



(12) 发明专利

(10) 授权公告号 CN 101971250 B

(45) 授权公告日 2012. 05. 09

(21) 申请号 200880127979. 1

(22) 申请日 2008. 09. 15

(30) 优先权数据

12/047, 344 2008. 03. 13 US

(85) PCT申请进入国家阶段日

2010. 09. 10

(86) PCT申请的申请数据

PCT/US2008/076341 2008. 09. 15

(87) PCT申请的公布数据

W02009/114035 EN 2009. 09. 17

(73) 专利权人 索尼爱立信移动通讯有限公司

地址 瑞典隆德

(72) 发明人 贾森·考恩

(74) 专利代理机构 北京三友知识产权代理有限公司

公司 11127

代理人 李辉 张旭东

(51) Int. Cl.

G10L 15/26 (2006. 01)

(56) 对比文件

US 6088671 A, 2000. 07. 11, 全文.

CN 1337817 A, 2002. 02. 27,

WO 03085641 A1, 2003. 10. 16,

US 5799279 A, 1998. 08. 25,

审查员 刘红梅

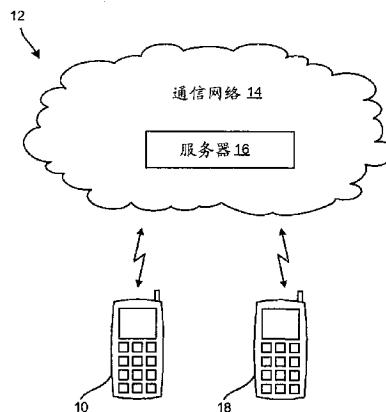
权利要求书 2 页 说明书 10 页 附图 2 页

(54) 发明名称

具有活动语音识别的移动电子设备

(57) 摘要

电子设备 (10、16) 使用语音识别针对能够引起动作的语音而分析语音通信。当检测到能够引起动作的语音时, 电子设备可以执行相应的功能, 包括: 在日志中存储信息, 或者向用户呈现一个或更多个程序、服务和 / 或控制功能。能够引起动作的语音可以是使用专家系统检测为针对程序的潜在命令或数据输入的预定命令和 / 或语音模式。



1. 一种在用户和另一方之间的电话会话期间主动识别语音的电子设备 (10、16), 该电子设备 (10、16) 包括控制电路 (30), 该控制电路 (30):

将所述电话会话转换为文本;

分析该文本以生成引起程序动作的程序命令或数据输入, 所述分析包括:

分析该文本以标识构成能够引起动作的文本的所述会话的一部分, 所述能够引起动作的文本是词、短语、句子结构、顺序或主题中的至少一个; 并且

将所述能够引起动作的文本与所述程序的能够引起动作的功能进行匹配并从所匹配的能够引起动作的文本和能够引起动作的功能中生成所述程序命令或数据输入; 并且

将所述程序命令或数据输入传送到所述程序以按照所述程序命令或数据输入进行动作; 并且

其中, 所述电子设备是移动电话, 所述电话会话是电话呼叫的一部分。

2. 根据权利要求 1 所述的电子设备, 其中, 所述控制电路还基于所述能够引起动作的语音而运行所述程序。

3. 根据权利要求 1 到 2 中任何一项所述的电子设备, 其中, 通过专家系统执行所述分析, 该专家系统在周围句子结构的语境中分析词和短语以检测所述能够引起动作的语音。

4. 根据权利要求 1 到 3 中任何一项所述的电子设备, 其中, 所述程序是因特网浏览器。

5. 根据权利要求 4 所述的电子设备, 其中, 使用所述能够引起动作的文本来将所述因特网浏览器导向特定因特网网页, 以访问相应的服务。

6. 根据权利要求 5 所述的电子设备, 其中, 所述服务是从绘制地图和方向服务、目录服务、天气预报服务、餐馆指南或电影列表服务中的一种选择的。

7. 根据权利要求 1 到 3 中任何一项所述的电子设备, 其中, 所述程序是生成电子邮件消息、即时消息、文本消息或多媒体消息之一的消息收发程序。

8. 根据权利要求 1 到 3 中任何一项所述的电子设备, 其中, 所述程序是联系人列表。

9. 根据权利要求 1 到 3 中任何一项所述的电子设备, 其中, 所述程序是用于存储约会条目的日程序。

10. 一种在移动电话 (10) 的用户和另一方之间的电话会话期间主动识别语音并响应于语音而动作的方法, 该方法包括以下步骤:

将所述电话会话转换为文本;

分析所述文本以生成引起所述移动电话的程序动作的程序命令或数据输入; 所述分析步骤包括:

分析所述文本以标识构成能够引起动作的文本的所述会话的一部分, 所述能够引起动作的文本是词、短语、句子结构、顺序或主题中的至少一个; 并且

将所述能够引起动作的文本与所述程序的能够引起动作的功能进行匹配并从所匹配的能够引起动作的文本和能够引起动作的功能中生成所述程序命令或数据输入; 以及

运行所述程序以按照所述程序命令或数据输入进行动作。

11. 根据权利要求 10 所述的方法, 其中, 所述程序是因特网浏览器。

12. 根据权利要求 11 所述的方法, 其中, 使用所述能够引起动作的文本来将所述因特网浏览器导向特定因特网网页, 以访问相应的服务。

13. 根据权利要求 12 所述的方法, 其中, 所述服务是从地图绘制和方向服务、目录服

务、天气预报服务、餐馆指南或电影列表服务中的一种选择的。

14. 根据权利要求 10 所述的方法,其中,所述程序是生成电子邮件消息、即时消息、文本消息或多媒体消息之一的消息收发程序。

15. 根据权利要求 10 所述的方法,其中,所述程序是联系人列表。

具有活动语音识别的移动电子设备

技术领域

[0001] 本发明的技术总体上涉及电子设备,更具体地涉及用于监控针对能够引起动作的语音的音频通信并且在检测到能够引起动作的语音时执行指定的功能和/或向电子设备的用户提供选项的系统和方法。

