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示意图；

[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

Petitioner Exhibit 1004, Page 814 of 1518

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。

[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，“i、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 ou1 ee en1 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；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

Petitioner Exhibit 1004, Page 815 of 1518

Petitioner Exhibit 1004, Page 816 of 1518

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“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。

[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 ei1”、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 ei2 ch oo ou1 ee en1 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，“i、p、h、o、n、e”。

Petitioner Exhibit 1004, Page 817 of 1518

[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。

[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

Petitioner Exhibit 1004, Page 818 of 1518

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 eng1”，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 eng1”，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“i、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“i、p、h、o、n、e”；随后，序列确定装置 206 根据该字母语音序列“i、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 ou1 ee en1 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”。

Petitioner Exhibit 1004, Page 819 of 1518

[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

Petitioner Exhibit 1004, Page 820 of 1518

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，“i、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 ou1 ee en1 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 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。

[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。

Petitioner Exhibit 1004, Page 821 of 1518

[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。

[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 ei1”、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 ou1 ee en1 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，“i、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。

Petitioner Exhibit 1004, Page 822 of 1518

Petitioner Exhibit 1004, Page 823 of 1518

[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

Petitioner Exhibit 1004, Page 824 of 1518

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“aa ai4 f eng1”，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 eng1”，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 eng1”，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“i、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

Petitioner Exhibit 1004, Page 825 of 1518

输入的字母语音序列“i、p、h、o、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“i、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 ou1 ee en1 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。此外,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。

