



(12) 发明专利

(10) 授权公告号 CN 102763159 B

(45) 授权公告日 2014.07.09

(21) 申请号 201180009581.X

(51) Int. Cl.

(22) 申请日 2011.01.11

G1OL 15/22(2006.01)

(30) 优先权数据

12/686,774 2010.01.13 US

G1OL 15/26(2006.01)

(85) PCT国际申请进入国家阶段日

2012.08.15

(56) 对比文件

CN 101228503 A, 2008.07.23,

WO 01/46946 A1, 2001.06.28,

US 5930751 A, 1999.07.27,

US 2006/0247931 A1, 2006.11.02,

US 2003/0171928 A1, 2003.09.11,

(86) PCT国际申请的申请数据

PCT/US2011/020825 2011.01.11

审查员 陈斌

(87) PCT国际申请的公布数据

W02011/088038 EN 2011.07.21

(73) 专利权人 苹果公司

地址 美国加利福尼亚州

(72) 发明人 阿拉姆·林达尔 波利卡尔波·伍德

(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 刘国伟

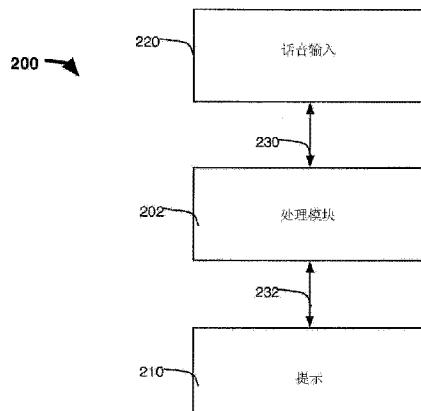
权利要求书3页 说明书11页 附图5页

(54) 发明名称

语音输入的处理

(57) 摘要

本发明是针对在提供提示时处理由电子装置接收的话音输入。明确地说，本发明是针对在监视话音输入的同时将提示序列提供给用户（例如，旁白提示）。当接收到所述话音输入时，可为所述话音输入识别特性时戳，且可将所述特性时戳和与所述已提供提示中的每一者相关联的周期或窗口进行比较。所述电子装置接着可确定对应于包含所述特性时戳的窗口的所述提示曾是所述用户曾希望将所述话音输入应用到的所述提示。所述装置可处理所述话音输入以提取用户指令，且将所述指令应用到所述已识别提示（例如，且执行与所述提示相关联的操作）。



1. 一种用于处理响应于提示而提供的话音输入的方法,其包括:

在具有至少一个处理器和存储器的电子装置处:

自动提供提示序列,其中每一提示与多个时间周期中的一相应时间周期相关联;

在提供所述提示序列中的提示的同时接收话音输入;

识别与所述已接收话音输入相关联的特性时戳;

识别所述多个时间周期中包含所述特性时戳的时间周期;以及

将所述已接收话音输入应用到所述提示序列中与所述已识别时间周期相关联的相应提示。

2. 根据权利要求 1 所述的方法,其进一步包括:

为每一提示界定初始时戳和最终时戳,其中所述初始时戳与所述最终时戳之间的周期组成与各自的所述提示相关联的所述时间周期。

3. 根据权利要求 2 所述的方法,其中:

所述初始时戳不同于对应于开始提供所述提示的时戳;且

所述最终时戳不同于对应于停止提供所述提示的时戳。

4. 根据权利要求 3 所述的方法,其中:

所述最终时戳是在对应于停止提供所述提示的所述时戳之后。

5. 根据权利要求 2 到 4 中任一权利要求所述的方法,其进一步包括:

界定最终时戳和初始时戳中的至少一者,使得与按顺序提供的提示相关联的时间周期重叠。

6. 根据权利要求 2 到 4 中任一权利要求所述的方法,其进一步包括:

确定每一提示的相对重要性;以及

基于所述提示的所述已确定相对重要性而改变每一提示的所述时间周期的长度。

7. 根据权利要求 6 所述的方法,其中改变进一步包括:

改变所述最终时戳超过对应于停止提供所述提示的所述时戳的量。

8. 根据权利要求 7 所述的方法,其进一步包括:

对所述提示进行排序,使得较不重要的提示在较重要的提示之间,以防止与所述较重要的提示相关联的所述时间周期重叠。

9. 根据权利要求 1 到 4 中任一权利要求所述的方法,其中识别所述特性时戳进一步包括:

识别期间接收所述话音输入的特性时戳。

10. 根据权利要求 9 所述的方法,其进一步包括:

为每一提示界定初始时戳和最终时戳,其中所述初始时戳与所述最终时戳之间的周期组成与所述提示相关联的所述时间周期;以及

识别初始时戳与最终时戳的组合,对于所述组合,所述特性时戳大于所述初始时戳但小于所述最终时戳。

11. 根据权利要求 1 到 4 中任一权利要求所述的方法,其中:

自动提供提示序列进一步包括自动提供话音输出提示序列,其中每一提示与一相异电子装置操作相关联。

12. 根据权利要求 1 到 4 中任一权利要求所述的方法,其中:

