提供到语音识别服务器 420 以用于处理。用户计算装置 200 可通过网络 410 将音频输入传输到语音识别服务器 420(14)。语音识别服务器可识别音频输入中包含的任何语音 (15)。响应于识别出语音,语音识别服务器 420 可确定用户 401 不再对着用户计算装置 200 说话,并且停止任何随后的语音识别 (16)。例如,用户 401 可说出并不对应于结构化命令或查询的字词,诸如无向的自然语言语音 406。语音识别服务器 420 也可分析语音的速度、仔细程度、音调变化或清晰度,以确定该语音并不指向用户计算装置 200,并且不应被处理成语音识别结果。

[0070] 其他类型的音频输入也可促使语音识别服务器 420 停止随后的语音识别。或者,语音识别服务器 420 可确定接收的音频输入并不包含语音。响应于接收到并不包含指向用户计算装置 200 的语音的一个或多个音频输入,语音识别服务器 420 可确定语音识别结果不应被生成,并且语音识别应停止。此外,音频输入可包含预先确定的休眠字,所述休眠字可以是用户 401 选择的。如果语音识别服务器 420 检测到休眠字,那么语音识别服务器 420 可停止对音频输入执行语音识别。此外,语音识别服务器 420 可确定多个用户 401 存在于用户计算装置 200 的附近 (例如,通过对用户计算装置 200 获得的多个音频输入执行说话人鉴别)。如果鉴别的用户 401 的数量满足阈值 (所述阈值可以是大于一的任何数量的用户 401),那么语音识别服务器 420 可确定用户计算装置 200 获得的任何音频输入不太可能被处理成语音识别结果。

[0071] 响应于确定用户 401 的语音并不指向用户计算装置 200(或由于上述任何其他原因,确定不应执行随后的语音识别),语音识别服务器 420 可通过网络 410 将停用指令传输到用户计算装置 200(17)。响应于接收到停用指令,用户计算装置 200 可停用网络接口模块 206 以及功率管理子系统 100 的一个或多个部件 (18),诸如应用处理模块 112、语音处理模块 110 和 / 或语音检测模块 108。其他条件也可促使语音识别服务器 420 将停用指令传输到用户计算装置 200。例如,返回到图 4A,如果语音识别服务器 420 确定唤醒字并未存在于在状态 (1) 处接收到的音频输入中,那么语音识别服务器 420 可将停用指令传输到用户计算装置 200。或者,语音识别服务器 420 可确定自从上次从用户计算装置 200 接收到包含语音的音频输入,已经过去了阈值时间量,并且可因此将停用指令传输到用户计算装置 200。可确定其他标准,用于将停用指令传输到用户计算装置 200。

[0072] 再次返回到图 4A,在接收到确定包含唤醒字的随后音频输入之后,用户计算装置 200 可启用功率管理子系统 100 的部件和网络接口模块 206,并且将音频输入传输到语音识别服务器 420。本文中示出的示例性操作因此可自己重复。

[0073] 图 4A、图 4B 和图 4C 中示出的示例性操作用于说明性目的。本文中示出的示例性操作中可省去一个或多个状态,或者可添加额外的状态。在一个特定的实例中,在传输语音识别服务器 420 将为其生成语音识别结果的音频输入之前,用户计算装置 200 无需从语音识别服务器 420 获得唤醒字的确认。此外,在停用网络接口模块 206 和 / 或功率管理子系统 100 的一个或多个部件诸如应用处理模块 112、语音处理模块 110 或语音检测模块 108 之前,用户计算装置 200 无需获得停用指令。事实上,功率管理子系统 100 可 (通过音频检测模块 106) 确定,自从用户计算装置 200 已获得能量水平满足能量水平阈值的音频输入之后,已经过去了至少阈值时间量。或者,用户计算装置 200 可 (通过语音检测模块 108) 确定,自从获得包含语音的音频输入之后,已经过去了至少阈值时间量。响应于确定已经过去

Petitioner Exhibit 1004, Page 759 of 1518

了阈值时间量，功率管理子系统 100 可使网络接口模块 206 停用，并且可停用其本身的一个或多个部件，如上文相对于图 3 所述。

[0074] 此外，功率管理子系统 100 可配置来识别用户 401 选择并说出的休眠字，采用大体类似于图 3 所示的识别唤醒字的方式。如果休眠字被功率管理子系统 100 检测到（例如，被语音处理模块 110 检测到），那么网络接口模块 206 和 / 或功率管理子系统 100 的一个或多个部件可被停用。同样地，如果用户计算装置 200 包含其自身装置上的语音识别能力，那么响应于检测到休眠字，它们可被停用。

[0075] 图 5 描绘出说明性用户接口 500，其可由用户计算装置 200 提供用于定制功率管理子系统 100 和用户计算装置 200 的操作。在一个实施例中，用户接口模块 212 通过用户接口 500 处理用户输入，并将其提供到功率管理子系统 100。

[0076] 能量水平阈值元件 502 可使得用户能够指定阈值能量水平，在该阈值能量水平，语音检测模块 108 应被启用，如图 3 中的框 308 所示。例如，如果用户计算装置 200 处于相对嘈杂的环境，或者如果用户计算装置 200 正在经历音频检测模块 106 确定的相当大量的“误报”，那么用户 401 可能会希望提高语音处理模块 108 被启用时所处的能量水平阈值。如果用户 401 处于相对安静的环境，或者如果用户计算装置 200 正在经历相当大量的漏报，那么用户 401 可能会希望降低语音检测模块 108 被启用时所处的能量水平阈值。如上所述，能量水平阈值可对应于音量阈值、强度阈值、振幅阈值或与音频输入相关的其他阈值。

[0077] 关键字置信度阈值元件 504 可使得用户能够指定阈值分数，在该阈值分数，语音处理模块 110 确定关键字存在。同样地，鉴别置信度阈值元件可使得用户能够指定阈值分数，在该阈值分数，语音处理模块 110 确定用户说出关键字。在一个实施例中，响应于语音处理模块 110 识别出唤醒字（例如，语音处理模块 110 确定满足阈值的分数，所述分数对应于语音中包含唤醒字的可能性），应用处理模块 112 和网络接口模块 206 被启用。在另一实施例中，响应于至少通过对应于唤醒字与用户相关联的可能性的阈值分数，语音处理模块 110 确定唤醒字与用户 401 相关联，应用处理模块 112 和网络接口模块 206 被启用。在另外的实施例中，响应于语音处理模块 110 至少通过阈值分数识别唤醒字并且至少通过阈值分数确定唤醒字与用户 401 相关联，应用处理模块 112 和网络接口模块 206 被启用。其他阈值元件可被提供来使得用户 401 能够设置个别阈值，用于启用功率管理子系统 100 的任何或所有个别部件。另外的阈值元件可被提供来使得用户能够指定说明性例行程序 300 的一个或多个框可被跳过的分数，大体如上文相对于图 3 所论述。

[0078] 用户接口 500 还可包含一个或多个计时元件 508A 和 508B。每个计时元件均可用来设置阈值时间间隔，在该阈值时间间隔，网络接口模块 206 和 / 或功率管理子系统 100 的一个或多个部件自动停用。参考计时元件 508A，如果功率管理子系统 100 确定自从用户计算装置 200 获得能量水平满足能量水平阈值的音频输入之后，已经过去了至少阈值时间间隔，那么网络接口模块 206 可自动停用，此外，功率管理子系统 100 的应用处理模块 112、语音处理模块 110、语音检测模块 108 也是如此。另外的计时元件也可用来设置阈值时间间隔，在该阈值时间间隔之后，语音识别服务器 420 将停用指令自动发送到网络接口模块 206 和功率管理子系统 100，大体如上文相对于图 4C 所论述。也可提供用于功率管理子系统 100 的其他模块的计时元件。

[0079] 继续参考图 5，用户 401 可选择唤醒字是否应由语音识别服务器 420 通过服务器确

Petitioner Exhibit 1004, Page 760 of 1518

认元件 510 来进行确认。在一些实施例中，在语音处理模块 110 检测到唤醒字之后，如果唤醒字的确认从语音识别服务器 420 接收到，应用处理模块 112 和网络接口模块 206 才保持启用。如果用户 401 要求唤醒字的服务器确认，那么随后获得的音频输入可以不传输到语音识别服务器 420，除非唤醒字被确认。然而，如上所述，未必要求确认。如果用户 401 并不要求唤醒字的服务器确认，那么用户计算装置 200 可将随后获得的一个或多个音频输入传输到语音处理模块 110，所述语音处理模块检测语音中的唤醒字和 / 或确定语音与用户 401 相关联。

[0080] 用户 401 也可选择是否要求用说话人鉴别元件 512 进行说话人鉴别。如果用户 401 要求说话人鉴别，那么语音处理模块 110 和 / 或语音识别服务器 420 可用来确定包含对应于唤醒字的语音的音频输入是否与用户 401 相关联。响应于语音处理模块 110 确定用户 401 就是语音的说话人，应用处理模块 112 和网络接口模块 206 可被启用。同样地，响应于从语音识别服务器 420 接收到用户 401 就是唤醒字的说话人的确认，网络接口模块 206 可保持启用状态。然而，如果用户 401 并不要求说话人鉴别，那么语音处理模块 110 和语音识别服务器 420 均无需鉴别说话人。

[0081] 用户接口 500 还可包含装置上的识别选择元件 514，其中用户 401 可选择用户计算装置 200 是否自己生成语音识别结果，或者音频输入是否发送到语音识别服务器 420 以用于处理成语音识别结果。如果用户计算装置 200 并不包含装置上的语音识别能力，那么装置上的识别选择元件 514 可任选地被禁用或停止。此外，如果电源 218 下降至低于阈值供电水平（例如，电池电量百分数），那么装置上的识别选择元件 514 可自动取消选择（并且装置上的语音识别能力自动禁用），因为处理单元 202 和 / 或应用处理模块 112 实施的装置上的语音识别能力可能会要求相对较大的电耗。

[0082] 唤醒字方框 516 和休眠字方框 518 可包括用户接口元件，借此，用户 401 可记录用户 401 说出的唤醒字或休眠字，并将其回放。当用户 401 记录唤醒字或休眠字时，网络接口模块 206 可自动启用，从而使得包含用户语音的音频输入可被提供到语音识别服务器 420。语音识别服务器 420 可返回记录的唤醒字或休眠字的抄本，从而用户可确定语音识别服务器 420 是否正确理解了记录的唤醒字或休眠字。或者，当用户 401 记录唤醒字或休眠字时，用户计算装置 200 的任何装置上的语音识别能力均可被启用，以对记录的用户 401 的语音进行转录。说出的唤醒字或休眠字的频谱指示也可由用户接口 500 提供。任选地，唤醒字方框 516 和休眠字方框 518 可包含唤醒字或休眠字的暗示，并且可指示用户 401 提供的唤醒字或休眠字的质量，所述质量可反映出唤醒字或休眠字会产生误报或漏报的可能性。关于提示关键字的进一步信息可见于 2012 年 11 月 6 日提交的标题为“唤醒字评估 (WAKE WORD EVALUATION)”的美国专利申请号 13/670,316。该申请案的公开内容以全文引用的方式并入本文中。

[0083] 出于说明的目的，已将本公开的各方面论述为硬件实施方式。然而，如上所述，功率管理子系统 100 可部分或完全由处理单元 202 实施。例如，功率管理子系统 100 的一些或全部功能可实施为处理单元 202 执行的软件指令。在一个特定的非限制性实例中，语音处理模块 110、应用处理模块 112 以及功率管理模块 120 的功能可实施为处理单元 202 执行的软件。因此，处理单元 202 可配置来响应于检测到唤醒字而选择性地启用和 / 或停用网络接口模块 206。可能还有其他的实施方式。

Petitioner Exhibit 1004, Page 761 of 1518

[0084] 根据实施例，本文所述的例行程序或算法中的任一个的某些动作、事件或功能可用不同的顺序来执行，并可被添加、合并或完全忽略（例如，并非所有描述的操作或事件都是实践算法所必需的）。此外，在某些实施例中，操作或事件可同时执行（例如，通过多线程处理、中断处理，或者多个处理器或处理器芯，或者在其他并行架构上），而非连续执行。

[0085] 结合本文公开的实施例描述的各种说明性逻辑块、模块、例行程序和算法步骤可实施为电子硬件、计算机软件，或是这两个的组合。为了清楚地说明硬件和软件的这种互换性，上文总体上就其功能描述了各种说明性部件、块、模块和步骤。此类功能是实施为硬件还是软件，取决于整个系统上施加的特定应用和设计限制。所述功能可用不同的方式实施，以便用于每个特定的应用，但此类实施决策不应被解释为脱离本公开的范围。

[0086] 除非另行说明，或者根据上下文可用其他方式理解，否则本文中所用的条件性语言，诸如“可”、“可以”、“可能”、“也许”、“例如”等，总体上意图传达以下内容：某些实施例包含（但其他实施例不包含）某些特征、元件和/或步骤。因此，这些条件性语言总体上并不意图暗示：特征、元件和/或步骤是一项或多项实施例无论如何都需要的，或者在有或没有作者输入或提示的情况下，一项或多项实施例必然包含决策逻辑，不论这些特征、元件和/或步骤均包含其中还是将在任何特定的实施例中执行。术语“包括”、“包含”、“具有”等是同义词，以开放的形式使用，并且并不排除额外的元件、特征、动作、操作等等。此外，术语“或”具有包含意义（而并非排除意义），因此（例如）在用于连接一系列元件时，术语“或”意味着这一系列元件中的一个、一些或所有元件。

[0087] 除非另行说明，否则结合性语言（诸如短语“X、Y和Z中的至少一个”）将在上下文中被理解为大体传达一个项目、术语等可以是X、Y或Z或者其组合。因此，这些结合性语言通常并不意图暗示某些实施例要求分别存在X中的至少一个、Y中的至少一个以及Z中的一个。

[0088] 虽然上文的具体实施方式已展示、描述并指出应用于各种实施例的新颖特征，但是应理解，在不脱离本发明的精神的情况下，可对所示装置或算法的形式和细节进行各种省略、取代和改变。应认识到，本文所述的本发明的某些实施例可采用不提供本文所述的所有特征和优点的方式实现，因为一些特征可独立于其他特征单独使用或实践。本文公开的本发明的范围由所附权利要求书、而非上述说明指示。权利要求书的范围涵盖其等效意义和范围内的所有变化。

[0089] 条款：

[0090] 1. 一种系统，其包括：

[0091] 音频输入模块，其配置来接收音频输入；

[0092] 与所述音频输入模块通信的音频检测模块，所述音频检测模块配置来确定所述音频输入的音量，并且至少部分基于所确定的音量，使语音检测模块启用；

[0093] 所述语音检测模块配置来确定指示所述音频输入包括语音的可能性的第一分数，并且至少部分基于所述分数，使唤醒字识别模块启用；

[0094] 所述唤醒字识别模块配置来确定指示所述音频输入包括唤醒字的可能性的第二分数；以及

[0095] 网络接口模块，其配置来将所获得的音频输入的至少一部分传输到远程计算装置。

Petitioner Exhibit 1004, Page 762 of 1518

[0096] 2. 根据条款 1 所述的系统，其中所述音频输入装置包括麦克风，所述音频检测模块包括第一数字信号处理器，所述语音检测模块包括第二数字信号处理器，并且所述唤醒字识别模块包括微处理器。

[0097] 3. 根据条款 1 所述的系统，其中：

[0098] 所述语音检测模块还配置来使用以下各者中的至少一个来确定所述第一分数：隐马尔可夫模型、高斯混合模型、多个频谱带中的能量或者多个频谱带中的信噪比；并且

[0099] 所述唤醒字识别模块还配置来使用以下各者中的至少一个来确定所述第二分数：应用处理模块、隐马尔可夫模型以及高斯混合模型。

[0100] 4. 根据条款 1 所述的系统，其中：

[0101] 所述唤醒字识别模块还配置来至少部分基于所述第一分数来使所述音频检测模块停用；并且

[0102] 所述唤醒字识别模块还配置来至少部分基于所述第二分数来使所述语音检测模块停用。

[0103] 5. 一种操作第一计算装置的计算机实施的方法，所述方法包括：

[0104] 接收音频输入；

[0105] 从所述音频输入确定一个或多个值，其中所述一个或多个值包括以下各者中的至少一个：

[0106] 指示所述音频输入的能量水平的第一值；或

[0107] 指示所述音频输入包括语音的可能性的第二值；

[0108] 至少部分基于所述一个或多个值，启用所述第一计算装置的第一模块；

[0109] 通过所述第一模块来执行操作，其中所述操作包括以下各者中的至少一个：

[0110] 确定所述音频输入包括唤醒字；

[0111] 对所述音频输入的至少一部分执行语音识别，以获得语音识别结果；或

[0112] 使所述音频输入的至少一部分传输到第二计算装置。

[0113] 6. 根据条款 5 所述的计算机实施的方法，其中：

[0114] 所述第一模块包括处理器，所述处理器可在低功率状态与高功率状态之间切换；并且

[0115] 所述处理器只有在处于所述高功率状态时才执行所述操作。

[0116] 7. 根据条款 6 所述的计算机实施的方法，其中启用所述第一模块包括将所述处理器从所述低功率状态切换到所述高功率状态。

[0117] 8. 根据条款 6 所述的计算机实施的方法，其还包括停用所述第一模块，其中停用所述第一模块包括将所述处理器从所述高功率状态切换到所述低功率状态。

[0118] 9. 根据条款 6 所述的计算机实施的方法，其中所述处理器包括数字信号处理器或微处理器中的至少一个。

[0119] 10. 根据条款 5 所述的计算机实施的方法，其中所述第一模块包括配置来由微处理器执行的软件模块。

[0120] 11. 根据条款 10 所述的计算机实施的方法，其中启用所述第一模块包括使所述微处理器执行所述软件模块。

[0121] 12. 根据条款 5 所述的计算机实施的方法，其中所述操作包括使所述音频输入的

Petitioner Exhibit 1004, Page 763 of 1518

至少一部分传输到所述第二计算装置，并且还包括从所述第二计算装置接收语音识别结果。

[0122] 13. 根据条款 12 所述的计算机实施的方法，其中所述语音识别结果包括以下各者中的至少一个：所述音频输入的至少一部分的抄本以及对包括在所述音频输入的至少一部分中的智能代理查询的响应。