Petitioner Exhibit 1004, Page 826 of 1518

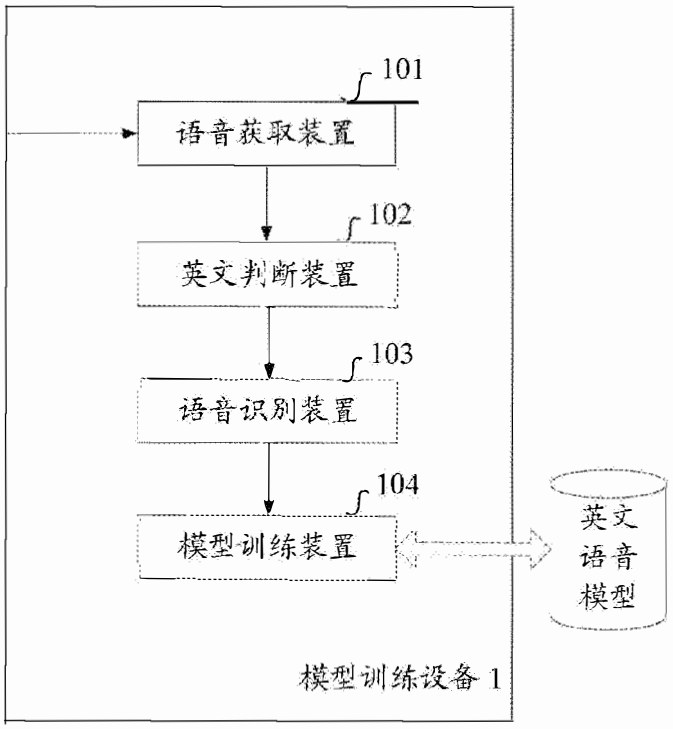


图 1

Petitioner Exhibit 1004, Page 827 of 1518

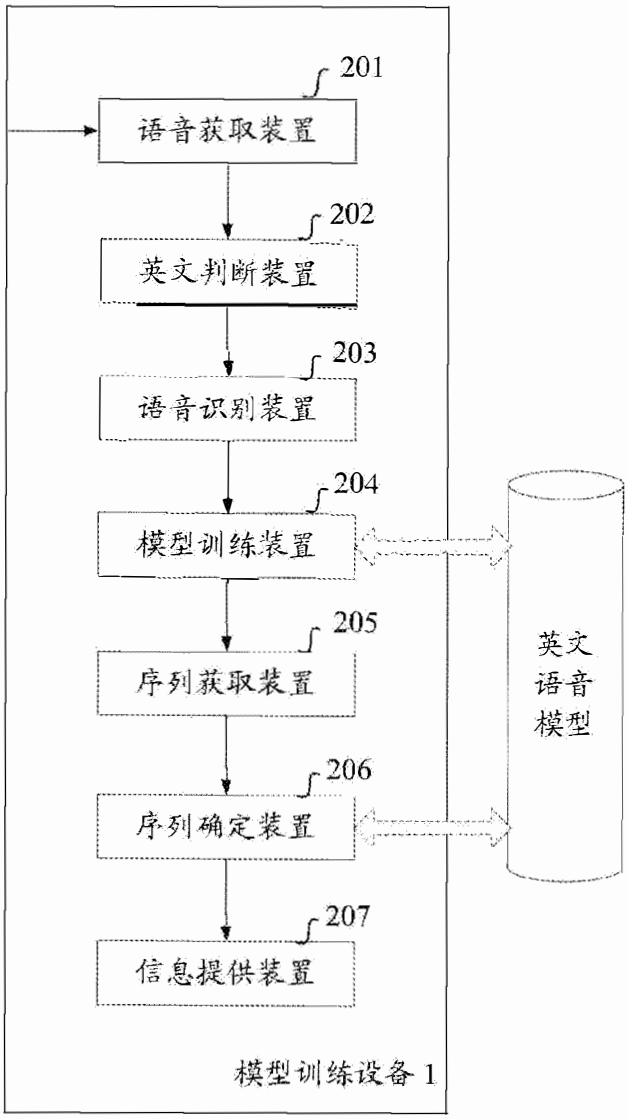


图 2

Petitioner Exhibit 1004, Page 828 of 1518

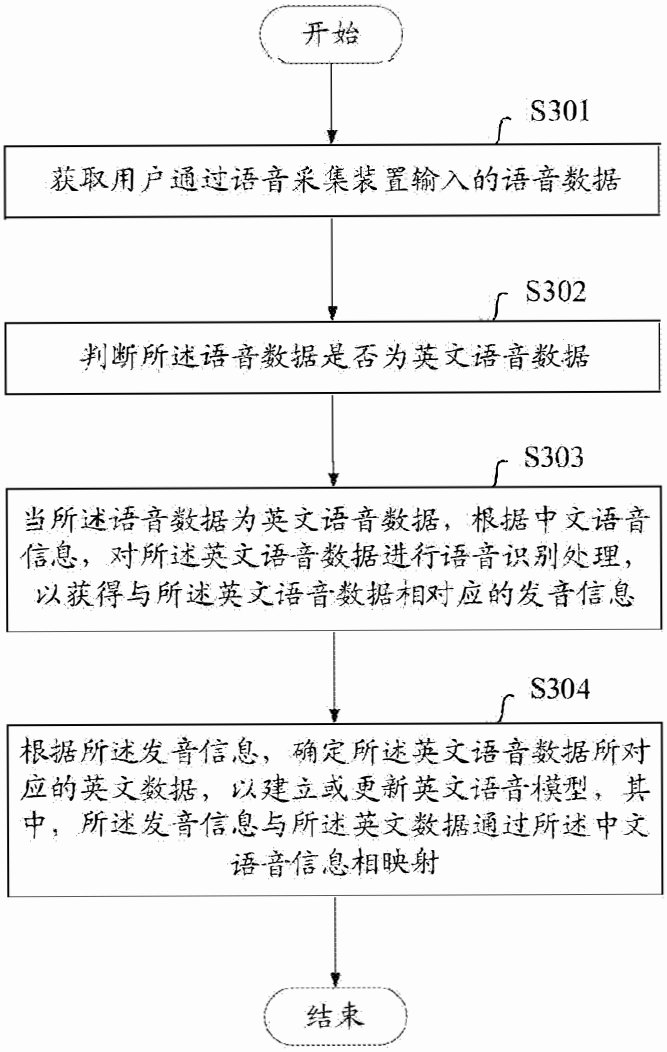


图 3

Petitioner Exhibit 1004, Page 829 of 1518

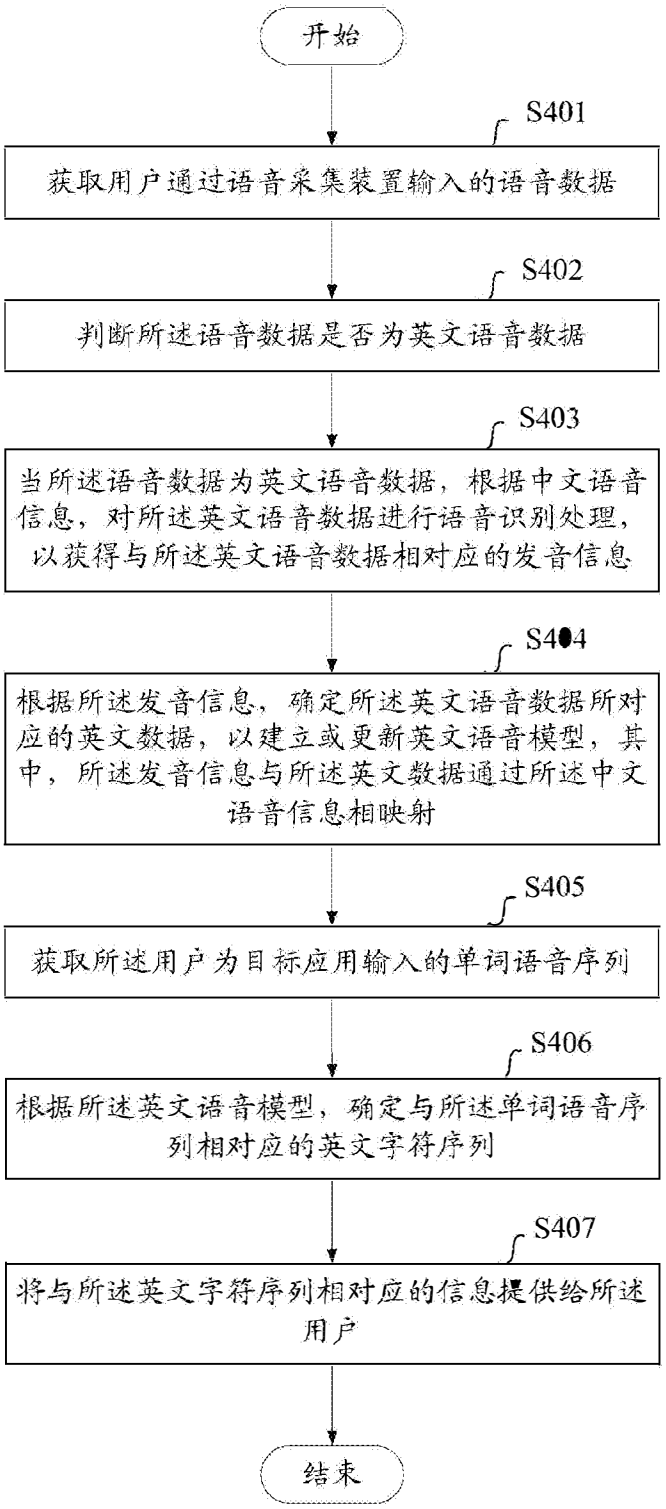


图 4

Petitioner Exhibit 1004, Page 830 of 1518

CORRECTED VERSION

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference **IEE160077PCT**	**FOR FURTHER ACTION**	see Form PCT/ISA/220 as well as, where applicable, item 5 below.
International application No. **PCT/CN2016/105343**	International filing date *(day/month/year)* **10 November 2016**	(Earliest) Priority Date *(day/month/year)*
Applicant **NUANCE SOFTWARE TECHNOLOGY (BEIJING) CO., LTD.**		

This international search report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This international search report consists of a total of **5** sheets.

☐ It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**

   a. With regard to the **language**, the international search was carried out on the basis of:

   ☑ the international application in the language in which it was filed.

   ☐ a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

   b. ☐ This international search report has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43.6*bis*(a)).

   c. ☐ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, see Box No. I.

2. ☐ **Certain claims were found unsearchable** (see Box No. II).

3. ☐ **Unity of invention is lacking** (see Box No. III).

4. With regard to the **title**,

   ☑ the text is approved as submitted by the applicant.

   ☐ the text has been established by this Authority to read as follows:

5. With regard to the **abstract**,

   ☐ the text is approved as submitted by the applicant.

   ☑ the text has been established, according to Rule 38.2, by this Authority as it appears in Box No. IV. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority.

