且产生用于激活语音助理单元 328 的激活信号。类似于图 4A,来自声音传感器 312 的声音检测器 314 的激活信号激活语音检测器 322 以接收且处理输入声音流。如果确定所接收的输入声音流是语音,那么语音检测器 322 产生激活信号(“开启”),所述激活信号连同输入声音流一起提供到语音激活单元 324。响应于来自语音检测器 322 的激活信号,语音激活单元 324 产生关键字分数和非关键字分数。基于由语音激活单元 324 确定的分数,语音激活控制单元 326B 控制语音激活单元 324。此外,语音激活控制单元 326B 基于所产生且发射到语音激活单元 324 的去激活信号的数目对确定输入声音流不包含目标关键字的次数计数。

[0059] 在此实施例中,如果去激活的次数大于或等于预定计数,那么语音激活控制单元 326B 可以将控制信号发射到声音传感器 312 的声音检测器 314 以去激活声音检测器 314。例如,当声音检测器 314 从语音激活控制单元 326B 接收控制信号时,声音检测器 314 可以在预定时间段(例如,10 秒)上去激活。在替代实施例中,如果语音激活单元 324 去激活的

次数大于或等于预定计数,那么声音传感器 312 可以在预定时间段上去激活。

[0060] 图 5 图示根据本发明的一个实施例的语音激活单元 324 的更详细框图,所述语音激活单元经配置以接收输入声音流且在所述输入声音流中检测目标关键字。语音激活单元 324 包含分段单元 510、特征提取器 520、分数确定单元 530 和关键字检测单元 540。分数确定单元 530 包含观察分数确定单元 550 和最大关键字分数确定单元 560。

[0061] 分段单元 510 接收来自话音检测器 322 的输入声音流并且将所接收的输入声音流分段成相等时间周期的多个顺序帧。特征提取器 520 从分段单元 510 依次接收经分段帧并且从所述帧中的每一者提取声音特征。在一个实施例中,特征提取器 520 可以使用梅尔倒频谱系数 (MFCC) 方法等任何合适的特征提取方法从所述帧提取声音特征。例如,在 MFCC 方法的情况下,从经分段帧中的每一者计算  $n$  维向量中的 MFCC 并且将所述向量用作声音特征。

[0062] 在分数确定单元 530 中,观察分数确定单元 550 从存储单元 318 依次接收声音特征并且接收目标关键字的状态信息。根据一个实施例,目标关键字的状态信息可以包含与目标关键字的多个部分相关联的多个状态以及概率模型(例如,概率函数),例如提供用于所述状态中的每一者的 GMM。如上文所描述,目标关键字可以划分成多个基本声音单元,且表示目标关键字的多个部分可以基于所述基本声音单元产生。在一些实施例中,目标关键字的状态还可以包含非关键字状态(例如,“填充物”状态),其不与表示目标关键字的多个部分中的任一者相关联。例如,在目标关键词包含预定数目的状态(所述状态中的每一者与音素等基本声音单元相对应)的情况下,非关键字状态可以表示除包含于目标关键字中的那些基本声音单元外的基本声音单元。

[0063] 当接收每个声音特征时,观察分数确定单元 550 基于声音特征和从存储单元 318 接收的状态信息确定目标关键字的每个状态的观察分数。在一个实施例中,通过根据用于相关联的状态的概率模型计算概率值来针对所接收的声音特征确定所述状态中的每一者的观察分数。概率值中的每一者可以用作相关联状态的观察分数。状态的高观察分数指示声音特征与状态的基本声音单元相对应的高概率。观察分数确定单元 550 将所接收的声音特征中的每一者的观察分数提供到最大关键字分数确定单元 560 以用于确定多个状态序列的关键字分数,所述状态序列对于目标关键字是可能的。

[0064] 最大关键字分数确定单元 560 接收声音特征中的每一者的观察分数且从存储单元 318 获得状态信息。状态信息可以包含目标关键字的多个状态,包含非关键字状态、所述状态当中的输入状态、以及从状态中的每一者到包含自身的下一状态的转移信息。输入状态可以表示在目标关键字的可能的状态序列中的每一者中非关键字状态转移到的第一状态。

[0065] 状态信息中的转移信息包含在目标关键字的可能的状态序列中的每一者中从所述状态中的每一者到下一状态的转移分数。转移分数可以表示在每个可能的状态序列中所述状态中的每一个转移到下一状态的概率值。转移分数还包含从非关键字状态到输入状态的转移分数。

[0066] 基于所接收的观察分数和转移分数,最大关键字分数确定单元 560 计算可能的状态序列中的每一者的关键字分数。在此情况下,状态序列可以从非关键字状态(即,起始状态)开始,因为非关键字状态在接收到输入声音流之前得到指派。由此,转移分数包含从非

关键字状态到输入状态的转移分数,并且还包含在状态序列中从非关键字状态到其自身的转移分数。以此方式,当从观察分数确定单元 550 接收每个声音特征的观察分数的集合时,最大关键字分数确定单元 560 将下一状态添加到每个状态序列且确定所更新的状态序列中的每一者的关键字分数。最大关键字分数确定单元 560 随后在所更新的状态序列的关键字分数当中选择最大关键字分数。在一个实施例中,可以通过使用维特比算法等任何合适的方法计算关键字分数来确定最大关键字分数。在确定最大关键字分数之后,最大关键字分数确定单元 560 将它提供给关键字检测单元 540。

[0067] 在从最大关键字分数确定单元 560 接收最大关键字分数后,关键字检测单元 540 基于最大关键字分数在输入声音流中检测目标关键字。例如,关键字检测单元 540 可以从存储单元 318 接收用于检测目标关键字的分数阈值,且如果最大关键字分数大于所接收的分数阈值,那么检测到目标关键字。在此情况下,分数阈值可以被设定为用于检测在所需置信度水平内的目标关键字的最小关键字分数。

[0068] 在一些实施例中,最大关键字分数确定单元 560 确定非关键字状态序列的非关键字分数。可以在所确定的可能的状态序列的关键字分数当中选择非关键字分数且将其提供到关键字检测单元 540。一旦检测到目标关键字,关键字检测单元 540 就产生并且提供激活信号以开启与目标关键字相关联的语音助理单元 328。

[0069] 基于关键字分数(例如,最大关键字分数)或非关键字分数,语音激活控制单元 326 确定输入声音流是否不包含目标关键字。在一个实施例中,语音激活控制单元 326 可以使用关键字分数和非关键字分数两者来改进检测准确性,尤其是当输入声音流包含可能影响关键字分数的噪声等环境声音时。在此实施例中,语音激活控制单元 326 可以确定指示在关键字分数和非关键字分数之间的差值的置信度分数。置信度分数指示从输入声音流检测到目标关键字的概率。如果置信度分数较高,那么它指示从输入声音流检测到目标关键字的高概率。另一方面,如果置信度分数较低,那么它指示从输入声音流检测到目标关键字的低概率。因此,如果置信度分数小于置信度分数阈值,那么语音激活控制单元 326 确定输入声音流不包含目标关键字。

[0070] 在另一实施例中,语音激活控制单元 326 在确定输入声音流不包含目标关键字时可以分析在某一时间段上的输入声音流。例如,如果用户说出目标关键字“hey”,那么当接收到仅音素“he”时,语音激活控制单元 326 可能初始地确定所接收的声音不是目标关键字。当语音激活控制单元 326 随后接收音素“y”时,它可以确定所接收的声音是目标关键字。因此,为了改进在确定输入声音流是否不包含目标关键字时的准确性,语音激活控制单元 326 基于非关键字分数的平均值确定输入声音流是否不包含目标关键字。

[0071] 语音激活控制单元 326 可以确定当前非关键字分数和总非关键字平均分数,且在一个实施例中确定当前非关键字分数和总非关键字平均分数之间的差值。当前非关键字分数可以是最近提取的声音特征计算出的非关键字分数。在另一实施例中,当前非关键字分数可以是多个最近提取的声音特征计算出的平均分数。总非关键字平均分数指示从在指定时间段上所提取的声音特征的全部计算出的平均分数,所述指定时间段可以周期性地复位。如果当前非关键字分数和总非关键字平均分数之间的差值大于平均非关键字分数阈值,那么语音激活控制单元 326 确定输入声音流不包含目标关键字。替代地,如果置信度分数小于置信度分数阈值且当前非关键字分数和总非关键字平均分数之间的差值大于平均

非关键字分数阈值,那么语音激活控制单元 326 可以确定输入声音流不包含目标关键字。

[0072] 图 6 图示根据一个实施例的输入声音流 600 的多个置信度分数 620、多个当前非关键字分数 640、多个当前关键字分数 650 以及多个总平均非关键字分数 660 的示范性曲线图 616 和 618。所接收的输入声音流 600 包含多个部分 602 到 614。如图所示,输入声音流 600 包含两个非关键字部分 602 和 606、三个关键字部分 604、610 和 614,以及两个沉默部分 608 和 612。如本文中所使用,术语“非关键字”是指除目标关键字、沉默和噪声外的所有声音。

[0073] 分数曲线图 616 图示输入声音流 600 的置信度分数 620 连同置信度分数阈值 630,所述置信度分数阈值是恒定的(如图 6 中通过粗体直线指示)。如所图示,与关键字部分 604、610 和 614 相对应的一些置信度分数 620 中超出置信度分数阈值 630。在这些情况下,语音激活控制单元 326 不确定输入声音流 600 不包含目标关键字。其它置信度分数 620 与非关键字部分 602 和 606 或沉默部分 608 和 612 相对应,且并未超出置信度分数阈值 630。因此,语音激活控制单元 326 可以确定与这些分数相对应的输入声音流 600 不包含目标关键字。

[0074] 分数曲线图 618 图示当前非关键字分数 640(如用黑色实线指示)、当前关键字分数 650(如用点线指示)、总平均非关键字分数 660(如用黑色粗体实线指示)。如带圆圈部分 670 中所图示,当前非关键字分数 640 中与非关键字部分 602 和 606 或沉默部分 612 相对应的一些超出总平均值非关键字分数 660 超过平均非关键字分数阈值。在这些情况下,语音激活控制单元 326 可以确定输入声音流 600 与不包含目标关键字的这些分数相对应。