[0123] 14. 根据条款 12 所述的计算机实施的方法，其还包括：

[0124] 至少部分基于所述一个或多个值，启用所述第一计算装置的第二模块，其中所述第二模块配置来实施语音识别应用；以及

[0125] 通过所述语音识别应用来处理所述语音识别结果。

[0126] 15. 一种装置，其包括：

[0127] 第一处理器，其配置来：

[0128] 确定一个或多个值，其中所述一个或多个值包括以下各者中的至少一个：指示音频输入的能量水平的第一值，或指示所述音频输入包括语音的可能性的第二值；以及

[0129] 至少部分基于所述一个或多个值，使第二处理器启用；

[0130] 所述第二处理器配置来执行操作，其中所述操作包括以下各者中的至少一个：

[0131] 确定所述音频输入包括唤醒字；

[0132] 对所述音频输入的至少一部分执行语音识别，以获得语音识别结果；或

[0133] 使所述音频输入的至少一部分传输到第二装置。

[0134] 16. 根据条款 15 所述的装置，其中所述第一处理器包括数字信号处理器或微处理器中的至少一个。

[0135] 17. 根据条款 15 所述的装置，其中所述第二处理器包括数字信号处理器或微处理器中的至少一个。

[0136] 18. 根据条款 15 所述的装置，其还包括配置来存储所述音频输入的存储器缓冲模块。

[0137] 19. 根据条款 18 所述的装置，其中配置来存储所述音频输入的所述存储器缓冲模块包括环形缓冲区。

[0138] 20. 根据条款 15 所述的装置，其还包括与所述第一处理器通信的音频输入模块，其中所述音频输入模块配置来获得所述音频输入。

[0139] 21. 一种系统，其包括：

[0140] 音频输入模块，其配置来获得音频输入；

[0141] 与所述音频输入模块通信的第一模块；以及

[0142] 与所述第一模块通信的第二模块；

[0143] 其中所述第一模块配置来：

[0144] 至少部分基于所述音频输入来确定一个或多个值，以及

[0145] 至少部分基于所述一个或多个值，使所述第二模块启用；并且

[0146] 其中所述第二模块配置来对所述音频输入的至少一部分执行语音识别。

[0147] 22. 根据条款 21 所述的系统，其中所述一个或多个值包括所述音频输入的音量。

[0148] 23. 根据条款 22 所述的系统，其中只有在所述音频输入的所述音量在至少阈值持续时间内满足音量阈值时，所述第二模块才被启用。

Petitioner Exhibit 1004, Page 764 of 1518

[0149] 24. 根据条款 21 所述的系统，其中所述一个或多个值包括所述音频输入包括语音的可能性。

[0150] 25. 根据条款 21 所述的系统，其中所述一个或多个值包括指示以下可能性的分数：所述音频输入包括唤醒字。

[0151] 26. 根据条款 25 所述的系统，其中所述一个或多个值还包括指示以下可能性的分数：所述唤醒字由与所述唤醒字相关联的用户说出。

[0152] 27. 根据条款 21 所述的系统，其中所述第二模块配置来通过为所述音频输入的至少一部分生成语音识别结果对所述音频输入的至少一部分执行语音识别。

[0153] 28. 根据条款 21 所述的系统，其中所述第二模块配置来通过以下操作对所述音频输入的至少一部分执行语音识别：

[0154] 将所述音频输入传输到远程计算装置；以及

[0155] 从所述远程计算装置接收所述音频输入的至少一部分的语音识别结果。

Petitioner Exhibit 1004, Page 765 of 1518

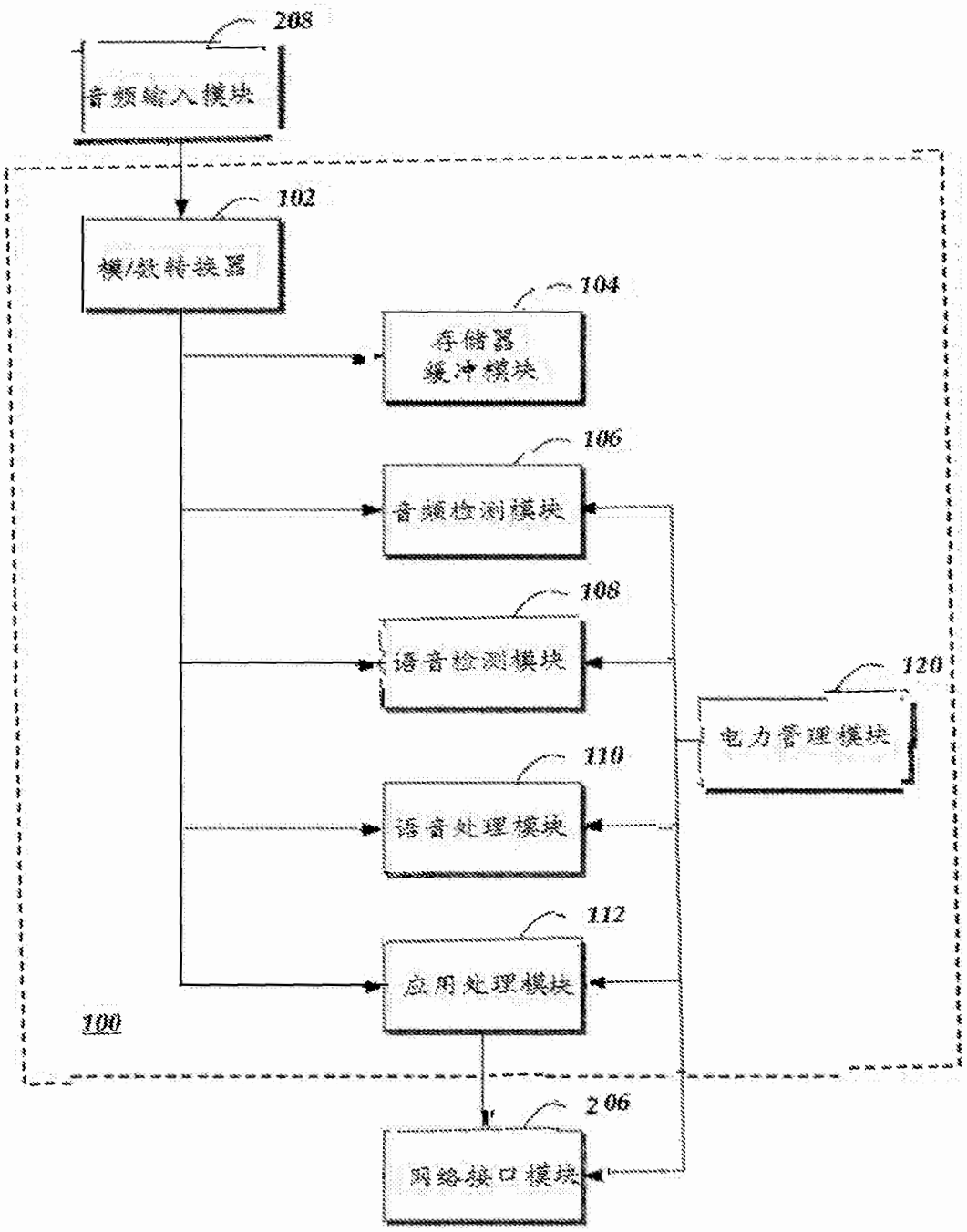


图 1

Petitioner Exhibit 1004, Page 766 of 1518

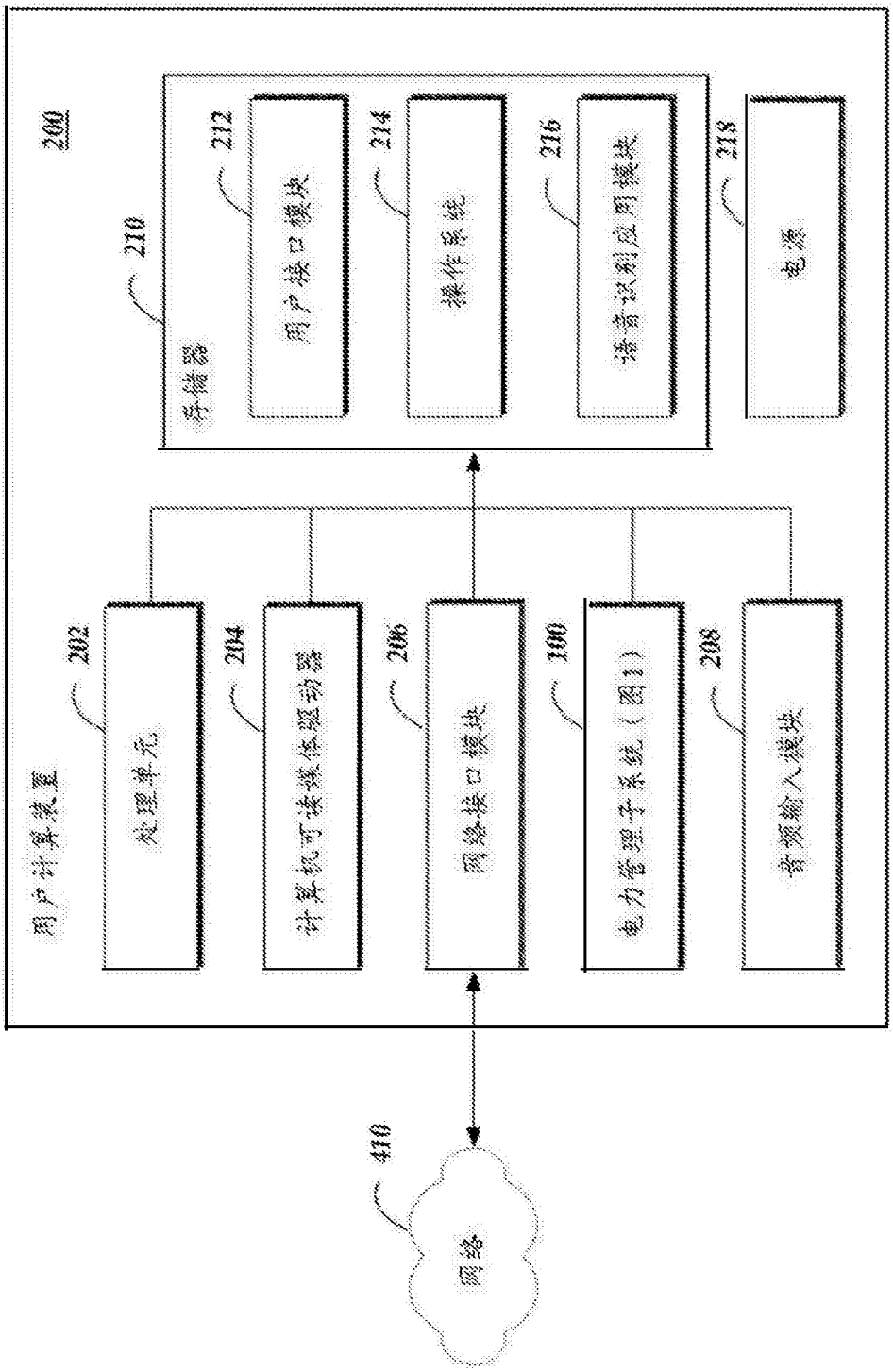


图 2

Petitioner Exhibit 1004, Page 767 of 1518

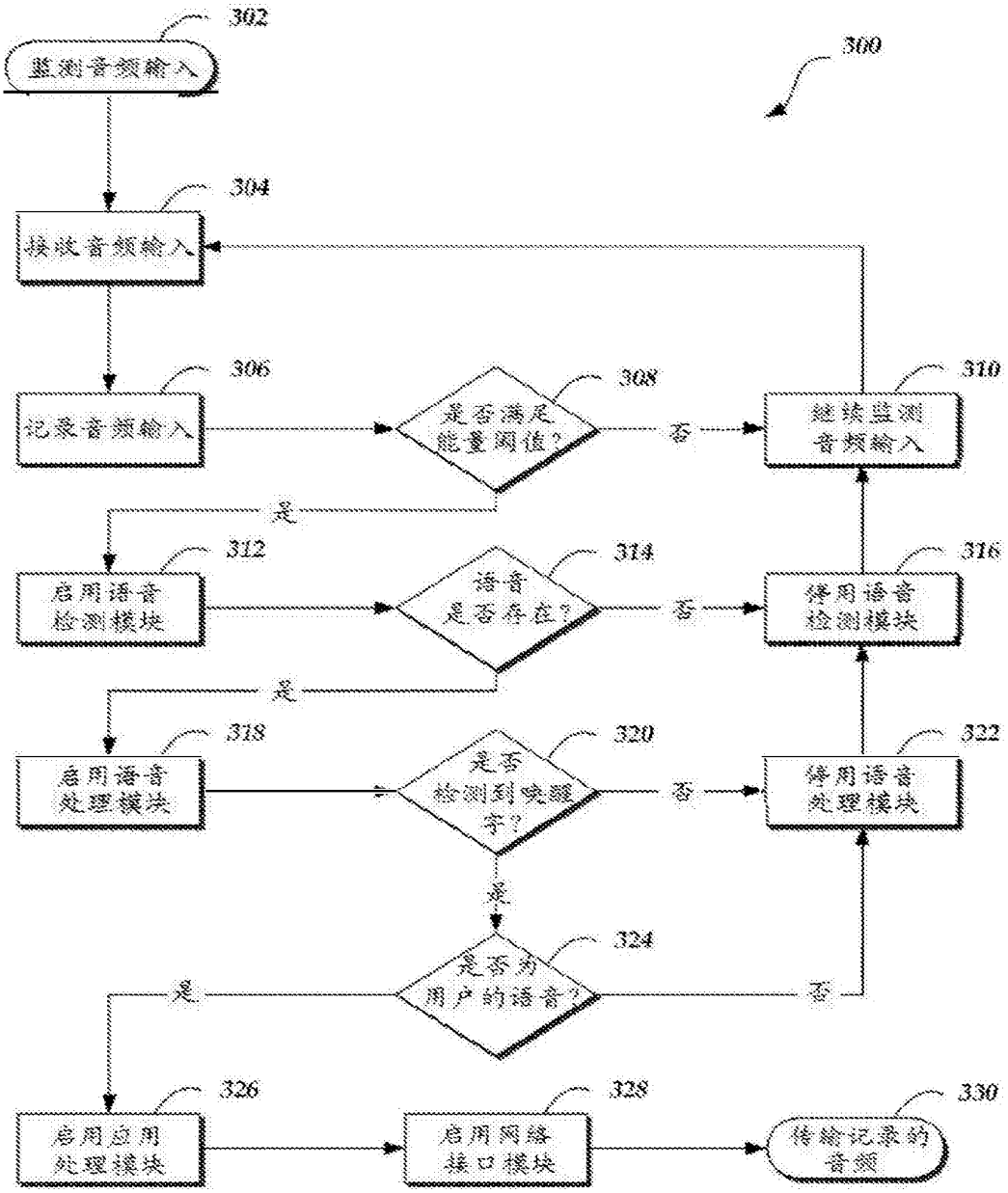


图 3

Petitioner Exhibit 1004, Page 768 of 1518

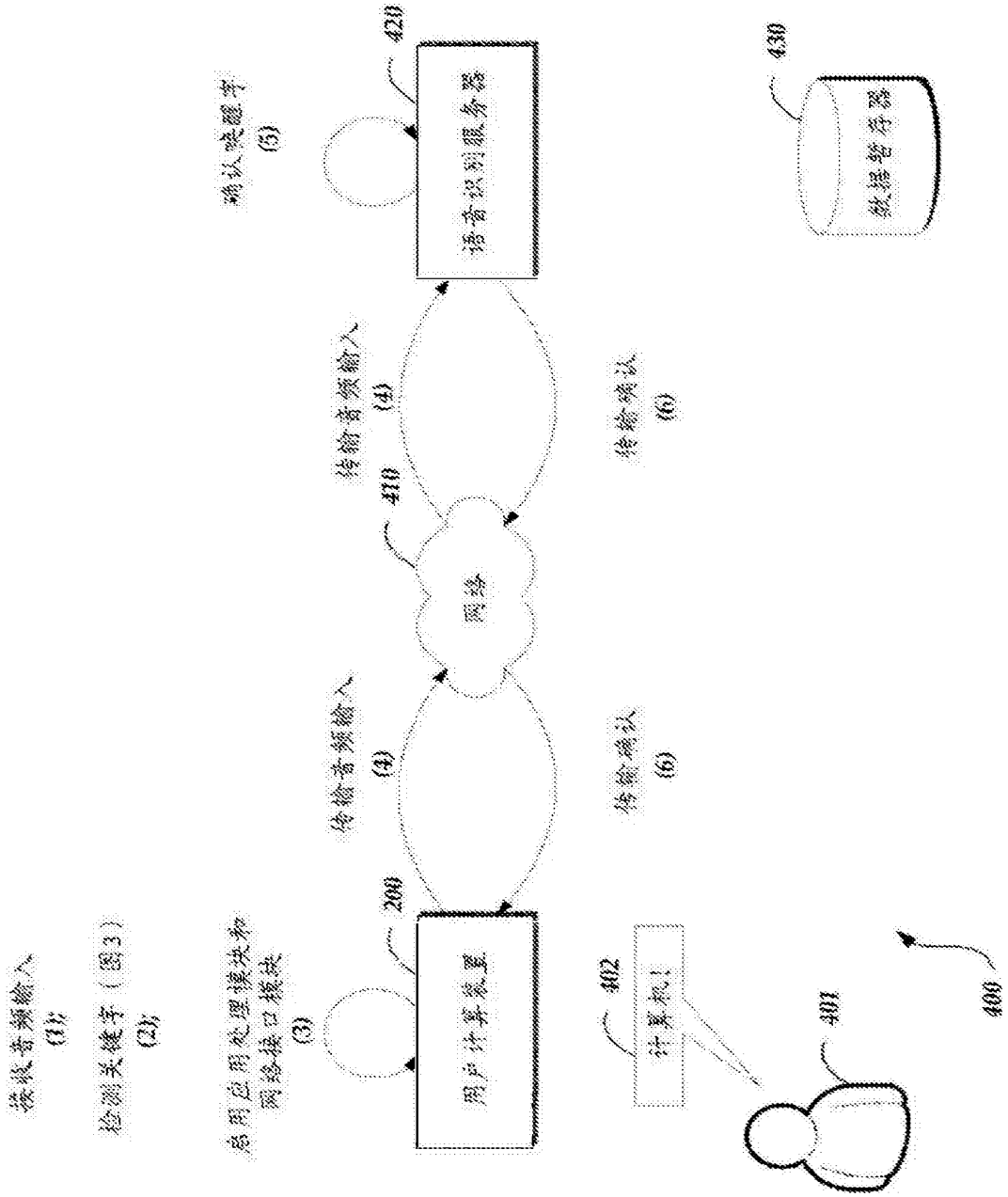


图 4A

Petitioner Exhibit 1004, Page 769 of 1518

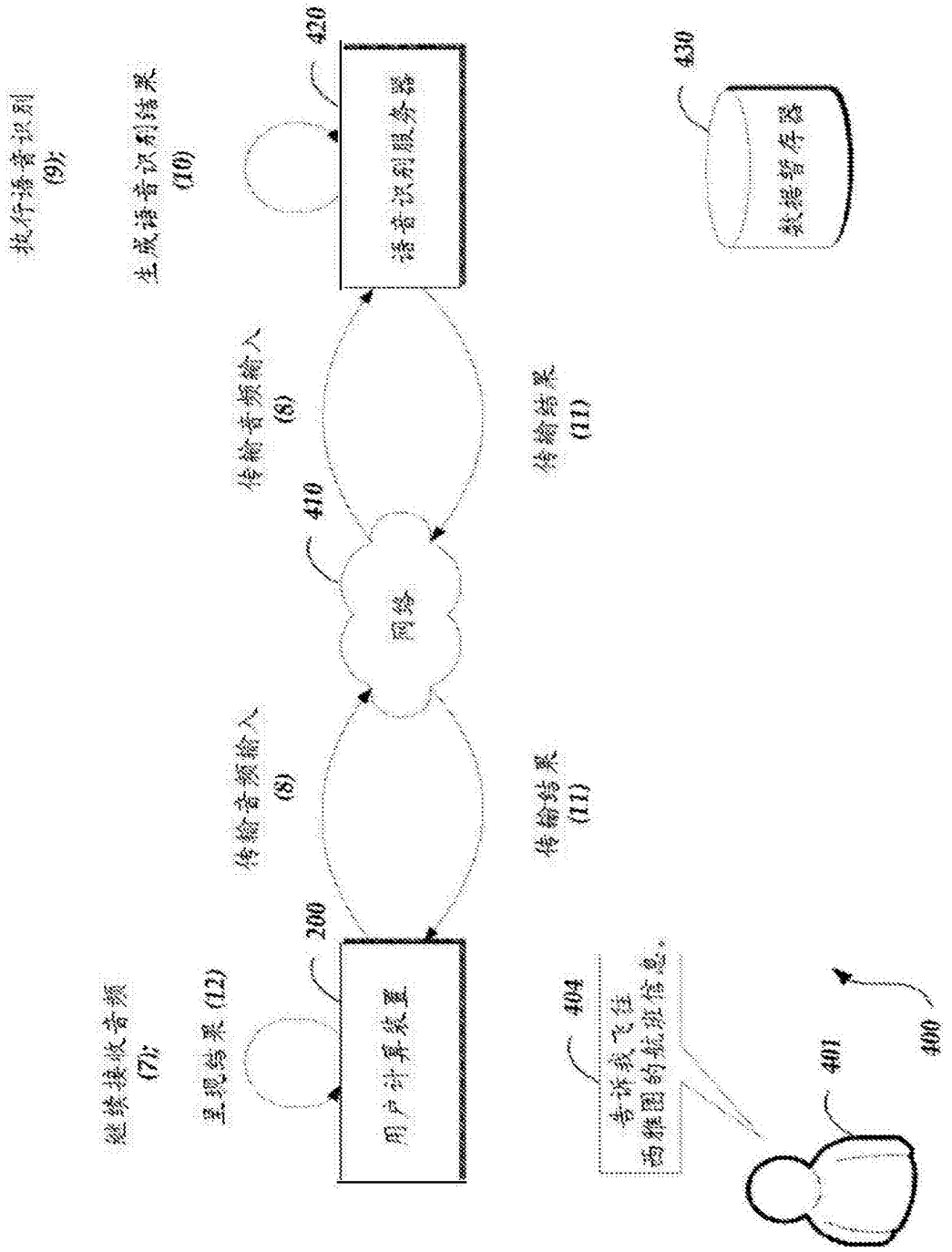


图 4B

Petitioner Exhibit 1004, Page 770 of 1518

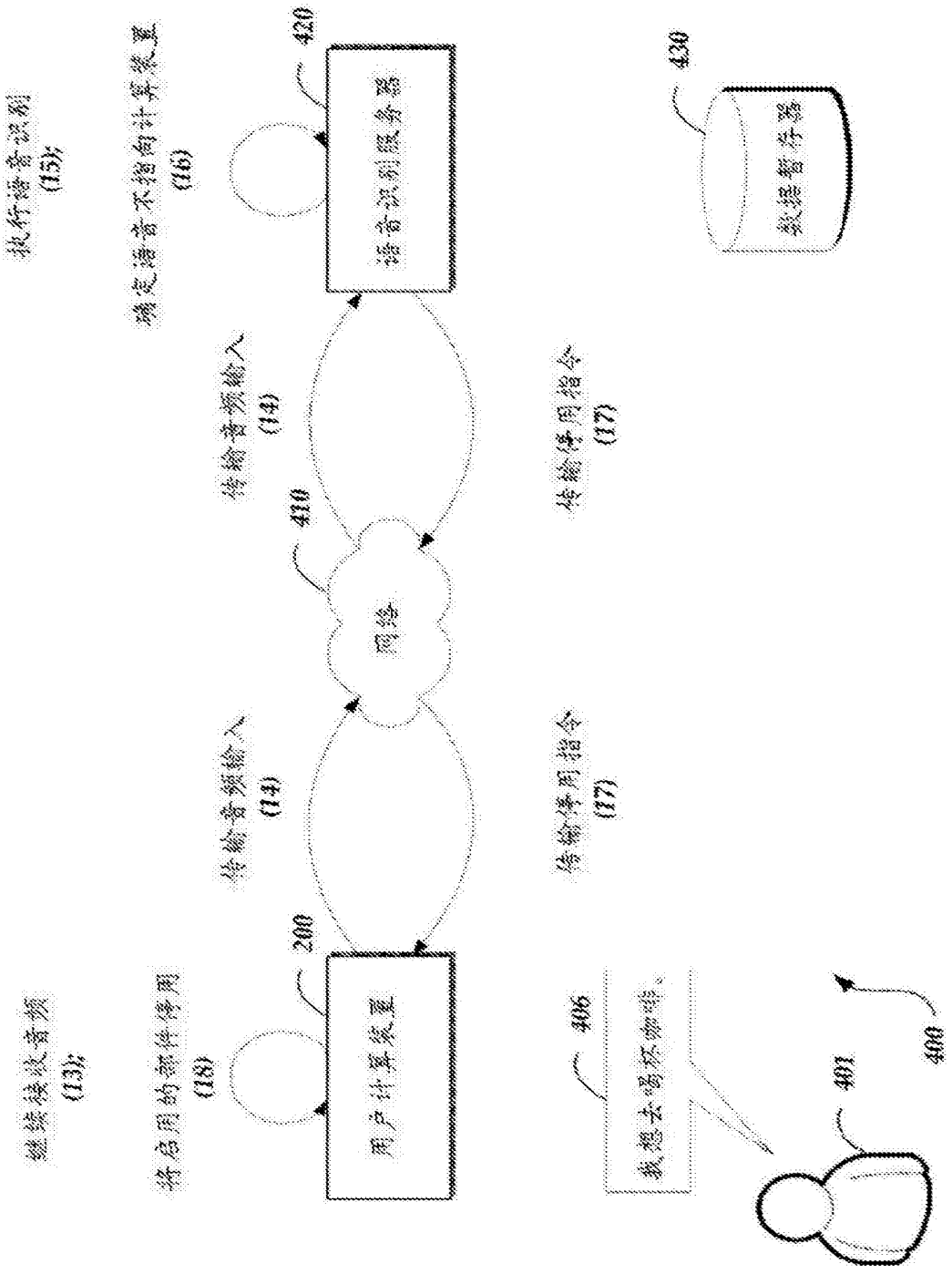


图 4C

Petitioner Exhibit 1004, Page 771 of 1518

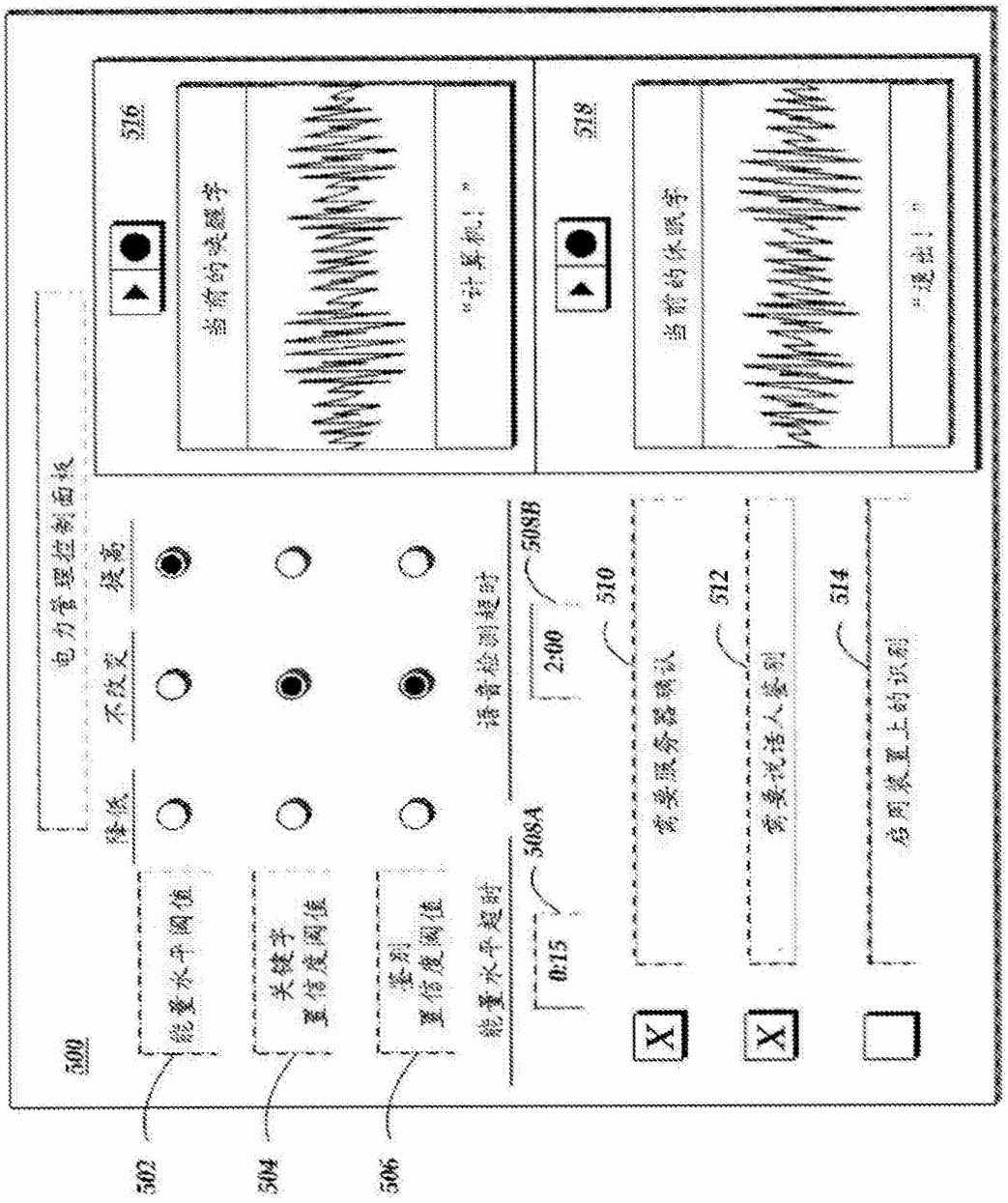


图 5

Petitioner Exhibit 1004, Page 772 of 1518

Publication number: CN103632668 A
Publication country: CHINA
Publication type: UNEXAMINED APPLIC. OPEN TO PUBLIC INSPECTION
Publication date: 20140312
Application number: CN201210299700
Application date: 20120821
Priority: CN201210299700201208 21CN20121299700;
Assignee: BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO ;
Assignee^{std}: BEIJING BAIDU NETCOM SCI AND TEC ;
Inventor^{std}: SU DAN ; WAN GUANGLU ;
International class^{8}: G10L15/06; G10L15/00;
Family members: CN103632668 A CN103632668 B

Title: METHOD AND APPARATUS FOR TRAINING ENGLISH VOICE MODEL BASED ON CHINESE VOICE INFORMATION

Title: 一种基于中文语音信息训练英文语音模型的方法与设备

Abstract:

The invention aims to provide a method and apparatus for training an English voice model based on Chinese voice information. Included are the steps of obtaining voice data inputted by a user through a voice collecting device; determining whether the voice data are English voice data; if so, conducting voice recognition processing for the English voice data based on the Chinese voice information to obtain corresponding pronunciation information to the English voice data; and determining corresponding English data to the English voice data based on the pronunciation information to create or update the English voice model. Compared with the prior art, on the basis of an original Chinese acoustic model, the method and apparatus for training the English voice model based on the Chinese voice information use the similar Chinese voice information and conduct voice recognition processing for the English voice data to obtain the corresponding pronunciation information to the English voice data, thereby achieving the creation or update of the English voice model and effectively controlling the size of the English voice model.

Abstract:

本发明的目的是提供一种用于基于中文语音信息训练英文语音模型的方法与设备；获取用户通过语音采集装置输入的语音数据；判断该语音数据是否为英文语音数据；当该语音数据为英文语音数据，根据中文语音信息，对该英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的发音信息；根据该发音信息，确定该英文语音数据所对应的英文数据，以建立或更新英文语音模型。与现有

The purpose of the present invention is to provide a method and apparatus based Chinese English voice model trained voice information; obtaining user data via voice input voice collecting device; determining whether or not the voice data to voice data in English; when the voice data into English voice data, according to Chinese voice messages, the English voice data for speech recognition processing, in order to obtain the English voice

技术相比，本发明在原有中文声学模型基础上，利用相似的中文语音信息，对英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的发音信息，实现对英文语音模型的建立或更新，有效地控制了英文语音模型的规模。

data corresponding pronunciation information; based on the pronunciation information, determined that the English voice data corresponding English data, in order to establish or Update English speech model. Compared with the prior art, the invention of the original Chinese acoustic model based on the similarity of Chinese voice information, data on the English voice for voice recognition processing to obtain the English voice data corresponding pronunciation information to realize the English voice create or update the model, to effectively control the size of the English voice model.

Claims:

1.一种用于基于中文语音信息训练英文语音模型的方法，其中，该方法包括以下步骤:a获取用户通过语音采集装置输入的语音数据；b判断所述语音数据是否为英文语音数据；c当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息；d根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射。

2.根据权利要求1所述的方法，其中，所述步骤c包括:-当所述语音数据为英文语音数据，根据单词模型所对应的中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的单词发音信息。

3.根据权利要求1所述的方法，其中，所述步骤c包括:-当所述语音数据为英文语音数据，根据字母模型所对应的中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的字母发音信息。

4.根据权利要求1至3中任一项所述的方法，其中，该方法还包括:-统计所述英文数据所对应的英文语音数据在所述语音数据中的出现次数；-根据所述出现次数，并结合所述语音数据的数量信息，确定所述英文数据的概率值；-根据所述概率值，更新所述英文语音模型。

5.根据权利要求4所述的方法，其中，该方法还包括:-当所述出现次数大于预定数量阈值，自所述英文语音模型中提取所述英文数据与所述英文语音数据；r根据所述英文数据与所述英文语音数据的映射关系，建立或更新与所述英文语音数据相对应的英文语音单元。

.

1 for a Chinese voice information speech model training method based on English, wherein the method comprises the steps of: a voice to obtain user data acquisition device through voice input; b judge the speech data Are the English voice data; c when the speech data is English voice data, according to Chinese voice information, data on the English voice for voice recognition processing, in order to obtain the English voice data corresponding pronunciation information; d according to the said pronunciation information, determine the English voice data corresponding English data, in order to establish or update the English voice model, wherein the pronunciation information and data through the English voice information is mapped to the Chinese.

.