6. With regard to the **drawings**,

   a. the figure of the **drawings** to be published with the abstract is Figure No. **2**

   ☑ as suggested by the applicant.

   ☐ as selected by this Authority, because the applicant failed to suggest a figure.

   ☐ as selected by this Authority, because this figure better characterizes the invention.

   b. ☐ none of the figures is to be published with the abstract.

Petitioner Exhibit 1004, Page 831 of 1518

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2016/105343

Box No. IV Text of the abstract (Continuation of item 5 of the first sheet)

A user device configured to perform wake-up word detection in a target language. The user device comprises at least one microphone(430) configured to obtain acoustic information from the environment of the user device, at least one computer readable medium(435) storing an acoustic model(150) trained on a corpus of training data(105) in a source language different than the target language, and storing a first sequence of speech units obtained by providing acoustic features(110) derived from audio comprising the user speaking a wake-up word in the target language to the acoustic model(150), and at least one processor(415,425) coupled to the at least one computer readable medium(435) and programmed to perform receiving, from the at least one microphone(430), acoustic input from the user speaking in the target language while the user device is operating in a low-power mode, applying acoustic features derived from the acoustic input to the acoustic model(150) to obtain a second sequence of speech units corresponding to the acoustic input, determining if the user spoke the wake-up word at least in part by comparing the first sequence of speech units to the second sequence of speech units, and exiting the low-power mode if it is determined that the user spoke the wake-up word.

Petitioner Exhibit 1004, Page 832 of 1518

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2016/105343

A. CLASSIFICATION OF SUBJECT MATTER

G10L 15/02(2006.01)i; G10L 15/06(2013.01)i; G10L 15/22(2006.01)i; G06F 1/32(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L 15/-; G06F 1/-

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, WPI, EPODOC, CNKI: sound+,voice?,speech+,audio?,acoustic+,model?,phoneme?,wak+,up,idle+,sleep+,stand+,by,low,power,activat+,trigger+,source?,target?,language?,native,nonnative

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2016189706 A1 (BROADCOM CORP.) 30 June 2016 (2016-06-30) description paragraphs [0030]-[0032], [0049]-[0070], [0086]-[0092], figures 2, 4, 5, 8	1-20
Y	CN 101650943 A (CHINESE ACAD INST. ACOUSTICS, ET AL.) 17 February 2010 (2010-02-17) description page 5 line 3 -page 6 line 9 , figures 1-3	1-20
A	CN 103021409 A (ANHUI USTC IFLYTEK CO., LTD.) 03 April 2013 (2013-04-03) the whole document	1-20
A	CN 105575395 A (ZTE CORP.) 11 May 2016 (2016-05-11) the whole document	1-20
A	US 2014365225 A1 (DSP GROUP) 11 December 2014 (2014-12-11) the whole document	1-20
A	CN 106098059 A (SHANGHAI JIAO TONG UNIVERSITY) 09 November 2016 (2016-11-09) the whole document	1-20

☑ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"●" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
13 July 2017	**21 September 2017**
Name and mailing address of the ISA/CN **STATE INTELLECTUAL PROPERTY OFFICE OF THE P.R.CHINA** **6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing 100088** **China** Facsimile No. **(86-10)62019451**	Authorized officer **KANG,Dandan** Telephone No. **(86-10)61648170**

Form PCT/ISA/210 (second sheet) (July 2009)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 103632668 A (BEIJING BAIDU NETCOM SCIENCE AND TECHNOLOGY CO., LTD.) 12 March 2014 (2014-03-12) the whole document	1-20

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2016/105343

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
US	2016189706	A1	30 June 2016	None			
CN	101650943	A	17 February 2010	None			
CN	103021409	A	03 April 2013	CN	103021409	B	24 February 2016
CN	105575395	A	11 May 2016	None			
US	2014365225	A1	11 December 2014	None			
CN	106098059	A	09 November 2016	None			
CN	103632668	A	12 March 2014	None			

Form PCT/ISA/210 (patent family annex) (July 2009)

PATENT COOPERATION TREATY

From the INTERNATIONAL SEARCHING AUTHORITY

To:
100037
China 8th Floor, Vantone New World Plaza, 2 Fuchengmenwai Street, Xicheng District, Beijing

CCPIT PATENT AND TRADEMARK LAW OFFICE

PCT

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43*bis*.1)

Date of mailing *(day/month/year)*	**04 August 2017**
Applicant's or agent's file reference	**FOR FURTHER ACTION** See paragraph 2 below
IEE160077PCT	

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/CN2016/105343	**10 November 2016**	

International Patent Classification (IPC) or both national classification and IPC
G10L 15/02(2006.01)i; G10L 15/06(2013.01)i; G10L 15/22(2006.01)i; G06F 1/32(2006.01)i

Applicant
NUANCE SOFTWARE TECHNOLOGY (BEIJING) CO., LTD.

1. This opinion contains indications relating to the following items:

- [x] Box No. I Basis of the opinion
- [ ] Box No. II Priority
- [ ] Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- [ ] Box No. IV Lack of unity of invention
- [x] Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement
- [ ] Box No. VI Certain documents cited
- [ ] Box No. VII Certain defects in the international application
- [ ] Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

If a demand for international preliminary examination is made, this opinion will be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notified the International Bureau under Rule 66.1*bis*(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Form PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

Name and mailing address of the ISA/	Date of completion of this opinion	Authorized officer
STATE INTELLECTUAL PROPERTY OFFICE OF THE P.R.CHINA **China 6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing 100088**	**31 July 2017**	**KANG,Dandan**
Facsimile No. **(86—10) 62019451**		Telephone No. **(86-10)61648170**

Form PCT/ISA/237 (Cover sheet) (July 2011)

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/CN2016/105343

Box No. I Basis of the opinion

1. With regard to the **language**, this opinion has been established on the basis of:
   - [✓] the international application in the language in which it was filed.
   - [ ] a translation of the international application into _____ which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).

2. [ ] This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43*bis*.1(a)).