背景技术

[0002] 移动无线电子设备变得越来越普及。例如,现在广泛使用移动电话、便携式媒体播放器和便携式游戏设备。另外,与特定类型的电子设备关联的功能已变得越来越多样化。举出几个示例,许多电子设备具有摄像机、文本消息收发性能、因特网浏览性能、电子邮件性能、视频回放性能、音频回放性能、图像显示性能以及免提耳机接口。

[0003] 尽管便携式电子设备可以向用户提供使用多个功能的能力,但是,当前便携式电子设备未提供在电话会话期间与多个功能交互的方便方式。例如,在呼叫期间用于访问非呼叫功能的用户接口常常是困难的,并且使用起来是耗时的。

发明内容

[0004] 为了在用户使用电子设备来执行电话呼叫(或其他音频通信)的同时提高用户与电子设备的功能交互的能力,本发明描述一种改进的电子设备,该电子设备分析针对用户的能够引起动作的语音的电话呼叫和/或会话中涉及的其他方的电话呼叫。当检测到能够引起动作的语音时,电子设备可以执行相应功能,这包括了在呼叫日志中存储信息、向用户提供一个或更多功能(如应用、服务和/或控制功能)、或一些其他动作。例如,能够引起动作的语音可以是预定的命令(如采用词或短语的形式)和/或使用专家系统而检测到的语音模式(如句子结构)。电子设备的操作和相应方法可以导致在电话呼叫或其他基于语音的通信(如,“一按即通”(push-to-talk)会话)期间和/或之后的改善的体验。例如,该系统和方法可以允许按照直观且简单的方式访问信息和服务。在会话期间可以容易获得的信息的示例性类型可以包括到目的地的方向、联系的电话号码、当前时间等。在本文件的随后部分中将更详细地描述多个其他示例性呼叫中的用户接口功能。

[0005] 根据本发明的一个方面,第一电子设备主动识别话音通信期间的语音。第一电子设备包括控制电路,该控制电路将话音通信转换为文本并分析该文本以检测能够引起程序动作的语音,能够引起动作的语音对应于能够引起程序动作的命令或数据输入。

[0006] 根据第一电子设备的一个实施方式,控制电路还基于能够引起动作的语音而运行程序。

[0007] 根据第一电子设备的一个实施方式,其中通过专家系统执行所述分析,该专家系统在周围句子结构的语境中分析词和短语以检测能够引起动作的语音。

[0008] 根据第一电子设备的一个实施方式,电子设备是服务器,并且服务器向客户机设备发送命令或数据输入,该客户机设备响应于命令或数据输入而运行程序。

[0009] 根据第一电子设备的一个实施方式,程序是因特网浏览器。

[0010] 根据第一电子设备的一个实施方式,使用能够引起动作的语音将因特网浏览器导向特定因特网网页,以访问相应服务。

[0011] 根据第一电子设备的一个实施方式,服务是从绘制地图和方向服务、目录服务、天气预报服务、餐馆指南或电影列表服务中的一种选择的。

[0012] 根据第一电子设备的一个实施方式,程序是生成电子邮件消息、即时消息、文本消息或多媒体消息之一的消息收发程序。

[0013] 根据第一电子设备的一个实施方式,程序是联系人列表。

[0014] 根据第一电子设备的一个实施方式,程序是用于存储约会条目的日历程序。

[0015] 根据第一电子设备的一个实施方式,程序控制电子设备的设置。

[0016] 根据第一电子设备的一个实施方式,电子设备是移动电话并且话音通信是电话呼叫。

[0017] 根据本发明的另一方面,第二电子设备主动识别话音通信期间的语音。第二电子设备包括:控制电路,该控制电路将话音通信转换为文本并分析该文本以检测能够引起动作的语音,能够引起动作的语音对应于在话音通信的结束之后对用户有价值的信息;和存储器,该存储器在会话日志中存储能够引起动作的语音。

[0018] 根据第二电子设备的一个实施方式,会话日志采用了包含对应于能够引起动作语音的文本的文本格式。

[0019] 根据第二电子设备的一个实施方式,会话日志采用了包含来自话音通信的对应于能够引起动作语音的音频数据的音频格式。

[0020] 根据第二电子设备的一个实施方式,能够引起动作的语音对应于姓名、电话号码、电子邮件地址、消息收发地址、街道地址、地点、到目的地的方向、日期、时间或者它们的组合中的至少一种。

[0021] 根据本发明的另一方面,提供了使用电子设备在话音通信期间主动识别语音并响应于语音而动作的第一方法,该方法包括以下步骤:将话音通信转换为文本;分析该文本以检测能够引起电子设备的程序动作的语音,能够引起动作的语音对应于能够引起程序动作的命令或数据输入;以及基于能够引起动作的语音运行程序。

[0022] 根据第一方法的一个实施方式,由专家系统执行分析,该专家系统在周围句子结构的语境中分析词和短语以检测能够引起动作的语音。

[0023] 根据第一方法的一个实施方式,在用户选择了运行程序的选项之后运行程序。

[0024] 根据第一方法的一个实施方式,程序是因特网浏览器。

[0025] 根据第一方法的一个实施方式,使用能够引起动作的语音将因特网浏览器导向特定因特网网页,以访问相应服务。

[0026] 根据第一方法的一个实施方式,服务是从绘制地图和方向服务、目录服务、天气预报服务、餐馆指南或电影列表服务中的一种选择的。

[0027] 根据第一方法的一个实施方式,程序是生成电子邮件消息、即时消息、文本消息或多媒体消息之一的消息收发程序。

[0028] 根据第一方法的一个实施方式,程序是联系人列表。

[0029] 根据第一方法的一个实施方式,程序是用于存储约会条目的日历程序。

[0030] 根据第一方法的一个实施方式,程序控制电子设备的设置。

[0031] 根据本发明的另一方面,提供了使用电子设备在话音通信期间主动识别语音并响应于语音而动作的第二方法,该方法包括以下步骤:将话音通信转换为文本;分析该文本以检测可以引起动作的语音,能够引起动作的语音对应于在话音通信的结束之后对用户有价值的信息;和在会话日志中存储能够引起动作的语音。