2 The method according to claim 1 or claim 2, wherein said step c comprises: - when the speech data is speech data in English, the word model based on voice information corresponding to the Chinese, English voice data to the voice recognition processing, in order to obtain the English voice data corresponding word pronunciation information.

.

3 method according to claim 1 or claim 2, wherein said step c comprises: - when the speech data is speech data in English, according to the letter model corresponding to the Chinese speech information, English voice data to the voice recognition processing, in order to obtain the

6.根据权利要求5所述的方法，其中，所述步骤r包括:-根据所述英文数据与所述英文语音数据的映射关系，采用插值计算方法，建立或更新所述英文语音单元。

7.根据权利要求5或6所述的方法，其中，该方法还包括:-根据所述英文语音数据及与该英文语音数据相对应的发音信息，更新所述英文语音单元。

8.根据权利要求1至7中任一项所述的方法，其中，该方法还包括:X获取所述用户为目标应用输入的单词语音序列；y根据所述英文语音模型，确定与所述单词语音序列相对应的英文字符序列；Z将与所述英文字符序列相对应的信息提供给所述用户。

9.根据权利要求8所述的方法，其中，所述目标应用为语音输入应用；其中，所述步骤z包括:-将所述英文字符序列作为语音输入候选词条提供给所述用户。

10.根据权利要求8所述的方法，其中，所述目标应用为语音搜索应用；其中，该方法还包括:-将所述英文字符序列作为所述用户的查询序列，获取与所述查询序列相对应的一个或多个搜索结果；其中，所述步骤Z包括:-将所述一个或多个搜索结果提供给所述用户。

11.根据权利要求8至10中任一项所述的方法，其中，所述步骤X包括:-获取所述用户为目标应用输入的字母语音序列；其中，所述步骤y包括:-根据所述英文语音模型，确定与所述字母语音序列相对应的英文字符序列。

12.一种用于基于中文语音信息训练英文语音模型的模型训练设备，其中，该设备包括:语音获取装置，用于获取用户通过语音采集装置输入的语音数据；英文判断装置，用于判断所述语音数据是否为英文语音数据；语音识别装置，用于当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息；模型训练装置，用于根 据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射。

13.根据权利要求12所述的模型训练设备，其中，所述语音识别装置用于:-当所述语音数据为英文语音数据，根据单词模型所对应的中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的单词发音信息。

English voice data information corresponding letter sounds.

4 according to the method as claimed in claims 1 3, wherein the method further comprises: - statistical data corresponding to the English, the English voice data in the voice The number of occurrences of the data; - according to the number of occurrences, and the amount of data combined with the voice message, determining the probability value data of English; - based on the probability value, updating the English speech model.

.

5 method according to claim 4, wherein the method further comprises: - when the number of occurrences is greater than a predetermined threshold amount, extracted from the English pronunciation of the English model data with the English voice data; r based on data mapping between the English and the English voice data, create or update data with the corresponding English voice English voice unit.

.

6 The method according to claim 5, wherein said step r comprises: - data according to the mapping relationship between the English and the English voice data, using the method of interpolation calculation, create or update the English voice unit.

7 5 or claim 6, wherein, wherein, the method further comprising: - According to the English voice data and voice data corresponding to the English pronunciation information update The English voice unit.

8 according to the method as claimed in any one of claims 1-7, wherein the method further comprises:. X Get the word speech sequence for the target application user input; y according to The English voice model to determine the sequence of words corresponding to the English voice character sequence; Z will provide the English character sequence information corresponding to the user.

.

9 The method according to claim 8, wherein the target application is speech input application;

14.根据权利要求12所述的模型训练设备，其中，所述语音识别装置用于:-当所述语音数据为英文语音数据，根据字母模型所对应的中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的字母发音信息。

15.根据权利要求12至14中任一项所述的模型训练设备，其中，该设备还包括模型更新装置，用于:-统计所述英文数据所对应的英文语音数据在所述语音数据中的出现次数；-根据所述出现次数，并结合所述语音数据的数量信息，确定所述英文数据的概率值；-根据所述概率值，更新所述英文语音模型。

16.根据权利要求15所述的模型训练设备，其中，该设备还包括:数据提取装置，用于当所述出现次数大于预定数量阈值，自所述英文语音模型中提取所述英文数据与所述英文语音数据；单元训练装置，用于根据所述英文数据与所述英文语音数据的映射关系，建立或更新与所述英文语音数据相对应的英文语音单元。

17.根据权利要求16所述的模型训练设备，其中，所述单元训练装置用于:-根据所述英文数据与所述英文语音数据的映射关系，采用插值计算方法，建立或更新所述英文语音单元。

18.根据权利要求16或17所述的模型训练设备，其中，该设备还包括单元更新装置，用于:-根据所述英文语音数据及与该英文语音数据相对应的发音信息，更新所述英文语音单元。

19.根据权利要求12至18中任一项所述的模型训练设备，其中，该设备还包括:序列获取装置，用于获取所述用户为目标应用输入的单词语音序列；序列确定装置，用于根据所述英文语音模型，确定与所述单词语音序列相对应的英文字符序列；信息提供装置，用于将与所述英文字符序列相对应的信息提供给所述用户。

20.根据权利要求19所述的模型训练设备，其中，所述目标应用为语音输入应用；其中，所述信息提供装置用于:-将所述英文字符序列作为语音输入候选词条提供给所述用户。

21.根据权利要求19所述的模型训练设备，其中，所述目标应用为语音搜索应用；其中，该设备还包括:搜索获取装置，用于将所述英文字符序列作为所述用户的查询序列，获取与所述查询序列相对应的一个或多个搜索结果；其中，所述信息提供装置用于:-将所述一个或多个搜索结果提供给所述用户。

wherein, z comprises the step of: - the sequence of characters as the English speech input Candidate entry provided to the user.

.

10 The method according to claim according to claim 8, wherein the target application is a voice search application; wherein the method further comprises: - the sequence of the English characters as the user the query sequence to obtain the query sequence corresponding to one or more search results; wherein Z comprises the steps of: - one or more search results to the user.

.

11 The method of any one of claims 8 to 10, according to claim, wherein X comprises the step of: - obtaining the sequence of the target user for the application of the letter speech input; wherein said step y include: - According to the English voice model to determine the sequence of letters corresponding English voice character sequence.

.

12 A model-based training devices Chinese English speech model training speech information, wherein the apparatus includes: a voice acquisition means for acquiring a voice user input via voice acquisition device transactions; English judging means for judging whether the data of the voice in English speech data; speech recognition means for, when the voice data for the English voice data, voice information based on Chinese, English voice data of the voice recognition processing to obtain with the English voice data corresponding pronunciation information; model training means for the pronunciation information, determine the English voice data corresponding English data, in order to create or update the English voice model, in which the said information relating to the English pronunciation of the Chinese voice information data through mapped.

.

13 in accordance with the model training apparatus according to claim 12, wherein said speech recognition means for: - when the speech data is speech data in English, according to the word model corresponding to Chinese voice message, the English voice data to voice recognition processing, in order to obtain the

22.根据权利要求19至21中任一项所述的模型训练设备，其中，所述序列获取装置用于:-获取所述用户为目标应用输入的字母语音序列；其中，所述序列确定装置用于:-根据所述英文语音模型，确定与所述字母语音序列相对应的英文字符序列。

English voice data corresponding word pronunciation information.

.

14 model training apparatus according to claim 12, wherein, wherein the speech recognition means for: - When the voice data into English voice data, based on the letter corresponding to the model Chinese voice message, the English voice data to voice recognition processing, in order to obtain the English voice data information corresponding letter sounds.

.