3. With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of a sequence listing filed or furnished:
   a. (means)
      - [ ] on paper
      - [ ] in electronic form
   b. (time)
      - [ ] in the international application as filed
      - [ ] together with the international application in electronic form
      - [ ] subsequently to this Authority for the purposes of search

4. [ ] In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

5. Additional comments:

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

International application No.
PCT/CN2016/105343

Box No. V **Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement**

1. Statement

Novelty (N)	Claims	1-20	YES
	Claims	None	NO
Inventive step (IS)	Claims	None	YES
	Claims	1-20	NO
Industrial applicability (IA)	Claims	1-20	YES
	Claims	None	NO

2. Citations and explanations :

[1] D1: US 2016189706 A1

[2] D2: CN 101650943 A

[3] D1 discloses a wake up phrase detection method and an electronic device(see description paragraphs [0030]-[0032], [0049]-[0070],[0086]-[0092], figures 2-5,8): electronic device may comprise a smart phone; FIG.4 depicts a training system 400, a user is prompted to utter their custom wake-up phrase(WUP), the utterances are captured as user training sequences 402(corresponding to "receiving acoustic input of a user speaking a wake-up word"); user training sequences 402 is provided to VAD 404, then to feature extraction component 406 to derive the feature vectors 408(corresponding to "acoustic features derived from the acoustic input"); the extracted features 408 are decoded by concatenated phoneme ASR engine 410 using phoneme model DB 414 which is implemented in a memory or storage medium and store phoneme models 206, and produce a phoneme transcription of the WUP, a Hidden Markov Model(HMM) is trained offline to generate a phoneme model 206(corresponding to "acoustic model"); based on the results produced by ASR 410, phoneme model concatenation component 416 determines the best phonemes models and concatenates them to produce an initial WUP model consist of the concatenated phoneme models(corresponding to "a sequence of speech units corresponding to the wake-up word spoken by the user" and "a first sequence of speech units" in claim 9 or "a second sequence of speech units" in claim 16); the WUP model used for WUP detector is stored in WUP model DB 418 which is implemented in a memory or storage medium ; FIG.5 depicts a user-specific wake up phrase detection system 500, a low-power processor is included and operates in a stand-by mode of the electronic device, receiving a wake-up phrase 514 as an input by a user(corresponding to "receiving acoustic input from the user speaking while the user device is operating in a low-power mode"), WUP 514 is provided to VAD 504 and then to feature extractor and WUP detector(WUPD) component 506 to detect the presence of the WUP in the captured audio by comparing the feature extracted to the WUP model stored in a non-volatile RAM; the presence of the WUP triggers the device to wake up, or come out of a stand-by mode or a sleep mode into a normal mode of operation(corresponding to "exiting the low-power mode if it is determined that the user spoke the wake-up word").

[4] D2 discloses a non-native speech recognition system and method (see description page 5 line 3 – page 6 line 9 , figures 1-3): a native language(corresponding to "the source language") model is provided for the native state decoding module, the non-native language(corresponding to "the target language") speech is decoded by the native state decoding module using the native language model, thus obtaining a native state decode sequence "b aa zh " (see figure 3).

[5] **Novelty and Inventive step**

[6] Claim 1 differs from D1 in: (1) the acoustic input of a user is spoken in a target language, and the acoustic model is trained in a source language different than the target language.

[7] Claim 5 differs from D1 in the above (1), and (2) at least one microphone configured to obtain acoustic information from the environment of the user device.

[8] Claim 9 differs from D1 in the above (1), (2) and (3) applying acoustic features derived from the acoustic input to the acoustic model to obtain a second sequence of speech units; determining if the user spoke the wake-up word at least in part by comparing the first sequence of speech units to the second sequence of speech units.

Petitioner Exhibit 1004, Page 838 of 1518

Box No. V **Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement**

[9] Claim 16 differs from D1 in the above (1) and (3).

[10] Therefore, claims 1-20 are novel (PCT Article 33(2)).

[11] The feature (1) has already been disclosed by D2. The person skilled in the art would easily acquire the technical feature of D2 to decode the acoustic input.

[12] The feature (2) of obtaining acoustic information by microphone is general solution in the art.

[13] As to the feature (3), it is common practice for the person skilled in the art to determine the wake-up word by comparing the first sequence of speech units to the second sequence of speech units which is obtained by applying acoustic features derived from the acoustic input to the acoustic model instead of the WUP model.

[14] Therefore, claims 1,5,9,16 do not involve an inventive step (PCT Article 33 (3)).

[15] Dependent claims 2-4,6-8,10-15,17-20 do not involve an inventive step (PCT Article 33 (3)).

[16] As to claims 2,6,10,17,4,8,13,18, it is common practice to adapt the acoustic model using the acoustic input and the sequence, and associate at least one task with the wake-up word and initiate performance of the task based on the wake-up detection.

[17] As to claims 3,7,11,12,14,15,19,20, those additional features have been disclosed by D1.

[18] **Industrial applicability**

[19] Claims 1-20 can find industrial applicability in the technical field of wake-up detecting (PCT Article 33 (4)).

Petitioner Exhibit 1004, Page 839 of 1518

DOCKET NO.: N0484.71090US01

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

First Named Inventor:	William F. Ganong, III
Application No.:	13/795,933
Confirmation No.:	2321
Filed:	March 12, 2013
For:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
Examiner:	H. C. Ho
Art Unit:	2644

Certificate of Electronic Filing under 37 C.F.R. § 1.8
I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being transmitted via the Office's electronic filing system in accordance with 37 C.F.R. § 1.6(a)(4).

Dated: May 22, 2019 Electronic Signature for MacAulay S. Rush: /MacAulay S. Rush/

MAIL STOP AMENDMENT
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

STATEMENT FILED PURSUANT TO THE DUTY OF DISCLOSURE UNDER 37 C.F.R. §§ 1.56, 1.97 AND 1.98

Sir:

Pursuant to the duty of disclosure under 37 C.F.R. §§ 1.56, 1.97 and 1.98, the undersigned requests consideration of this Information Disclosure Statement.

PART I: Compliance with 37 C.F.R. § 1.97

This Information Disclosure Statement has been filed before the mailing of a first Office Action after the filing of a Request for Continued Examination under 37 C.F.R. § 1.114.

No fee or certification is required.

PART II: Information Cited

The undersigned hereby makes of record in the above-identified application the information listed on the attached form PTO-1449 (modified PTO/SB/08). The order of presentation of the references should not be construed as an indication of the importance of the references.

The undersigned hereby makes the following additional information of record in the above-identified application.