[0032] 根据第二方法的一个实施方式,会话日志采用了包含对应于能够引起动作的语音的文本的文本格式。

[0033] 根据第二方法的一个实施方式,会话日志采用了包含来自话音通信的对应于能够引起动作的语音的音频数据的音频格式。

[0034] 根据第二方法的一个实施方式,能够引起动作的语音对应于姓名、电话号码、电子邮件地址、消息收发地址、街道地址、地点、到目的地的方向、日期、时间或者它们的组合中的至少一种。

[0035] 参照下面的描述和附图,将清楚本发明的这些和进一步的特征。在描述和附图中,具体公开了本发明的特定实施方式,以指示可采用本发明原理的一些方式,但是应当理解本发明的范围并不因此受到限制。相反,本发明包括落入所附的权利要求的精神和条款内的全部变化、修改和等同例。

[0036] 针对一个实施方式描述和/或例示的特征可以按照相同方式或类似方式在一个或更多个其他实施方式中使用,和/或与其他实施方式的特征相结合地或替代其他实施方式的特征而使用。

附图说明

[0037] 图 1 是示例性电子设备可以与另一电子设备通信的通信系统的示意图;

[0038] 图 2 是图 1 的示例性电子设备的示意框图;和

[0039] 图 3 是表示使用图 1 的电子设备的活动语音识别的示例性方法的流程图。

具体实施方式

[0040] 现在将参考附图来描述实施方式,其中在全部图中使用相似标号来表示相似部件。应理解的是,这些图可能不按比例绘制。

[0041] 在本文件中,实施方式主要是在移动电话的背景中进行描述。但应理解,移动电话的示例性背景不是可以使用所公开的系统和方法的方面的唯一操作环境。因此,在该文件中描述的技术可以应用于任何类型的合适电子设备,这些电子设备的示例包括移动电话、媒体播放器、游戏装置、计算机、寻呼机、通信机、电子记事簿、个人数字助理(PDA)、智能电话、便携式通信设备等。

[0042] 首先参考图 1 和图 2,电子设备 10 可以配置为作为通信系统 12 的一部分来操作。系统 12 可以包括具有服务器 16(或多个服务器)的通信网络 14,服务器 16 用于管理电子设备 10 发起的呼叫以及去往电子设备 10 的呼叫、向电子设备 10 发送数据并执行任何其他支持功能。电子设备 10 可以经由传输介质(未示出)与通信网络 14 交换信号。传输介质可以是任何适当的装置或组件,这包括例如通信塔(如,蜂窝通信塔)、无线接入点、卫星等。网络 14 可以支持多个电子设备和其他类型的终端用户装置的通信活动。如将理解的,服务器 16 可以配置为用于执行服务器功能的通常的计算机系统并可以包括处理器和存储

器,该处理器配置为执行包含逻辑指令的软件,该软件实现服务器 16 的功能,该存储器存储这样的软件。

[0043] 电子设备 10 可以对另一电子设备 (这被称为第二电子设备或远程电子设备 18) 发起呼叫或接收来自它的呼叫。在例示的实施方式中,远程电子设备 18 是另一移动电话,但可以是另一类型的装置,该装置能够允许远程电子设备 18 的用户与电子设备 10 的用户进行话音通信。此外,在电子设备 10 与远程电子设备 18 之间的通信可以是除了电话呼叫以外的话音通信的形式,诸如“一按即通”会话或来源于装置 10、18 中的任一个的话音消息。

[0044] 远程电子设备 18 示出为由通信网络 14 服务,将理解的是,远程电子设备 18 可以由诸如蜂窝服务提供商、卫星服务提供商、因特网语音协议 (VoIP) 服务提供商、常规有线电话系统 (如,简单的老式电话系统或 POTS) 等的不同通信网络来服务。如指出的,电子设备 10 还可以通过一个或多个这些类型的网络来工作。

[0045] 在描述用于监控话音通信的技术之前,将描述当被实现为移动电话时的电子设备 10 的示例性构造。在例示的实施方式中,电子设备 10 描述为提供和执行呼叫辅助功能 20,呼叫辅助功能 20 至少实现一些公开的监控和用户接口特征。在其他实施方式中,呼叫辅助功能 20 可以由服务器 16 提供。在该实施方式中,服务器 16 可以处理去往电子设备 10 或从电子设备 10 接收的话音数据,并且向电子设备 10 发送相应的控制和数据消息以调用描述的用户接口特征。

[0046] 在例示的实施方式中,电子设备 10 包括呼叫辅助功能 20。呼叫辅助功能 10 配置为针对能够引起动作的语音监控在电子设备 10 的用户与远程电子设备 18 的用户之间的话音通信。基于检测的能够引起动作的语音,呼叫辅助功能 20 向用户提供接口功能。能够引起动作的语音可以是这样的语音,该语音可以用作对程序的控制输入或数据输入。另外,能够引起动作的语音可以是对用户具有信息价值的语音。下面将更详细地描述呼叫辅助功能 20 的另外的详情和操作。

[0047] 呼叫辅助功能 20 可以实现为驻留在电子设备 10 中并由电子设备 10 执行的可执行代码。在一个实施方式中,呼叫辅助功能 20 可以是存储在计算机或机器可读介质上的程序。呼叫辅助功能 20 可以是独立软件应用或者形成执行与电子设备 10 有关的附加任务的软件应用的一部分。