在提供所述相应提示的同时接收所述话音输入包含在显示视觉提示、输出音频提示或提供触觉提示的同时接收所述话音输入。

13. 一种用于处理话音输入的电子设备，其包括：

用于自动提供提示序列的装置，其中每一提示与多个时间周期中的一相应时间周期相关联；

用于在提供所述提示序列中的提示的同时接收话音输入的装置；

用于识别与所述已接收话音输入相关联的特性时戳的装置；

用于识别所述多个时间周期中包含所述特性时戳的时间周期的装置；以及

用于将所述已接收话音输入应用到所述提示序列中与所述已识别时间周期相关联的相应提示的装置。

14. 根据权利要求 13 所述的电子设备，其进一步包括：

用于为每一提示界定初始时戳和最终时戳的装置，其中所述初始时戳与所述最终时戳之间的周期组成与各自的所述提示相关联的所述时间周期。

15. 根据权利要求 14 所述的电子设备，其中：

所述初始时戳不同于对应于开始提供所述提示的时戳；且

所述最终时戳不同于对应于停止提供所述提示的时戳。

16. 根据权利要求 15 所述的电子设备，其中：

所述最终时戳是在对应于停止提供所述提示的所述时戳之后。

17. 根据权利要求 14-16 中任一权利要求所述的电子设备，其进一步包括：

用于界定最终时戳和初始时戳中的至少一者，使得与按顺序提供的提示相关联的时间周期重叠的装置。

18. 根据权利要求 14-16 中任一权利要求所述的电子设备，其进一步包括：

用于确定每一提示的相对重要性的装置；以及

用于基于所述提示的所述已确定相对重要性而改变每一提示的所述时间周期的长度的装置。

19. 根据权利要求 18 所述的电子设备，其中用于改变的装置进一步包括：

用于改变所述最终时戳超过对应于停止提供所述提示的所述时戳的量的装置。

20. 根据权利要求 19 所述的电子设备，其进一步包括：

用于对所述提示进行排序，使得较不重要的提示在较重要的提示之间，以防止与所述较重要的提示相关联的所述时间周期重叠的装置。

21. 根据权利要求 13 到 16 中任一权利要求所述的电子设备，其中识别所述特性时戳进一步包括：

用于识别期间接收所述话音输入的特性时戳的装置。

22. 根据权利要求 21 所述的电子设备，其进一步包括：

用于为每一提示界定初始时戳和最终时戳的装置，其中所述初始时戳与所述最终时戳之间的周期组成与所述提示相关联的所述时间周期；以及

用于识别初始时戳与最终时戳的组合的装置，对于所述组合，所述特性时戳大于所述初始时戳但小于所述最终时戳。

23. 根据权利要求 13 到 16 中任一权利要求所述的电子设备，其中：

用于自动提供提示序列的装置进一步包括用于自动提供话音输出提示序列的装置，其中每一提示与一相异电子装置操作相关联。

24. 根据权利要求 13 到 16 中任一权利要求所述的电子设备，其中：

用于在提供所述相应提示的同时接收所述话音输入的装置包含用于在显示视觉提示、输出音频提示或提供触觉提示的同时接收所述话音输入的装置。

话音输入的处理

技术领域

[0001] 本发明是针对响应于连续提供的提示而处理由电子装置接收的话音输入。明确地说，本发明是针对识别特定接收的话音输入与之相关联的提示。

背景技术

[0002] 许多电子装置提供用户可接入的大量特征或操作。可用特征或操作的数目常常可超过使用电子装置的输入接口可得到的输入的数目。为了允许用户接入未特定联系到特定输入（例如，不与按键序列或按钮按下相关联的输入，例如可从苹果（Apple）公司购得的iPod上的MENU按钮）的电子装置操作，电子装置可提供具有可选择选项的菜单，其中所述选项与电子装置操作相关联。举例来说，电子装置可（例如）响应于从输入接口（例如，MENU按钮）中接收到与具有可选择选项的菜单相关联的输入而在显示器上显示所述菜单。

[0003] 因为菜单通常显示在电子装置显示器上，所以可能要求用户看着所述显示器以选择特定选项。这有时可能是不合意的。举例来说，如果用户希望节约电力（例如，在便携式电子装置中），那么要求所述电子装置显示菜单并移动由用户导览的突出显示区以提供选择可能需要可通过不驱动显示器而节省的电力。作为另一实例，如果用户处于暗环境中，且显示器不包含背光，那么用户可能不能够辨别菜单的已显示选项。作为又一实例，如果用户是盲人或视力受损，那么用户可能不能够查看已显示菜单。

[0004] 为了克服此问题，一些系统可允许用户通过话音来提供指令。明确地说，电子装置可包含用于检测用户所说的词语的音频输入电路。所述装置的处理电路接着可处理所述词语以向所述电子装置识别对应指令，且执行所述对应指令。为了处理已接收话音输入，电子装置可确定话音输入的内容以及对应于所述内容的指令。

[0005] 然而，在一些情况下，可响应于由装置所提供的提示而接收话音输入。举例来说，可在电子装置提供描述特定可选择选项的话音提示或旁白时提供话音输入。作为另一实例，可在电子装置依序显示一系列可选择选项时提供话音输入。由于接收整个话音输入、处理话音输入以及确定话音输入的内容所需要的时间，可在第一提示已结束之后且在提供第二提示时处理和理解用户响应于第一提示而提供的特定话音输入。所述装置因而可难以确定哪一提示与已接收话音输入相关联。

发明内容

[0006] 本发明是针对用于在提供对应于可选择选项的提示序列时处理已接收话音输入的系统和方法。明确地说，本发明是针对识别特定提示以与已接收话音输入相关联。

[0007] 电子装置可将提示序列提供给用户，其中每一提示与用户可选择的电子装置操作相关联。可使用任何合适方法来提供提示，包含（例如）作为已显示提示、音频提示或触觉提示。当提供用户感兴趣的提示时，用户可将输入提供给装置以指导所述装置执行与所述提示相关联的操作。

[0008] 用户可使用任何合适方法将输入提供给装置。在一些实施例中，用户可提供话音

输入。因为话音输入可要求整体上接收特定持续时间且接着处理特定持续时间以确定其内容,所以可在接收和处理话音输入所花费的时间提供若干提示。接着可要求电子装置确定已提供提示中的哪一者与话音输入有关。在一些实施例中,电子装置可界定与提示中的每一者相关联的输入窗口或持续时间,使得输入窗口或持续时间可指定期间已接收话音输入与对应提示有关的特定时间帧。输入窗口或持续时间可具有任何合适长度(例如,提供提示的时间量),且可从提供提示的时间偏移(例如,输入窗口在提示后偏移了2秒)。在一些情况下,不同提示可(例如)基于每一提示的选择的相对重要性或可能性或基于特定提示的长度(例如,装置提供提示的持续时间)而具有可变的输入窗口大小或长度。

[0009] 为了使话音输入与提示有关,电子装置可使特性时间与已接收话音输入相关联。接着可将特性时间和提示的输入窗口进行比较,以确定哪一输入窗口包含特性时间。电子装置接着可确定或推断对应于包含特性时间的输入窗口的提示曾是用户感兴趣的提示。特性时间可包含期间曾接收到话音输入的任何合适时间或时间范围。举例来说,特性时间可包含曾接收到话音输入的初始时间、从初始时间偏移的时间,或任何其它合适时间。

[0010] 在一些实施例中,从连续提供的提示起的输入窗口或持续时间可重叠(例如,如果所述提示中的一者较重要且具有扩大的输入窗口)。如果与话音输入相关联的特性时间包含在重叠的输入窗口或持续时间中,那么电子装置可识别一个或一个以上额外特性时间以与话音输入相关联。电子装置接着可选择包含原始特性时间以及一个或一个以上额外特性时间两者的特定输入窗口和对应提示。

[0011] 电子装置可处理已接收话音输入以提取与所述话音输入相关联的指令。电子装置接着可将已提取指令应用到与对应于已接收话音输入的提示相关联的一个或一个以上装置操作。在一些实施例中,可由从已提取指令确定的变量或值(例如,用以充当界定新媒体播放列表的种子的媒体项目)来表征或修改装置操作。在一些实施例中,已处理话音输入可改为或另外用以识别所述话音输入与之相关联的特定提示(例如,指导装置执行与已提供提示相关联的特定操作的话音输入)。

附图说明

[0012] 在考虑以下结合附图进行的详细描述后,本发明的上述和其它特征、本发明的性质和各种优点随即将更明显,在所述附图中:

[0013] 图1是根据本发明的一个实施例的电子装置的示意图;

[0014] 图2是根据本发明的一个实施例的用于处理随着按顺序提供提示而接收的话音输入的说明性系统的示意图;

[0015] 图3是根据本发明的一个实施例的用于提供提示并接收话音输入的说明性时间线的示意图;

[0016] 图4是根据本发明的一个实施例的具有关联周期的提示的示意图;

[0017] 图5是根据本发明的一个实施例的待提供的说明性提示序列的示意图;

[0018] 图6是根据本发明的一个实施例的说明性提示序列和待处理的话音输入的示意图;

[0019] 图7是根据本发明的一个实施例的用于处理对应于提示的话音输入的说明性过程的流程图;

[0020] 图 8 是根据本发明的一个实施例的用于处理对应于提示的话音输入的说明性过程的流程图；以及

[0021] 图 9 是根据本发明的一个实施例的用于为提示界定输入窗口的说明性过程的流程图。

具体实施方式

[0022] 电子装置可操作以接收由用户提供的话音输入来控制电子装置操作。在一些情况下，所提供的话音输入可对应于来自电子装置的提示，包含（例如）依序提供的一系列提示中的一者。

[0023] 电子装置可使用任何合适方法提示用户与所述装置交互。在一些实施例中，电子装置可提供一个或一个以上提示，所述提示各自与一装置操作或指令相关联，用户可选择所述提示以指导所述装置执行操作。举例来说，电子装置可提供用于控制媒体重放的旁白提示。作为另一实例，电子装置可提供列出电子装置可启动的应用程序的已显示提示。每一提示可被提供达特定持续时间，且随后由队列中的下一提示替换。