The undersigned would like to bring to the Examiner's attention the following co-pending applications that may contain subject matter related to this application:

Serial No.	Filing Date	Inventor(s)	Docket No.
*16/308,849	12-11-2018	Pfeffinger et al.	N0484.71178US01
*16/348,689	05-09-2019	Ren et al.	N0484.71177US00

*A copy of this reference is not provided as the Office has waived the requirement under 37 C.F.R. § 1.98(a)(2)(iii) for submitting a copy of a cited U.S. patent application if it is scanned to the Image File Wrapper system and is available on Private PAIR .

The above-identified co-pending applications may include subject matter that is technically-related to subject matter of the present application and may contain claims that recite elements similar to those recited in claims of the present application. In handling prosecution of the above-identified applications, the Examiners for these other applications presumably will search for relevant prior art and, in some circumstances, may take positions about the prior art and/or the interpretation of the claims that may be material to the examination of the present application.

Accordingly, the Examiner in the present application is requested to review the file histories of the above-identified co-pending applications to see whether there is any information that the Examiner believes may be relevant to the prosecution of the present application. It is assumed that the Examiner has access to the file histories of the above-identified applications. However, if the Examiner would like copies of any or all of the information included in the file histories of the above-identified applications, the Examiner is asked to contact the undersigned representative.

The undersigned would like to bring to the Examiner's attention the enclosed search reports or other communications from corresponding or related International or Foreign National Applications:

Serial No.	Date of Mailing	Type(s) of Communication	Docket No.

Petitioner Exhibit 1004, Page 841 of 1518

PCT/US2016/037495	**12-05-2016**	**International Search Report and Written Opinion**	**N0484.71178WO00**
PCT/CN2016/105343	**09-21-2017**	**International Search Report and Written Opinion**	**N0484.71177WO00**
PCT/US2016/037495	**12-27-2018**	**International Preliminary Report on Patentability**	**N0484.71178WO00**

PART III: Remarks

Documents cited anywhere in the Information Disclosure Statement are enclosed unless otherwise indicated. It is respectfully requested that:

1. The Examiner consider completely the cited information, along with any other information, in reaching a determination concerning the patentability of the present claims;

2. The enclosed form PTO-1449 (modified PTO/SB/08) be signed by the Examiner to evidence that the cited information has been fully considered by the United States Patent and Trademark Office during the examination of this application;

3. The citations for the information be printed on any patent which issues from this application.

By submitting this Information Disclosure Statement, the undersigned makes no representation that a search has been performed, of the extent of any search performed, or that more relevant information does not exist.

By submitting this Information Disclosure Statement, the undersigned makes no representation that the information cited in the Statement is, or is considered to be, material to patentability as defined in 37 C.F.R. § 1.56(b).

By submitting this Information Disclosure Statement, the undersigned makes no representation that the information cited in the Statement is, or is considered to be, in fact, prior art as defined by 35 U.S.C. § 102.

Notwithstanding any statements by the undersigned, the Examiner is urged to form his or her own conclusion regarding the relevance of the cited information.

Petitioner Exhibit 1004, Page 842 of 1518

An early and favorable action is hereby requested.

The Director is hereby authorized to charge any deficiency or credit any overpayment in the fees occasioned by the filing of this Information Disclosure Statement to our Deposit Account No. 23/2825 under Docket No. N0484.71090US01 from which the undersigned is authorized to draw.

Respectfully submitted,

/Andrew J. Tibbetts/

By: _____

Andrew J. Tibbetts, Reg. No. 65,139
Wolf, Greenfield & Sacks, P.C.
600 Atlantic Avenue
Boston, Massachusetts 02210-2206
Telephone: (617) 646-8000

Docket No.: N0484.71090US01
Date: May 22, 2019

Electronic Acknowledgement Receipt

EFS ID:	36086082
Application Number:	13795933
International Application Number:	
Confirmation Number:	2321
Title of Invention:	METHODS AND APPARATUS FOR DETECTING A VOICE COMMAND
First Named Inventor/Applicant Name:	William F. Ganong
Customer Number:	91337
Filer:	Andrew J. Tibbetts/MacAulay Rush
Filer Authorized By:	Andrew J. Tibbetts
Attorney Docket Number:	N0484.71090US01
Receipt Date:	22-MAY-2019
Filing Date:	12-MAR-2013
Time Stamp:	14:44:54
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Transmittal Letter	N048471090US01-TRN-AJT.pdf	75773 ab49e7b1358cb4cd3c27e3d9e0b829eb220124d2	no	1

Warnings:

Information:					
2		N048471090US01-IDS-AJT.pdf	146902 de7cd3a6fba4f07eeb89ed652b76964153b63477	yes	5
	Multipart Description/PDF files in .zip description				
	Document Description		**Start**	**End**	
	Transmittal Letter		1	4	
	Information Disclosure Statement (IDS) Form (SB08)		5	5	
Warnings:					
Information:					
3	Other Reference-Patent/App/Search documents	N048471177WO00_ISRWO_Mailed_09-21-2017.pdf	355918 f95cdf82350592183c72d020d22e5ce76e874819	no	9
Warnings:					
Information:					
4	Other Reference-Patent/App/Search documents	N048471178WO00_IPRP_Mailed_12-27-2018.pdf	193960 3350bf4bdec543970a0ba4ed58112b9df530b6f9	no	6
Warnings:					
Information:					
5	Other Reference-Patent/App/Search documents	N048471178WO00_ISRWO_Mailed_12-05-2016.pdf	231787 fe8e860e79879 5f77354fedb3935c5970719174b	no	7
Warnings:					
Information:					
6	Foreign Reference	EP2899955.pdf	301349 6db924a5d95d09a049b9795c987a1ea56ff194e3	no	17
Warnings:					
Information:					
7	Foreign Reference	WO14066192.pdf	258117 400ca0a10e0462f415754853a6bc1f729aa89ca3	no	26
Warnings:					
Information:					