[0048] 如下面将变得明显的,呼叫辅助功能 20 可以与其他软件程序 22 交互,该软件程序 22 由电子设备 10 存储和执行。为了图的简单,该其他程序 22 不单独标识。将理解的是,这里提到的程序 22 是代表性的,并且不是呼叫辅助功能 20 与之可以交互的程序 22 的穷举性列表。一个示例性程序 22 可以是设置控制功能。例如,呼叫辅助功能 20 的输出可以被输入到电子设备 10 的设置控制功能,以控制扬声器音量、显示亮度或其他可设置参数。作为另一示例,来自呼叫辅助功能 20 的输出可以被输入到因特网浏览器以使用由因特网服务器提供的服务来调用搜索。示例性服务可以包括但不限于一般的因特网搜索引擎、电话目录、天气预报服务、餐馆指南、地图绘制和向导服务、电影列表服务等。作为另一示例,呼叫辅助功能 20 可以与联系人列表数据库交互以搜索先前存储的信息或存储在话音通信期间获得的新信息。另一示例性程序 22 包括日历功能、时钟功能、消息收发功能 (如,电子邮件功能、即时消息收发功能、文本消息功能、多媒体消息功能等)、或任何其他适当的功能。

[0049] 电子设备 10 可以包括显示器 24。显示器 24 向用户显示信息,诸如操作状态、时

间、电话号码、联系信息、各种菜单、针对各种程序的图形用户界面 (GUI) 等。显示的信息使用户能够利用电子设备 10 的各种特征。显示器 24 还可以用于可视地显示由电子设备 10 接收的内容和 / 或从电子设备 10 的存储器 26 获得的内容。显示器 24 可以用于向用户呈现图像、视频和其他图形, 诸如照片、移动电视内容和与游戏关联的视频。

[0050] 键区 28 提供各种用户输入操作。例如, 键区 28 可以包括字母数字键, 以允许输入诸如电话号码、电话列表、联系人信息、备忘录、文本等的字母数字信息。另外, 键区 28 可以包括特殊功能键, 诸如用于发起或应答呼叫的“呼叫发送”键、和用于结束或“挂断”呼叫的“呼叫结束”键。特殊功能键还可以包括菜单导航和选择键, 以有助于在显示器 24 上显示的菜单中进行导航。例如, 可以具有点击装置和 / 或导航键以从用户接受方向输入。特殊功能键可以包括音频视频内容回放键以开始、停止和暂停回放、跳过或重复音轨 (track) 等。与移动电话关联的其它键可以包括音量键、音频静音键、开 / 关电源键、网络浏览器启动键、摄像机键等。键或键类似功能也可以实现为与显示器 24 关联的触摸屏。另外, 显示器 24 以及键区 28 可以彼此结合地使用以实现软键功能。

[0051] 电子设备 10 可以包括使电子设备 10 能够与通常可以是另一移动电话或固定电话的被叫 / 主叫设备 (如, 远程电子设备 18) 建立呼叫和 / 或交换信号的呼叫电路。但是, 被叫 / 主叫设备不必为另一个电话, 而可以是诸如因特网网络服务器、内容提供服务器等的一些其它装置。呼叫可以采取任何适当的形式。例如, 呼叫可以是在蜂窝电路交换网络上建立的常规呼叫或者是在蜂窝网络的分组交换能力或在诸如 WiFi (例如, 基于 IEEE 802. 11 标准的网络)、WiMax (例如, 基于 IEEE 802. 16 标准的网络) 等另选的分组交换网络上建立的因特网语音协议 (VoIP) 呼叫。另一个示例包括在蜂窝网络或另选网络上建立的视频使能呼叫。

[0052] 电子设备 10 可以被配置为产生、发送、接收和 / 或处理诸如文本消息、即时消息、电子邮件消息、多媒体消息、图像文件、视频文件、音频文件、铃声、流音频、流视频、数据馈送 (包括播客和内容聚合 (RSS :reallysimple syndication) 数据馈送)、因特网内容等的的数据。注意, 文本消息通常被一些人称为“SMS”, 这表示简单的消息服务。SMS 是用于交换文本消息的典型标准。类似地, 多媒体消息常被一些人称为“MMS”, 这表示多媒体消息服务。MMS 是用于交换多媒体消息的典型标准。对数据进行处理可以包括在存储器 26 中存储数据, 执行应用以允许用户与数据交互, 显示与数据关联的视频和 / 或图像内容, 输出与数据关联的音频声音等。

[0053] 继续参考图 2, 电子设备 10 可以包括被配置为对电子设备 10 的功能和操作执行整体控制的主控制电路 30。控制电路 30 可以包括诸如中央处理单元 (CPU)、微控制器或微处理器的处理装置 32。处理装置 32 执行存储在控制电路 30 内的存储器 (未示出) 和 / 或在诸如存储器 26 的单独存储器中的代码, 以执行电子设备 10 的操作。例如, 存储器 26 可以是缓冲器、闪存、硬盘驱动器、可移除介质、易失性存储器、非易失性存储器、随机存取存储器 (RAM) 或其它适当装置中的一种或更多种。在通常的设置中, 存储器 26 可以包括用于长期数据存储的非易失性存储器 (如, NAND 或 NOR 架构的闪存) 和充当控制电路 30 的系统存储器的易失性存储器。易失性存储器例如可以是利用同步动态随机存取存储器 (SDRAM) 实现的 RAM。存储器 26 可以与控制电路 30 通过数据总线交换数据。也可以具有存储器 26 与控制电路 30 之间的附带的控制线路和地址总线。

[0054] 处理装置 32 可以执行实现呼叫辅助功能 20 和程序 22 的代码。对于计算机编程、并特别是在针对移动电话或其他电子设备的应用编程的领域中的普通技术人员来说,如何对电子设备 10 编程以操作和执行与呼叫辅助功能 20 关联的逻辑功能是显而易见的。因此,为了简明的目的而省略了具体的编程代码的细节。另外,尽管根据实施方式由处理装置 23 执行呼叫辅助功能 20,但这样的功能还可以经由专用硬件或固件或者硬件、固件和 / 或软件的某种组合来执行。