[0024] 响应于检测到针对用户所需要的操作的提示，用户可提供指导装置执行与当前提示相关联的操作或指令的话音输入。归因于话音输入的长度以及处理话音输入所需要的时间，电子装置可在提供序列中的后续提示的同时终结处理输入。为了防止装置不正确地确定在话音输入处理结束时提供的提示与话音输入相关联，电子装置可界定与每一提示相关联的一个或一个以上时戳或时间范围。当话音输入被起初提供或由特定时戳完成或在特定时间范围（例如，如由话音输入的特性时间设置）内时，电子装置可使话音输入与对应提示相关联。明确地说，可将时戳或时间范围界定成使得在提示结束之后处理的话音输入仍可与前一提示相关联。

[0025] 每一提示可与时戳或时间范围的任何合适组合相关联。举例来说，提示可与延长超出期间提供所述提示的时间的时间范围相关联。在一些情况下，与特定提示相关联的时戳和时间范围可基于用户选择提示的历史记录、提示的类型或用户的话音输入或所述提示的任何其它特性而动态地改变。

[0026] 图 1 是根据本发明的一个实施例的电子装置的示意图。电子装置 100 可包含处理器 102、存储装置 104、存储器 106、输入接口 108 以及输出接口 110。在一些实施例中，可组合或省略电子装置组件 100 中的一者或一者以上（例如，组合存储装置 104 与存储器 106）。在一些实施例中，电子装置 100 可包含未组合或包含于图 1 所示的组件中的其它组件（例如，通信电路、定位电路、检测装置环境的感测电路、电源或总线），或图 1 所示的组件的若干例子。为了简单起见，图 1 中仅展示所述组件中的每一组件中的一者。

[0027] 处理器 102 可包含操作以控制电子装置 100 的操作和性能的任何处理电路或控制电路。举例来说，处理器 102 可用以运行操作系统应用程序、固件应用程序、媒体重放应用程序、媒体编辑应用程序，或任何其它应用程序。在一些实施例中，处理器可驱动显示器，且处理从用户接口中接收的输入。

[0028] 存储装置 104 可包含（例如）一个或一个以上存储媒体，所述存储媒体包含硬盘驱动器、固态驱动器、快闪存储器、永久存储器（例如 ROM）、任何其它合适类型的存储组件，或其任何组合。存储装置 104 可存储（例如）媒体数据（例如，音乐和视频文件）、应用程

序数据（例如，用于实施装置 100 上的功能）、固件、用户偏好信息（例如，媒体重放偏好）、验证信息（例如，与经授权用户相关联的数据库）、生活方式信息（例如，饮食偏好）、锻炼信息（例如，由锻炼监视设备获得的信息）、交易信息（例如，诸如信用卡信息等信息）、无线连接信息（例如，可使电子装置 100 能够建立无线连接的信息）、预订信息（例如，跟踪播客或电视放映或用户预订的其它媒体的信息）、联系人信息（例如，电话号码和电子邮件地址）、日历信息，以及任何其它合适数据或其任何组合。

[0029] 存储器 106 可包含高速缓存存储器、半永久存储器（例如 RAM），和 / 或用于临时存储数据的一种或一种以上不同类型的存储器。在一些实施例中，存储器 106 也可用于存储用以操作电子装置应用程序的数据，或可存储在存储装置 104 中的任何其它类型的数据。在一些实施例中，存储器 106 和存储装置 104 可被组合为单个存储媒体。

[0030] 输入接口 108 可将输入提供到电子装置的输入 / 输出电路。输入接口 108 可包含任何合适输入接口，例如按钮、小键盘、拨号盘、点按式选盘或触摸屏。在一些实施例中，电子装置 100 可包含电容性感测机构，或多触摸电容性感测机构。在一些实施例中，输入接口可包含用于接收用户的话音输入的麦克风或其它音频输入接口。输入接口可包含用于将对应于话音输入的已接收模拟信号转换为可经处理和分析以识别特定词语或指令的数字信号的模 / 数转换器。

[0031] 输出接口 110 可包含用于提供音频输出、视觉输出或其它类型的输出（例如，嗅觉、味觉或触觉输出）的一个或一个以上接口。举例来说，输出接口 110 可包含构建到电子装置 100 中的一个或一个以上扬声器（例如，单声道或立体声扬声器），或操作以耦合到音频输出机构的音频连接器（例如，音频插孔或适当的蓝牙连接）。输出接口 110 可操作以使用有线或无线连接将音频数据提供给耳机、头戴式耳机或耳塞。作为另一实例，输出接口 110 可包含用于提供用户可见的显示的显示电路（例如，屏幕或投影系统）。显示器可包含并入在电子装置 100 中的屏幕（例如，LCD 屏幕）、用于在远离电子装置 100 的表面上提供内容显示的可移动显示器或投影系统（例如视频投影仪），或任何其它合适显示器。输出接口 110 可与输入 / 输出电路（未图示）介接以将输出提供给装置的用户。

[0032] 在一些实施例中，电子装置 100 可包含操作以提供数据传送路径的总线，所述数据传送路径用于向控制处理器 102、存储装置 104、存储器 106、输入接口 108、输出接口 110 以及包含于电子装置中的任何其它组件、从控制处理器 102、存储装置 104、存储器 106、输入接口 108、输出接口 110 以及包含于电子装置中的任何其它组件中或在控制处理器 102、存储装置 104、存储器 106、输入接口 108、输出接口 110 以及包含于电子装置中的任何其它组件之间传送数据。

[0033] 用户可使用任何合适方法来与电子装置交互。在一些实施例中，用户可使用触摸输入接口（例如键盘、按钮、鼠标或触敏表面）的一个或一个以上手指来提供输入。在一些实施例中，用户可改为或另外通过以特定方式摇晃或移动电子装置（例如，使得输入接口的运动感测组件检测用户移动）来提供输入。在一些实施例中，用户可改为或另外将话音输入提供给电子装置。举例来说，用户可向嵌入在电子装置中或连接到电子装置的麦克风讲话。

[0034] 用户可在任何合适时间将话音输入提供给电子装置。在一些实施例中，电子装置可连续地监视话音输入（例如，当所述装置不处于休眠模式时，或在所有时间）。在一些实

施例中，电子装置可响应于进入话音输入的用户输入或指令而监视话音输入。举例来说，用户可选择按钮或选项，或以使得传感器检测到用户希望被提供话音输入（例如，近程传感器检测到用户已将电子装置放到用户的嘴边）的方式放置所述装置。在一些实施例中，电子装置可在一个或一个以上特定应用程序或进程正在所述装置上运行时监视用户输入。举例来说，电子装置可在媒体重放应用程序、话音控制应用程序、搜索应用程序或任何其它合适应用程序中监视话音输入。

[0035] 在一个实施方案中，电子装置可将可选择提示提供给用户，且可响应于所述提示而监视话音输入或其它类型的输入。电子装置可提供任何合适类型的提示，包含（例如）视觉提示（例如，提供于显示器上）、音频提示（例如，由音频输出接口输出）、触觉提示（例如，使用所述装置内的振动机构）或任何其它合适类型的提示中的一者或一者以上。举例来说，不包含视觉或显示输出接口的电子装置（例如，可从苹果公司购得的 iPod Shuffle）可提供音频菜单，音频菜单包含各自与一装置操作相关联的一连串提示。在一个实施方案中，音频菜单可包含用于创建新播放列表、选择现有播放列表、根据艺术家、专辑或标题来选择媒体项目的音频提示，或与控制不具有显示器的装置上的媒体重放有关的任何其它指令或操作。由用户提供的提示可以特定速率自动循环，使得每一提示被提供达特定持续时间（例如，对应于提示的内容的话音输出所需要的持续时间）。