[0075] 图 7 图示根据本发明的一个实施例的一种方法 700 的流程图,所述方法用于基于由语音激活单元 324 计算的关键字分数和非关键字分数去激活移动装置 310 的语音激活单元 324。移动装置 310 在 710 处通过声音传感器 312 接收输入声音流。随后,在 720 处,声音传感器 312 的声音检测器 314 和话音检测器 322 确定输入声音流是否是超出阈值声音强度的话音。如果输入声音流是超出阈值声音强度的话音,那么在 730 处,话音检测器 322 激活语音激活单元 324。然而,如果输入声音流不是超出阈值声音强度的话音,那么在 710 处,声音传感器 312 的声音检测器 314 和话音检测器 322 再次接收随后的输入声音流。

[0076] 在 740 处,语音激活单元 324 可以将输入声音流分段成多个声音帧且从每个声音帧提取声音特征。在提取声音特征之后,语音激活单元 324 从声音特征计算关键字分数和非关键字分数且将关键字分数和非关键字分数发射到语音激活控制单元 326。基于由语音激活单元 324 计算的关键字分数和非关键字分数,语音激活控制单元 326 在 750 处确定关键字分数和非关键字分数是否指示非目标关键字。如果语音激活控制单元 326 不能确定关键字分数和非关键字分数指示非目标关键字,那么语音激活单元 324 在 760 处接收随后的输入声音流。随后,语音激活单元 324 可以在 740 处计算随后的输入声音流的关键字分数和非关键字分数,直到语音激活单元 324 在 750 处检测到关键字分数和非关键字分数指示非目标关键字。如果语音激活控制单元 326 在 750 处确定关键字分数和非关键字分数指示非目标关键字(即,输入声音流指示非目标关键字),那么语音激活控制单元 326 在 770 处去激活语音激活单元 324。

[0077] 图 8 图示根据本发明的一个实施例的一种方法 800 的流程图,所述方法用于通过语音激活控制单元 326 去激活语音激活单元 324 以终止接收输入声音流和处理正被接收的

输入声音流。可以了解,在方法 800 的 810 到 840 处的操作以与在方法 700 的 740 到 770 处的操作相同的方式执行,且因此省略其描述。在于 840 处去激活语音激活单元 324 之后,语音激活控制单元 326 在 850 处增加去激活的次数。例如,移动装置 310 可以具有用于存储计数值的存储单元 318。计数值的初始值可以设为“0”。如果语音激活控制单元 326 将去激活信号发射到语音激活单元 324 且语音激活单元 324 去激活,那么语音激活控制单元 326 可以对计数值加“1”。

[0078] 在 860 处,语音激活控制单元 326 确定去激活的次数是否等于或超出预定值。如果去激活的次数等于或超出预定值,那么在 870 处,语音激活控制单元 326 发射控制信号以控制话音检测器 322 以防止在随后的输入声音流指示话音时开启语音激活单元 324。因此,语音激活控制单元 326 可以去激活语音激活单元 324 以终止接收输入声音流且处理正被接收的输入声音流。在将控制信号发射到话音检测器 322 之后,语音激活控制单元 326 在 890 处复位去激活的次数。然而,如果去激活的次数小于预定值,那么在 880 处,语音激活控制单元 326 发射控制信号以控制话音检测器 322 以在随后的输入声音流指示话音时开启语音激活单元 324。因此,在语音激活单元 324 的去激活之后,如果话音检测器 322 从随后的输入声音流检测到话音,那么话音检测器 322 可以开启语音激活单元 324。

[0079] 图 9 图示根据本发明的一个实施例的一种方法 900 的流程图,所述方法用于在去激活语音激活单元 324 以终止接收输入声音流和处理正被接收的输入声音流之后再激活所述语音激活单元 324。可以了解,在方法 900 的 910 到 970 处的操作以与在方法 700 的 710 到 770 处的操作相同的方式执行,且因此省略其描述。在 970 处已经去激活语音激活单元 324 之后,语音激活控制单元 326 在 980 处对去激活的次数计数。在 990 处,语音激活控制单元 326 确定去激活的次数是否等于或超出预定值。如果去激活的次数小于预定值,那么话音检测器 322 在 910 处接收随后的输入声音流。然而,如果去激活的次数等于或超出预定值,那么话音检测器 322 在 995 处确定是否检测到在所述输入声音流和随后的输入声音流之间的背景改变。如果检测到背景改变,那么话音检测器 322 处理随后的输入声音流以确定随后的输入声音流是否指示话音。

[0080] 图 10 图示根据本发明的一个实施例的一种方法 1000 的详细流程图,所述方法用于检测输入声音流的背景改变以在去激活的次数等于或超出预定值时再激活语音激活单元 324。如上文在图 9 中所描述,语音激活控制单元 326 在 990 处确定去激活的次数是否等于或超出预定值。如果去激活的次数不超出预定值,那么话音检测器 322 在 910 处接收随后的输入声音流。然而,如果去激活的数目等于或超出预定值,那么话音检测器 322 在 1010 处接收随后的输入声音流。在 1020 处,话音检测器 322 确定在先前接收的输入声音流和随后的输入声音流之间的沉默的持续时间。如果沉默的持续时间不大于预定时间段,那么话音检测器 322 重复地接收随后的输入声音流。然而,如果沉默的持续时间大于预定时间段(即,检测到输入声音的背景改变),那么话音检测器 322 在 920 处确定所接收的随后的输入声音是否是随后的输入声音流的话音。

[0081] 图 11 图示根据本发明的另一实施例的一种方法 1100 的详细流程图,所述方法用于检测输入声音的背景改变以在去激活的次数等于或超出预定值时再激活语音激活单元 324。如上文在图 9 中所描述,语音激活控制单元 326 在 990 处确定去激活的次数是否等于或超出预定值。如果去激活的次数不超出预定值,那么话音检测器 322 在 910 处接收随后

的输入声音流。然而,如果去激活的数目等于或超出预定值,那么话音检测器 322 在 1110 处接收随后的输入声音流。在 1120 处,话音检测器 322 检测在所捕获的随后的输入声音流的背景中的改变。如果未检测到在背景中的改变,那么话音检测器 322 可以重复地接收随后的输入声音流。然而,如果检测到在背景中的改变(即,检测到输入声音的背景改变),那么话音检测器 322 在 920 处确定所接收的随后的输入声音是否是随后的输入声音流的话音。

[0082] 图 12A 图示出根据本发明的一个实施例的来自用于确定平均非关键字分数阈值的模拟的多条线 1210a、1220a 和 1230a 的曲线图 1200a。在此实施例中,利用没有噪声的输入声音流执行模拟。曲线图 1200a 的 x 轴表示平均非关键字分数阈值,而 y 轴表示随平均非关键字分数阈值而变的线 1210a、1220a 和 1230a 的概率。线 1210a 指示当输入声音流实际上包含目标关键字时语音激活控制单元 326 错误地确定输入声音流不包含目标关键字的“假警报”的概率。

[0083] 假警报是不希望的,因为可能需要由用户重复输入目标关键字。因此,可能需要将假警报的概率设为近零。在假警报线 1210a 中,当平均非关键字分数阈值大于“2”时,假警报的概率接近近零。因此,在语音激活控制单元 326 中的平均非关键字分数阈值可以设定成“2”的值或更大。

[0084] 线 1220a 指示当输入声音流实际上不包含目标关键字时语音激活控制单元 326 错误地确定输入声音流包含目标关键字的“失误”的概率。另一方面,线 1230a 指示当输入声音流不包含目标关键字时由语音激活控制单元 326 正确地检测到非关键字的概率。如图所示,“失误”线 1220a 示出“失误”的概率初始地平坦近零,但在平均非关键字分数阈值“2”之后大大增加。相反,线 1230a 指示正确地检测到非关键字的概率靠近概率值“1”,但当平均非关键字分数达到值“2”时开始大大减小。考虑到线 1210a、1220a 和 1230a,平均非关键字分数阈值可以设在“2”和“3”之间以用于最优性能。

[0085] 图 12B 图示出根据本发明的另一实施例的来自用于确定平均非关键字分数阈值的模拟的多条线 1210b、1220b 和 1230b 的曲线图 1200b。在此实施例中,利用具有噪声的输入声音流执行模拟。类似于图 12A,线 1210b 指示当输入声音流实际上包含目标关键字时语音激活控制单元 326 错误地确定输入声音流不包含目标关键字的“假警报”的概率。此外,线 1220b 指示当输入声音流实际上不包含目标关键字时语音激活控制单元 326 错误地确定输入声音流包含目标关键字的“失误”的概率。另一方面,线 1230b 指示当输入声音流不包含目标关键字时由语音激活控制单元 326 正确地检测到非关键字的概率。如图所示,当平均非关键字分数阈值大于“3”时,假警报线 1210b 接近近零。另外,“失误”线 1220b 示出“失误”的概率初始地平坦近零,但在平均非关键字分数阈值“2”之后大大增加。相反,线 1230b 指示正确地检测到非关键字的概率靠近概率值“1”,但当平均非关键字分数阈值达到值“2”时开始大大减小。考虑到线 1210b、1220b 以及 1230b,平均非关键字分数阈值可以设在“3”和“4”之间以用于最优性能(即,用于减小假警报的概率)。如上文所描述,可以基于移动装置的背景信息(例如,嘈杂的背景)调整平均非关键字分数阈值。

[0086] 图 13 是根据本发明的一个实施例的具有无线通信能力的示范性移动装置 1300 的框图。移动装置 1300 可以是蜂窝式电话、终端、手机、个人数字助理(PDA)、无线调制解调器、无绳电话等等。无线通信系统可以是码分多址(CDMA)系统、全球移动通信系统(GSM)

系统、宽带 CDMA (W-CDMA) 系统、长期演进 (LTE) 系统、LTE 高级系统等等。

[0087] 移动装置 1300 可能能够经由接收路径和发射路径提供双向通信。在接收路径上, 由基站发射的信号被天线 1312 接收, 并且被提供到接收器 (RCVR) 1314。接收器 1314 调节且数字化所接收信号, 并且将经调节和数字化的信号提供到数字区段 1320 用于进一步处理。在发射路径上, 发射器 (TMTR) 从数字区段 1320 接收待发射的数据, 处理并调节所述数据, 且产生经调制信号, 所述经调制信号经由天线 1312 发射到基站。接收器 1314 和发射器 1316 是支持 CDMA、GSM、W-CDMA、LTE、LTE 高级等的收发器的部分。