[0055] 电子设备 10 可以包括耦合到无线电电路 36 的天线 34。无线电电路 36 包括用于经由天线 34 发送和接收信号的射频发送器和接收器。无线电电路 36 可以被配置为在通信系统 12 中工作并可以用于发送和接收数据和 / 或音频视频内容。用于与网络 14 交互的接收器类型包括但不限于全球移动通信系统 (GSM)、码分多址 (CDMA)、宽带 CDMA (WCDMA)、通用分组无线电服务 (GPRS)、WiFi、WiMAX 等以及这些标准的高级版本。将会理解的是,天线 34 和无线电电路 36 可以代表一个或一个以上的无线电收发器。

[0056] 电子设备 10 还包括用于处理由无线电电路 36 发送和从无线电电路 36 接收的音频信号的声音信号处理电路 38。与声音处理电路 38 耦合的是使用户能够经由电子设备 10 收听和说话的扬声器 40 和麦克风 42。无线电电路 36 和声音处理电路 38 均耦合到控制电路 30 以执行整体操作。音频数据可以从控制电路 30 传递到声音信号处理电路 38 以向用户进行回放。音频数据例如可以包括由存储器 26 存储并由控制电路 30 获取的音频文件的音频数据,或从移动无线电服务接收的诸如流音频数据形式的音频数据。声音处理电路 38 可以包括任何适当的缓冲器、解码器、放大器等。

[0057] 通过将视频数据转换为用于驱动显示器 24 的视频信号的视频处理电路 44,显示器 24 可以耦合到控制电路 30。视频处理电路 44 可以包括任何适当的缓冲器、解码器、视频数据处理器等。视频数据可以由控制电路 30 产生,从存储在存储器 26 中的视频文件中获取,从无线电电路 38 接收到的输入视频数据流得到,或者通过任何其它适当方法获得。

[0058] 电子设备 10 还可以包括一个或更多个输入 / 输出 (I/O) 接口 46。I/O 接口 46 可以是典型的移动电话 I/O 接口的形式,并且可以包括一个或更多个电连接器。正如典型的那样,I/O 接口 46 可以用于将电子设备 10 耦合到电池充电器以对电子设备 10 内的电源单元 (PSU) 48 的电池充电。另外或另选地,I/O 接口 46 可以用于将电子设备 10 连接到具有与电子设备 10 的有线接口的头戴式耳机组件 (例如,个人免提 (PHF) 装置)。此外,I/O 接口 46 可以用于经由数据线将电子设备 10 连接到个人计算机或其它装置以交换数据。当连接到车载电源适配器或电源插座适配器时,电子设备 10 可以经由 I/O 接口 46 接收工作电力。在缺乏外部电源时,PSU 48 可以供电以使电子设备 10 工作。

[0059] 电子设备 10 可以包括用于拍摄数字图片和 / 或电影的摄像机 50。对应于图片和 / 或电影的图像和 / 或视频文件可以存储在存储器 26 中。

[0060] 电子设备 10 还可以包括位置数据接收器 52,诸如全球定位系统 (GPS) 接收器、伽利略卫星系统接收器等。位置数据接收器 52 可以用于电子设备 10 的位置的确定。

[0061] 电子设备 10 还可以包括诸如红外收发器和 / 或 RF 接口 (如,蓝牙接口) 的局域无线接口 54,用于建立与附件、另一移动无线电终端、计算机或另一装置的通信。例如,在头戴式耳机组件具有相应的无线接口的实施方式中,局域无线接口 54 可以可操作地将电子设备 10 耦合到头戴式耳机组件 (例如,PHY 装置)。

[0062] 另外参考图 3, 例示的是用于实现在涉及电子设备 10 的话音通信期间有效地识别和作用于语音的示例性方法的逻辑操作。例如通过执行呼叫辅助功能 20 的实施方式而执行该示例性方法。因而, 图 3 的流程图可以被视为由电子设备 10 执行的方法的描述步骤。在其他实施方式中, 一些步骤可以由服务器 16 执行。

[0063] 尽管图 3 示出执行功能逻辑块的特定顺序, 但相对于示出的顺序可以改变执行这些块的顺序。另外, 两个或更多个连续示出的块可以同时地执行或部分同时地执行。还可以省略特定的块。

[0064] 在一个实施方式中, 如果用户使用头戴式耳机装置 (例如, PHY) 或扬声器电话功能来参与话音通信, 结合图 3 描述的功能可以工作地最佳。按照该方式, 电子设备 10 不需要保持靠着用户的头, 使得用户在通信期间可以观看显示器 24 并且 / 或者可以操作键区 28。

[0065] 将理解的是, 这些操作可以应用于输入的音频数据 (如, 来自远程电子设备 18 的用户的语音)、输出的音频数据 (如, 来自电子设备 10 的用户的语音)、或输入和输出的音频数据二者。

[0066] 逻辑流可以开始于块 56, 其中, 可以对电子设备 10 是否当前正在用于诸如电话会话、“一按即通”通信或话音消息回放的音频 (如, 话音) 通信进行确定。如果电子设备 10 当前没有进行音频通信, 逻辑流可以等待, 直到音频通信开始。如果在块 56 中做出肯定性确定, 逻辑流则进行到块 58。

[0067] 在例示的实施方式中, 音频通信示出为在电子设备 10 的用户与远程装置 18 的用户之间的、在这两个装置之间建立的电话呼叫期间的会话。在块 58, 该会话可以针对能够引起动作的语音的存在而被监控。例如, 可以使用语音识别来将包含了各装置 10 和 18 的用户的话音模式的音频信号转换为文本。针对可以充当电子设备 10 调用特定动作的命令或提示的预定的词或短语而分析该文本, 如在下面将更详细描述。另外, 专家系统可以分析该文本以标识词、短语、句子结构、顺序和其他口头信息, 以标识据此可采取动作的谈话的一部分。在一个实施方式中, 专家系统可以实现为评价谈话的主题, 并将该信息与可以在谈话期间或之后辅助用户的电子设备 10 的程序和功能进行匹配。为此, 专家系统可以包含匹配在谈话的周围语音的语境中取得的特定词和 / 或短语的一组匹配规则, 以将这些词和短语与电子设备的能够引起动作的功能匹配。例如, 与吃、餐馆、方向、地点、天气、或其他话题的问题有关的句子结构可以提示专家系统标识能够引起动作的语音。另外, 与这些或其他话题有关的信息陈述可以提示专家系统来标识能够引起动作的语音。作为示例, 信息陈述可以用“我的地址是...”来开始。