[0036] 用户可使用任何合适方法来提供选择提示中的一者的输入。在一些实施例中，用户可使用装置的输入接口（例如按钮或触敏表面）来提供输入。用户可通过与输入接口交互（例如，执行示意动作或按下按钮）来提供输入。当输入较短以使得电子装置可在提示的持续时间内接收和处理输入时，用户可较容易地选择提示，并接收指示恰当提示曾被选择的反馈。

[0037] 由用户提供的一些输入可要求接收和处理较长的时间量。举例来说，接收和处理话音输入所需要的持续时间可长于接收和处理按钮按下或加速计输出所需要的持续时间。明确地说，持续时间可长得致使电子装置可在起初接收到话音输入时提供第一提示，且在最终处理话音输入时提供第二提示。电子装置接着可需要确定第一提示和第二提示中的哪一者与已接收话音输入相关联。

[0038] 图 2 是根据本发明的一个实施例的用于处理随着按顺序提供提示而接收的话音输入的说明性系统的示意图。系统 200 可包含处理模块 202，处理模块 202 经由路径 230 和 232 而连接到提示 210 和话音输入 220。处理模块 202 可包含在电子装置（例如，电子装置 100，图 1）中作为硬件、固件和软件的任何合适组合。举例来说，处理模块 202 可被提供为指导控制电路或处理器的操作的代码。处理模块 202 可依序将一系列提示 210 提供给装置的用户（例如，使用输出接口）。举例来说，响应于进入菜单的用户请求，处理模块 202 可识别与涉及所述菜单的指令或操作有关或对应的一组提示，且可指导输出接口提供所述提示。可以任何合适形式提供提示，包含（例如）作为视觉提示（例如，所显示的可选择选项）、音频提示（例如，旁白选项）、触觉提示（例如，对应于消息的振动），或任何其它形式。

[0039] 处理模块 202 可识别待提供的任何合适数目个提示，包含（例如）根据电子装置可用的内容而确定的数目。举例来说，处理模块 202 可针对存储在装置上的每一播放列表或针对存储在装置上的媒体项目的每一艺术家提供提示。可使用任何合适方法来提供提示。举例来说，可依序提供个别提示，使得在特定时间仅提供单个提示。或者，处理模块 202

可同时提供若干提示。在一些实施例中，处理模块 202 可提供提示 210，使得在不同时刻提供一个或一个以上不同提示。明确地说，处理模块 202 可重覆循环不同组提示 210（例如，重覆循环个别提供的提示，或重覆循环所提供的多组提示），使得用户可在不同时间选择不同提示。

[0040] 当用户检测到感兴趣提示被提供时，用户可将话音输入 220 提供给处理模块 202。话音输入 220 可具有任何合适内容，包含（例如）指示感兴趣提示的选择的内容。处理模块 202 可接收话音输入 220，且处理话音输入以识别所述输入的特定词语或短语。处理模块 202 可使用任何合适方法来处理话音输入，包含（例如）通过将已接收话音输入 220 与已知词语库进行比较，以及确定已识别库词语或短语的组合的含义。通过处理话音输入 220，处理模块 202 可识别用户感兴趣的对应提示 210，且执行对应于所述提示的操作或提供对应于所述提示的指令。

[0041] 如上文所论述，因为可花费时间来检测、接收（例如，记录以供处理）和处理话音输入，所以处理模块可在用户感兴趣的提示已被另一提示替换之后终结处理话音输入。图 3 是根据本发明的一个实施例的用于提供提示并接收话音输入的说明性时间线的示意图。时间线 300 可包含描绘时间推移的时间轴 302。在适当时间，电子装置（例如，处理模块）可依序提供提示 310、312、314 和 316。提示 310、312、314 和 316 可包含任何合适类型的提示，包含（例如）个别音频提示、已显示提示的集合，或任何其它提示。描绘提示 310、312、314 和 316 的方框中的每一者的长度可提供期间提供提示的持续时间的指示（例如，用于重放对应于音频提示的音频剪辑的时间）。当用户听到感兴趣提示时，用户可将话音输入 320 提供给装置。表示话音输入 320 的方框可指示用于检测和接收话音输入的持续时间（例如，部分 322），以及用于处理话音输入并确定所述输入的内容的持续时间（例如，部分 324）。从时间线 300 的实例可看出，话音输入 320 可同提示 312、314 和 316 重叠。此外，话音输入 320 的部分 322 仅同提示 312 和 314 重叠，且话音输入 320 的部分 324 仅同提示 314 和 316 重叠。另外，话音输入 322 在提示 310 结束之后不久开始。因此，话音输入 320 可合理地应用到提示 310、312、314 和 316 中的任一者。因此，处理模块可需要用于确保话音输入与对应提示恰当地相关联的系统或程序。

[0042] 为了确保话音输入与适当提供的提示相关联，可使每一提示与界定周期或输入窗口的时序信息相关联。如果在所述周期期间接收到话音输入，那么话音输入将对应于提示。可使用任何合适方法使一周期与每一提示相关联。图 4 是根据本发明的一个实施例的具有关联周期的提示的示意图。提示 400 可具有任何合适持续时间，包含（例如）由时间线 410 上的时戳 412 和 414 界定的持续时间。可基于提示的类型或基于由提示提供的信息而选择持续时间。举例来说，时戳 412 与 414 之间的持续时间对用户来说可至少长得足以阅读和理解书面或图形提示。作为另一实例，可将时戳 412 与 414 之间的持续时间选择成使得所述持续时间对于话音输入至少足够长以供完全听到特定指令（例如，至少长得足以重放对应于话音输出提示的整个音频剪辑）。在一些实施例中，可将时戳 412 与 414 之间的持续时间选择为长于使用户理解提示所需要的最小值，以向用户提供较长的输入窗口或周期来提供选择输入（例如，选择话音输入）。

[0043] 提示 400 可与期间将假定已检测话音输入与提示 400 有关的输入窗口或周期 420 相关联。周期 420 可同时戳 412 与 414 之间的持续时间的某一部分或全部重叠。举例来说，

周期 420 可与提示 400 的持续时间匹配。在一些实施例中，周期 420 可延长超出提示 400 的开始和结束中的一者或两者。因为可连续提供若干提示，所以可将周期 420 界定成使得其不同与另一提示相关联的周期重叠，或同所述周期最低程度地重叠。在提示 400 的实例中，周期 420 可由时戳 422 且由时戳 424 界定，时戳 422 是在时戳 412 与 414 之间（例如，在提供提示 400 时的周期期间），时戳 424 是在时戳 414 之后（例如，当不再提供提示 400）。时戳 412 与 422 之间的持续时间可和时戳 414 与 424 之间的持续时间实质上相同，使得当在提示 400 后接有后续提示时，与所述后续提示相关联的周期或输入窗口将仅在时戳 424 时开始，而不在时戳 414 时开始（例如，限制与提示 400 和后继提示相关联的输入窗口之间的重叠）。