15 model training apparatus according to claim 12 to 14 in any one of, wherein the apparatus further comprises a model updating means for: - the statistical data corresponding to the English in English, the number of occurrences in the speech data in the speech data; - according to the number of occurrences, and the combined information of the number of voice data, the data determining the probability value in English; - based on the probability value, updating the English speech model.

.

16 model training apparatus according to claim 15, wherein, the apparatus further comprising: data extracting means for a predetermined number of times greater than the threshold value, from the English occurs when the English speech model to extract the data with the English voice data; unit training device for data according to the mapping between the English and the English voice data, create or update data with the corresponding English voice English voice unit.

.

17 in accordance with the model training apparatus according to claim 16, wherein said means for training means: - data according to the mapping relationship between the English and the English voice data, using interpolation method, create or update the English voice unit.

.

18 model training device according to claim 16 or 17, wherein the apparatus further includes means updating means for: - according to the English,

and the English voice data and voice data corresponding pronunciation information, updating the English voice unit.

.

19 model training apparatus according to claim 12 to 18, wherein any one, wherein the apparatus further comprises: a sequence acquisition means for acquiring the user input for the target application word speech sequence; sequence determination means for the English voice model to determine the sequence of words corresponding to the English voice character sequence; information providing means for providing the English character sequence corresponding information to the user.

.

20 in accordance with the model training apparatus according to claim 19, wherein the target application is speech input application; wherein said information providing means for: - the English characters Voice input sequence as a candidate entry provided to the user.

21 in accordance with the model training apparatus according to claim 19, wherein the target application is a voice search application; wherein, the apparatus further comprising: acquisition means searches for the English character sequence as the query sequence to the user, access to the query sequence corresponding to one or more search results; wherein the information providing means for: - the one or more search results to the said user.

.

22 model training apparatus according to claim 19 to 21 in any one of, wherein said sequence acquisition means for: - obtaining the target application user inputs the letter Voice sequence; wherein the means for determining the sequence: - According to the English voice model to determine the sequence of letters corresponding English voice character sequence.

Description:

一种基于中文语音信息训练英文语音模型的方法与设备

A method and apparatus based Chinese English speech training speech information model

技术领域

[0001] 本发明涉及自然语言处理技术领域，尤其涉及一种用于基于中文语音信息训练英文语音模型的技术。

背景技术

[0002] 现有的训练英文语音模型的方式一般利用现有的英语语音数据库，包括语音数据及其标注，在发音单元集合中加入英文发音单元，利用英语数据进行训练。现有技术的问题在于，采用该种方式训练英文语音模型需要一定规模的英语语音数据，且难以对加入模型中的英语词的规模进行控制。

[0003] 因此，如何克服上述困难，在控制规模的同时，实现对英文语音模型的训练，成为本领域技术人员亟需解决的问题之一。

发明内容

[0004] 本发明的目的是提供一种用于基于中文语音信息训练英文语音模型的方法与设备。

[0005] 根据本发明的一个方面，提供了一种用于基于中文语音信息训练英文语音模型的方法，其中，该方法包括以下步骤:

[0006] a获取用户通过语音采集装置输入的语音数据；

[0007] b判断所述语音数据是否为英文语音数据；

[0008] c当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息；

[0009] d根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射。

[0010] 根据本发明的另一方面，还提供了一种用于基于中文语音信息训练英文语音模型的模型训练设备，其中，该设备包括:

[0011] 语音获取装置，用于获取用户通过语音采集装置输入的语音数据；

[0012] 英文判断装置，用于判断所述语音数据是否为英文语音数据；

FIELD

Invention relates to the field of natural language processing technology, in particular, relates to a voice-based Chinese information technology training English voice model

[0001] The present.

BACKGROUND

[0002] conventional way English voice model training in general use existing English speech database, including data and voice annotation, join English pronunciation unit in pronunciation unit set, use of English training data. Problems of the prior art is that, using this way of training the English voice model requires a certain amount of English voice data, and it is difficult to join model to control the size of the English word.

[0003] Therefore, how to overcome these difficulties, at the same time control the size, to achieve the English speech model training, become one skilled in the need to solve.

Summary

[0004] The present invention is to provide a method and apparatus for Chinese-based voice information training English voice model.

[0005] In accordance with one aspect of the invention, there is provided a voice message for Chinese training English voice model-based approach, wherein the method comprises the following steps:

[0006] a voice to obtain user data acquisition device through voice input;

[0007] b determines whether the data is the voice English voice data;

[0008] c when the speech data is English voice data, according to Chinese voice message, the English voice data to voice recognition processing, in order to obtain the English voice data corresponding pronunciation information;

[0009] d according to the pronunciation information, determine the English voice data corresponding English data, in order to establish or update the English voice model, wherein the

[0013] 语音识别装置，用于当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息；

[0014] 模型训练装置，用于根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射。

[0015] 与现有技术相比，本发明在原有中文声学模型基础上，利用相似的中文语音信息，对英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的发音信息，实现对英文语音模型的建立或更新，有效地控制了英文语音模型的规模；进一步地，本发明还可与语音识别相结合，支持用户以按字母及按单词两种方式输入英文语音序列，提升了用户的使用体验。

附图说明

[0016] 通过阅读参照以下附图所作的对非限制性实施例所作的详细描述，本发明的其它特征、目的和优点将会变得更明显:

[0017] 图1示出根据本发明一个方面的用于基于中文语音信息训练英文语音模型的设备不意图；

[0018] 图2示出根据本发明一个优选实施例的用于基于中文语音信息训练英文语音模型的设备示意图；

[0019] 图3示出根据本发明另一个方面的用于基于中文语音信息训练英文语音模型的方法流程图；

[0020] 图4示出根据本发明一个优选实施例的用于基于中文语音信息训练英文语音模型的方法流程图。

[0021] 附图中相同或相似的附图标记代表相同或相似的部件。

具体实施方式

[0022] 下面结合附图对本发明作进一步详细描述。

[0023] 图1示出根据本发明一个方面的用于基于中文语音信息训练英文语音模型的设备示意图。模型训练设备1包括语音获取装置101、英文判断装置102、语音识别装置103和模型训练装置104。

[0024] 其中，语音获取装置101获取用户通过语音采集装置输入的语音数据。具体地，用户或大量用

pronunciation information with the English voice information data by the Chinese mapped.

[0010] According to another aspect of the present invention, there is provided a method for Chinese speech information in English speech model training model based training device, wherein the device comprises:

[0011] speech acquisition means for acquiring user voice data acquisition device through voice input;

[0012] English judging means for judging whether or not the voice data is voice data in English;

[0013] speech recognition means for, when the voice data is voice data in English, according to Chinese voice message, the English voice data to voice recognition processing, in order to obtain data corresponding to the English voice pronunciation information;

[0014] model training apparatus, according to the pronunciation information for determining the English voice data corresponding English data, to establish or update the English speech model, wherein the pronunciation information by the English data The Chinese voice information is mapped.

[0015] Compared with the prior art, the invention of the original Chinese acoustic model based on the similarity of Chinese voice information, data on the English voice for voice recognition processing to obtain data corresponding to the English voice pronunciation information to achieve the establishment or update English voice model, effectively control the size of the English voice model; further, the present invention can also be combined with voice recognition that allows users to press two ways to enter the English word speech sequence alphabetical and to enhance the user experience.

Brief Description

[0016] by reading the following detailed description taken with the drawings made of non-limiting example with reference to other features, objects and advantages of the present invention will become more apparent:

[0017] FIG. 1 shows a device-based Chinese English voice model trained voice information is

Petitioner Exhibit 1004, Page 780 of 1518

户通过与语音采集装置的交互，输入了语音数据，语音获取装置101通过调用该语音采集装置所提供的应用程序接口(API)，获取该用户或大量用户输入的语音数据；或者，用户或大量用户通过与语音采集装置的交互，输入了语音数据，并将该语音数据存入语音数据库中，语音获取装置101通过调用该语音数据库所提供的应用程序接口(API)，从该语音数据库中获取用户或大量用户输入的语音数据。

[0025] 本领域技术人员应能理解上述获取语音数据的方式仅为举例，其他现有的或今后可能出现的获取语音数据的方式如可适用于本发明，也应包含在本发明保护范围以内，并在此以引用方式包含于此。

[0026] 英文判断装置102判断所述语音数据是否为英文语音数据。具体地，英文判断装置102根据语音获取装置101所获取的用户通过语音采集装置输入的语音数据，例如通过声学模型匹配等方式，判断该语音数据是否为英文语音数据。

[0027] 本领域技术人员应能理解上述判断英文语音数据的方式仅为举例，其他现有的或今后可能出现的判断英文语音数据的方式如可适用于本发明，也应包含在本发明保护范围以内，并在此以引用方式包含于此。

[0028] 语音识别装置103当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息。具体地，当英文判断装置102判断该语音数据为英文语音数据，语音识别装置103根据中文语音信息，如中文声韵母信息，对该英文语音数据进行语音识别处理，例如，当英文判断装置判断该语音数据为英文语音数据时，语音识别装置103采用所有可能的中文声韵母信息去识别该英文语音数据，以确定该英文语音数据的发音信息。

[0029] 例如，用户通过与语音采集装置的交互，输入了语音数据单词“iphone”，语音获取装置101通过调用该语音采集装置所提供的应用程序接口(API)，获取了该语音数据“iphone”；随后，英文判断装置102采用声学模型匹配的方式，判断该语音数据为英文语音数据；随后，语音识别装置103采用所有可能的中文声韵母信息，对该英文语音数据“iphone”进行强制对齐，确定该英文语音数据“iphone”的发音信息为“aa ai4 f engl”。在此，aa、ai4、f、engl等即为中文声韵母信息，其中的数字表示该韵母信息所对应的声调。本领域技术人员应能理解，在此，中文声韵母信息及声调的表示仅为示例，而不应对本发明产生任何限制。

not intended that the present invention according to one aspect;

[0018] Figure 2 illustrates an example of a preferred embodiment of the present invention for the device schematic Chinese English speech voice information model-based training;

[0019] Figure 3 shows a flowchart of a method of training Chinese English speech voice information model-based aspect of the present invention, according to another;

[0020] Figure 4 shows an example of a flowchart of a method for training Chinese English speech voice information model-based accordance with a preferred embodiment of the present invention.

[0021] In the drawings the same or similar reference numerals represent the same or similar parts.

DETAILED DESCRIPTION

[0022] below in conjunction with the accompanying drawings of the present invention will be further described in detail.

[0023] FIG. 1 shows a schematic diagram of the device based on the Chinese model of voice information English voice training with an aspect of the present invention is used in accordance with. Model training device 1 comprises a voice acquisition means 101, judging means 102 in English, the speech recognition apparatus 103 and the model training device 104.

[0024] wherein the voice acquisition means 101 acquires voice data acquisition device user via voice input. Specifically, the user or a large number of users to interact with the voice collecting device, the input voice data, voice acquisition means 101 by calling the voice collection application interface means provided (API), to obtain the user's voice input or a large number of user data; Alternatively, the user or a large number of users to interact with the voice collecting device, the input voice data and voice data stored in the voice database, the speech acquires the application program interface (API) means 101 by calling the speech provided by the database, Get voice data user or a large number of user input from the speech database.

[0030] 优选地，语音识别装置103还可以根据单词模型所对应的中文语音信息、或字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的单词发音信息、或字母发音信息。例如，对于“iphone”，若用户输入语音采集装置的为按单词读的“iphone”，则语音识别装置103确定其对应的采用中文语音信息表示的发音信息为“aa ai4 f engl”;若用户输入语音采集装置的为按字母读的“iphone”，即，“1、p、h、o、n、e”，则语音识别装置103确定其对应的采用中文语音信息表示的发音信息为“aa ai4 p i4ei2 ch oo oul ee enl ii i4，，。

[0031] 本领域技术人员应能理解上述识别发音信息的方式仅为举例，其他现有的或今后可能出现的识别发音信息的方式如可适用于本发明，也应包含在本发明保护范围以内，并在此以引用方式包含于此。

[0032] 模型训练装置104根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射。具体地，模型训练装置104根据该语音识别装置103所识别获得的发音信息，确定该英文语音数据所对应的英文数据，如根据发音信息“aa ai4 f engl”，确定该发音信息对应的英文数据为“iphone”，再将该发音信息与该英文数据的映射关系存入英文语音模型，以实现对该英文语音模型的建立或更新，其中，该发音信息与该英文数据通过该中文语音信息相映射。

[0033] 在此，英文数据为一具体的英文单词，如“iphone”等，而英文语音数据为该具体的英文单词所对应的语音数据，如“iphone”所对应的语音数据，通过语音获取装置101所获取。

[0034] 本领域技术人员应能理解上述建立或更新英文语音模型的方式仅为举例，其他现有的或今后可能出现的建立或更新英文语音模型的方式如可适用于本发明，也应包含在本发明保护范围以内，并在此以引用方式包含于此。

[0035] 在此，本发明在原有中文声学模型基础上，利用相似的中文语音信息，对英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的发音信息，实现对英文语音模型的建立或更新，有效地控制了英文语音模型的规模。

[0036] 优选地，模型训练设备1的各个装置之间是持续不断工作的。具体地，语音获取装置101获取用户通过语音采集装置输入的语音数据；英文判断装置102判断所述语音数据是否为英文语音数据；

[0025] skilled in the art should be able to understand the way the acquired speech data is only an example, other existing or future way to obtain voice data that may occur as applicable to the present invention, it should be included in the present invention Within the scope of protection, and in this way is incorporated herein by reference.

[0026] English judging means 102 judges whether the data is the voice English voice data. In particular, the English voice collecting judging means 102 by means of input speech data according to the voice acquisition unit 101 acquired the user, for example, by matching acoustic models, etc., it is determined whether or not the voice data is voice data in English.

[0027] skilled in the art should be able to understand the judgment of the English way of example only voice data, other available data to determine the English voice mode or may occur in the future as applicable to the present invention, it should also be included Within the scope of the present invention and hereby incorporated herein by reference.

[0028] Speech recognition device 103 when the speech data is English voice data, according to Chinese voice message, the English voice data to voice recognition processing, in order to obtain the English voice data corresponding pronunciation information. Specifically, when the English judgment means 102 judges that the voice data is English voice data, the voice recognition device 103 according to Chinese voice information, such as Chinese vowel sound information, the English voice data for speech recognition processing, for example, when the judgment means judges that the English When the voice data into English voice data, the voice recognition device 103 uses all possible Chinese vowel sound information to identify the English voice data, voice data to determine if the English pronunciation information.

[0029] For example, the user interacts with the voice collecting device, voice data input word "iphone", a voice interface device 101 to get the application by calling the voice collecting apparatus is provided (API), to obtain the voice Data "iphone"; then judging means 102 English acoustic models using matching methods to determine if the voice data is the English voice data; subsequently, the voice recognition device

语音识别装置103当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息；模型训练装置104根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射。在此，本领域技术人员应理解“持续”是指模型训练设备1的各装置分别不断地进行语音数据的获取、英文语音数据的判断、英文语音数据的识别处理及英文语音模型的建立或更新，直至该模型训练设备1在较长时间内停止获取语音数据。

[0037] 优选地，语音识别装置103当所述语音数据为英文语音数据，根据单词模型所对应的中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的单词发音信息。具体地，当该英文判断装置102判断语音获取装置101所获取的语音数据为英文语音数据，语音识别装置103采用单词模型所对应的中文语音信息，如“aaai4 f engl”等，对所述英文语音数据进行语音识别处理，如强制对齐处理，采用所有可能的中文声韵母信息，对该英文语音数据进行强制对齐，以获得与所述英文语音数据相对应的单词发音信息。

[0038] 在此，该单词模型中存储有大量单词所对应的中文语音信息，该单词模型可以位于模型训练设备1中，或者位于与该模型训练设备1通过网络相连接的第三方设备中。

[0039] 优选地，语音识别装置103当所述语音数据为英文语音数据，根据字母模型所对应的中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的字母发音信息。具体地，当该英文判断装置102判断语音获取装置101所获取的语音数据为英文语音数据，语音识别装置103根据字母模型所对应的中文语音信息，如二十六个英文字母所对应的中文语音信息，如字母“a”对应的中文语音信息“aa eil”、字母“b”对应的中文语音信息“b i4”、字母“c”对应的中文语音信息“s ei4”等，对该英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的字母发音信息。

[0040] 优选地，语音识别装置103对于英文判断装置102判断为英文语音数据为语音数据，优先采用单词模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，当该语音识别装置103采用单词模型无法识别出该英文语音数据的发音信息时，再采用字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的字母发音信息。相似地，该语音识别装置103也可以优先采用字母模型所对应的中

103 uses all possible Chinese vowel sound information, the English voice data "iphone" forced alignment, determine the English voice data "iphone" pronunciation information "aa ai4 f engl". Here, aa, ai4, f, engl and other information that is Chinese vowel sound, which indicates that the vowel digital information corresponding to the tone. The skilled in the art will appreciate that in this, Chinese vowel sound and tonal representation of information are examples only, and not to deal with the present invention, any restrictions.

[0030] Preferably, the speech recognition device 103 may be based on word models corresponding to speech information in Chinese, or Chinese letter model corresponding to the voice message, the English voice data for voice recognition processing, in order to obtain the English voice data corresponding word pronunciation information, or to pronounce the letter information. For example, for "iphone", if the user enters a voice collection device for "iphone", the words read by a voice recognition device 103 to determine the corresponding voice information indicating the use of Chinese pronunciation information "aa ai4 f engl"; if the user input voice collecting means is read by the letter "iphone", i.e., "1, p, h, ?, n, e", the speech recognition apparatus 103 determines the corresponding voice information expressed in Chinese pronunciation information "aa ai4 p i4ei2 ch oo oul ee enl ii i4,,.

[0031] skilled in the art should be able to understand the way information is the identification pronunciation of example only, other available information the way for future identification pronunciation or may occur as applicable to the present invention, it should be included in the present invention Within the scope of protection, and in this way is incorporated herein by reference.

[0032] Model training apparatus 104 according to the pronunciation information, determine the English voice data corresponding English data, to establish or update the English speech model, wherein the pronunciation information of the data by the English Chinese voice information is mapped. Specifically, the model is based on 103 training device 104 recognizes the voice recognition unit pronunciation information, determined that the English voice data corresponding English data, such as "aa ai4 f

文语音信息，对该英文语音数据进行语音识别处理，当该语音识别装置103采用字母模型无法识别出该英文语音数据的发音信息时，再采用单词模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的单词发音信息。在此，优先采用单词模型还是优先采用字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，可以是该模型训练设备1预先设定的，也可以应事件而进行调整。

[0041] 例如，对于语音获取装置101所获取的语音数据“iphone”，英文判断装置102判断该语音数据“iphone”为英文语音数据，语音识别装置103优先采用单词模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，但无法识别出该英文语音数据的发音信息；随后，该语音获取装置101再采用字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，获得与该英文语音数据相对应的字母发音信息“aa ai4 p i4 ei2ch oo oul ee enl ii i4”，则可知该语音获取装置101所获取的语音数据为按字母读的“iphone”，即，“1、p、h、o、η、e”。

[0042] 更优选地，语音识别装置103分别采用单词模型所对应的中文语音信息、字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，并分别对语音识别处理所获得的结果进行打分，确定打分较高的发音信息为该英文语音数据的发音信息。

[0043] 在此，由于用户在输入英文单词时有时按单词读有时按字母读，本发明对每个需要识别的英文单词，利用相似的单词模型对应的中文语音信息、字母模型对应的中文语音信息，确定其对应的发音信息，提升了语音识别的准确性。

[0044] 在一优选实施例中，该模型训练设备1还包括模型更新装置(未示出)。该模型更新装置统计所述英文数据所对应的英文语音数据在所述语音数据中的出现次数；根据所述出现次数，并结合所述语音数据的数量信息，确定所述英文数据的概率值；根据所述概率值，更新所述英文语音模型。具体地，当大量用户通过语音采集装置输入语音数据，其中，多个用户所输入的英文语音数据为相同的英文语音数据，语音获取装置101获取该大量用户输入的大量语音数据；英文判断装置102判断该大量语音数据为英文语音数据；随后，语音识别装置103根据中文语音信息，对该大量英文语音数据进行语音识别处理，以获得与该大量英文语音数据相对应的发音信息，在此，由于该多个用户输入的为相同的英文语音数据，因此，该语音识别装置103所获得的关于该多个用户输入的英文语音数据所对

engl", determining the pronunciation information corresponding English data based on pronunciation information the "iphone", then the mapping of the English pronunciation information and data stored in the English speech model, in order to achieve the establishment or update English voice model, wherein the information associated with the English pronunciation of the Chinese voice information data by phase mappings.