[0068] 在块 58 之后, 逻辑流可以进行到块 60, 在块 60, 可以确定立即的能够引起动作的语音是否已被识别。立即的能够引起动作的语音可以是预定的命令、词或短语, 电子设备 10 使用这样的命令、词或短语来引起相应的响应。例如, 如果用户说出短语“启动网络浏览器”, 则在块 60 做出肯定性确定并且可以启动浏览器程序。作为另一示例, 用户可以说出短语“音量增大”, 以使电子设备 10 通过增加扬声器音量而进行响应, 使得用户可以更好地收听远程电子设备 18 的用户。按照该方式, 用户可以说出预定的词或短语以启动程序 22 中的一个、显示特定信息 (如, 一天中的时间、日期、联系人列表条目等)、开始记录谈话、结束记录谈话或采取可以与口头命令关联的任何其他动作, 这期间电子设备 10 实际上在进行与远程电子设备 18 的呼叫。

[0069] 如果立即的能够引起动作的语音未在块 60 中识别,逻辑流可以进行到块 62。在块 62 中。可以确定是否识别了任何能够引起动作的语音。块 62 的结果可以基于专家系统执行的分析,如与块 58 结合地描述的。作为示例,如果用户发出诸如“什么”、“你说什么”、“原谅我”、“对不起”“请您重复”的陈述,专家系统可以从这些短语中提取突出的词以确定用户在理解远程装置 18 的用户时是有困难的。在该情况下,专家系统可以将用户的语音关联到电子设备 10 的音量控制。

[0070] 作为另一示例,如果用户开始讨论与如何到达特定目的地有关的方向,专家系统可以将语音与通过因特网 web 浏览器程序 22 可使用的地图绘制服务关联。同理,与吃或餐馆有关的语音(如,用户之一说“哪里是吃的好地点”或“你想去哪里吃正餐”)可以变得与餐馆指南和 / 或使用因特网 web 浏览器 22 或其他程序 22 可访问的地图绘制服务关联。另一其他语音可以与诸如电影列表、目录(如,住宅电话列表,通常称为“白页”,和 / 或商业电话列表,通常称为“黄页”)、天气预报服务等其他服务关联。如将理解的,专家系统可以试图识别语音,此时可以收集信息以辅助一个用户或这两个用户。该类型的语音的识别可以与因特网 web 浏览器或其他信息收集工具关联。根据可确定的详情的程度,语音可以与诸如上述搜索引擎、地图绘制服务、天气预报服务、餐馆指南、电影列表、电话目录等之一的特定服务或特定的因特网网页关联。

[0071] 其他语音可以导致语音与用于执行任务的应用关联。例如,语音可以调用电子设备 10 的联系人列表程序 22 的搜索。例如,如果用户说“让我找乔的电话号码”,电子设备可以打开用户的联系人列表并搜索与名字“乔”关联的电话号码。作为另一示例,如果用户讨论何时进行人员会晤或者何时安排随后的电话呼叫,语音可以与日历功能关联,并且日历功能可以被显示给用户以容易查阅。其他语音可以与诸如电子邮件功能、即时消息收发功能、文本消息功能或多媒体消息功能的消息收发程序 22 关联。作为示例,如果用户说“我用电子邮件将该图片发给你”,对电子邮件功能和 / 或照片查看功能的关联将被进行。根据从语音获得的信息量,特定的照片可以自动附着到电子邮件消息和 / 或电子邮件消息可以使用来自用户的联系人列表的存储的电子邮件地址而被自动寻址。

[0072] 在其他情况下,用户之一可以口头上向其他用户提供有价值的信息,诸如电话号码、街道地址、方向、电子邮件地址、会晤的日期和时间或其他信息。专家系统可以配置为通过信息的格式识别信息的传送。例如,系列数字可以表示电话号码。其他语音可以指示街道地址(例如,与词汇街道、道路、干道、林荫路之一结合地使用的数字)。其他信息可以是电子邮件地址、即时消息地址、方向(如,包含一个或更多个词汇转、直走、左、右、高速等的指示)或其他信息。当该类型的语音被识别时,电子设备 10 可以存储该信息。通过存储转换后的语音的文本日志、存储包含音频通信自身的音频文件以供用户将来回放、或这些存储技术的二者,可以存储信息。

[0073] 在块 62 的肯定性确定之后,逻辑流可以进行到块 64。在块 64,可以从语音中提取信息的项。示例性的信息的项在上面已描述,并且可以包括但不限于街道地址、人的名字、地点、电影姓名、日期和 / 或时间、电话号码、电子邮件地址、或来自会话的任何其他可识别信息。如将描述的,该信息可以被输入程序 22 之一中,以进一步处理。从其他源可以收集另外的信息。例如,可以获得标识电子设备 10 和 / 或远程电子设备 18 的位置的位置信息。例如,位置信息可以格式化为 GPS 位置数据。例如,可以使用位置信息来向电子设备 10 的

用户和 / 或远程装置 18 的用户提供到特定目的地的方向。

[0074] 逻辑流可以进行到块 66, 其中被标识为对用户具有潜在用途的信息可以被存储在会话日志中, 如指示的, 信息可以按照文本格式、音频格式或者文本和音频格式二者来存储。

[0075] 在块 68, 可以标识程序 22, 程序 22 基于检测的能够引起动作的语音而对用户可以是有用的。标识的程序 22 可以是与上述语音关联的程序, 诸如可以将识别的能够引起动作的语音接受为输入的程序。如指示的, 程序可以包括因特网网络浏览器或其他信息收集工具、电子邮件消息程序或其他消息收发程序、联系人列表数据库、日历功能、时钟功能、电子设备 10 的设置控制功能、或任何其他可适用的应用。另外, 可以作用于能够引起动作的语音的程序 22 的标识可以包括特定功能、特征、服务或使用标识的程序可访问的因特网网页的标识。