8	Foreign Reference	CN101650943_pdf.pdf	446560 696a77664c44cae493f525217b0fdb1adc07aa77	no	27
Warnings:					
Information:					
9	Foreign Reference	CN103021409_pdf.pdf	1809820 da3f9a390cfdf5bf61d64b85648897dfeb7e1893	no	23
Warnings:					
Information:					
10	Foreign Reference	CN103632668_pdf.pdf	1526633 d8724490028e3d0d5069d1f038b0c25acb01c39a	no	58
Warnings:					
Information:					
11	Foreign Reference	CN104575504_pdf.pdf	846845 69a705a21a018535986296a3b955087baf07cd34	no	30
Warnings:					
Information:					
12	Foreign Reference	CN105575395_pdf.pdf	2748082 01b5dca48363567d728c5e234fea771896f301a6	no	52
Warnings:					
Information:					
13	Foreign Reference	CN106098059_pdf.pdf	922481 04d20b20ff6b80e872ca107287ecb1a93912d846	no	24
Warnings:					
Information:					
		Total Files Size (in bytes):	9864227		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

FORM PTO-1449/A and B (modified PTO/SB/08)	APPLICATION NO.: 13/795,933	ATTY. DOCKET NO.: N0484.71090US01
INFORMATION DISCLOSURE STATEMENT BY APPLICANT	FILING DATE: March 12, 2013	CONFIRMATION NO.: 2321
	FIRST NAMED INVENTOR: William F. Ganong, III	
Sheet 1 of 1	GROUP ART UNIT: 2644	EXAMINER: H. C. Ho

U.S. PATENT DOCUMENTS

Examiner's Initials #	Cite No.	U.S. Patent Document Number	Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication or Issue of Cited Document MM-DD-YYYY
		9,940,936	B2	Sejnoha et al.	04-10-2018
		2016-0189706	A1	Zopf et al.	06-30-2016
		2014-0365225	A1	Haiut	12-11-2014
		2014-0249817	A1	Hart et al.	09-04-2014

FOREIGN PATENT DOCUMENTS

Examiner's Initials #	Cite No.	Foreign Patent Document Office/ Country	Number	Kind Code	Name of Patentee or Applicant of Cited Document	Date of Publication of Cited Document MM-DD-YYYY	Translation (Y/N)
		CN	101650943	A	Chinese Acad Inst Acoustics	02-17-2010	Y
		CN	103021409	A	Anhui Ustc Iflytek Co	04-03-2013	Y
		CN	103632668	A	Beijing Baidu Netcom Sci and Tec.	03-12-2014	Y
		CN	104575504	A	Univ Shanghai	04-29-2015	Y
		CN	105575395	A	Zte corp	05-11-2016	Y
		CN	106098059	A	Univ Shanghai Jiaotong	11-09-2016	Y
		EP	2 899 955	A1	Huawei Device	07-29-2015	
		WO	2014/066192	A1	Microsoft Corp.	05-01-2014	

OTHER ART — NON PATENT LITERATURE DOCUMENTS

Examiner's Initials #	Cite No	Include name of the author (in CAPITAL LETTERS), title of the article (when appropriate), title of the item (book, magazine, journal, serial, symposium, catalog, etc.), date, page(s), volume-issue number(s), publisher, city and/or country where published.	Translation (Y/N)
		International Search Report and Written Opinion for International Application No. PCT/CN2016/105343 mailed September 21, 2017 (N0484.71177WO00).	
		International Search Report and Written Opinion for International Application No. PCT/US2016/037495 mailed December 05, 2016 (N0484.71178WO00).	
		International Preliminary Report on Patentability for International Application No. PCT/US2016/037495 mailed December 27, 2018 (N0484.71178WO00).	

[NOTE – No copies of U.S. patents, published U.S. patent applications, or pending, unpublished patent applications stored in the USPTO's Image File Wrapper (IFW) system, are included. See 37 CFR § 1.98 and 1287OG163. Copies of all other patent(s), publication(s), unpublished, pending U.S. patent applications, or other information listed are provided as required by 37 CFR § 1.98 unless 1) such copies were provided in an IDS in an earlier application that complies with 37 CFR § 1.98, and 2) the earlier application is relied upon for an earlier filing date under 35 U.S.C. § 120.]

EXAMINER:	DATE CONSIDERED:

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to Applicant.

7243442.1

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 May 2014 (01.05.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/066192 A1

(51) International Patent Classification:
G10L 15/24 (2013.01) *G06K 9/62* (2006.01)
G06F 3/01 (2006.01) *G10L 15/22* (2006.01)
G06K 9/00 (2006.01) *A63F 13/00* (2014.01)

(21) International Application Number:
PCT/US2013/065793

(22) International Filing Date:
18 October 2013 (18.10.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
13/662,293 26 October 2012 (26.10.2012) US

(71) Applicant: **MICROSOFT CORPORATION** [US/US]; One Microsoft Way, Redmond, Washington 98052-6399 (US).

(72) Inventors: **KAPUR, Jay**; c/o Microsoft Corporation, LCA - International Patents, One Microsoft Way, Redmond, Washington 98052-6399 (US). **TASHEV, Ivan**; c/o Microsoft Corporation, LCA - International Patents, One Microsoft Way, Redmond, Washington 98052-6399 (US). **SELTZER, Mike**; c/o Microsoft Corporation, LCA - International Patents, One Microsoft Way, Redmond, Washington 98052-6399 (US). **HODGES, Stephen Edward**; c/o Microsoft Corporation, LCA - International Patents, One Microsoft Way, Redmond, Washington 98052-6399 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,

[Continued on next page]

(54) Title: AUGMENTING SPEECH RECOGNITION WITH DEPTH IMAGING

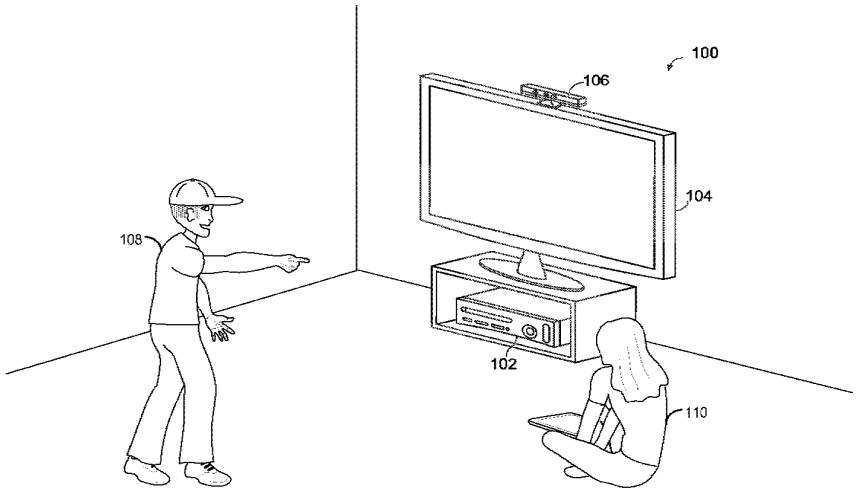


FIG. 1

(57) Abstract: Embodiments related to the use of depth imaging to augment speech recognition are disclosed. For example, one disclosed embodiment provides, on a computing device, a method including receiving depth information of a physical space from a depth camera, receiving audio information from one or more microphones, identifying a set of one or more possible spoken words from the audio information, determining a speech input for the computing device based upon comparing the set of one or more possible spoken words from the audio information and the depth information, and taking an action on the computing device based upon the speech input determined.