[0033] In this case, the English data is a particular English word, such as "iphone" and so on, but the English voice data for specific English word corresponding to the voice data, such as the "iphone" corresponding to the voice data, The acquisition means 101 acquire voice.

[0034] skilled in the art should be understood that the above to create or update the way of example only English voice model, or other existing way to create or update the English voice model may arise in the future as applicable to the present invention, It should also be included within the scope of the present invention and hereby incorporated herein by reference.

[0035] In this case, the present invention is in the original Chinese acoustic model based on the similarity of Chinese voice information, data on the English voice for voice recognition processing to obtain the English voice data corresponding pronunciation information, to achieve create or update English voice model, effectively control the size of the English voice model.

[0036] Preferably, between each device model training apparatus 1 is continued work. Specifically, the voice acquisition means 101 acquires the user through voice data acquisition device voice input; if the English judgment means 102 judges that the voice data is English voice data; a voice recognition device 103 when the speech data is English voice data, according to Chinese Speech information on the English voice data for voice recognition processing to obtain the data corresponding to the English voice pronunciation information; model training apparatus 104 according to the pronunciation information, determine the English voice data corresponding English data, to English create or update a speech model, wherein the pronunciation information and data through the English voice information is mapped to the Chinese. In this case,

应的发音信息相同或相似；接着，模型训练装置104根据该多个相同或相似的发音信息，确定该多个英文语音数据所对应的相同的英文数据；模型更新装置统计该相同的英文数据所对应的英文语音数据在该语音数据中的出现次数；接着，该模型更新装置再根据该英文语音数据的出现次数，并结合该语音数据的数量信息，确定该英文数据的概率值，如将该出现次数除以该大量语音数据的数量信息，获得该英文数据的概率值；该模型更新装置再将该英文数据及其对应的概率值存入该英文语音模型中，以实现对该英文语音模型的更新。

[0045] 优选地，该模型训练设备1还包括数据提取装置(未示出)和单元训练装置(未示出)。该数据提取装置当所述出现次数大于预定数量阈值，自所述英文语音模型中提取所述英文数据与所述英文语音数据；该单元训练装置根据所述英文数据与所述英文语音数据的映射关系，建立或更新与所述英文语音数据相对应的英文语音单元。具体地，该模型训练设备1可以设置一个预定数量阈值，该预定数量阈值用来判断英文数据所对应的英文语音数据在该语音数据中的出现次数是否超过该值，当该出现次数大于该预定数量阈值时，数据提取装置自该英文语音模型中提取该英文数据与该英文语音数据；该单元训练装置根据该英文数据与该英文语音数据的映射关系，为该英文语音数据新建一个英文语音单元，以实现对该英文语音单元的建立或更新。

[0046] 更优选地，该单元训练装置根据所述英文数据与所述英文语音数据的映射关系，采用插值计算方法，建立或更新所述英文语音单元。具体地，该单元训练装置根据该英文数据与该英文语音数据的映射关系，为该英文数据新建一英文语音单元，再例如通过将该中文发音信息所对应的中文语音单元的模型参数复制给该英文语音单元，然后在英文数据上对齐，得到统计量，再与该复制所获得的参数进行插值计算，以实现对该英文语音单元的建立或更新。在此，所述中文语音单元包括为每个中文发音信息，如中文声母或韵母信息所建立的发音单元，如“s”、“ei4”等中文声韵母信息所对应的中文语音单元。

[0047] 更优选地，该模型训练设备1还包括单元更新装置(未示出)。该单元更新装置根据所述英文语音数据及与该英文语音数据相对应的发音信息，更新所述英文语音单元。具体地，在单元训练装置为英文数据新建一英文语音单元之后，单元更新装置不断地根据用户输入的英文语音数据，及由语音识别装置103识别所获得的与该英文语音数据相对应的发音信息，训练该英文语音单元。

the skilled artisan will appreciate that "persistent" means each device model training equipment are constantly acquire voice data and establish judgment English voice data recognition processing and English voice English voice data model or updating until the model training device 1 in a long time I stop getting voice data.

[0037] Preferably, the speech recognition device 103 when the speech data is English voice data, according to the word model corresponding to the Chinese voice message, the English voice data to voice recognition processing to obtain the English and speech data corresponding word pronunciation information. Specifically, when the English voice judging means 102 determines device 101 acquired by the speech data acquired for the English voice data, the voice recognition device 103 using the word model corresponding to the Chinese voice message, such as "aaai4 f engl", etc., on the English voice data for voice recognition processing, such as forced alignment treatment, using all possible Chinese vowel sound information, the mandatory English speech data aligned to obtain the corresponding English voice data word pronunciation information.

[0038] In this case, the word model stored in a large number of words corresponding to the Chinese voice message, the word model in the model training device 1, or in a third party through a network connected with the model training device 1 devices.

[0039] Preferably, the speech recognition device 103 when the speech data is English voice data, based on the letter corresponding to the Chinese model of voice information, data on the English voice for voice recognition processing to obtain the English and Voice data information corresponding letter sounds. Specifically, when the English voice judging means 102 determines device 101 acquired by the speech data acquired for the English voice data, the voice recognition device 103 according to the letter corresponding to the Chinese model of voice information, such as twenty-six the letters corresponding to the Chinese pronunciation information, such as the letter "a" corresponding to the Chinese voice message "aa eil", the letter "b" corresponding Chinese voice message "b i4", the letter "c" corresponding to the Chinese voice messages "s ei4" and so on, the English voice Data for voice recognition

[0048] 在此，本发明对出现此处大于预定阈值的英文语音数据，为该英文语音数据及英文数据新建英文语音单元，并对该英文语音单元不断地进行训练，以进一步提升该英文语音数据的发音信息的准确性。

[0049] 图2示出根据本发明一个优选实施例的用于基于中文语音信息训练英文语音模型的设备示意图。该模型训练设备1还包括序列获取装置205、序列确定装置206和信息提供装置207。以下参照图2对该优选实施例进行详细描述:具体地，语音获取装置201获取用户通过语音采集装置输入的语音数据；英文判断装置202判断所述语音数据是否为英文语音数据；语音识别装置203当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息；模型训练装置204根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射；序列获取装置205获取所述用户为目标应用输入的单词语音序列；序列确定装置206根据所述英文语音模型，确定与所述单词语音序列相对应的英文字符序列；信息提供装置207将与所述英文字符序列相对应的信息提供给所述用户。其中，语音获取装置201、英文判断装置202、语音识别装置203和模型训练装置204与图1所示对应装置相同或相似，故此处不再赘述，并通过引用的方式包含于此。

[0050] 其中，序列获取装置205获取所述用户为目标应用输入的单词语音序列。具体地，用户例如通过与用户设备的语音采集装置的交互，在目标应用中输入单词语音序列，序列获取装置205例如通过调用该用户设备所提供的应用程序接口(API)等方式，获取该用户为目标应用所输入的单词语音序列。

[0051] 序列确定装置206根据所述英文语音模型，确定与所述单词语音序列相对应的英文字符序列。具体地，序列确定装置206根据该序列获取装置205所获取的单词语音序列，通过语音识别处理，确定该单词语音序列的发音信息，进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该单词语音序列相对应的英文字符序列。例如，用户通过与用户设备的语音采集装置的交互，输入了单词语音序列“iphone”，序列获取装置205例如通过调用该用户设备所提供的应用程序接口(API)等方式，获取了该用户输入的单词语音序列“iphone”；随后，序列确定装置206根据该单词语音序列“iphone”，通过语音识别处理，确定该单词语音序列的发音信息“aa ai4 f engl”，进而，根据该发音

processing, in order to obtain the English voice data information corresponding letter sounds.

[0040] Preferably, the speech recognition apparatus 103 for English English determining means 102 determines that the voice data is voice data, priority to Chinese word models corresponding to speech information, the English voice data for voice recognition processing, when the the speech recognition apparatus using the word model 103 can not recognize the English pronunciation information of voice data, and then using the letter model corresponding to the speech information in Chinese, English speech data to the speech recognition process, to obtain the corresponding English voice data pronunciation information letters. Similarly, the speech recognition device 103 may be preferably used in Chinese letter model corresponding to the voice information, the English voice data for voice recognition processing, when the use of the voice recognition device 103 recognizes that the letters can not model speech data in English pronunciation information When, then using word models corresponding to speech information in Chinese, English speech data to the speech recognition process, to obtain the corresponding English voice data word pronunciation information. Here, the word models using priority or precedence letter model corresponding to the speech information in Chinese, English speech data to the speech recognition processing, the model may be set in advance training device 1 may also be adjusted to be an event.

[0041] For example, the voice data acquired by the speech device 101 acquired "iphone", the English judgment means 102 judges that the speech data "iphone" for the English voice data, the voice recognition device 103 preferentially uses the word model corresponding Chinese voice messages, the English voice data for speech recognition processing, but does not recognize the English pronunciation of voice data information; Subsequently, the voice acquisition means 101 and then using the letter corresponding to the Chinese model of voice information, the English voice data for voice recognition processing, voice data obtained with the English pronunciation of letters corresponding to information "aa ai4 p i4 ei2ch oo oul ee enl ii i4", it is understood that the speech data acquired by the speech device 101 is acquired by letter read "iphone", That is, "1, p, h, ?, eta, e".

信息，在该英文语音模型中进行匹配查询，确定与该单词语音序列相对应的英文字符序列“iphone”。

[0052] 信息提供装置207将与所述英文字符序列相对应的信息提供给所述用户。例如，当该目标应用为语音输入应用，该信息提供装置207通过调用诸如JSP、ASP或PHP等动态网页技术，将该英文字符序列作为语音输入候选词条提供给该用户；当该目标应用为语音搜索应用；该信息提供装置207通过调用诸如JSP、ASP或PHP等动态网页技术，将根据该英文字符序列搜索获得的一个或多个搜索结果提供给该用户。

[0053] 优选地，该目标应用为语音输入应用；其中，该信息提供装置207将所述英文字符序列作为语音输入候选词条提供给所述用户。例如，当该目标应用为语音输入应用，用户通过与用户设备的语音采集装置的交互，在语音输入应用中输入了单词语音序列“iphone”，序列获取装置205例如通过调用该用户设备所提供的应用程序接口(API)等方式，获取了该用户输入的单词语音序列“iphone”；随后，序列确定装置206根据该单词语音序列“iphone”，通过语音识别处理，确定该单词语音序列的发音信息“aa ai4 f engl”，进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该单词语音序列相对应的英文字符序列“iphone”;进而，信息提供装置207通过调用诸如JSP、ASP或PHP等动态网页技术，将该英文字符序列“iphone”作为语音输入候选词条提供给该用户，以供该用户选择输入。

[0054] 优选地，该目标应用为语音搜索应用；其中，该模型训练设备1还包括搜索获取装置(未示出)。该搜索获取装置将所述英文字符序列作为所述用户的查询序列，获取与所述查询序列相对应的一个或多个搜索结果；其中，该信息提供装置207将所述一个或多个搜索结果提供给所述用户。具体地，当该目标应用为语音搜索应用，序列确定装置206根据该英文语音模型，确定与该单词语音序列相对应的英文字符序列；搜索获取装置根据该序列确定装置206所确定的英文字符序列，将该英文字符序列作为该用户的查询序列，进行匹配查询，获取与该查询序列相对应的一个或多个搜索结果；随后，信息提供装置207通过调用诸如JSP、ASP或PHP等动态网页技术，将该一个或多个搜索结果提供给该用户。

[0055] 例如，当该目标应用为语音搜索应用，用户通过与用户设备的语音采集装置的交互，在语音输入应用中输入了单词语音序列“iphone”，序列获取装置205例如通过调用该用户设备所提供的应用程序接口(API)等方式，获取了该用户输入的单词语音序列“iphone”；随后，序列确定装置206根据该单词

[0042] More preferably, the speech recognition apparatus 103 word models respectively corresponding to the voice information Chinese, Chinese letter model corresponding to the voice message, the English voice data for voice recognition processing, voice recognition processing for respectively The results obtained for scoring, scoring higher pronunciation information to determine that the English pronunciation of voice data information.

[0043] In this case, since the user when entering the English word sometimes word by word and by letter read read sometimes, the present invention requires the identification of each English word, use a similar word model corresponding Chinese voice messages, letters, correspondence model The Chinese voice information to determine its corresponding pronunciation information, improve the accuracy of speech recognition.

[0044] In a preferred embodiment, the apparatus further includes a model training model updating means (not shown). The statistical model updating means of said data corresponding to the English English number of occurrences of the voice data in the speech data; according to the number of occurrences, and the amount of data combined with the voice message, the probability of determining the data value of the English; based on the probability value, updating the English voice model. In particular, when a large number of user input via voice speech data acquisition device, wherein a plurality of the English voice data input by the user to the same English voice data, the voice acquisition means 101 acquires the number of users of a large number of input voice data; English judging means 102 Analyzing the bulk voice data into English voice data; then, a voice recognition device 103 according to Chinese voice message, the voice data of a large number of English speech recognition processing, in order to obtain the large number of English speech data corresponding pronunciation information, in this case, since The plurality of user inputs to the same English voice data, therefore, the English voice data on the plurality of user input of the voice recognition device 103 to obtain information corresponding to the same or similar pronunciation; then, based on the model training device 104 a plurality of the same or similar pronunciation information, identify same English voice data of the plurality of data corresponding English; model updating means statistical data of

语音序列“iphone”，通过语音识别处理，确定该单词语音序列的发音信息“aa ai4 f engl”，进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该单词语音序列相对应的英文字符序列“iphone”;进而，搜索获取装置根据该序列确定装置206所确定的英文字符序列，将该英文字符序列“iphone”作为该用户的查询序列，进行匹配查询，获取与该查询序列相对应的一个或多个搜索结果；随后，信息提供装置207通过调用诸如JSP、ASP或PHP等动态网页技术，将与该英文字符序列“iphone”相对应的一个或多个搜索结果提供给该用户，供该用户浏览。

[0056] 更优选地，序列获取装置205获取所述用户为目标应用输入的字母语音序列；其中，所述序列确定装置206根据所述英文语音模型，确定与所述字母语音序列相对应的英文字符序列。具体地，用户例如通过与用户设备的语音采集装置的交互，在目标应用中输入字母语音序列，序列获取装置205例如通过调用该用户设备所提供的应用程序接口(API)等方式，获取该用户为目标应用所输入的字母语音序列；随后，序列确定装置206根据该序列获取装置205所获取的字母语音序列，通过语音识别处理，确定该字母语音序列的发音信息，进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该字母语音序列相对应的英文字符序列。例如，用户通过与用户设备的语音采集装置的交互，输入了字母语音序列“1、p、h、o、n、e”，序列获取装置205例如通过调用该用户设备所提供的应用程序接口(API)等方式，获取了该用户输入的字母语音序列“1、p、h、o、n、e”;随后，序列确定装置206根据该字母语音序列“1、p、h、o、n、e”，通过语音识别处理，确定该字母语音序列的发音信息“aa ai4 p i4 ei2 ch oo oul ee enl ii i4”,进而，根据该发音信息,在该英文语音模型中进行匹配查询，确定与该字母语音序列相对应的英文字符序列“iphone”。

[0057] 相似地，信息提供装置207将与该英文字符序列相对应的信息提供给所述用户。接上例，当该目标应用为语音输入应用，该信息提供装置207通过调用诸如JSP、ASP或PHP等动态网页技术，将该英文字符序列“iphone”作为语音输入候选词条提供给该用户，以供该用户选择输入；当该目标应用为语音搜索应用；搜索获取装置根据该序列确定装置206所确定的英文字符序列，将该英文字符序列“iphone”作为该用户的查询序列，进行匹配查询，获取与该查询序列相对应的一个或多个搜索结果；随后，信息提供装置207通过调用诸如JSP、ASP或PHP等动态网页技术，将与该英文字符序

the same English voice data corresponding to the number of occurrences of English in the speech data; then, The model updating means then according to the number of occurrences of the English voice data, combined with the amount of information of the voice data, to determine the probability that the value of the English data, such as the emergence of a large number of voice data divided by the amount of information, access to the English data The probability value; the model data updating means then the English and the corresponding probability values are stored in the English speech model in order to achieve the English voice model update.

[0045] Preferably, the apparatus 1 further comprises a model training data extracting means (not shown) and the unit training device (not shown). The data extracting means when the number of occurrences is greater than a predetermined threshold amount, the English data and extract the data from the English voice English voice model; the unit training apparatus according to the English and the English voice data mapping data relationship, create or update data with the corresponding English voice English voice unit. Specifically, the model training device 1 may set a predetermined threshold amount, the predetermined threshold amount used to determine the English data corresponding English voice data the number of occurrences in the speech data exceeds this value, when the number of occurrences is greater than the predetermined When the threshold number, data extraction means extracts the data associated with the English voice English voice data from the English model; the unit training device mapping relation to the English voice data based on the English data, create a new unit for the English voice English voice data in order to achieve the English phonetic units built or updated.

[0046] More preferably, the unit training device data according to the mapping relation between the English and the English voice data, using interpolation method, create or update the English voice unit. Specifically, the unit training device mappings between data based on the English, and the English voice data, a new English voice unit data for the English, then by the model parameters such as Chinese pronunciation information corresponding Chinese phonetic units copied to the English speech unit, and then on the English data alignment, obtain statistics, and then with the

列“iphone”相对应的一个或多个搜索结果提供给该用户，供该用户浏览。

[0058] 在此，本发明还可与语音识别相结合，支持用户以按字母及按单词两种方式输入英文语音序列，提升了用户的使用体验。

[0059] 图3示出根据本发明另一个方面的用于基于中文语音信息训练英文语音模型的方法流程图。

[0060] 在步骤S301中，模型训练设备1获取用户通过语音采集装置输入的语音数据。具体地，用户或大量用户通过与语音采集装置的交互，输入了语音数据，在步骤S301中，模型训练设备1通过调用该语音采集装置所提供的应用程序接口(API)，获取该用户或大量用户输入的语音数据；或者，用户或大量用户通过与语音采集装置的交互，输入了语音数据，并将该语音数据存入语音数据库中，在步骤S301中，模型训练设备1通过调用该语音数据库所提供的应用程序接口(API)，从该语音数据库中获取用户或大量用户输入的语音数据。

[0061] 本领域技术人员应能理解上述获取语音数据的方式仅为举例，其他现有的或今后可能出现的获取语音数据的方式如可适用于本发明，也应包含在本发明保护范围以内，并在此以引用方式包含于此。

[0062] 在步骤S302中，模型训练设备1判断所述语音数据是否为英文语音数据。具体地，在步骤S302中，模型训练设备1根据在步骤S301中所获取的用户通过语音采集装置输入的语音数据，例如通过声学模型匹配等方式，判断该语音数据是否为英文语音数据。

[0063] 本领域技术人员应能理解上述判断英文语音数据的方式仅为举例，其他现有的或今后可能出现的判断英文语音数据的方式如可适用于本发明，也应包含在本发明保护范围以内，并在此以引用方式包含于此。

[0064] 在步骤S303中，模型训练设备1当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息。具体地，当在步骤S302中，模型训练设备1判断该语音数据为英文语音数据，在步骤S303中，模型训练设备1根据中文语音信息，如中文声韵母信息，对该英文语音数据进行语音识别处理，例如，当在步骤S302中，模型训练设备1判断该语音数据为英文语音数据时，在步骤S303中，模型训练设备1采用所有可能的中文

parameters of the copy obtained by interpolating calculations to achieve the establishment or update the English voice unit. Here, the Chinese pronunciation for each speech unit includes Chinese information, such as Chinese initial or final information established pronunciation unit, such as "s", "ei4" and other Chinese vowel sound information corresponding Chinese phonetic units.