[0076] 在块 68 之后, 或者在块 60 中的肯定性确定之后, 逻辑流可以进行到块 70, 在块 70, 可以向用户呈现程序 22 的列表, 程序 22 可以基于被检测的能够引起动作的语音而对用户是有用的。该列表可以具体地标识可执行的程序、服务和 / 或与能够引起动作的语音具有逻辑关系的控制功能。向用户显示的项可以是可选择的, 使得用户可以选择显示的选项, 以快速地访问关联的程序、服务或控制功能。在一些情况下, 能够引起动作的语音可以对应于可以在没有用户交互的情况下执行的特征。在该情况下, 可以省略基于能够引起动作的语音向用户呈现选项, 并且适当的程序 22 可以自动地被调用, 以执行与能够引起动作的语音以及任何关联的提取的信息相对应的动作。

[0077] 在块 70 之后, 逻辑流可以进行到块 72, 其中确定用户是否选择显示的选项。如果用户选择显示的选项, 逻辑流可以进行到块 74, 其中与选择的选项关联的程序 22 运行, 以执行相应的任务。这些相应的任务可以包括但不限于执行控制动作 (如, 调节音量设置)、从联系人列表条目搜索并取得信息、在联系人列表条目中存储信息、开始生成消息、与日历功能的交互、启动因特网网络浏览器和浏览具体服务 (如, 餐馆指南、地图绘制服务、电影列表、天气预报服务、电话目录等)、执行因特网搜索。在块 74 之后, 逻辑流可以进行到块 76, 其中如果适当的话, 来自在块 74 运行的程序 22 的输出可以被显示给用户。例如, 来自地图绘制的交互地图的指示可以显示在显示器 24 上。

[0078] 在块 62 或块 72 之一中的否定性确定之后, 或者在块 76 之后, 逻辑流可以进行到块 78。在块 78 中, 可以确定音频通信是否已结束。如果没有结束, 逻辑流可以返回到块 58, 以继续针对另外的能够引起动作的语音而监控音频通信。如果在块 78 中已确定会话结束了, 逻辑流可以进行到块 80。

[0079] 在块 80, 可以确定用户是否已选择一选项, 以针对音频通信打开会话日志。如指示的, 会话日志可以采用文本格式和 / 或音频格式。在一个实施方式中, 只要检测到能够引起动作的语音以提示会话日志的存储, 用户就可以被提供这样的机会: 当音频通信完成之后或在音频通信期间打开和查阅日志。另外, 历史的会话日志可以为了用户在将来的某一时间的参考而被存储。

[0080] 如果用户不启动会话日志, 逻辑流可以返回到块 56 以等待另一音频通信的开始。如果用户在块 80 的确启动通信日志, 逻辑流可以进行到块 82, 其中用户可以查阅所存储的信息。例如, 用户可以读取存储的文本以获取诸如方向、地址、电话号码、人的名字、电子邮

件地址等的信息。如果用户查阅包含音频通信的记录的音频文件,用户可以收听关注的信息。在一个实施方式中,通信日志可以存储与全部音频通信有关的信息。在其他实施方式中,会话日志可以包含与被发现具有能够引起动作的语音成分的音频通信的部分有关的文本和 / 或音频信息。在块 82 之后,逻辑流可以返回到块 56,以等待另一音频通信开始。

[0081] 在前面的描述中,关于英语语言给出了描述的功能的示例。将理解的是,主要通过专家系统的规则进行的语言分析可以适于英语之外的语言。例如,通过检测与方向共同使用的短语和词并通过分析包含这些词和短语的句子结构,针对从一个位置到另一位置的方向可以监控会话,而不管作为基础的语言如何。接着,驾驶或其他行驶方向可以从话音通信提取,并且为了将来的使用而可以存储所提取的信息。同理,地址可以从会话中提取并作为对地图绘制服务的输入来使用,以获得到该位置的方向以及周围区域的地图。

[0082] 上述技术可以向用户提供使用与电子设备之间的方便的接口,该电子设备 10 可以在电话呼叫或其他话音通信期间使用。该技术允许用户使用来自话音通信的有关信息来与电子设备交互。