[0044] 可使用任何合适方法来界定每一提示 400 的输入窗口或周期 420 的长度和位置。在一些实施例中，可基于提示的开始和结束而界定持续时间。举例来说，每一周期可在从提示的开始起的特定持续时间（例如，在开始之后 5 秒，或在提示的 2% 已被提供之后）开始，且在从提示的结束起的特定持续时间（例如，在提示的结束时、在当前或下一提示的持续时间的 2% 之后，或在 5 秒之后）结束。可使用初始时戳和最终时戳而为处理模块界定周期，初始时戳和最终时戳两者均可与提示相关联。

[0045] 在一些实施例中，输入窗口或周期 420 的长度和位置可基于输入窗口或周期 420 与之相关联的特定提示而变化。明确地说，可将一些提示确定为较重要或较可能由用户选择。与那些提示相关联的周期因而可长于与较不重要的提示或较不可能被选择的提示相关联的周期。举例来说，与较可能被选择的提示相关联的周期可在或较靠近提示的开始时开始、可进一步延长超出提示的结束，或此两者。

[0046] 电子装置可使用任何合适方法来确定提示选择的重要性或可能性。在一些实施例中，电子装置可提示用户提供最感兴趣的操作类型的指示，或用户很可能选择的特定提示。或者或另外，电子装置可从与所述装置的过去用户交互确定用户通常选择的特定提示，或用户提供给所述装置的提示或指令类型（例如，创建用户在不同情形下选择的提示的历史简档）。在一些实施例中，电子装置可识别使用所述装置的若干用户中的每一者，且确定所述若干用户中的每一者感兴趣的提示。

[0047] 在一些实施例中，可基于提示的相对重要性或基于与每一提示相关联的周期的长度而确定提示的次序。因为当提示周期延长超过提示的结束时，提示周期固有地限制开始点，且因此限制与后续提示相关联的周期的持续时间。因此，可能需要将较不重要的提示放在由装置提供的较重要的提示之间。图 5 是根据本发明的一个实施例的待提供的说明性提示序列的示意图。序列 500 可包含沿着时间线 501 连续提供的提示 502、504、506 和 508。在序列 500 中，提示 504 和 508 可比提示 502 和 506 重要。提示中的每一者可分别与对应周期 512、514、516 和 518 相关联。如图 5 所示，对应于较重要或相关的提示 514 和 518 的周期 514 和 518 可实质上长于对应于较不重要或相关的提示 502 和 506 的周期 512 和 516。明确地说，周期 512 可实质上在提示 502 的结束时结束，而周期 514 可在提示 504 的结束以及进入提示 506 的显著部分之后结束。周期 516 可在提示 506 之后不久结束（例如，延长进入期间提供提示 508 的时间的较短量），而周期 518 可延长超出提示 508 的结束。在序列 500 的实例中，周期 516 和 518 可部分地重叠。通过将较不重要的提示 506 放在提示 504 与 508 之间，周期 514 和 518 两者可分别延长超出提示 504 和 508 的持续时间，且减小周期

516 的持续时间。如果提示 506 在提示 504 与 508 之间尚不可用,那么周期 514 和 518 中的一者或两者可能已被要求较小以适应彼此,或可能已显著地重叠。

[0048] 在一些实施例中,电子装置可改为或另外通过调整提供提示的时间长度来间接控制与所述提示相关联的周期的持续时间。举例来说,电子装置可将每一周期界定为与提示的持续时间匹配或对应(例如,所述周期与提示开始和结束时间匹配,或从开始和结束时间稍微偏移),且改变每一提示的持续时间以增加或减小期间已接收输入将对应于已提供提示的周期。然而,此方法可提供用户体验,其中一些提示可被急冲或加速,而其它提示被抽出。

[0049] 一旦已确定与每一提示相关联的周期或输入窗口,电子装置(例如,处理模块)就可确定话音输入的哪一或哪些部分将用作识别话音输入所对应的对应提示的时戳。图 6 是根据本发明的一个实施例的说明性提示序列和待处理的话音输入的示意图。序列 600 可包含沿着时间线 602 依序提供的提示 610、612、614 和 616。每一提示可分别与一对应周期或输入窗口 620、622、624 和 626 相关联,在所述周期或输入窗口期间,已检测话音输入与对应提示相关联。话音输入 630 可在序列 600 被提供时予以提供,且可包含对应于由电子装置检测和记录话音输入的已检测部分 632,以及对应于对已检测话音输入进行处理以确定用户的输入的内容的处理部分 634。

[0050] 在一些情况下,可在提供若干相异提示时发生话音输入 630。在图 6 的特定实例中,话音输入 630 在期间提供提示 612 的时戳 640 时开始,且在期间提供提示 616 的时戳 646 时结束。因此,话音输入 630 在期间曾提供提示 614 的整个周期期间持续。此外,因为对于提示 610 的周期 620 延长到期间提供提示 612 的时间中,所以话音输入 630 曾在周期 620、622、624 和 626 期间被提供。电子装置可使用任何合适方法来确定使话音输入 630 与所述周期中的哪一者相关联。在一些实施例中,电子装置起初可确定话音输入是否同若干周期重叠。如果所述输入同若干周期重叠,那么电子装置可审查话音输入的内容,且尝试基于话音输入内容而确定将使话音输入与之相关联的特定提示。举例来说,电子装置可确定话音输入内容是否调出所述提示中的一者的指令或操作(例如,“播放播放列表 3”,当术语“播放列表 3”包含在所述提示的一者中或包含在与所述提示中的一者相关联的元数据中时)。作为另一实例,电子装置可处理话音输入以确定指令是否与任何提示有关(例如,指令代替地为不与提示有关的任意命令,例如“关机”)。

[0051] 在一些情况下,电子装置可改为或另外从话音输入 630 中选择特定的特性时戳以与整个话音输入相关联。在一些情况下,电子装置可改为或另外界定时间范围或持续时间,以表征曾接收到话音输入 630 的时间。举例来说,电子装置可选择时戳 640 或时戳 646(例如,话音输入的开始或结束)。或者,电子装置可选择时戳 644,时戳 644 指示用户提供的话音输入的结束(例如,已检测部分 632 的结束)。作为又一实例,电子装置可从所述装置检测到用户提供的输入时的周期内选择时戳 642。时戳 642 可对应于在用户提供的输入期间的任何合适时间,包含(例如)输入的中间(例如,时戳 640 与 644 之间的中途,或时戳 640 与 646 之间的中途)、从话音输入的开始或结束起的预定时间(例如,在用户开始讲话之后 2 秒,或进入已接收话音输入的 10%)、当接收到关键词或短语时(例如,当曾接收到指令关键字(例如“播放”、“暂停”或“跳过”)时),或在话音输入 630 内的任何其它合适时间。

[0052] 一旦电子装置已选择特定时戳以与话音输入相关联,电子装置就可确定包含所述

时戳的提示周期或输入窗口,且接着确定对应于所述周期或输入窗口的提示。如果若干重叠的周期或输入窗口包含所述时戳,那么电子装置可选择第二或替代时戳以应用到话音输入。电子装置接着可选择对应于其中含有第二时戳的周期的提示。在一些情况下,电子装置可改为或另外比较同若干周期或输入窗口中的每一者或同对应提示重叠的话音输入 630(或部分 632 和 634) 的量(例如,分别同提示 610 和 612 的周期 620 和 622 重叠的话音输入 630 的量)。