[0088] 数字区段 1320 包含各种处理、接口和存储器单元, 例如, 调制解调器处理器 1322、精简指令集计算机 / 数字信号处理器 (RISC/DSP) 1324、控制器 / 处理器 1326、内部存储器 1328、广义音频编码器 1332、广义音频解码器 1334、图形 / 显示器处理器 1336, 和 / 或外部总线接口 (EBI) 1338 等。调制解调器处理器 1322 执行对数据发射和接收的处理, 例如, 编码、调制、解调和解码。RISC/DSP 1324 执行无线装置 1300 的通用和专用处理。控制器 / 处理器 1326 控制数字区段 1320 内的各种处理和接口单元的操作。内部存储器 1328 存储用于数字区段 1320 内的各种单元的数据和 / 或指令。

[0089] 广义音频编码器 1332 执行对来自音频源 1342、麦克风 1343 等的输入信号的编码。广义音频解码器 1334 执行对经译码音频数据的解码, 并且将输出信号提供到扬声器 / 耳机 1344。应注意, 广义音频编码器 1332 和广义音频解码器 1334 不一定需要与音频源、麦克风 1343 和扬声器 / 耳机 1344 介接, 并且因此未在所述移动装置 1300 中示出。图形 / 显示器处理器 1336 执行针对呈现给显示器单元 1346 的图形、视频、图像和文本的处理。EBI 1338 促进在数字区段 1320 与主存储器 1348 之间的数据传送。

[0090] 数字区段 1320 利用一或多个处理器、DSP、微处理器、RISC 等来实施。还可在一或多个专用集成电路 (ASIC) 和 / 或一些其它类型的集成电路 (IC) 上制造数字区段 1320。

[0091] 一般来说, 本文中描述的任何装置指示各种类型的装置, 例如, 无线电话、蜂窝电话、膝上型计算机、无线多媒体装置、无线通信个人计算机 (PC) 卡、PDA、外部或内部调制解调器、通过无线信道通信的装置等等。装置可以具有各种名称, 例如, 存取终端 (AT)、存取单元、订户单元、移动台、客户端装置、移动单元、移动电话、移动设备、远程站、远程终端、远程单元、用户装置、用户设备、手持式装置等。本文中描述的任何装置均可具有用于存储指令和数据的存储器以及硬件、软件、固件或其组合。

[0092] 本文中所描述的技术通过各种装置来实施。例如, 这些技术用硬件、固件、软件或其组合来实施。所属领域的技术人员将进一步了解, 结合本文中的揭示内容描述的各种说明性逻辑块、模块、电路和算法步骤可以实施为电子硬件、计算机软件或两者的组合。为清楚说明硬件与软件的此可互换性, 上文已大体上关于其功能性而描述了各种说明性组件、块、模块、电路及步骤。此类功能性是实施为硬件还是软件取决于特定应用及施加于整个系统的设计约束。所属领域的技术人员可针对每一特定应用以不同方式实施所描述的功能性, 但此类实施决策不应被解释为引起本发明的范围的偏离。

[0093] 对于硬件实施方案, 用于执行技术的处理单元在一或多个 ASIC、DSP、数字信号处理装置 (DSPD)、可编程逻辑装置 (PLD)、现场可编程门阵列 (FPGA)、处理器、控制器、微控制器、微处理器、电子装置、经设计以执行本文中所描述的功能的其它电子单元、计算机或其组合内实施。

[0094] 因此,结合本文中的揭示内容描述的各种说明性逻辑块、模块和电路可以用通用处理器、DSP、ASIC、FPGA 或经设计以执行本文所述的功能的其它可编程逻辑装置、离散门或晶体管逻辑、离散硬件组件或其任何组合来实施或执行。通用处理器可以是微处理器,但在替代方案中,处理器可以是任何常规处理器、控制器、微控制器或状态机。处理器还可以实施为计算装置的组合,例如,DSP 与微处理器的组合、多个微处理器、结合 DSP 核心的一或多个微处理器,或任何其它此类配置。

[0095] 如果用软件实施,则可将功能作为一或多个指令或代码存储在计算机可读媒体上或经由计算机可读媒体发射。计算机可读媒体包含计算机存储媒体与通信媒体两者,所述通信媒体包含促进将计算机程序从一处传送到另一处的任何媒体。存储媒体可以为可由计算机存取的任何可用媒体。作为举例且并不限于此,此类计算机可读媒体可包括 RAM、ROM、EEPROM、CD-ROM 或其它光盘存储装置、磁盘存储装置或其它磁性存储装置,或可用于携载或存储呈指令或数据结构形式的所需程序代码且可由计算机存取的任何其它媒体。此外,任何连接都恰当地称为计算机可读媒体。例如,如果使用同轴电缆、光纤电缆、双绞线、数字订户线(DSL)或例如红外线、无线电及微波等无线技术从网站、服务器或其它远程源发射软件,则同轴电缆、光纤电缆、双绞线、DSL 或例如红外线、无线电及微波等无线技术包含于媒体的定义中。如本文中所使用,磁盘和光盘包含压缩光盘(CD)、激光光盘、光学光盘、数字多功能光盘(DVD)、软性磁盘和蓝光光盘,其中磁盘通常以磁性方式再现数据,而光盘用激光以光学方式再现数据。上述各项的组合也应包含在计算机可读媒体的范围内。

[0096] 提供本发明的先前描述以使得所属领域的技术人员能够制造或使用本发明。所属领域的技术人员将易于明白对本发明的各种修改,且本文中界定的一般原理在不脱离本发明的精神或范围的情况下应用于其它变体。因此,本发明并不意图限于本文中所描述的实例,而应符合与本文中所揭示的原理及新颖特征相一致的最广泛范围。

[0097] 尽管称示范性实施方案利用在一或多个独立计算机系统的背景下的当前揭示的标的物的方面,但所述标的物不受如此限制,而是可结合任何计算环境(例如网络或分布式计算环境)来实施。此外,当前揭示的标的物的方面可以在多个处理芯片或装置中或跨越多个处理芯片或装置实施,且可类似地跨越多个装置实现存储。此类装置可以包含 PC、网络服务器和手持式装置。

[0098] 尽管已经以特定地针对结构特征和/或方法动作的语言来描述标的物,但应理解,所附权利要求书中所界定的标的物未必限于上文所描述的具体特征或动作。确切地说,揭示上文所描述的具体特征和动作作为实施权利要求书的实例形式。