[0083] 尽管已示出和描述了特定实施方式,但本领域的其他技术人员在阅读和理解本说明书后将理解的是,可以出现落入所附权利要求的范围内的等同物和变型例。

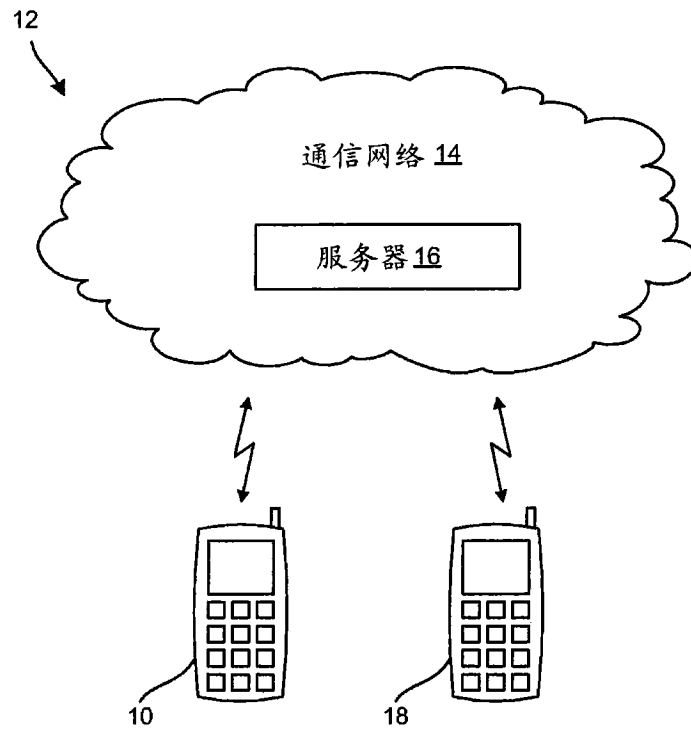


图 1

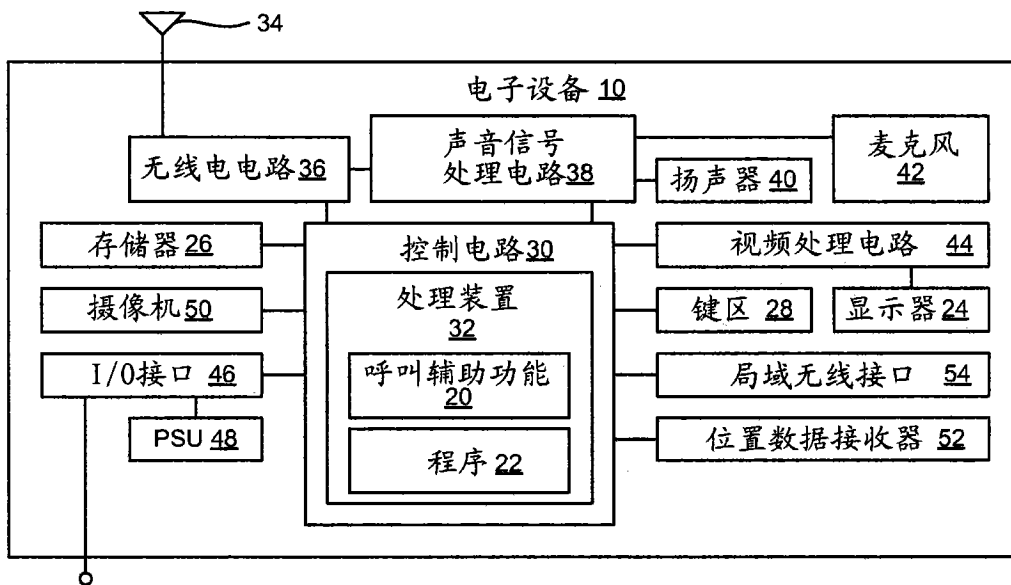


图 2

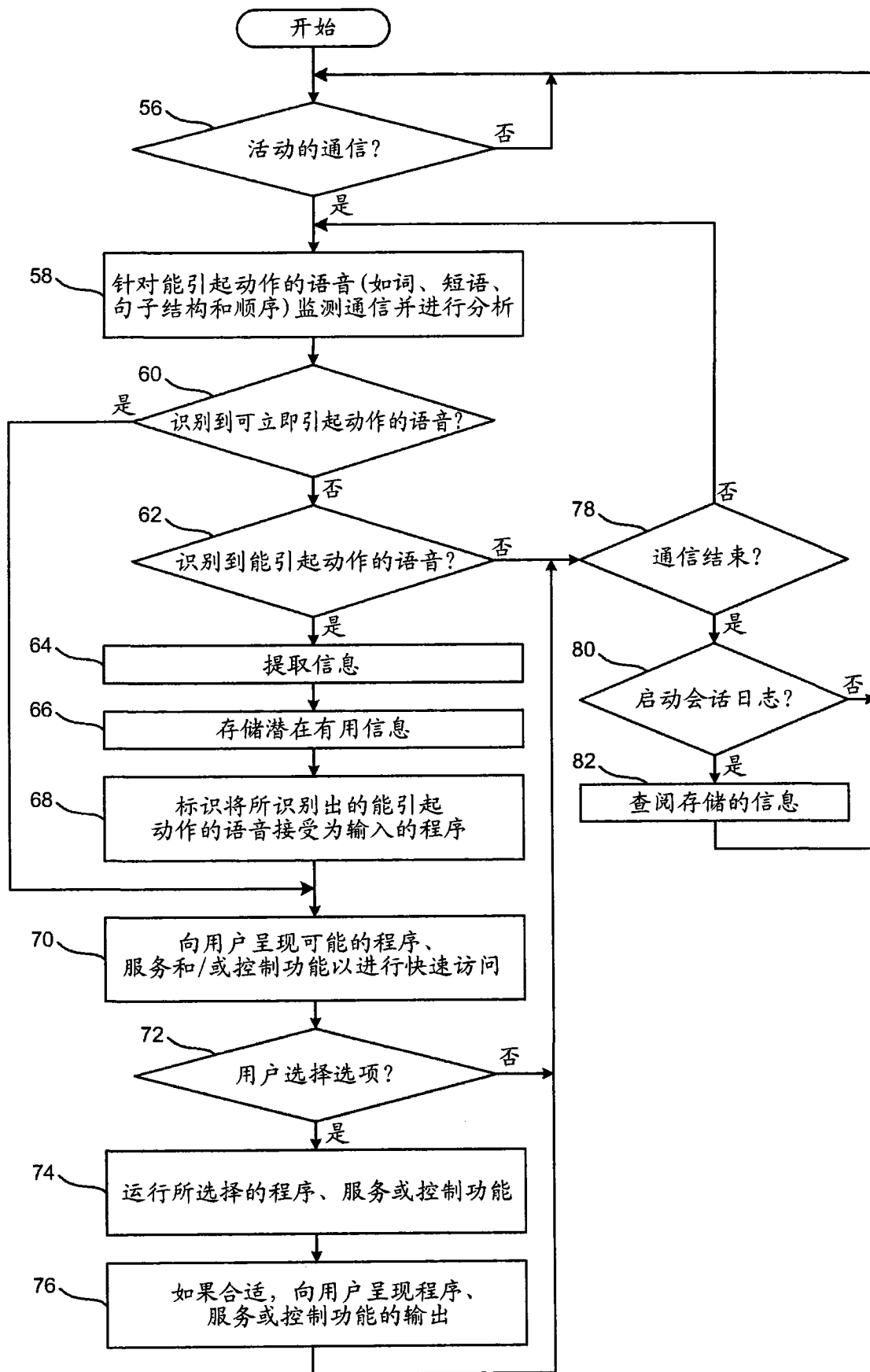


图 3