[0053] 一旦已识别特定提示,就可从已处理话音输入的内容中提取指令,且可将所述指令应用到所述特定提示。举例来说,如果指令包含“选择”指令,那么可执行与特定提示相关联的操作或进程。作为另一实例,如果指令包含“选择下一个”或“回到上一个”指令,那么电子装置可执行涉及提供提示(例如,且提供上一提示)或涉及实施与不同于所识别的特定提示的提示相关联的操作或进程的操作或进程(例如,改为执行来自下一提示的操作)。作为又一实例,指令可提供用于执行与提示相关联的特定操作的一个或一个以上变量或值(例如,提供媒体项目以充当用于产生新播放列表的种子)。一旦已接收到指令且执行对应操作,电子装置就可退出其中提供提示的模式(例如,假如所述指令不与提供提示序列有关)。然而,在一些实施例中,电子装置可改为或另外在确定话音输入是否对应于已提供提示之前处理话音输入以识别指令。明确地说,电子装置起初可确定话音输入指令是否与提示中的一者有关(例如,话音输入为“选择这个”),且如果话音输入对应于一提示,那么电子装置起初可仅确定哪一提示与所述输入相关联。

[0054] 图 7 是根据本发明的一个实施例的用于处理对应于提示的话音输入的说明性过程的流程图。过程 700 可在步骤 702 处开始。在步骤 704 处,电子装置可确定是否曾提供提示。举例来说,电子装置可确定是否已启用用于提供提示的模式(例如,用户是否已接入旁白菜单模式)。如果电子装置确定尚未提供提示,那么过程 700 可移动到步骤 706 并结束。

[0055] 如果在步骤 704 处电子装置改为确定提示被提供,那么过程 700 可移动到步骤 708。在步骤 708 处,电子装置可依序将提示提供给用户。举例来说,电子装置可重覆循环一组提示,其中并非所有提示均同时被提供。明确地说,电子装置可依序提供一系列旁白提示。在步骤 710 处,电子装置可确定曾接收到还是正在接收话音输入。举例来说,电子装置可确定输入接口(例如,麦克风)是否已检测到对应于话音输入的信号。如果电子装置确定尚未接收到或未在接收话音输入,那么过程 700 可返回到步骤 708,且继续依序提供提示。如果在步骤 710 处电子装置改为确定话音输入曾被或正在被接收,那么过程 700 可移动到步骤 712。

[0056] 在步骤 712 处,电子装置可识别与已接收话音输入相关联的特性时戳。举例来说,电子装置可识别曾接收到话音输入的开始时间、话音输入曾结束的结束时间、曾处理话音输入的时间,或期间曾提供或处理话音输入的任何其它合适时间。时戳可包含任何合适的时间度量,包含(例如)装置时间、相对于一个或一个以上提示的时间,或可返回与已接收提示有关的任何其它时间。在步骤 714 处,电子装置可识别对应于已提供提示中包含特性时戳的提示的时间周期。举例来说,电子装置可识别与已提供提示中的每一者相关联的时间周期或输入窗口,且将所述时间周期或输入窗口的范围和所述特性时戳进行比较。在步骤 716 处,电子装置可确定是否曾识别出若干时间周期。举例来说,电子装置可确定特性时

戳是否属于与已接收提示相关联的时间周期或输入窗口中的若干者（例如，如果时间周期或输入窗口重叠）。如果电子装置确定特性时戳仅属于一个时间周期，那么过程 700 可移动到步骤 718。在步骤 718 处，电子装置可处理话音输入以提取指令。举例来说，电子装置可识别话音输入的特定词语或短语（例如，通过与词典进行比较），且识别与已识别词语或短语相关联的指令。过程 700 接着可移动到步骤 724。

[0057] 如果在步骤 716 处电子装置改为确定特性时戳属于若干时间周期，那么过程 700 可移动到步骤 720。在步骤 720 处，电子装置可识别与已接收话音输入相关联的额外特性时戳。举例来说，电子装置可选择在期间曾提供话音输入的时间范围内的另一时戳。在一些情况下，电子装置可改为或另外识别特性时间范围以与话音输入相关联。在步骤 722 处，电子装置可识别已提供提示的包含原始特性时戳和额外时戳的时间周期。举例来说，电子装置可识别与已提供提示中的每一者相关联的时间周期或输入窗口，且将所述时间周期或输入窗口的范围与所述特性时戳和额外时戳进行比较。过程 700 接着可移动到上文所描述的步骤 718。然而，在一些实施例中，过程 700 可返回到步骤 716 以确定若干时间周期是否仍与原始特性时戳和额外特性时戳相关联。如果识别出若干时间周期，那么过程 700 可返回到步骤 720，在步骤 720 处，电子装置可识别又一额外特性时戳。

[0058] 在步骤 724 处，电子装置可将已提取指令应用到对应于已识别时间周期的提示。举例来说，电子装置可执行选择特定提示或提供执行与特定提示相关联的操作所需要的一个或一个以上变量（例如，提供媒体项目以充当用于产生新播放列表的种子）的指令。过程 700 接着可在步骤 706 处结束。

[0059] 图 8 是根据本发明的一个实施例的用于处理对应于提示的话音输入的说明性过程的流程图。过程 800 可在步骤 802 处开始。在步骤 804 处，电子装置可提供提示序列，所述提示各自与一时间周期相关联。举例来说，电子装置可依序显示或提供用于若干装置选项的音频输出。在步骤 806 处，电子装置可接收话音输入。举例来说，所述装置的输入接口可接收话音输入。在步骤 808 处，电子装置可识别与话音输入相关联的特性时间。举例来说，电子装置可识别期间曾在接收或处理话音输入的特定时间。在步骤 810 处，电子装置可识别包含所述特性时间的时间周期。举例来说，电子装置可识别特性时间所属的特定时间周期或窗口。在步骤 812 处，电子装置可将话音输入应用到与已识别时间相关联的提示。举例来说，电子装置可从话音输入中提取指令，且将所述指令应用到所述提示。过程 800 接着可在步骤 814 处结束。

[0060] 图 9 是根据本发明的一个实施例的用于为提示界定输入窗口的说明性过程的流程图。过程 900 可在步骤 902 处开始。在步骤 904 处，电子装置可识别多个提示以依序提供给用户。可使用任何合适方法来提供提示，包含（例如）使用视觉、音频或触觉提示。在步骤 806 处，电子装置可相对于用于提供提示的开始时间和结束时间中的至少一者界定偏移量。举例来说，电子装置可将经分配用于提供提示的时间的持续时间或百分比界定为偏移量。在步骤 908 处，电子装置可确定界定用于提供的输入窗口的边界的初始时间和最终时间，其中初始时间和最终时间中的至少一者从开始时间和结束时间偏移了已界定偏移量。举例来说，用于确定哪些话音输入与已提供提示相关联的输入窗口可由从期间曾提供提示的开始时间和结束时间偏移的初始时戳和最终时戳界定（例如，输入窗口比曾提供提示的时间晚五秒）。过程 900 接着可在步骤 912 处结束。

[0061] 尽管本文关于个人计算装置而描述本发明的实施例中的许多实施例,但应理解,本发明不限于个人计算应用,而是通常适用于其它应用。

[0062] 本发明的实施例优选地由软件实施,但也可在硬件或硬件与软件的组合中实施。还可将本发明的实施例体现为计算机可读媒体上的计算机可读代码。计算机可读媒体为可存储可在此后由计算机系统读取的数据的任何数据存储装置。计算机可读媒体的实例包含只读存储器、随机存取存储器、CD-ROM、DVD、磁带,以及光学数据存储装置。计算机可读媒体也可分布在网络耦合计算机系统上,使得计算机可读代码以分布式方式予以存储和执行。

[0063] 现在已知或日后想出的对如所属领域的技术人员所看到的所主张标的物的非实质性改变被明确预期为同等地在所附权利要求书的范围内。因此,所属领域的技术人员现在或日后已知的明显替换被界定为在已界定要素的范围内。