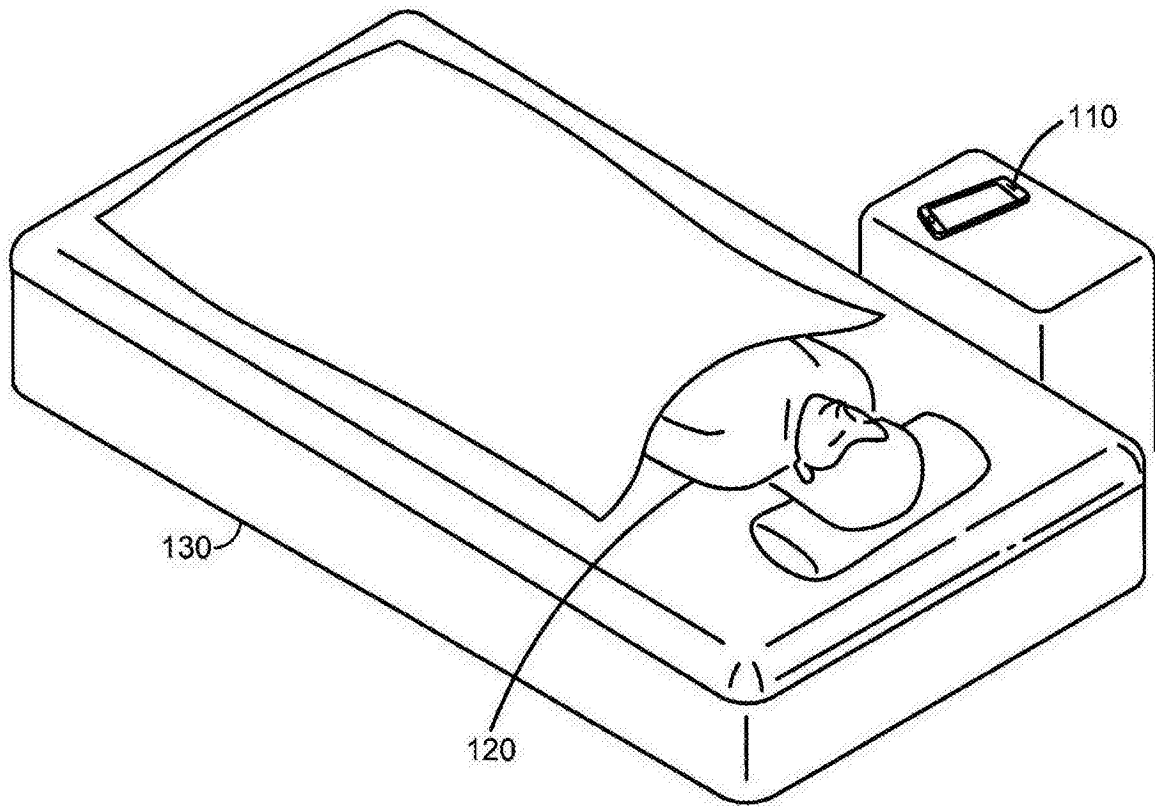


图 1

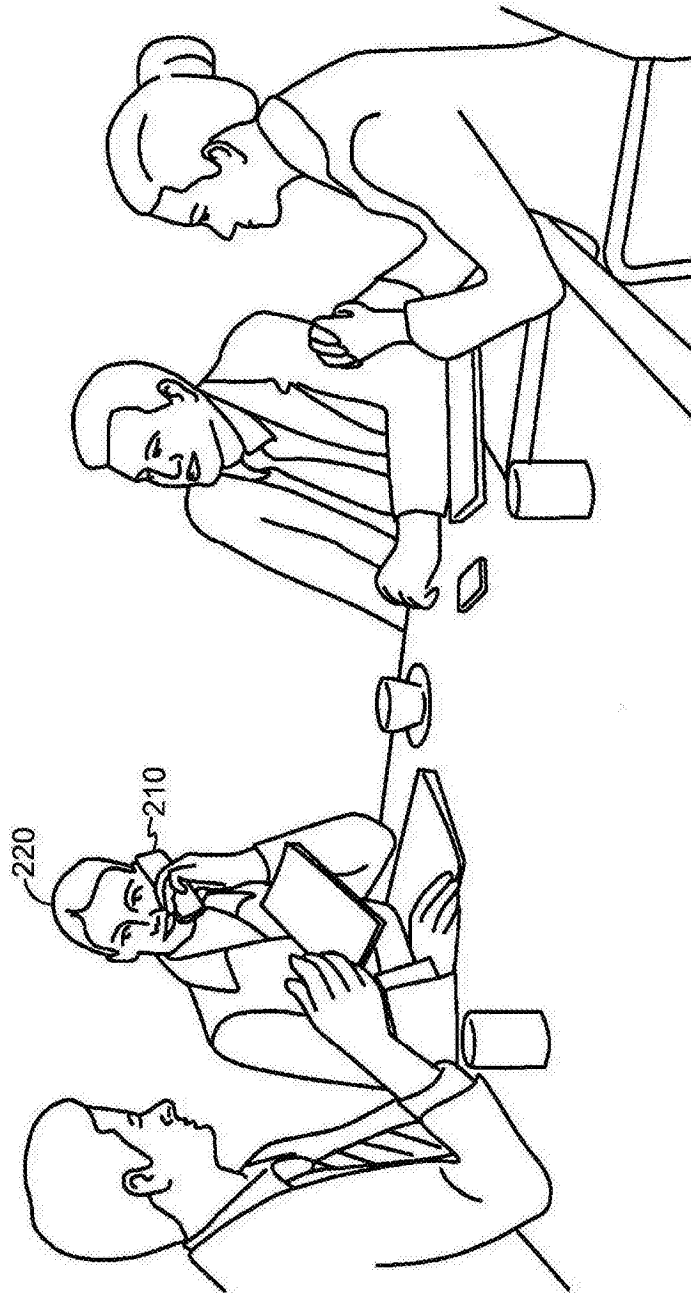


图 2

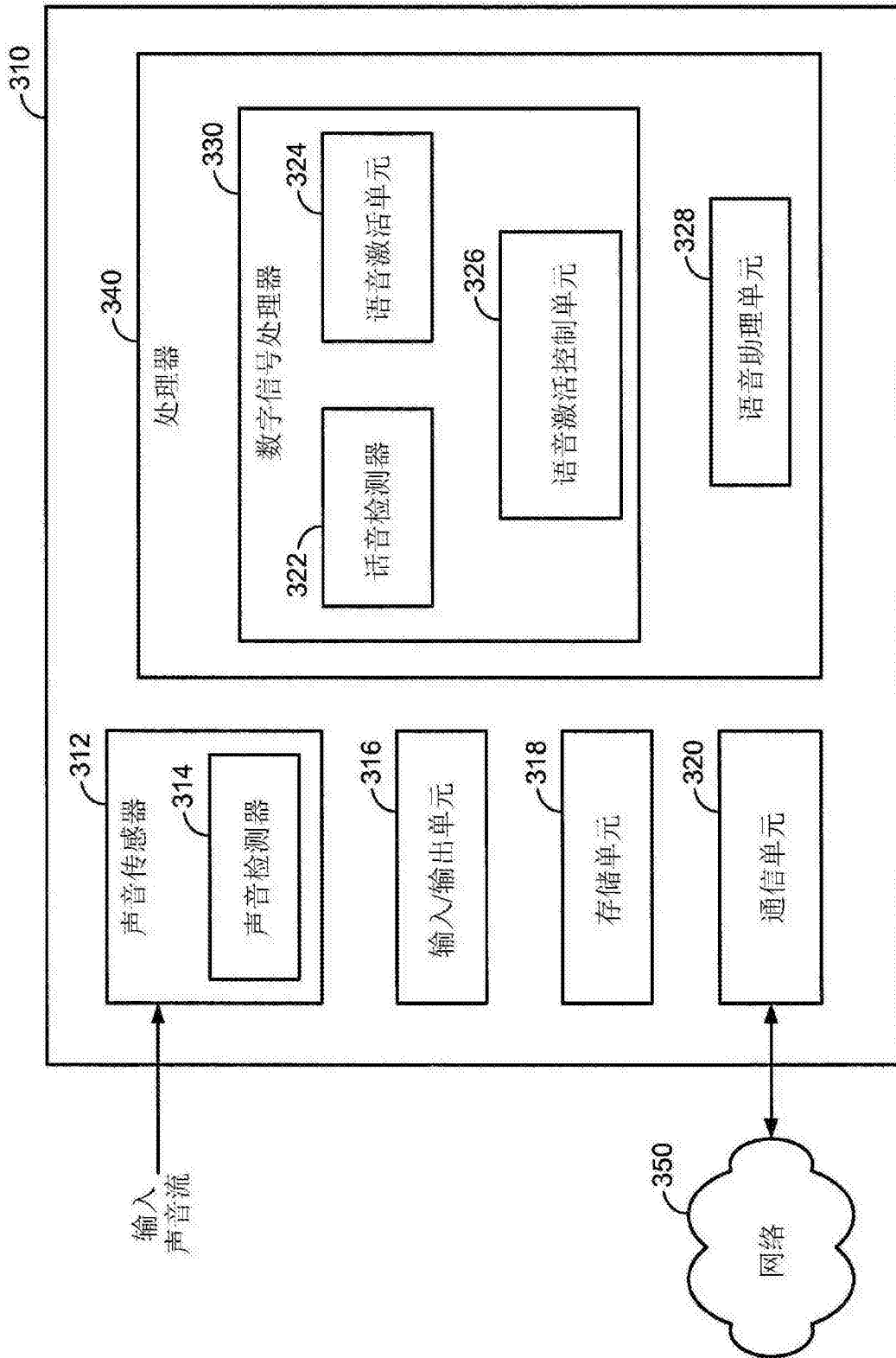


图 3