[0047] More preferably, the model training apparatus 1 further comprises a unit updating means (not shown). The unit updating means according to the English voice data and voice data corresponding to the English pronunciation information, updating the English voice unit. Specifically, after the unit training device a new English voice unit data is in English, means constantly updating means according to the English voice user input data, and 103 recognized by the voice recognition device obtained with the English voice data corresponding pronunciation information training the English voice unit.

[0048] In this case, the present invention is greater than a predetermined threshold value appears here in English, voice data, voice and data for English and English New English voice unit data, and the English phonetic units constantly training to further enhance The accuracy of the English pronunciation of voice data information.

[0049] Figure 2 shows a schematic diagram of the device based on the Chinese model of voice information in accordance with English voice training example for a preferred embodiment of the present invention. The model training apparatus 1 further comprises a sequence acquisition means 205, determining means 206 and the sequence information providing apparatus 207. Referring now to Figure 2 of this preferred embodiment is described in detail: specifically, the voice acquisition means 201 acquires voice data through the voice collecting user input means; whether the speech data judging means 202 judges English English speech data; speech recognition means 203 When the voice data is voice data in English, according to Chinese voice information, data on the English voice for voice recognition processing, in order to obtain the English voice data corresponding pronunciation information; model training device 204 information based on the pronunciation, determine the English voice data corresponding English data, in order to establish or update the English voice model,

声韵母信息去识别该英文语音数据，以确定该英文语音数据的发音信息。

[0065] 例如，用户通过与语音采集装置的交互，输入了语音数据单词“iphone”，在步骤S301中，模型训练设备1通过调用该语音采集装置所提供的应用程序接口(API)，获取了该语音数据“iphone”;随后，在步骤S302中，模型训练设备1采用声学模型匹配的方式，判断该语音数据为英文语音数据；随后，在步骤S303中，模型训练设备1采用所有可能的中文声韵母信息，对该英文语音数据“iphone”进行强制对齐，确定该英文语音数据“iphone”的发音信息为“aa ai4f engl”。在此，aa、ai4、f、engl等即为中文声韵母信息,其中的数字表示该韵母信息所对应的声调。本领域技术人员应能理解，在此，中文声韵母信息及声调的表示仅为示例，而不应对本发明产生任何限制。

[0066] 优选地，在步骤S303中，模型训练设备1还可以根据单词模型所对应的中文语音信息、或字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的单词发音信息、或字母发音信息。例如，对于“iphone”，若用户输入语音采集装置的为按单词读的“iphone”，则在步骤S303中，模型训练设备1确定其对应的采用中文语音信息表示的发音信息为“aa ai4 fengl”;若用户输入语音采集装置的为按字母读的“iphone”，即，“1、p、h、o、n、e”，则在步骤S303中，模型训练设备1确定其对应的采用中文语音信息表示的发音信息为“aa ai4 pi4 ei2 ch oo oul ee enl iii4”。

[0067] 本领域技术人员应能理解上述识别发音信息的方式仅为举例，其他现有的或今后可能出现的识别发音信息的方式如可适用于本发明，也应包含在本发明保护范围以内，并在此以引用方式包含于此。

[0068] 在步骤S304中，模型训练设备1根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射。具体地，在步骤S304中，模型训练设备1根据在步骤S303中所识别获得的发音信息，确定该英文语音数据所对应的英文数据，如根据发音信息“aa ai4 fengl”，确定该发音信息对应的英文数据为“iphone”，再将该发音信息与该英文数据的映射关系存入英文语音模型，以实现对该英文语音模型的建立或更新，其中，该发音信息与该英文数据通过该中文语音信息相映射。

wherein the pronunciation information and data through the English voice information is mapped to the Chinese; sequence acquisition means 205 acquires the target users word voice input sequence applications; sequence determination means 206 according to the English voice model to determine the sequence of words corresponding to the English voice character sequence; information providing device 207 will provide the English character sequence information corresponding to the said user. Wherein, the voice acquisition means 201, judging means 202 in English, the speech recognition apparatus corresponding to 203 in FIG. 1 and FIG model training apparatus 204 same as or similar means, it is not mentioned here, and is incorporated herein by reference.

[0050] wherein the sequence acquisition means 205 acquires the user word voice input sequence of the target application. Specifically, for example, by the user to interact with the voice collecting device user device, enter the word voice in the target application, sequence acquisition means 205, for example by calling the user equipment provided by the application program interface (API), etc., to obtain the user Target applications for the word voice input sequence.

[0051] sequence determination means 206 according to the English voice model to determine the sequence of words corresponding to the English speech sequence of characters. Specifically, the sequence determining means 206 to obtain the sequence of words based on the speech unit sequence 205 acquired through voice recognition processing, determines the word sequence of speech pronunciation information, and then, based on the pronunciation information, the match query model in the English speech determining the word sequence corresponding to the English voice character sequence. For example, the user interacts with the voice collecting device user device, enter the word speech sequence "iphone", sequence acquisition means 205, for example by calling the user equipment provides an application program interface (API), etc., to get the user input word speech sequence "iphone"; then, the sequence determining means 206 according to the word speech sequence "iphone", through voice recognition processing, determines the word sequence of speech pronunciation information "aa ai4 f engl", and then, based on the pronunciation

[0069] 在此，英文数据为一具体的英文单词，如“iphone”等，而英文语音数据为该具体的英文单词所对应的语音数据，如“iphone”所对应的语音数据，通过语音获取装置101所获取。

[0070] 本领域技术人员应能理解上述建立或更新英文语音模型的方式仅为举例，其他现有的或今后可能出现的建立或更新英文语音模型的方式如可适用于本发明，也应包含在本发明保护范围以内，并在此以引用方式包含于此。

[0071] 在此，本发明在原有中文声学模型基础上，利用相似的中文语音信息，对英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的发音信息，实现对英文语音模型的建立或更新，有效地控制了英文语音模型的规模。

[0072] 优选地，模型训练设备1的各个步骤之间是持续不断工作的。具体地，在步骤S301中，模型训练设备1获取用户通过语音采集装置输入的语音数据；在步骤S302中，模型训练设备1判断所述语音数据是否为英文语音数据；在步骤S303中，模型训练设备1当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息；在步骤S304中，模型训练设备1根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射。在此，本领域技术人员应理解“持续”是指模型训练设备1的各步骤分别不断地进行语音数据的获取、英文语音数据的判断、英文语音数据的识别处理及英文语音模型的建立或更新，直至该模型训练设备1在较长时间内停止获取语音数据。

[0073] 优选地，在步骤S303中，模型训练设备1当所述语音数据为英文语音数据，根据单词模型所对应的中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的单词发音信息。具体地，当在步骤S302中，模型训练设备1判断在步骤S301中所获取的语音数据为英文语音数据，在步骤S303中，模型训练设备1采用单词模型所对应的中文语音信息，如“aa ai4 f engl”等，对所述英文语音数据进行语音识别处理，如强制对齐处理，采用所有可能的中文声韵母信息，对该英文语音数据进行强制对齐，以获得与所述英文语音数据相对应的单词发音信息。

[0074] 在此，该单词模型中存储有大量单词所对应的中文语音信息，该单词模型可以位于模型训练设备1中，~~或者位~~于与该模型训练设备1通过网络相连接的第三方设备中。

information, match queries in the English voice model, determined voice the word sequence corresponding English character sequence "iphone".

[0052] information providing device 207 will be provided to the user and the English character sequence corresponding information. For example, when the target application for voice input application, the information providing device 207 such as dynamic web technology by calling JSP, ASP or PHP, etc., the English character sequence as a candidate for entry speech input provided to the user; when the target applications voice search application; the information providing device 207 by calling the dynamic web technology such as JSP, ASP or PHP, etc., will be provided to the user in accordance with one or more of the search results for that search English character sequence obtained.

[0053] Preferably, the target applications for voice input applications; wherein the information providing device 207 as the English voice input candidate character sequence entries available to the user. For example, when the target application for voice input application, users interact with the voice of the user equipment acquisition device, enter the voice input applications, the word speech sequence "iphone", sequence acquisition means 205, for example by calling the user equipment provided Application Program Interface (API), etc., to obtain the sequence of the word voice input by the user "iphone"; then the sequence determination means 206 voice sequence information based on the word "iphone", through voice recognition processing, determine the sequence of the word voice pronunciation "aa ai4 f engl", and then, according to the pronunciation information, the English voice model to match the query to determine the sequence of words corresponding to the English voice character sequence "iphone"; In addition, the information providing device 207, such as by calling JSP, ASP or PHP and other dynamic web technology, the English character sequence "iphone" candidate for entry as a voice input provided to the user, for the user to select the input.

[0054] Preferably, the target application for the voice search application; wherein the model training apparatus 1 also includes a search acquiring device (not shown). The search acquisition means the English character sequence

[0075] 优选地，在步骤S303中，模型训练设备1当所述语音数据为英文语音数据，根据字母模型所对应的中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的字母发音信息。具体地，当在步骤S302中，模型训练设备1判断在步骤S301中所获取的语音数据为英文语音数据，在步骤S303中，模型训练设备1根据字母模型所对应的中文语音信息，如二十六个英文字母所对应的中文语音信息，如字母“a”对应的中文语音信息“aa eil”、字母“b”对应的中文语音信息“b i4”、字母“c”对应的中文语音信息“s ei4”等，对该英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的字母发音信息。

[0076] 优选地，在步骤S303中，模型训练设备1对于在步骤S302中判断为英文语音数据为语音数据，优先采用单词模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，当在步骤S303中，模型训练设备1采用单词模型无法识别出该英文语音数据的发音信息时，再采用字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的字母发音信息。相似地，在步骤S303中，模型训练设备1也可以优先采用字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，当在步骤S303中，模型训练设备1采用字母模型无法识别出该英文语音数据的发音信息时，再采用单词模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的单词发音信息。在此，优先采用单词模型还是优先采用字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，可以是该模型训练设备1预先设定的，也可以应事件而进行调整。

[0077] 例如，对于在步骤S301中，模型训练设备1所获取的语音数据“iphone”，在步骤S302中，模型训练设备1判断该语音数据“iphone”为英文语音数据，在步骤S303中，模型训练设备1优先采用单词模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，但无法识别出该英文语音数据的发音信息；随后，在步骤S301中，模型训练设备1再采用字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，获得与该英文语音数据相对应的字母发音信息“aa ai4 p i4 ei2 ch oo oul ee enl ii i4”,则可知在步骤S301中所获取的语音数据为按字母读的“iphone”，即，“1、p、h、o、n、e”。

[0078] 更优选地，在步骤S303中，模型训练设备1分别采用单词模型所对应的中文语音信息、字母模型所对应的中文语音信息，对该英文语音数据进行语音识别处理，并分别对语音识别处理所获得的结

as the user's query sequence to obtain the query sequence corresponding to one or more search results; wherein the information providing device 207 the one or more search results provided to the user. Specifically, when the target application is a voice search applications, sequence determination means 206 according to the English voice model to determine the word speech sequence corresponding English character sequences; search acquisition means 206 according to the determined English character sequence determines the sequence, the English character sequences as the query sequence for the user, to match the query, get the query sequence corresponding to one or more search results; subsequently, the information providing device 207 such as dynamic web technology by calling JSP, ASP or PHP, etc., the one or more search results to the user.

[0055] For example, when the target application for the voice search application, users interact with the voice collecting device user equipment, applications in voice input Voice input word sequence "iphone", sequence acquisition means 205, for example by calling The equipment provided by the user application program interface (API), etc., to get the word speech sequence of the user input "iphone"; then the sequence determination means 206 according to the word speech sequence "iphone", through voice recognition process, determining the voice pronounce the word sequence information "aa ai4 f engl", and then, according to the pronunciation information, in the English voice model to match the query to determine the sequence of words corresponding to the English voice character sequence "iphone"; thus, search Being apparatus according to the sequence determined by determining the sequence of characters in English 206 device, the English character sequence "iphone" as the query sequence for the user, to match the query, get the query sequence corresponding to one or more search results; subsequently, information providing device 207 by calling the dynamic web technology such as JSP, ASP or PHP, etc., will be provided to the English character sequence "iphone" corresponding to one or more search results to the user, for the user to browse.

[0056] More preferably, the sequence acquisition means 205 acquires the target application users voice input letter sequence; wherein, said

果进行打分，确定打分较高的发音信息为该英文语音数据的发音信肩、O

[0079] 在此，由于用户在输入英文单词时有时按单词读有时按字母读，本发明对每个需要识别的英文单词，利用相似的单词模型对应的中文语音信息、字母模型对应的中文语音信息，确定其对应的发音信息，提升了语音识别的准确性。

[0080] 在一优选实施例中，该方法还包括步骤S308(未示出)。在步骤S308中，模型训练设备1统计所述英文数据所对应的英文语音数据在所述语音数据中的出现次数；根据所述出现次数，并结合所述语音数据的数量信息，确定所述英文数据的概率值；根据所述概率值，更新所述英文语音模型。具体地，当大量用户通过语音采集装置输入语音数据，其中，多个用户所输入的英文语音数据为相同的英文语音数据，在步骤S301中，模型训练设备1获取该大量用户输入的大量语音数据；在步骤S302中，模型训练设备1判断该大量语音数据为英文语音数据；随后，在步骤S303中，模型训练设备1根据中文语音信息，对该大量英文语音数据进行语音识别处理，以获得与该大量英文语音数据相对应的发音信息，在此，由于该多个用户输入的为相同的英文语音数据，因此，在步骤S303中，模型训练设备1所获得的关于该多个用户输入的英文语音数据所对应的发音信息相同或相似；接着，在步骤S304中，模型训练设备1根据该多个相同或相似的发音信息，确定该多个英文语音数据所对应的相同的英文数据；在步骤S308中，模型训练设备1统计该相同的英文数据所对应的英文语音数据在该语音数据中的出现次数；接着，在步骤S308中，模型训练设备1再根据该英文语音数据的出现次数，并结合该语音数据的数量信息，确定该英文数据的概率值，如将该出现次数除以该大量语音数据的数量信息，获得该英文数据的概率值；该在步骤S308中，模型训练设备1再将该英文数据及其对应的概率值存入该英文语音模型中，以实现对该英文语音模型的更新。

[0081] 优选地，该方法还包括步骤S309(未示出)和步骤S310(未示出)。在步骤S309中，当所述出现次数大于预定数量阈值，模型训练设备1自所述英文语音模型中提取所述英文数据与所述英文语音数据；在步骤S310中，模型训练设备1根据所述英文数据与所述英文语音数据的映射关系，建立或更新与所述英文语音数据相对应的英文语音单元。具体地，该模型训练设备1可以设置一个预定数量阈值，该预定数量阈值用来判断英文数据所对应的英文语音数据在该语音数据中的出现次数是否超过该值，当该出现次数大于该预定数量阈值时，在步骤S309中，模型训练设备1自该英文语音模型中提取

sequence determining means 206 according to the English speech model, determining the phase sequence of letters speech corresponding English character sequence. Specifically, for example, by the user to interact with the voice collecting device user device, enter the letters speech sequence in the target application, sequence acquisition means 205, for example by calling the user equipment provides an application program interface (API), etc., to obtain the user the letter sequence of the target speech input application; Subsequently, sequence determination means 206 acquires speech sequence of letters apparatus 205 according to the sequence obtained by the speech recognition processing for determining the sequence of letters voice pronunciation information, and then, based on the pronunciation information, match queries in the English speech model, determining the sequence of letters corresponding English voice character sequence. For example, user interaction with the voice collecting device user's device, voice input letter sequence "1, p, h, o, n, e", the sequence to obtain the application by calling the interface device 205 such as provided by the User Equipment (API), etc., to obtain the letters of the user input speech sequence "1, p, h, o, n, e"; then, the sequence determining means 206 based on the letter speech sequence "1, p, h, o, n, e ", through voice recognition processing, it is determined that the sequence of letters voice pronunciation information" aa ai4 p i4 ei2 ch oo oul ee enl ii i4 ", and then, based on the pronunciation information, English voice model in the matching query, determining The letter corresponding to a sequence of English voice character sequence "iphone".

[0057] Similarly, the information providing device 207 will be provided to the user with the English character sequence corresponding information. Connect embodiment, when the target application for voice input application, the information providing device 207 such as dynamic web technology by calling JSP, ASP or PHP, etc., the English character sequence "iphone" as a voice input candidate terms provided to the user, for the user to select the input; if the target application is a voice search application; search acquiring apparatus according to the English character sequence determined by the sequence determination means 206, the English character sequence "iphone" as the query sequence for the user, to match the query, Get in touch with the

该英文数据与该英文语音数据；在步骤S310中，模型训练设备1根据该英文数据与该英文语音数据的映射关系，为该英文语音数据新建一个英文语音单元，以实现对该英文语音单元的建立或更新。

[0082] 更优选地，在步骤S310中，模型训练设备1根据所述英文数据与所述英文语音数据的映射关系，采用插值计算方法，建立或更新所述英文语音单元。具体地，在步骤S310中，模型训练设备1根据该英文数据与该英文语音数据的映射关系，为该英文数据新建一英文语音单元，再例如通过将该中文发音信息所对应的中文语音单元的模型参数复制给该英文语音单元，然后在英文数据上对齐，得到统计量，再与该复制所获得的参数进行插值计算，以实现对该英文语音单元的建立或更新。在此，所述中文语音单元包括为每个中文发音信息，如中文声母或韵母信息所建立的发音单元，如“s”、“ei4”等中文声韵母信息所对应的中文语音单元。

[0083] 更优选地，该方法还包括步骤S311 (未示出)。在步骤S311中，模型训练设备1根据所述英文语音数据及与该英文语音数据相对应的发音信息，更新所述英文语音单元。具体地，在步骤S310中，模型训练设备1为英文数据新建一英文语音单元之后，在步骤S311中，模型训练设备1不断地根据用户输入的英文语音数据，及由在步骤S303中识别所获得的与该英文语音数据相对应的发音信息，训练该英文语音单元。

[0084] 在此，本发明对出现此处大于预定阈值的英文语音数据，为该英文语音数据及英文数据新建英文语音单元，并对该英文语音单元不断地进行训练，以进一步提升该英文语音数据的发音信息的准确性。

[0085] 图4示出根据本发明一个优选实施例的用于基于中文语音信息训练英文语音模型的方法流程图。以下参照图4对该优选实施例进行详细描述:具体地，在步骤S401中，模型训练设备1获取用户通过语音采集装置输入的语音数据；在步骤S402中，模型训练设备1判断所述语音数据是否为英文语音数据；在步骤S403中，模型训练设备1当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息；在步骤S404中，模型训练设备1根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射；在步骤S405中，模型训练设备1获取所述用户为目标应用输入的单词语音序列；在步骤S406中，模型训练设备1根据所述英文语音模型，确定与所述单词语音序列相对应的英文

query sequence corresponding to one or more search results; subsequently, the information providing device 207 such as dynamic web technology by calling JSP, ASP or PHP, etc., will work with the English character sequence "iphone" corresponding to one or more Results provided to the user, for the user to browse.

[0058] In this case, the present invention can also be combined with voice recognition that allows users to press letters and press Enter word English speech sequence in two ways to improve the user experience.

[0059] Figure 3 shows a flowchart of a method Chinese voice information model-based training English voice to another aspect of the present invention is used in accordance with.

[0060] In step S301, the model training device 1 acquires the user voice data input through the voice collecting apparatus. Specifically, the user or a large number of users to interact with the voice collecting device, the input voice data, in step S301, the model training device 1 by calling the application program interface that provides voice collecting apparatus (API), to obtain the user or a large number of voice data input by the user; or, users or large number of users through an interactive voice acquisition device, the input voice data and voice data stored in the voice database, in step S301, the model training device 1 by calling the voice database application program interface provided by (API), to obtain a large number of users or user voice data input from the speech database.

[0061] skilled in the art should be able to understand the way the acquired speech data is only an example, other existing or future way to obtain voice data that may occur as applicable to the present invention, it should be included in the present invention Within the scope of protection, and in this way is incorporated herein by reference.

[0062] In step S302, the speech model training apparatus 1 determines whether the data is the English voice data. Specifically, in step S302, the model based on the training apparatus 1 in step S301 the user acquired by the speech data input voice collecting means, e.g., by matching acoustic models, etc., it is determined whether the speech data is speech data in English.