[0064] 出于说明而非限制的目的而呈现本发明的上述实施例。

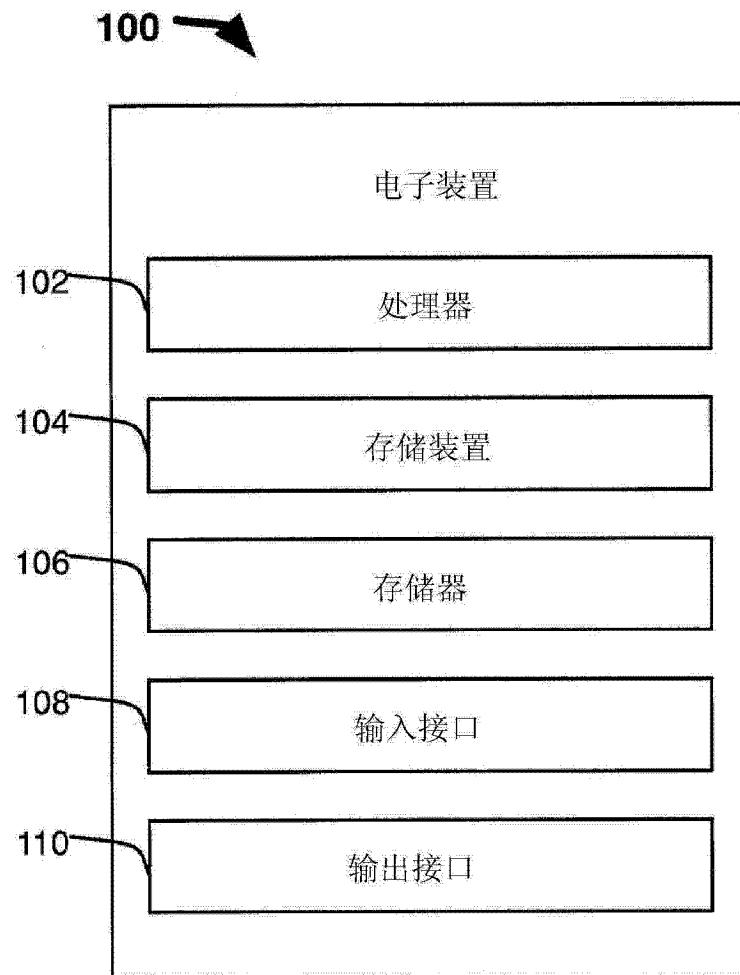


图 1

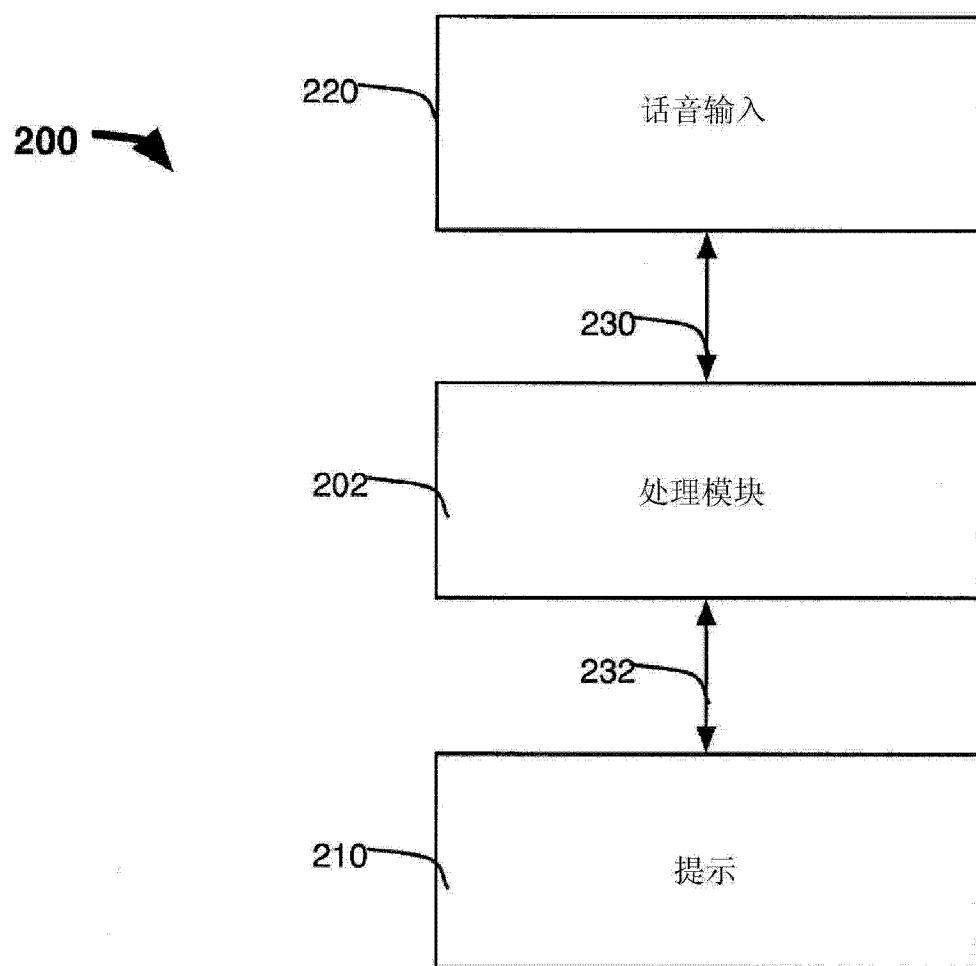


图 2

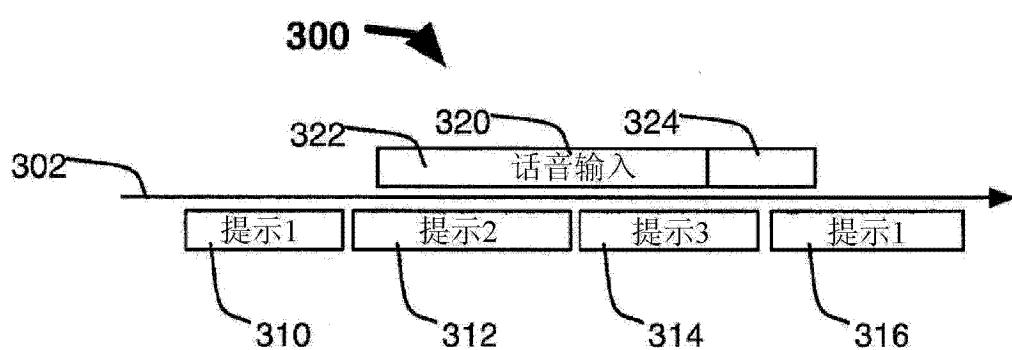


图 3