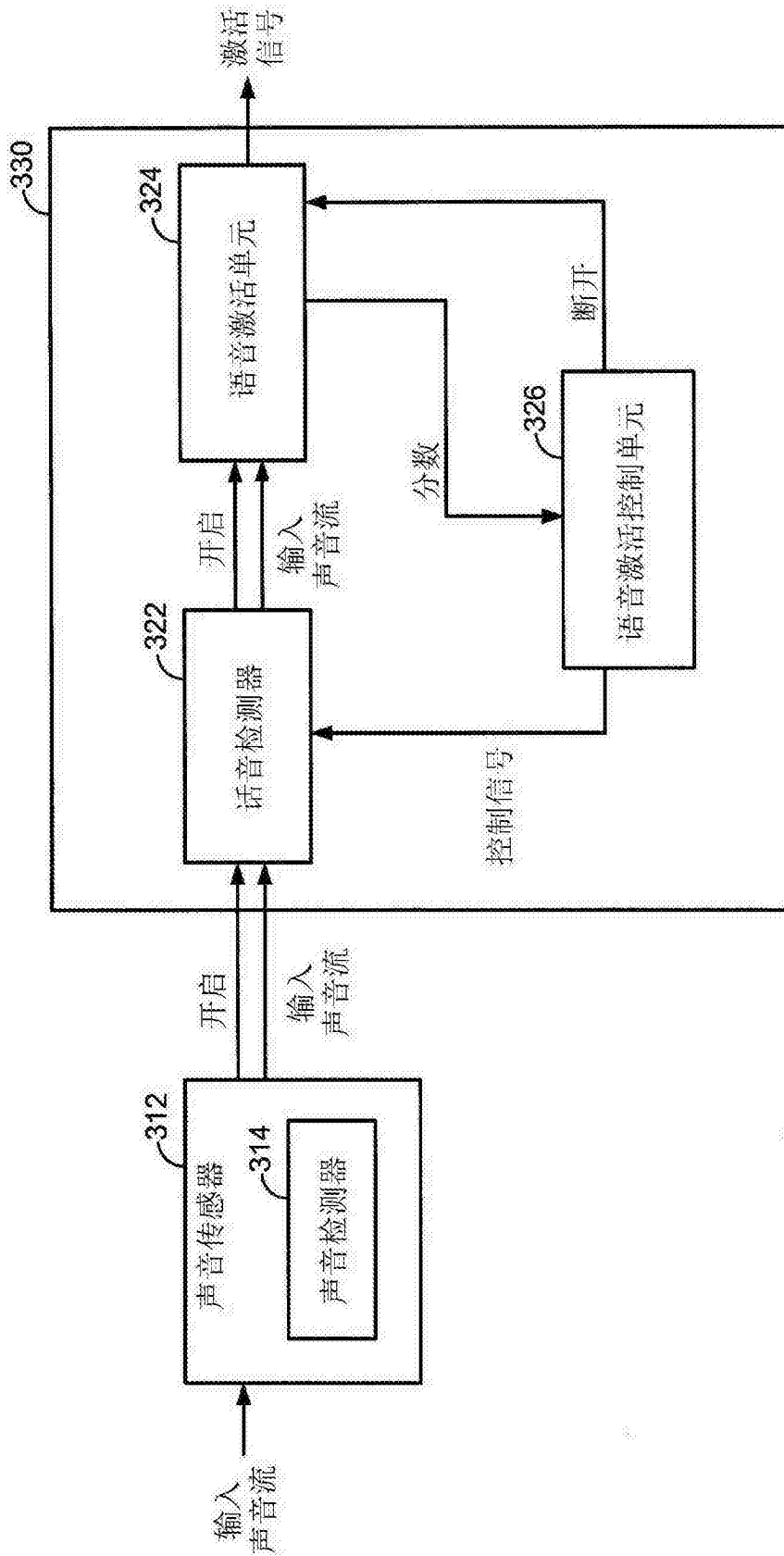


图 4A

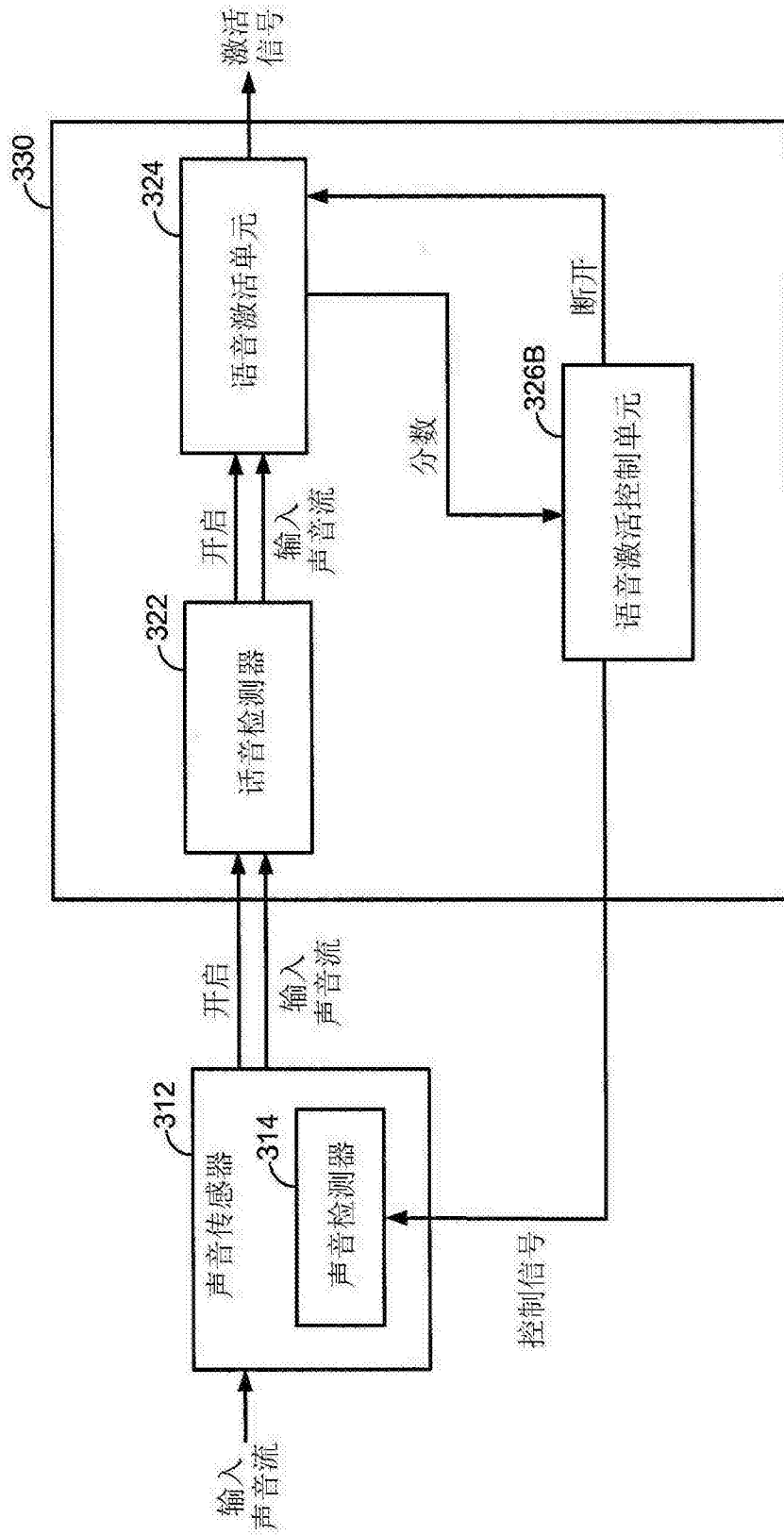


图 4B

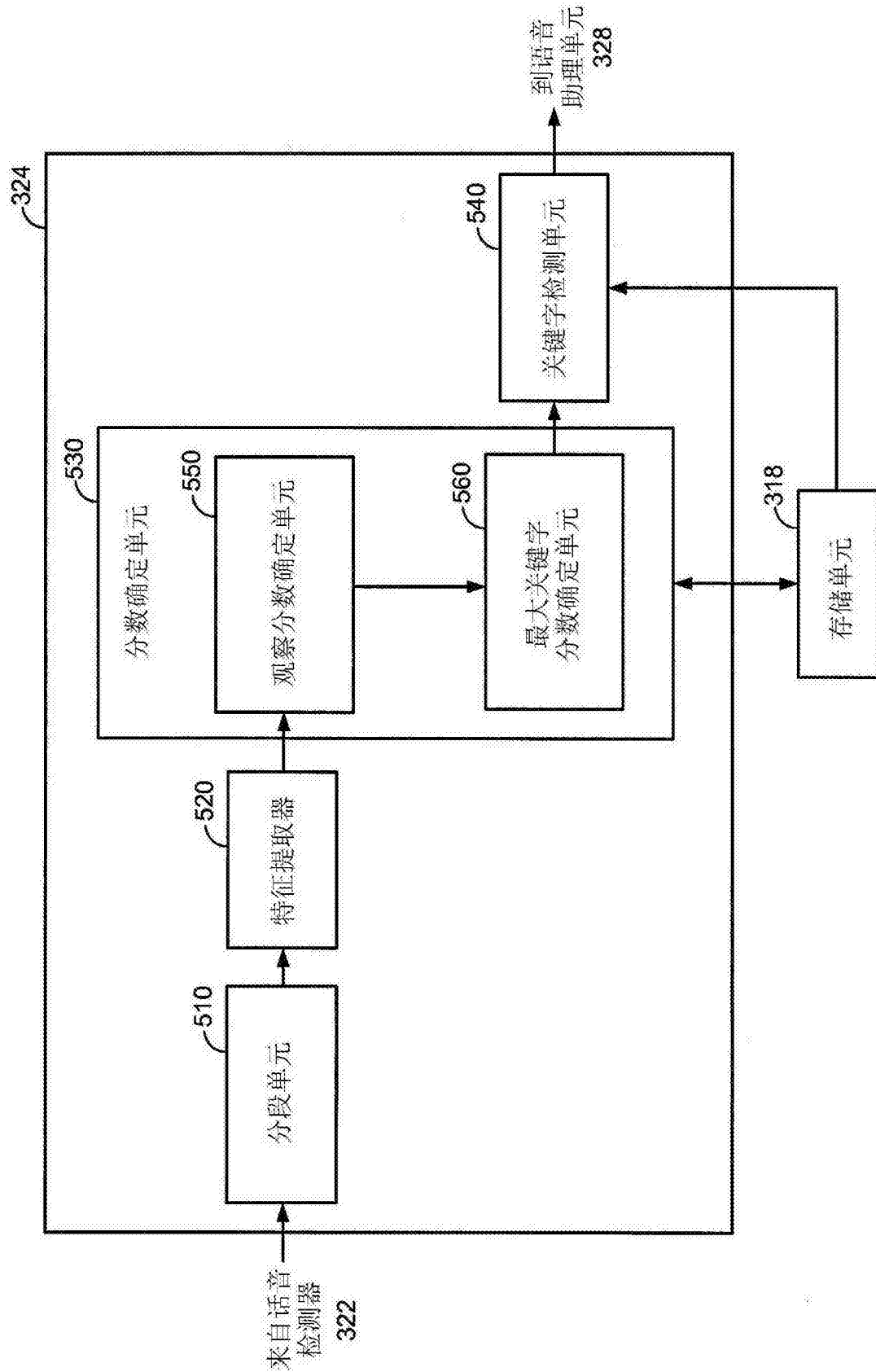


图 5