Petitioner Exhibit 1004, Page 794 of 1518

字符序列；在步骤S407中，模型训练设备1将与所述英文字符序列相对应的信息提供给所述用户。其中，步骤S401-S404与图3所示对应步骤相同或相似，故此处不再赘述，并通过引用的方式包含于此。

[0086] 在步骤S405中，模型训练设备1获取所述用户为目标应用输入的单词语音序列。具体地，用户例如通过与用户设备的语音采集装置的交互，在目标应用中输入单词语音序列，在步骤S405中，模型训练设备1例如通过调用该用户设备所提供的应用程序接口(API)等方式，获取该用户为目标应用所输入的单词语音序列。

[0087] 在步骤S406中，模型训练设备1根据所述英文语音模型，确定与所述单词语音序列相对应的英文字符序列。具体地，在步骤S406中，模型训练设备1根据在步骤S405中所获取的单词语音序列，通过语音识别处理，确定该单词语音序列的发音信息，进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该单词语音序列相对应的英文字符序列。例如，用户通过与用户设备的语音采集装置的交互，输入了单词语音序列“iphone”，在步骤S405中，模型训练设备1例如通过调用该用户设备所提供的应用程序接口(API)等方式，获取了该用户输入的单词语音序列“iphone”；随后，在步骤S406中，模型训练设备1根据该单词语音序列“iphone”，通过语音识别处理，确定该单词语音序列的发音信息“aaai4 f engl”，进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该单词语音序列相对应的英文字符序列“iphone”。

[0088] 在步骤S407中，模型训练设备1将与所述英文字符序列相对应的信息提供给所述用户。例如，当该目标应用为语音输入应用，在步骤S407中，模型训练设备1通过调用诸如JSP、ASP或PHP等动态网页技术，将该英文字符序列作为语音输入候选词条提供给该用户；当该目标应用为语音搜索应用；在步骤S407中，模型训练设备1通过调用诸如JSP、ASP或PHP等动态网页技术，将根据该英文字符序列搜索获得的一个或多个搜索结果提供给该用户。

[0089] 优选地，该目标应用为语音输入应用；其中，在步骤S407中，模型训练设备1将所述英文字符序列作为语音输入候选词条提供给所述用户。例如，当该目标应用为语音输入应用，用户通过与用户设备的语音采集装置的交互，在语音输入应用中输入了单词语音序列“iphone”，在步骤S405中，模型训练设备1例如通过调用该用户设备所提供的应用程序接口(API)等方式，获取了该用户输入的单词语音序列“iphone”;随后，在步骤S406中，模型训练

[0063] skilled in the art should be able to understand the judgment of the English way of example only voice data, other available data to determine the English voice mode or may occur in the future as applicable to the present invention, it should also be included Within the scope of the present invention and hereby incorporated herein by reference.

[0064] In step S303, the model training device 1 when the speech data is English voice data, according to Chinese voice message, the English voice data to voice recognition processing, in order to obtain the data with English voice corresponding pronunciation information. In particular, when in step S302, the model training device 1 determines the speech data into English voice data, in step S303, the model training apparatus according to Chinese voice information, such as Chinese vowel sound information, the English voice data for voice recognition treatment, for example, when in step S302, the model training device 1 determines the speech data into English voice data, in step S303, the model training device 1 uses all possible Chinese vowel sound information to identify the English voice data to determine The English pronunciation of voice data information.

[0065] For example, the user interacts with the voice collecting device, voice data input word "iphone", in step S301, the model training device 1 by calling the voice collecting apparatus provided application programming interface (API) obtain the speech data "iphone"; Subsequently, in step S302, the model uses an acoustic model training device 1 matching methods to determine if the voice data is the English voice data; Subsequently, in step S303, the model uses all the training equipment 1 Possible Chinese vowel sound information, the English voice data "iphone" to force alignment, determine the English voice data "iphone" pronunciation information "aa ai4f engl". Here, aa, ai4, f, engl and other information that is Chinese vowel sound, which indicates that the vowel digital information corresponding to the tone. The skilled in the art will appreciate that in this, Chinese vowel sound and tonal representation of information are examples only, and not to deal with the present invention, any restrictions.

[0066] Preferably, in the step S303, the model training device 1 may also be based on a word model corresponding to the speech information in

设备1根据该单词语音序列“iphone”，通过语音识别处理，确定该单词语音序列的发音信息“aa ai4 f engl”，进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该单词语音序列相对应的英文字符序列“ iphone”;进而，在步骤S407中，模型训练设备1通过调用诸如JSP、ASP或PHP等动态网页技术，将该英文字符序列“iphone”作为语音输入候选词条提供给该用户，以供该用户选择输入。

[0090] 优选地，该目标应用为语音搜索应用；其中，该方法还包括步骤S412 (未示出)。在步骤S412中，模型训练设备1将所述英文字符序列作为所述用户的查询序列，获取与所述查询序列相对应的一个或多个搜索结果；其中，在步骤S407中，模型训练设备1将所述一个或多个搜索结果提供给所述用户。具体地，当该目标应用为语音搜索应用，在步骤S406中，模型训练设备1根据该英文语音模型，确定与该单词语音序列相对应的英文字符序列；在步骤S412中，模型训练设备1根据在步骤S406中所确定的英文字符序列，将该英文字符序列作为该用户的查询序列，进行匹配查询，获取与该查询序列相对应的一个或多个搜索结果；随后，在步骤S407中，模型训练设备1通过调用诸如JSP、ASP或PHP等动态网页技术，将该一个或多个搜索结果提供给该用户。

[0091] 例如，当该目标应用为语音搜索应用，用户通过与用户设备的语音采集装置的交互，在语音输入应用中输入了单词语音序列“iphone”，在步骤S405中，模型训练设备1例如通过调用该用户设备所提供的应用程序接口(API)等方式，获取了该用户输入的单词语音序列“iphone”；随后，在步骤S406中，模型训练设备1根据该单词语音序列“ iphone”，通过语音识别处理，确定该单词语音序列的发音信息“aa ai4 f engl”，进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该单词语音序列相对应的英文字符序列“iphone”;进而，在步骤S412中，模型训练设备1根据在步骤S406中所确定的英文字符序列，将该英文字符序列“iphone”作为该用户的查询序列，进行匹配查询，获取与该查询序列相对应的一个或多个搜索结果；随后，在步骤S407中，模型训练设备1通过调用诸如JSP、ASP或PHP等动态网页技术，将与该英文字符序列“iphone”相对应的一个或多个搜索结果提供给该用户，供该用户浏览。

[0092] 更优选地，在步骤S405中，模型训练设备1获取所述用户为目标应用输入的字母语音序列；其中，在步骤S406中，模型训练设备1根据所述英文语音模型，确定与所述字母语音序列相对应的英文字符序列。具体地，用户例如通过与用户设备的语

Chinese, or Chinese letter model corresponding to the voice message, the English voice data for voice recognition processing, to get the English word voice data corresponding pronunciation information, or to pronounce the letter information. For example, for "iphone", if the user input voice collecting device is read by the word "iphone", then in step S303, the model training device 1 to determine the corresponding voice information indicating the use of Chinese pronunciation information "aa ai4 fengl "; if the user input device for voice collecting alphabetical read" iphone ", namely," 1, p, h, ?, n, e ", then in step S303, the model training device 1 to determine the corresponding adopt Chinese Voice messages indicated pronunciation information "aa ai4 pi4 ei2 ch oo oul ee enl iii4".

[0067] skilled in the art should be able to understand the way information is the identification pronunciation of example only, other available information the way for future identification pronunciation or may occur as applicable to the present invention, it should be included in the present invention Within the scope of protection, and in this way is incorporated herein by reference.

[0068] In step S304, the model training apparatus according to the pronunciation information, determine the English voice data corresponding English data, in order to establish or update the English voice model, wherein the pronunciation information and the English data by the Chinese voice information is mapped. Specifically, the English data corresponding English voice data in step S304, the model training device 1 in step S303, based on the identified information received pronunciation, identified as "aa ai4 fengl", determined in accordance with the pronunciation information corresponding pronunciation information English data "iphone", then the mapping of the English pronunciation information and data stored in the English speech model, in order to achieve the establishment or update English voice model, wherein the information associated with the English pronunciation of the Chinese data voice information is mapped.

[0069] In this case, the English data is a particular English word, such as "iphone" and so on, but the English voice data for specific English word corresponding to the voice data, such as the

音采集装置的交互，在目标应用中输入字母语音序列，在步骤S405中，模型训练设备1例如通过调用该用户设备所提供的应用程序接口(API)等方式，获取该用户为目标应用所输入的字母语音序列；随后，在步骤S406中，模型训练设备1根据在步骤S405中所获取的字母语音序列，通过语音识别处理，确定该字母语音序列的发音信息，进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该字母语音序列相对应的英文字符序列。例如，用户通过与用户设备的语音采集装置的交互，输入了字母语音序列“1、P、h、o、n、e”，在步骤S405中，模型训练设备1例如通过调用该用户设备所提供的应用程序接口(API)等方式，获取了该用户输入的字母语音序列“1、P、h、0、n、e”；随后，在步骤S406中，模型训练设备1根据该字母语音序列“1、P、h、o、n、e”，通过语音识别处理，确定该字母语音序列的发音信息“aa ai4 pi4 ei2 ch oo oul ee enl ii i4”,进而，根据该发音信息，在该英文语音模型中进行匹配查询，确定与该字母语音序列相对应的英文字符序列“iphone”。

[0093] 相似地，在步骤S407中，模型训练设备1将与该英文字符序列相对应的信息提供给所述用户。接上例，当该目标应用为语音输入应用，在步骤S407中，模型训练设备1通过调用诸如JSP、ASP或PHP等动态网页技术，将该英文字符序列“iphone”作为语音输入候选词条提供给该用户，以供该用户选择输入；当该目标应用为语音搜索应用；在步骤S412中，模型训练设备1根据在步骤S406中所确定的英文字符序列，将该英文字符序列“iphone”作为该用户的查询序列，进行匹配查询，获取与该查询序列相对应的一个或多个搜索结果；随后，在步骤S407中，模型训练设备1通过调用诸如JSP、ASP或PHP等动态网页技术，将与该英文字符序列“iphone”相对应的一个或多个搜索结果提供给该用户，供该用户浏览。

[0094] 在此，本发明还可与语音识别相结合，支持用户以按字母及按单词两种方式输入英文语音序列，提升了用户的使用体验

[0095] 本发明的软件程序可以通过处理器执行以实现上文所述步骤或功能。同样地，本发明的软件程序(包括相关的数据结构)可以被存储到计算机可读记录介质中，例如，RAM存储器，磁或光驱动器或软磁盘及类似设备。另外，本发明的一些步骤或功能可采用硬件来实现，例如，作为与处理器配合从而执行各个功能或步骤的电路。

[0096]另外，本发明的一部分可被应用为计算机程序产品，例如计算机程序指令，当其被计算机执行时，通过该计算机的操作，可以调用或提供根据本

"iphone" corresponding to the voice data, The acquisition means 101 acquire voice.

[0070] skilled in the art should be understood that the above to create or update the way of example only English voice model, or other existing way to create or update the English voice model may arise in the future as applicable to the present invention, It should also be included within the scope of the present invention and hereby incorporated herein by reference.

[0071] In this case, the present invention is in the original Chinese acoustic model based on the similarity of Chinese voice information, data on the English voice for voice recognition processing to obtain the English voice data corresponding pronunciation information, to achieve create or update English voice model, effectively control the size of the English voice model.

[0072] Preferably, between each step model training apparatus 1 is continued work. Specifically, in step S301, the model training device 1 acquires the user via voice input speech data acquisition device; in step S302, the speech model training apparatus 1 determines whether the data is the English voice data; in step S303, the model train device 1 when the speech data is English voice data, according to Chinese voice information, data on the English voice for voice recognition processing, in order to obtain the English voice data corresponding pronunciation information; in step S304, the Model Training Equipment 1 according to the pronunciation information, determine the English voice data corresponding English data, in order to establish or update the English voice model, wherein the pronunciation information and data through the English voice information is mapped to the Chinese. In this case, the skilled artisan will appreciate that "persistent" refers to the steps of model training device 1 constantly acquire voice data respectively, set up to judge the English voice data recognition processing and English voice English voice data model or updating until the model training device 1 in a long time I stop getting voice data.

[0073] Preferably, in step S303, the model training device 1 when the speech data is English voice data, according to the word model corresponding to the Chinese voice message, the English voice data to voice recognition processing, in order to

发明的方法和/或技术方案。而调用本发明的方法的程序指令，可能被存储在固定的或可移动的记录介质中，和/或通过广播或其他信号承载媒体中的数据流而被传输，和/或被存储在根据所述程序指令运行的计算机设备的工作存储器中。在此，根据本发明的一个实施例包括一个装置，该装置包括用于存储计算机程序指令的存储器和用于执行程序指令的处理器，其中，当该计算机程序指令被该处理器执行时，触发该装置运行基于前述根据本发明的多个实施例的方法和/或技术方案。

[0097] 对于本领域技术人员而言，显然本发明不限于上述示范性实施例的细节，而且在不背离本发明的精神或基本特征的情况下，能够以其他的具体形式实现本发明。因此，无论从哪一点来看，均应将实施例看作是示范性的，而且是非限制性的，本发明的范围由所附权利要求而不是上述说明限定，因此旨在将落在权利要求的等同要件的含义和范围内的所有变化涵括在本发明内。不应将权利要求中的任何附图标记视为限制所涉及的权利要求。此夕卜，显然“包括”一词不排除其他单元或步骤，单数不排除复数。装置权利要求中陈述的多个单元或装置也可以由一个单元或装置通过软件或者硬件来实现。第一，第二等词语用来表示名称，而并不表示任何特定的顺序。

obtaining voice data corresponding to the English word pronunciation information. Specifically, when the step S302, the voice data model training apparatus determines the acquired in step S301 English voice data, in step S303, the model uses the word model training device 1 corresponding to the Chinese voice message, such as "aa ai4 f engl "and so on, the English voice data for speech recognition processing, such as mandatory alignment process, using all possible Chinese vowel sound information, the English voice data to force alignment for the English voice data corresponding to word pronunciation information.

[0074] In this case, the word model stored in a large number of words corresponding to the Chinese voice message, the word model in the model training device 1, or in a third party through a network connected with the model training device 1 devices.

[0075] Preferably, in step S303, the model training device 1 when the speech data is English voice data, based on the letter corresponding to the Chinese model of voice information, data on the English voice for voice recognition processing, in order to obtaining voice data corresponding to the English pronunciation of letters information. Specifically, when the step S302, the voice data model training apparatus determines the acquired in step S301 English voice data, in step S303, the model training device 1 according to the letter corresponding to the Chinese model of voice information, such as the twenty six letters of the alphabet that corresponds to Chinese voice information, such as the letter "a" corresponding to the Chinese voice message "aa eil", the letter "b" corresponding Chinese voice message "b i4", the letter "c" corresponding to the Chinese voice message " s ei4 "and so on, the English voice data for speech recognition processing, in order to obtain the English voice data information corresponding letter sounds.

[0076] Preferably, in step S303, the model training device 1 is determined in step S302 for the English voice data as voice data, give priority to the word model corresponding to the Chinese voice message, the English voice data for voice When the recognition process, when in step S303, the model uses the word model training apparatus can not recognize the English pronunciation of voice data information, and then using the letter corresponding to the Chinese model of voice

information, the English voice data for speech recognition processing, in order to get the English voice data information corresponding to the pronunciation of letters. Similarly, in step S303, the model training device 1 may be given priority letter model corresponding to the Chinese voice message, the English voice data for speech recognition processing, when in step S303, the model training apparatus does not recognize the use of letter model When the pronunciation of the English voice data information, and then using the word model corresponding to the Chinese voice message, the English voice data for speech recognition processing, in order to obtain the English voice data corresponding word pronunciation information. Here, the word models using priority or precedence letter model corresponding to the speech information in Chinese, English speech data to the speech recognition processing, the model may be set in advance training device 1 may also be adjusted to be an event.

[0077] For example, in step S301, the model train equipment acquired 1 speech data "iphone", in step S302, the model training device 1 determines the speech data "iphone" for the English voice data, in step S303, the model training device 1 priority model using the corresponding Chinese word speech information, the English voice data for speech recognition processing, but does not recognize the English pronunciation of voice data information; then, in step S301, the model training equipment 1 and then using the letter corresponding to the Chinese model of voice information, the English voice for voice recognition processing data obtained with the English voice data information corresponding to pronounce the letter "aa ai4 p i4 ei2 ch oo oul ee enl ii i4", then Voice data showed acquired in step S301 is alphabetical read "iphone", namely, "1, p, h, ?, n, e".

[0078] More preferably, in step S303, the model training device 1 word models respectively corresponding to the voice information Chinese, Chinese letter model corresponding to the voice message, the English voice data for voice recognition processing, and Speech recognition processing results were obtained for scoring, scoring higher pronunciation information to determine the pronunciation for the English letter shoulder voice data, ?

[0079] In this case, due to user input English word sometimes word by word reading Sometimes alphabetical read, the present invention requires the identification of each English word, use a similar word model corresponding Chinese voice messages, letters, Chinese voice information corresponding to the model to determine the corresponding pronunciation information, improve the accuracy of speech recognition.

[0080] In a preferred embodiment, the method further comprises the step S308 (not shown). In step S308, a statistical model of the English training device data corresponding to the number of occurrences English voice data in the voice data; based on the number of occurrences, combined with the amount of information of the voice data, to determine the English the probability value of the data; based on the probability value, updating the English voice model. Specifically, when a large number of users via voice input speech data acquisition device, wherein a plurality of the English voice data input by the user to the same English voice data, in step S301, the model of the large number of training device 1 acquires a large number of voice data input by the user; in step S302, the model training device 1 judges that a lot of voice data into English voice data; Subsequently, in step S303, the model training apparatus according to Chinese voice information, a large number of English speech data to the speech recognition process to obtain The large number of English speech data corresponding pronunciation information, in this case, since the plurality of user input to the same English voice data, and therefore, in step S303, the model obtained training device 1 is about the plurality of user input in English voice data corresponding to the same or similar pronunciation information; then, in step S304, the model training device 1 according to the plurality of the same or similar pronunciation information, identify same English data of the plurality of data corresponding to the English voice; in step S308, a statistical model training equipment that same data in English English voice data corresponding to the number of occurrences in the speech data; then, in step S308, the model number of occurrences of re-training device 1 according to the English voice data, and combined with the number of messages in the voice data to determine the probability of the value of the English data, such as the emergence of a large number of voice data divided by the amount of information, the probability of obtaining

the value of the data in English; this in step S308, the model training device 1 again The English data and its corresponding probability values are stored in the voice model in English, in order to achieve the English voice model update.

[0081] Preferably, the method further comprises the step S309 (not shown) and step S310 (not shown). In step S309, when the number of occurrences is greater than a predetermined threshold number, model in English speech model training equipment to extract a data from said in English with the English voice data; in step S310, the model based on the training apparatus data mapping between the English and the English voice data, create or update data with the corresponding English voice English voice unit. Specifically, the model training device 1 may set a predetermined threshold amount, the predetermined threshold amount used to determine the English data corresponding English voice data the number of occurrences in the speech data exceeds this value, when the number of occurrences is greater than the predetermined When the threshold number, in step S309, the model training device 1 extracts the English data associated with the English voice data from the English voice model; step S310 in the model training device 1 mapping relationship with the English voice data based on the English data, create a new unit for the English voice English voice data, in order to achieve the English phonetic units built or updated.

[0082] More preferably, in step S310, the model training device 1 according to data mapping between the English and the English voice data, using interpolation method, create or update the English voice unit. Specifically, in step S310, the model training device 1 according to the English data mapping relation to the English voice data, a new English voice unit data for the English, and then, for example by the Chinese pronunciation information corresponding Chinese phonetic units copy the model parameters to the English speech unit, and then align the English data, obtain statistics, and then with the parameters of the copy obtained by interpolating calculations to achieve the establishment or update the English voice unit. Here, the Chinese pronunciation for each speech unit includes Chinese information, such as Chinese initial or final information established pronunciation unit, such as "s", "ei4"

and other Chinese vowel sound information corresponding Chinese phonetic units.