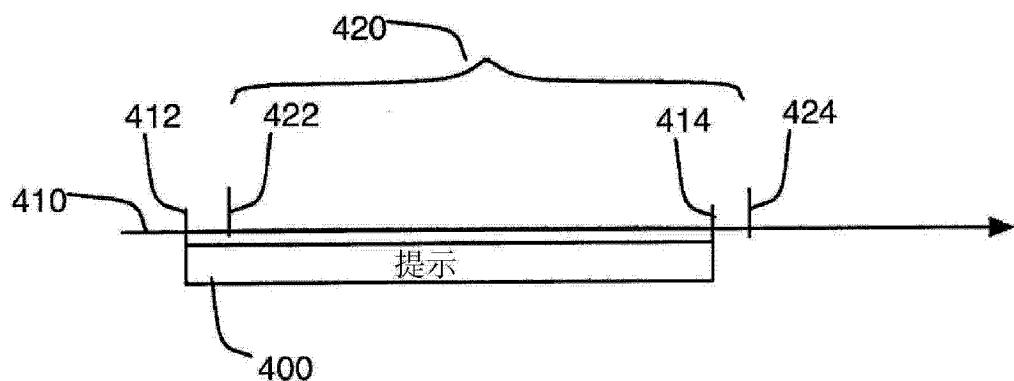


图 4

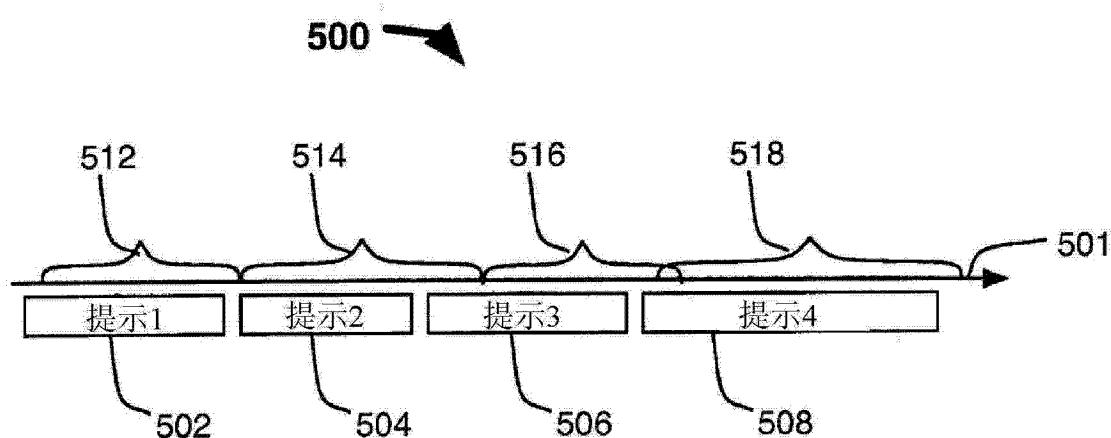


图 5

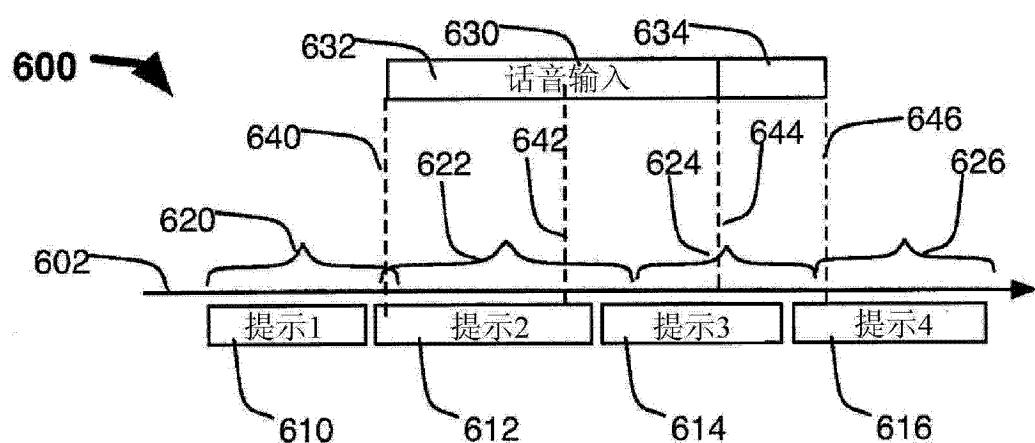


图 6

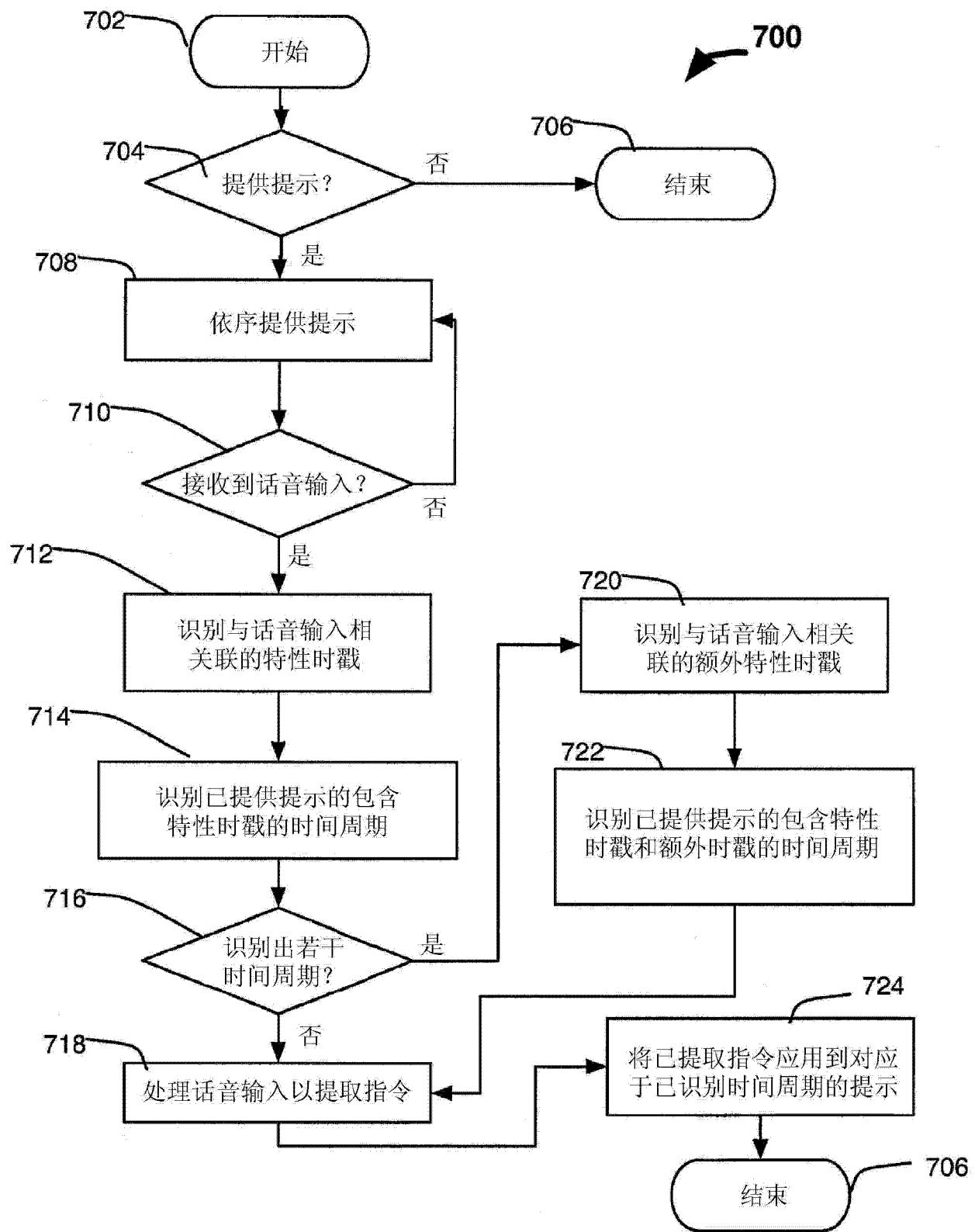


图 7