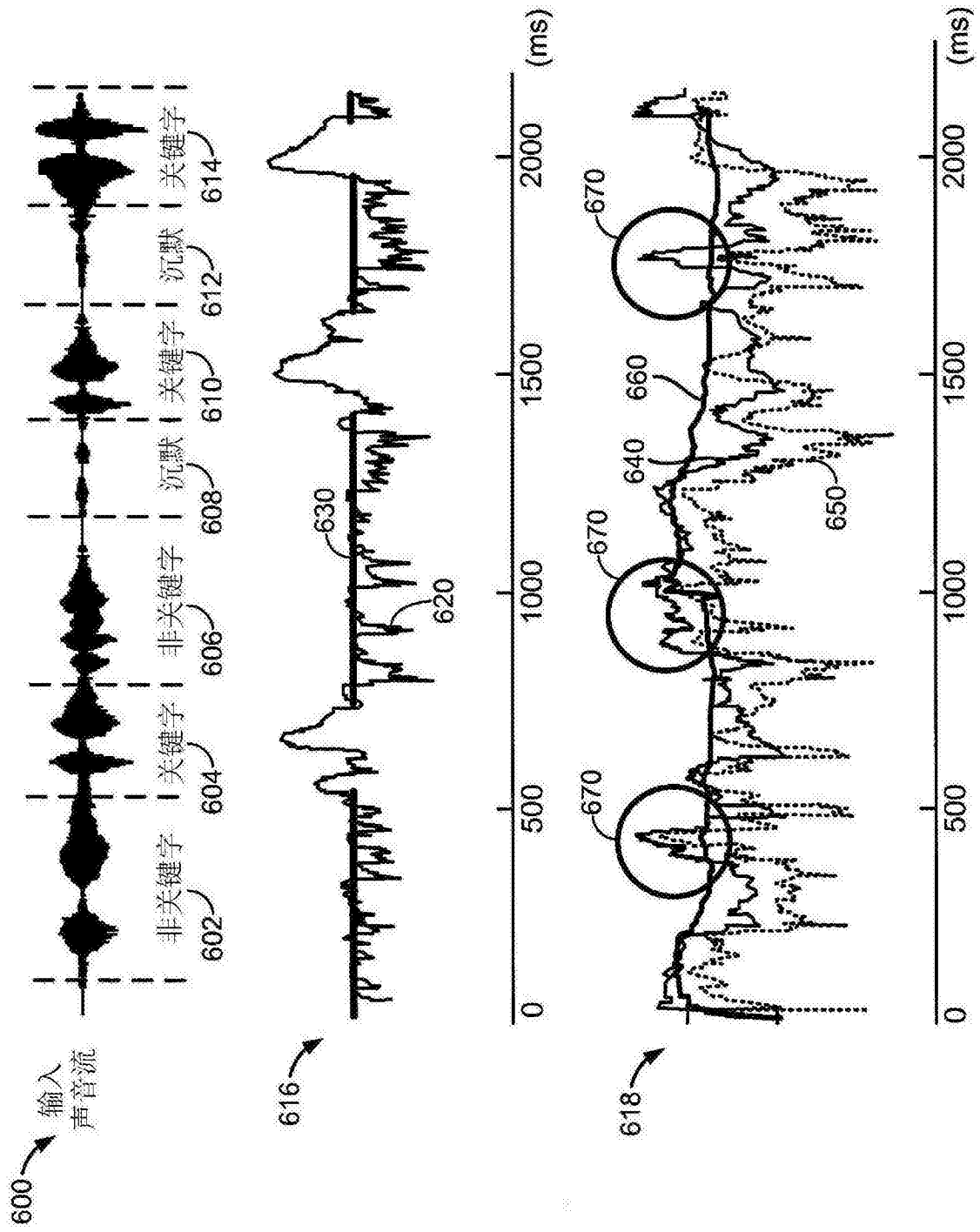


图 6

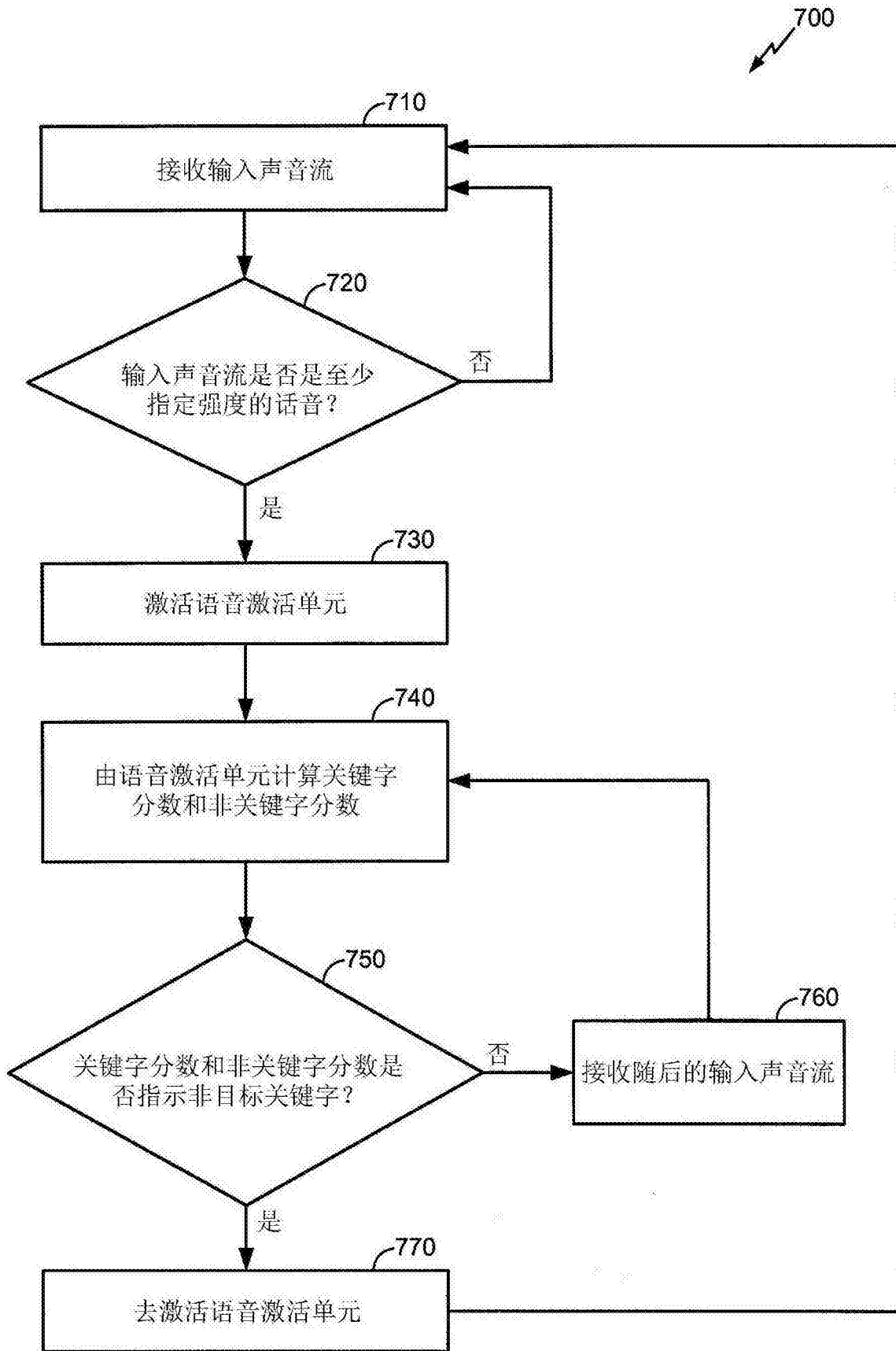


图 7

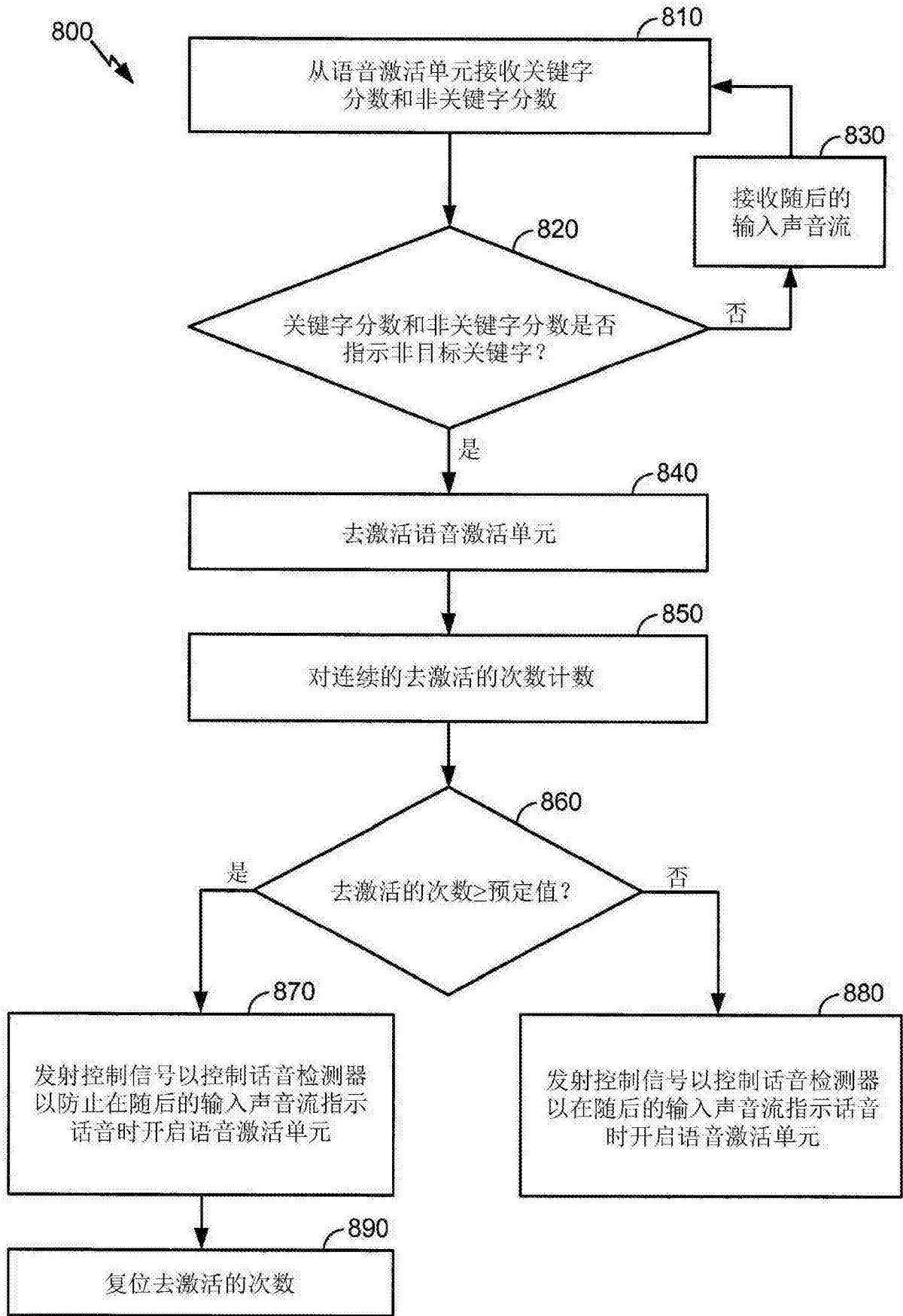


图 8