[0083] More preferably, the method further comprises the step S311 (not shown). In step S311, the model training apparatus according to the English voice data and voice data corresponding to the English pronunciation information, updating the English voice unit. After Specifically, in step S310, the model training device 1 is a new English voice data in English units, in step S311, the model continuously training device 1 according to the English voice user input data, and in step S303 is identified by the acquired English voice data with the corresponding pronunciation information, training the English voice unit.

[0084] In this case, the present invention is greater than a predetermined threshold value appears here in English, voice data, voice and data for English and English New English voice unit data, and the English phonetic units constantly training to further enhance The accuracy of the English pronunciation of voice data information.

[0085] Figure 4 shows an example of a flowchart of a method for training Chinese English speech voice information model-based accordance with a preferred embodiment of the present invention. Referring now to Figure 4 of the preferred embodiment described in detail: specifically, in step S401, the model training device 1 acquires voice data through the voice collecting user input means; in the step S402, the speech model training apparatus 1 determines the data Are the English voice data; step S403, the model training apparatus when the voice data is voice data in English, according to Chinese voice information, data on the English voice for voice recognition processing, in order to obtain the data with English voice corresponding pronunciation information; in step S404, the model training apparatus according to the pronunciation information, determine the English voice data corresponding English data, in order to establish or update the English voice model, wherein the information relating to the English pronunciation data by the Chinese voice information is mapped; in step S405, the model training device 1 acquires the user word voice input sequence of the target application; in step S406, the model training apparatus according to the English voice model to determine the The word speech sequence corresponding to a

sequence of characters in English; in step S407, the model will provide a training device with the English character sequence information corresponding to the user. Wherein the step of S401-S404 of FIG. 3 correspond to the same or similar steps, it is not mentioned here, and incorporated herein by reference.

[0086] In step S405, the model training device 1 acquires the user application for the target word voice input sequence. Specifically, the user, for example user equipment through interactive voice collecting device, enter the word speech sequence in the target application, in step S405, a model training equipment such as an application by calling the user equipment provided interface (API), etc. way, to get the user to the word voice input sequence of the target application.

[0087] In step S406, the model training apparatus according to the English voice model to determine the sequence of words corresponding to the English speech sequence of characters. Specifically, in step S406, the model training device 1 according to the word speech sequence in step S405 acquired, through voice recognition processing, determine the sequence of the pronunciation of the word voice information, and thus, according to the pronunciation information, in the English speech model In matching the query to determine the word speech sequence corresponding English character sequence. For example, a user with the user interaction device audio capture device, enter the word speech sequence "iphone", in step S405, a model training equipment such as an application program interface (API) calls by other means provided by the user equipment, Get the word speech sequence of the user input "iphone"; Subsequently, in step S406, the model training device 1 according to the word speech sequence "iphone", through voice recognition processing, determine the sequence of the word voice pronunciation information "aaai4 f engl "Furthermore, according to the pronunciation information, matching the query in this English voice model, determined voice the word sequence corresponding English character sequence" iphone ".

[0088] In step S407, the model training device 1 will be provided to the user and the English character sequence corresponding information. For example, when the target application to use

voice input, in step S407, the model training device 1 by calling the dynamic web technology such as JSP, ASP or PHP, etc., the English character sequence as a candidate for entry speech input provided to the user; when The target application for the voice search application; in step S407, the model training device 1 by calling the dynamic web technology such as JSP, ASP or PHP, etc., will be provided to the user in accordance with one or more of the search results for that search English character sequence obtained.

[0089] Preferably, the target applications for voice input applications; wherein, in step S407, the model training device 1 as the English voice input candidate character sequence entries available to the user. For example, when the target application for voice input application, the user through an interactive voice collecting device user devices with voice input application input speech sequence word "iphone", in step S405, the model training device 1 for example, by calling the user equipment provided by the application program interface (API), etc., to get the word speech sequence of the user input "iphone"; Subsequently, in step S406, the model training device 1 according to the word speech sequence "iphone", through voice recognition process to determine the sequence of the word voice pronunciation information "aa ai4 f engl", and then, according to the pronunciation information, in the English voice model to match the query to determine the sequence of words corresponding to the English voice character sequence "iphone"; Further, in step S407, the model training device 1 by calling the dynamic web technology such as JSP, ASP or PHP, etc., the English character sequence "iphone" as a voice input candidate terms provided to the user, for the user to select the input.

[0090] Preferably, the target application for the voice search application; wherein the method further comprises the step S412 (not shown). In step S412, a model training apparatus the English character sequence as the query sequence to the user, access to the query sequence corresponding to one or more search results; wherein, in step S407, the model training device 1 the one or more search results to the user. Specifically, when the target application is a voice search applications, in step S406, the model training device 1 according to the English voice model to determine the word

speech sequence corresponding English character sequences; in step S412, the model training device 1 according to English character sequence determined in step S406, the English character sequence as a query sequence to the user to match the query, get the query sequence corresponding to one or more search results; Subsequently, in step S407, the model Training Equipment 1 by calling dynamic web technology such as JSP, ASP or PHP, etc., the one or more search results to the user.

[0091] For example, when the target application is interactive voice acquisition device voice search applications, users and user devices, enter the word in the speech input speech sequence applications "iphone", in step S405, the model train for example, by calling the device a user equipment provided by the application program interface (API), etc., to get the word speech sequence of the user input "iphone"; Subsequently, in step S406, the model training device 1 according to the word speech sequence " iphone ", through voice recognition processing, determine the sequence of the word voice pronunciation information" aa ai4 f engl ", and then, according to the pronunciation information, the English voice model to match the query to determine the sequence of words corresponding to the English voice character sequence "iphone"; and then, in step S412, the model training device 1 according to the English character sequence determined in step S406, the English character sequence "iphone" as the query sequence for the user, to match the query, access and A Subsequently, in step S407, the model training device 1 by calling the dynamic web technology such as JSP, ASP or PHP, etc., will work with the English character sequence "iphone" corresponding; the query sequence corresponding to one or more search results or more search results to the user, for the user to browse.

[0092] More preferably, in step S405, the model training device 1 acquires the user-letter sequence of the target application voice input; wherein, in step S406, the model training apparatus according to the English speech model, determining the sequence of letters corresponding English voice character sequence. Specifically, the user, for example user equipment through interactive voice collecting device, enter the letters speech sequence in the target application, in step S405, a model training equipment such as an application

by calling the user equipment provided interface (API), etc. way, to get the user to the letter sequence of the target application voice input; Subsequently, in step S406, the model training device 1 according to the letter sequence in step S405 voice acquired by the voice recognition processing, determine the sequence of voice pronounce the letter information, and thus, according to the pronunciation information, the match query in the English speech model, determining the sequence of letters corresponding English voice character sequence. For example, user interaction with the voice collecting device user's device, voice input letter sequence "1, P, h, ?, n, e", in step S405, the model training device 1 for example by calling in the user equipment is provided The application program interface (API) and other ways to get the user input speech sequence of letters "1, P, h, 0, n, e"; Subsequently, in step S406, the model training device 1 according to the letter speech sequence " 1, P, h, o, n, e ", through voice recognition processing, it is determined that the sequence of letters voice pronunciation information" aa ai4 pi4 ei2 ch oo oul ee enl ii i4 ", and then, based on the pronunciation information in the English voice model to match the query to determine the sequence of letters corresponding English voice character sequence "iphone".

[0093] Similarly, in step S407, the model training device 1 will be provided to the user with the English character sequence corresponding information. Connected cases, while providing the target application for voice input applications, in step S407, the model training device 1 by calling the dynamic web technology such as JSP, ASP or PHP, etc., the English character sequence "iphone" as a voice input candidate entry to the user, for the user to select the input; if the target application is a voice search application; in step S412, the model training device 1 according to the English character sequence determined in step S406, the English character sequence "iphone" as The user's query sequence, to match the query, get the query sequence corresponding to one or more search results; Subsequently, in step S407, the model training device 1 by calling the dynamic web technology such as JSP, ASP or PHP, etc., will the English character sequence "iphone" corresponding to one or more search results to the user, for the user to browse.

[0094] In this case, the present invention can also be combined with voice recognition that allows users to press letters and press Enter word English speech sequence in two ways to improve the user experience

[0095] software program of the present invention may be executed to implement the above steps or functions by a processor. Similarly, the software program of the present invention (including associated data structures) can be stored in the computer-readable recording medium, for example, RAM memory, magnetic or optical drive or floppy disk and similar devices. In addition, some steps or functions of the present invention may be implemented in hardware, such as the processor with the circuit to perform various functions or steps.

[0096] In addition, the present invention can be applied as part of a computer program product, such as computer program instructions which, when executed by a computer, by operating the computer, you can call or provide the method according to the invention and / or Technical solutions. The method of the present invention is calling program instructions may be stored in fixed or removable recording medium, and / or are transmitted by radio or other signal bearing medium, data flow, and / or stored in accordance with the working memory of said computer program instructions to run the device. Here, according to one embodiment of the present invention comprises an apparatus embodiment, the apparatus includes a memory for storing computer program instructions and a processor for executing program instructions, wherein, when the computer program instructions being executed by the processor, trigger The device is running according to the aforementioned embodiments of the present invention a plurality-based methods and / or technology programs.

The case of

[0097] skilled artisan, the present invention is clearly not limited to the details of the above-described exemplary embodiment, but without departing from the spirit or essential characteristics of the present invention, it is possible to achieve this in other specific forms invention. Therefore, whether from what point of view, should be considered as an exemplary embodiment, but not limiting, the scope of the

present invention is defined by the appended claims rather than the foregoing description define, therefore intended to fall claim All changes which come within the meaning and range of equivalents of the elements within the present invention covers. The claims should not be seen as the right to restrict any reference signs covered requirements. Bu this evening, apparently "comprising" does not exclude other elements or steps, singular does not exclude a plurality. A plurality of units or devices recited in the claims means may be formed of a single unit or be implemented by means of software or hardware. First, second, etc. words used to indicate the name, but does not indicate any particular order.

Citing: CN101593173 ; CN101593173 A ; CN101650943 ; CN101650943 A ; CN101826325 ; CN101826325 A ; CN102193643 ; CN102193643 A ; JP2006092994 ; JP2006092994 A2 ;

Cited by: CN104951100 A ; CN105070289 A ; CN105244027 A ; WO18086033 A1 ;

(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 103632668 A
(43)申请公布日 2014.03.12

(21)申请号 201210299700.6

(22)申请日 2012.08.21

(71)申请人 北京百度网讯科技有限公司
地址 100085 北京市海淀区上地十街 10 号百度大厦 2 层

(72)发明人 苏丹 万广鲁

(74)专利代理机构 北京汉昊知识产权代理事务所(普通合伙) 11370
代理人 罗朋 周建华

(51)Int.Cl.
G10L 15/06(2013.01)
G10L 15/00(2013.01)

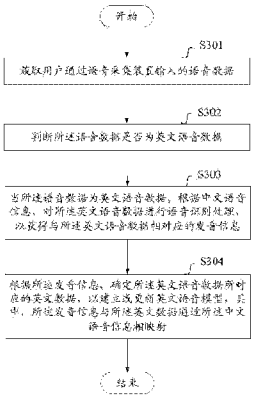
权利要求书3页 说明书14页 附图4页

(54)发明名称

一种基于中文语音信息训练英文语音模型的方法与设备

(57)摘要

本发明的目的是提供一种用于基于中文语音信息训练英文语音模型的方法与设备;获取用户通过语音采集装置输入的语音数据;判断该语音数据是否为英文语音数据;当该语音数据为英文语音数据,根据中文语音信息,对该英文语音数据进行语音识别处理,以获得与该英文语音数据相对应的发音信息;根据该发音信息,确定该英文语音数据所对应的英文数据,以建立或更新英文语音模型。与现有技术相比,本发明在原有中文声学模型基础上,利用相似的中文语音信息,对英文语音数据进行语音识别处理,以获得与该英文语音数据相对应的发音信息,实现对英文语音模型的建立或更新,有效地控制了英文语音模型的规模。



CN 103632668 A

1. 一种用于基于中文语音信息训练英文语音模型的方法，其中，该方法包括以下步骤：

a 获取用户通过语音采集装置输入的语音数据；

b 判断所述语音数据是否为英文语音数据；

c 当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息；

d 根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射。

2. 根据权利要求 1 所述的方法，其中，所述步骤 c 包括：

- 当所述语音数据为英文语音数据，根据单词模型所对应的中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的单词发音信息。

3. 根据权利要求 1 所述的方法，其中，所述步骤 c 包括：

- 当所述语音数据为英文语音数据，根据字母模型所对应的中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的字母发音信息。

4. 根据权利要求 1 至 3 中任一项所述的方法，其中，该方法还包括：

- 统计所述英文数据所对应的英文语音数据在所述语音数据中的出现次数；

- 根据所述出现次数，并结合所述语音数据的数量信息，确定所述英文数据的概率值；

- 根据所述概率值，更新所述英文语音模型。

5. 根据权利要求 4 所述的方法，其中，该方法还包括：

- 当所述出现次数大于预定数量阈值，自所述英文语音模型中提取所述英文数据与所述英文语音数据；

r 根据所述英文数据与所述英文语音数据的映射关系，建立或更新与所述英文语音数据相对应的英文语音单元。

6. 根据权利要求 5 所述的方法，其中，所述步骤 r 包括：

- 根据所述英文数据与所述英文语音数据的映射关系，采用插值计算方法，建立或更新所述英文语音单元。

7. 根据权利要求 5 或 6 所述的方法，其中，该方法还包括：

- 根据所述英文语音数据及与该英文语音数据相对应的发音信息，更新所述英文语音单元。

8. 根据权利要求 1 至 7 中任一项所述的方法，其中，该方法还包括：

x 获取所述用户为目标应用输入的单词语音序列；

y 根据所述英文语音模型，确定与所述单词语音序列相对应的英文字符序列；

z 将与所述英文字符序列相对应的信息提供给所述用户。

9. 根据权利要求 8 所述的方法，其中，所述目标应用为语音输入应用；

其中，所述步骤 z 包括：

- 将所述英文字符序列作为语音输入候选词条提供给所述用户。

10. 根据权利要求 8 所述的方法，其中，所述目标应用为语音搜索应用；

其中，该方法还包括：

- 将所述英文字符序列作为所述用户的查询序列，获取与所述查询序列相对应的一个

Petitioner Exhibit 1004, Page 810 of 1518

或多个搜索结果；

其中，所述步骤 z 包括：

- 将所述一个或多个搜索结果提供给所述用户。

11. 根据权利要求 8 至 10 中任一项所述的方法，其中，所述步骤 x 包括：

- 获取所述用户为目标应用输入的字母语音序列；

其中，所述步骤 y 包括：

- 根据所述英文语音模型，确定与所述字母语音序列相对应的英文字符序列。

12. 一种用于基于中文语音信息训练英文语音模型的模型训练设备，其中，该设备包括：

语音获取装置，用于获取用户通过语音采集装置输入的语音数据；

英文判断装置，用于判断所述语音数据是否为英文语音数据；

语音识别装置，用于当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息；

模型训练装置，用于根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射。

13. 根据权利要求 12 所述的模型训练设备，其中，所述语音识别装置用于：

- 当所述语音数据为英文语音数据，根据单词模型所对应的中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的单词发音信息。

14. 根据权利要求 12 所述的模型训练设备，其中，所述语音识别装置用于：

- 当所述语音数据为英文语音数据，根据字母模型所对应的中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的字母发音信息。

15. 根据权利要求 12 至 14 中任一项所述的模型训练设备，其中，该设备还包括模型更新装置，用于：

- 统计所述英文数据所对应的英文语音数据在所述语音数据中的出现次数；

- 根据所述出现次数，并结合所述语音数据的数量信息，确定所述英文数据的概率值；

- 根据所述概率值，更新所述英文语音模型。

16. 根据权利要求 15 所述的模型训练设备，其中，该设备还包括：

数据提取装置，用于当所述出现次数大于预定数量阈值，自所述英文语音模型中提取所述英文数据与所述英文语音数据；

单元训练装置，用于根据所述英文数据与所述英文语音数据的映射关系，建立或更新与所述英文语音数据相对应的英文语音单元。

17. 根据权利要求 16 所述的模型训练设备，其中，所述单元训练装置用于：

- 根据所述英文数据与所述英文语音数据的映射关系，采用插值计算方法，建立或更新所述英文语音单元。

18. 根据权利要求 16 或 17 所述的模型训练设备，其中，该设备还包括单元更新装置，用于：

- 根据所述英文语音数据及与该英文语音数据相对应的发音信息，更新所述英文语音单元。

Petitioner Exhibit 1004, Page 811 of 1518

19. 根据权利要求 12 至 18 中任一项所述的模型训练设备，其中，该设备还包括：

序列获取装置，用于获取所述用户为目标应用输入的单词语音序列；

序列确定装置，用于根据所述英文语音模型，确定与所述单词语音序列相对应的英文字符序列；

信息提供装置，用于将与所述英文字符序列相对应的信息提供给所述用户。

20. 根据权利要求 19 所述的模型训练设备，其中，所述目标应用为语音输入应用；

其中，所述信息提供装置用于：

- 将所述英文字符序列作为语音输入候选词条提供给所述用户。

21. 根据权利要求 19 所述的模型训练设备，其中，所述目标应用为语音搜索应用；

其中，该设备还包括：

搜索获取装置，用于将所述英文字符序列作为所述用户的查询序列，获取与所述查询序列相对应的一个或多个搜索结果；

其中，所述信息提供装置用于：

- 将所述一个或多个搜索结果提供给所述用户。

22. 根据权利要求 19 至 21 中任一项所述的模型训练设备，其中，所述序列获取装置用于：

- 获取所述用户为目标应用输入的字母语音序列；

其中，所述序列确定装置用于：

- 根据所述英文语音模型，确定与所述字母语音序列相对应的英文字符序列。

Petitioner Exhibit 1004, Page 812 of 1518

一种基于中文语音信息训练英文语音模型的方法与设备

技术领域

[0001] 本发明涉及自然语言处理技术领域，尤其涉及一种用于基于中文语音信息训练英文语音模型的技术。

背景技术

[0002] 现有的训练英文语音模型的方式一般利用现有的英语语音数据库，包括语音数据及其标注，在发音单元集合中加入英文发音单元，利用英语数据进行训练。现有技术的问题在于，采用该种方式训练英文语音模型需要一定规模的英语语音数据，且难以对加入模型中的英语词的规模进行控制。

[0003] 因此，如何克服上述困难，在控制规模的同时，实现对英文语音模型的训练，成为本领域技术人员亟需解决的问题之一。

发明内容

[0004] 本发明的目的是提供一种用于基于中文语音信息训练英文语音模型的方法与设备。

[0005] 根据本发明的一个方面，提供了一种用于基于中文语音信息训练英文语音模型的方法，其中，该方法包括以下步骤：

[0006] a 获取用户通过语音采集装置输入的语音数据；

[0007] b 判断所述语音数据是否为英文语音数据；

[0008] c 当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息；

[0009] d 根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射。

[0010] 根据本发明的另一方面，还提供了一种用于基于中文语音信息训练英文语音模型的模型训练设备，其中，该设备包括：

[0011] 语音获取装置，用于获取用户通过语音采集装置输入的语音数据；

[0012] 英文判断装置，用于判断所述语音数据是否为英文语音数据；

[0013] 语音识别装置，用于当所述语音数据为英文语音数据，根据中文语音信息，对所述英文语音数据进行语音识别处理，以获得与所述英文语音数据相对应的发音信息；

[0014] 模型训练装置，用于根据所述发音信息，确定所述英文语音数据所对应的英文数据，以建立或更新英文语音模型，其中，所述发音信息与所述英文数据通过所述中文语音信息相映射。

[0015] 与现有技术相比，本发明在原有中文声学模型基础上，利用相似的中文语音信息，对英文语音数据进行语音识别处理，以获得与该英文语音数据相对应的发音信息，实现对英文语音模型的建立或更新，有效地控制了英文语音模型的规模；进一步地，本发明还可与语音识别相结合，支持用户以按字母及按单词两种方式输入英文语音序列，提升了用户的

Petitioner Exhibit 1004, Page 813 of 1518