WO 2014/066192 A1

EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

— *with international search report (Art. 21(3))*

AUGMENTING SPEECH RECOGNITION WITH DEPTH IMAGING

BACKGROUND

[0001[Computerized speech recognition seeks to identify spoken words from audio information, such as from audio signals received via one or more microphones. However, ambiguities may arise in identifying spoken words in the audio information. Further, the context of the spoken words, for example, whether the spoken words were intended to be a speech input to a computing device, may not be easily determined from such audio information.

SUMMARY

[0002[Embodiments related to the use of depth imaging to augment speech recognition are disclosed. For example, one disclosed embodiment provides, on a computing device, a method including receiving depth information of a physical space from a depth camera, receiving audio information from one or more microphones, identifying a set of one or more possible spoken words from the audio information, determining a speech input for the computing device based upon comparing the set of one or more possible spoken words from the audio information and the depth information, and taking an action on the computing device based upon the speech input determined.

[0003[This Summary is provided to introduce a selection of concepts in a simplified form that are further described below in the Detailed Description. This Summary is not intended to identify key features or essential features of the claimed subject matter, nor is it intended to be used to limit the scope of the claimed subject matter. Furthermore, the claimed subject matter is not limited to implementations that solve any or all disadvantages noted in any part of this disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

[0004[FIG. 1 shows a schematic example of a speech recognition environment according to an embodiment of the disclosure.

[0005[FIG. 2 is a flow chart illustrating a method for recognizing speech according to an embodiment of the disclosure.

[0006[FIG. 3 is a flow chart illustrating a method for recognizing speech according to another embodiment of the disclosure.

[0007[FIG. 4 is a flow chart illustrating a method for recognizing speech according to a further embodiment of the disclosure.

Petitioner Exhibit 1004, Page 851 of 1518

[0008[FIG. 5 schematically shows a non-limiting computing system.

DETAILED DESCRIPTION

[0009[Computerized speech recognition may pose various challenges. For example, pronunciation of individual words, accent, sharpness, tone, imperfections/impediments, and other variables of human speech may differ widely between users. Additionally, reverberation and/or noise and other unwanted sounds (e.g., loudspeakers, vacuum cleaners, etc.) in the room in which the words are spoken may hinder speech recognition. Further, the context in which the recognized words are spoken may impact such factors as whether a recognized speech segment was intended as a speech input.

[0010[Accordingly, embodiments are disclosed that relate to augmenting a speech recognition process with literal and/or contextual information identified in depth information received from a depth camera. For example, in some embodiments, movements of the speaker's mouth, tongue, and/or throat may be identified from the depth information and used to confirm the identity of possible spoken words identified via audio data, identify words not detected in audio data, etc. Additionally, in some embodiments, gestures, postures, etc. performed by the speaker may be identified from the depth information and used to place the identified words into a desired context, such as confirming that the identified spoken words were intended as an input to a computing device. The term "speech recognition" as used herein may include word recognition, speaker recognition (e.g. which of two or more users in an environment is speaking), semantic recognition, emotion recognition, and/or the recognition of any other suitable aspect of speech in a use environment.

[0011[FIG. 1 shows a non-limiting example of a speech recognition environment 100. In particular, FIG. 1 shows a computing system 102 in the form of an entertainment console that may be used to play a variety of different games, play one or more different media types, and/or control or manipulate non-game applications and/or operating systems. FIG. 1 also shows a display device 104 such as a television or a computer monitor, which may be used to present media content, game visuals, non-game computing content, etc., to users.

[0012[The speech recognition environment 100 further includes a capture device 106 in the form of a depth camera that visually monitors or tracks objects and users within an observed scene. Capture device 106 may be operatively connected to the computing system 102 via one or more interfaces. As a non-limiting example, the computing system 102 may include a universal serial bus to which the capture device 106 may be connected. Capture device 106 may be used to recognize, analyze, and/or track one or more human subjects and/or

Petitioner Exhibit 1004, Page 852 of 1518

objects within a physical space, such as user 108. In one non-limiting example, capture device 106 may include an infrared light source to project infrared light onto the physical space and a depth camera configured to receive infrared light. Capture device also may comprise other sensors, including but not limited to two-dimensional image sensor(s) (e.g. a visible light camera such as an RGB image sensor and/or a grayscale sensor) and one or more microphones (e.g. a directional microphone array). While depicted as providing input to an entertainment console, it will be understood that a depth camera may be used to provide input relevant to speech recognition for any suitable computing system, and may be used in non-gaming environments.

[0013[In order to image objects within the physical space, the infrared light source may emit infrared light that is reflected off objects in the physical space and received by the depth camera. Based on the received infrared light, a depth map of the physical space may be constructed. Capture device 106 may output the depth map derived from the infrared light to computing system 102, where it may be used to create a representation of the physical space imaged by the depth camera. The capture device may also be used to recognize objects in the physical space, monitor movement of one or more users, perform gesture recognition, etc. Virtually any depth finding technology may be used without departing from the scope of this disclosure. Example depth finding technologies are discussed in more detail with reference to FIG. 5.

[0014[FIG. 1 also shows a scenario in which capture device 106 tracks user 108 so that the movements of the user may be interpreted by computing system 102. In particular, movements of the mouth, tongue, and/or throat of user 108 may be monitored to determine if the user 108 is speaking. If user 108 is speaking, audio information received by computing system 102 (e.g. via one or more microphones incorporated into capture device 106 and/or located external to capture device 106) may be analyzed to recognize one or more of the words spoken by the user. The mouth, tongue, and/or throat movements also may be used to augment the process of identifying the spoken words, for example by confirming that the identified words were spoken, adding additional identified spoken words, etc.