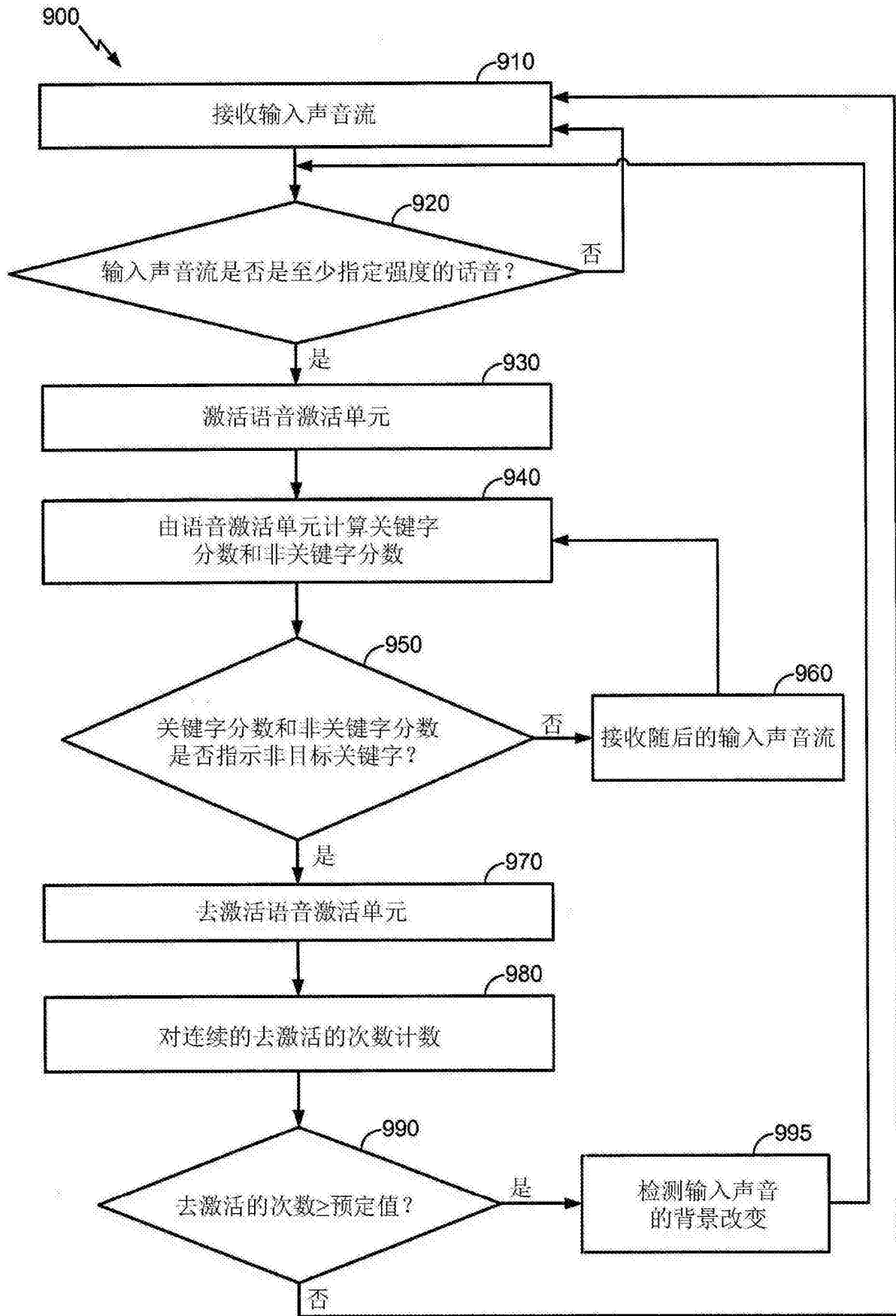


图 9

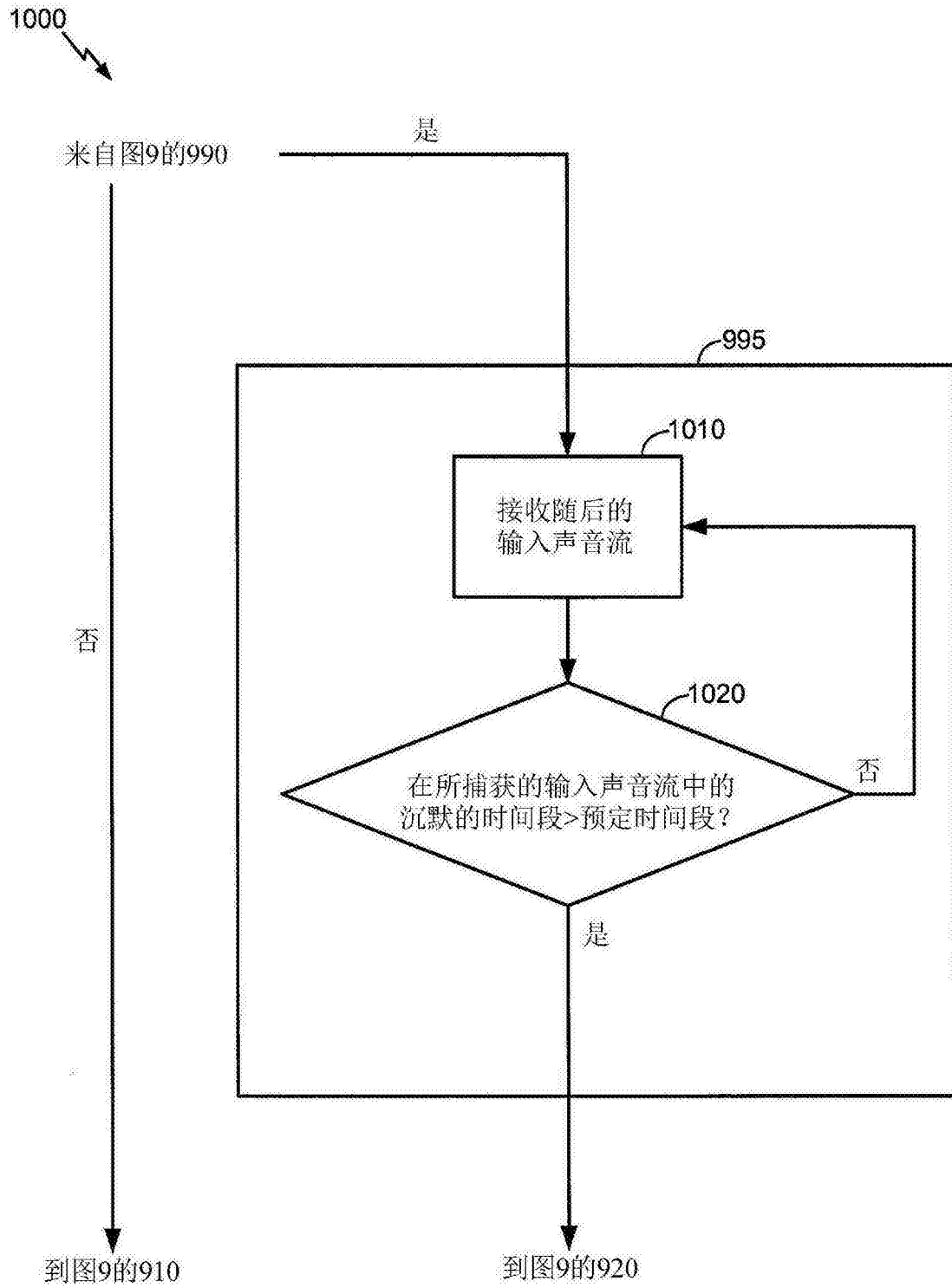


图 10

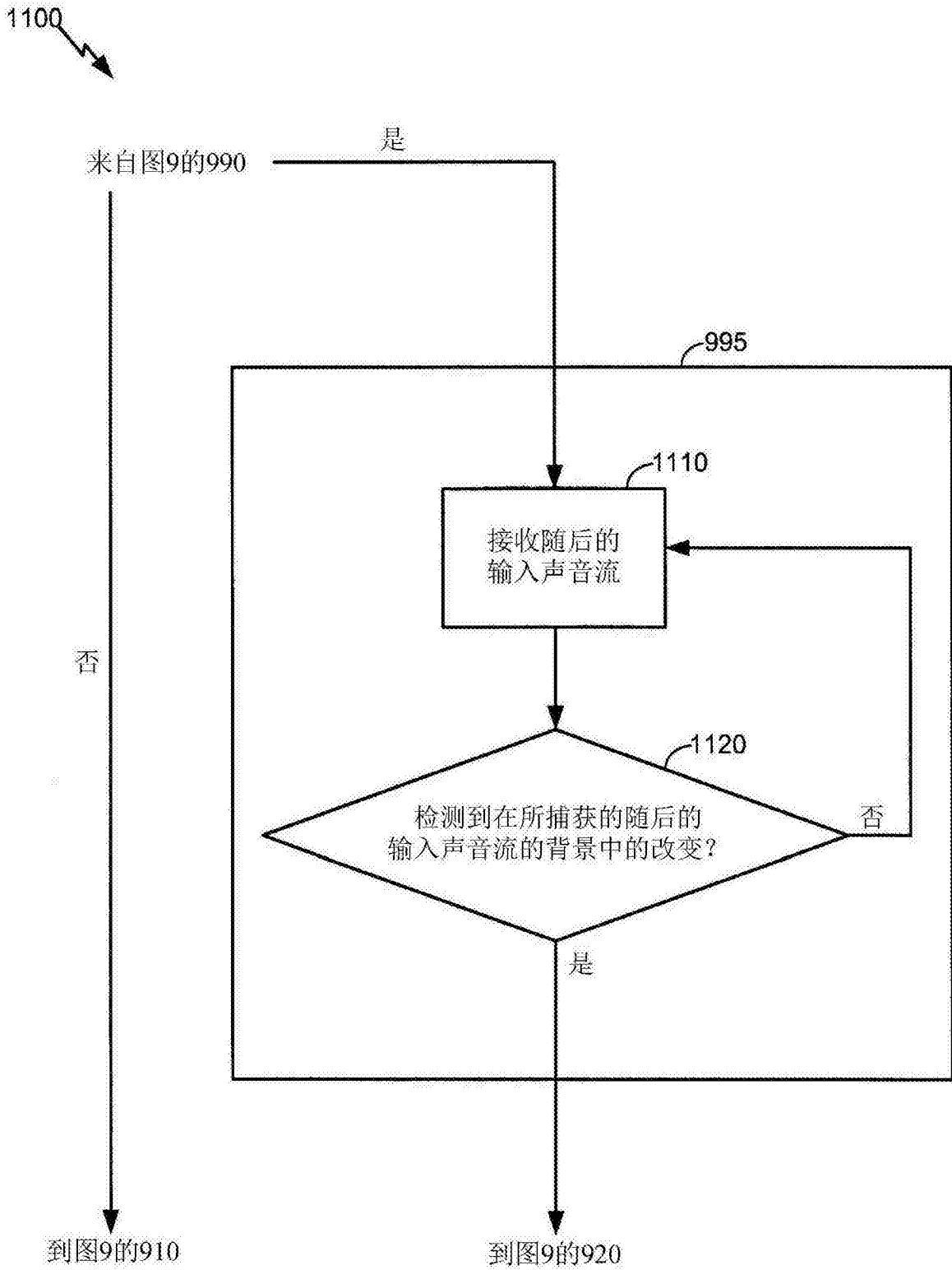


图 11

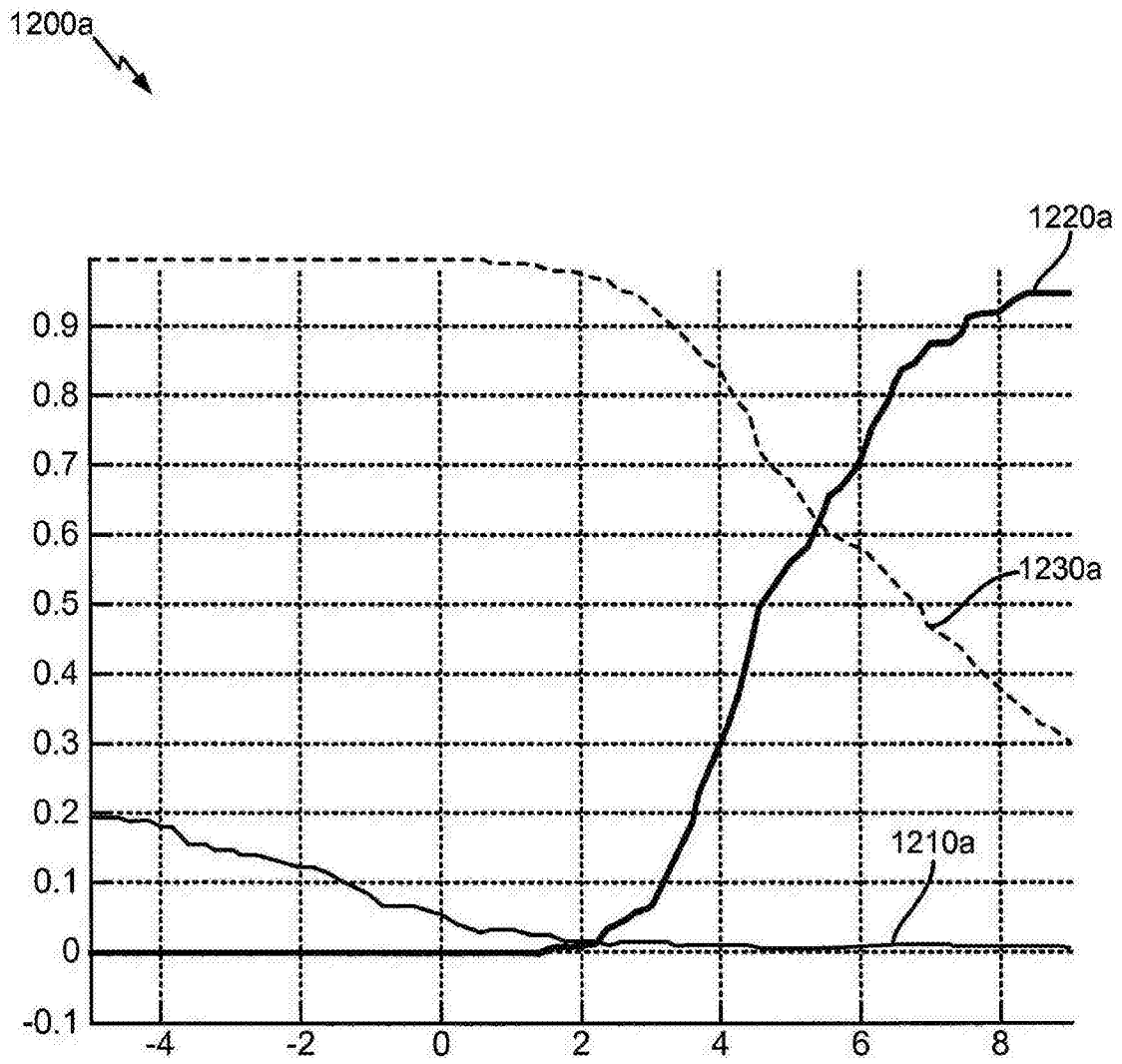


图 12A

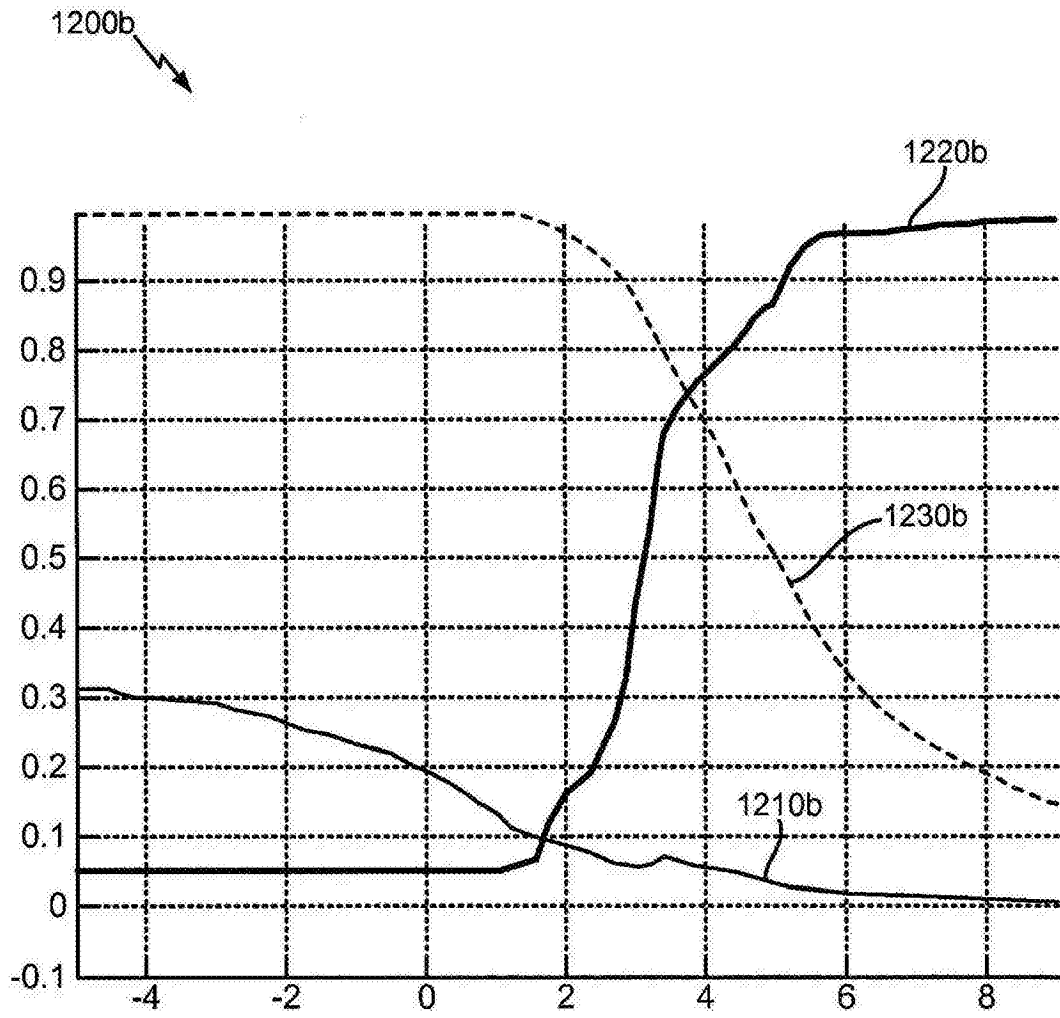


图 12B

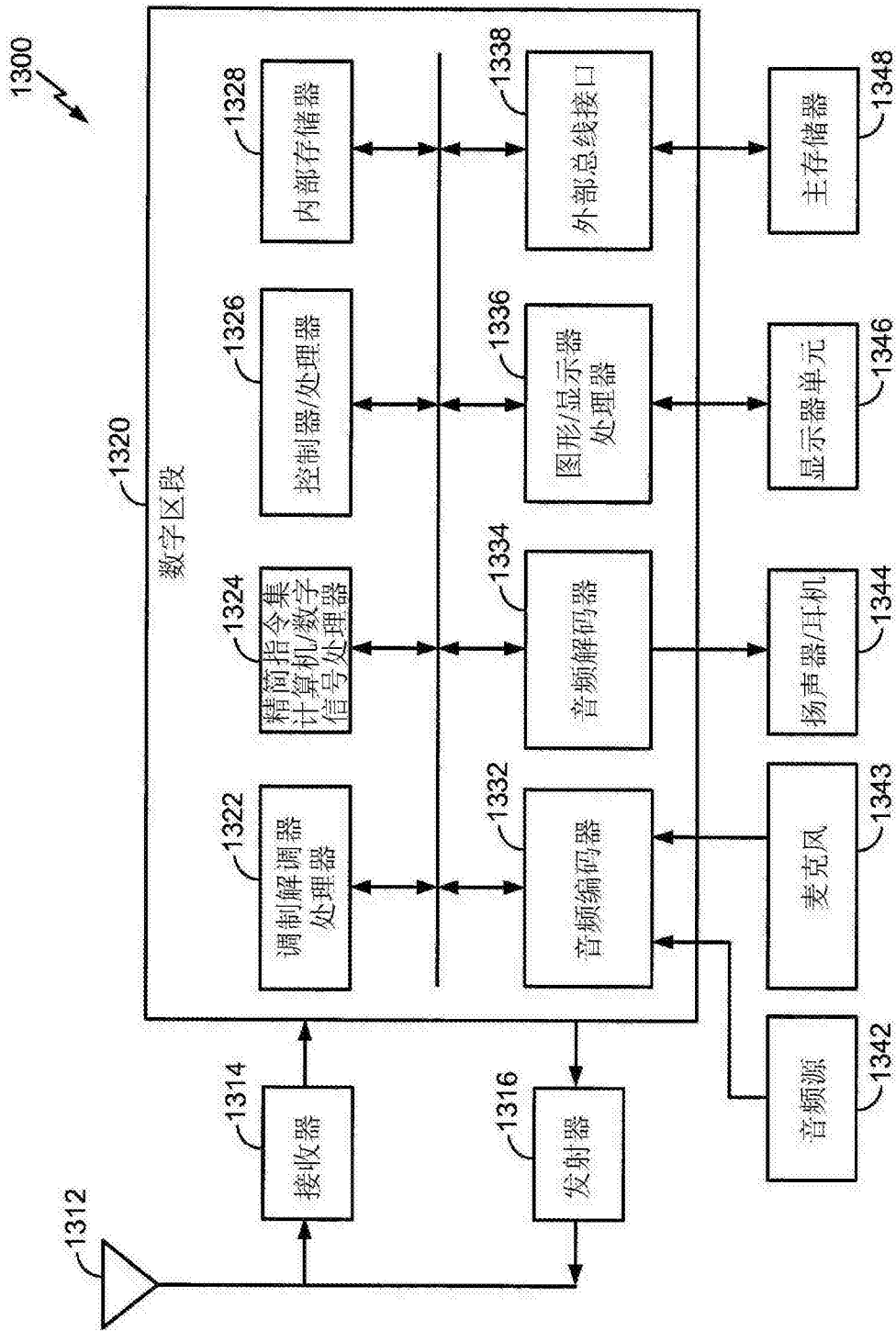


图 13