[0015[Information from the capture device may also be used to determine various contextual elements of the identified spoken words. For example, if additional users are present in the physical space, such as user 110, the user from which the spoken words were received may be distinguished from other users by comparing the spoken words to the mouth/throat/tongue movements of one or more users in the physical space. Further, facial recognition, speaker identification (e.g. based on the user's height, body shape, gait, etc.),

Petitioner Exhibit 1004, Page 853 of 1518

and/or other suitable techniques further may be used to determine the identity of the person speaking. The relative positions and/or orientations of one or more users in a room also may be tracked to help determine whether a speaker is making a speech input. For example, if a user is not facing the capture device when speaking, it may be determined that the user is not speaking to the system. Likewise, where multiple users are visible by the capture device, whether a user is facing the capture device may be used as information to identify which person made a speech input.

[0016[Furthermore, once one or more users have been identified, the one or more users may be tracked (via the capture device, for example). This may help to facilitate the efficient matching of future recognized speech to identified speakers, and therefore to quickly identify which speech recognition model/parameters to use for a particular user (e.g. to tune the speech recognition for that user).

[0017[Further, gestures performed by user 108 identified via information from capture device 106 may be used to identify contextual information related to identified spoken words. For example, if user 108 is speaking with the intent to control computing system 102 via voice commands, user 108 may perform one or more gestures and/or postures, deliberate or otherwise, that may indicate this intent. Examples include, but are not limited to, pointing toward display device 104, looking at computing system 102 or display device 104 while speaking, or performing a specific gesture that is associated with a recognized user input. Thus, by identifying the gesture performed by user 108 as well as identifying the spoken words, a determination of the intent of the user to control the computing device may be made. Likewise, if user 108 is looking at another user, gesturing toward another user, etc., while speaking, an intent to control the computing device may not be inferred in some embodiments.

[0018[Other types of contextual information likewise may be determined from the information received from capture device 106. For example, in some embodiments, an emotional state of user 108 when speaking may be determined by facial and/or body features, postures, gestures, etc., of user 108 from depth information. As yet another example, objects in the imaged physical space may be identified and used to distinguish ambiguous words. For example, compound words such as "quarterback" may be difficult to distinguish from the individual words ("quarter" and "back") that make up the compound word. Therefore, in the case of such ambiguities, depth image data of the physical space may be used to detect objects, actions, etc., that may provide context to help determine the actual word or words spoken. In the specific example of "quarterback," depth image data

Petitioner Exhibit 1004, Page 854 of 1518

may be analyzed to determine the presence of objects and/or other contextual clues to help disambiguate these terms, such as money in a user's hand, football-related objects (e.g. is the user seated in front of the television watching a football game), etc. Such information also may be used in some instances to help disambiguate homonyms, such as "ate" and "eight."

[0019[Computing system 102 also may be configured to communicate with one or more remote computing devices, not shown in FIG. 1. For example, computing system 102 may receive video content directly from a broadcaster, third party media delivery service, or other content provider. Computing system 102 may also communicate with one or more remote services via the Internet or another network, for example in order to analyze the received audio and/or image data, perform the speech recognition, etc. While the embodiment depicted in FIG. 1 shows computing system 102, display device 104, and capture device 106 as separate elements, in some embodiments one or more of the elements may be integrated into a common device.

[0020[FIG. 2 shows a flow diagram depicting an embodiment of a method 200 for recognizing speech of a user. Method 200 may be performed by a computing device configured to receive and process audio and depth information, such as information received from capture device 106.

[0021[At 202, method 200 includes receiving depth information from a depth camera. As explained above, the depth information may be used to construct a depth map of the imaged physical space including one or more users. Additionally, image information from a visible light camera may also be received. At 204, method 200 includes receiving audio information acquired via one or more microphones, which may include directional microphones in some embodiments. At 206, one or more possible spoken words are identified from the audio information. The one or more possible spoken words may be identified by the computing device using any suitable speech recognition processes.

[0022[At 208, method 200 includes determining a speech input for the computing device based on the one or more possible spoken words and the depth information. The speech input may comprise a command that indicates an action to be performed by the computing device, content intended to be displayed on a display device and/or recorded by a computing device, and/or any other suitable speech input.

[0023[The identified possible spoken words and the depth information may be utilized in any suitable manner to determine the speech input. For example, as indicated at 210, movements of the user's mouth, tongue and/or throat may be utilized to determine possible

Petitioner Exhibit 1004, Page 855 of 1518

sounds and/or words spoken by the user. These identified possible sounds/words may then be used to disambiguate any potentially ambiguous possible spoken words from the audio information, and/or to increase a certainty of word identifications, as described in more detail below.

[0024[Similarly, in some embodiments, mouth, tongue and/or throat movements may be used to independently determine a set of possible spoken words. This set of possible spoken words may similarly be compared to the set of possible spoken words determined from the audio information to help disambiguate any uncertainty in the correct identification of words from the audio information, to add any potential missed words to the audio data, etc.

[0025[As mentioned above, the depth information also may be used to identify contextual elements related to the possible speech segments, as indicated at 212. Any suitable contextual elements may be identified. Examples of such contextual elements may include, but are not limited to, an identity of the user, an emotion of the user, a gesture performed by the user, one or more physical objects in the physical space of the user, etc. The contextual elements identified from the depth information may be used to confirm a speech input identified from the audio information, disambiguate any ambiguous possible spoken words (e.g. compound words, homonyms, etc.), place the speech input into a desired context, utilize a directional microphone system to isolate that speaker from others in the environment, tune the speech recognition based on known speech attributes of the identified user, and/or for any other suitable purposes.

[0026[Continuing with FIG. 2, method 200 comprises, at 214, taking an action on the computing device based on upon the speech input. For example, an action indicated by a command speech input may be performed, text content corresponding to the spoken words may be displayed on the display device, etc. Further, in some embodiments, the text content may be tagged with an emotional state, such that words may have a different appearance depending upon the user's detected emotional state when the words were spoken.

[0027[FIG. 3 shows a flow diagram depicting an embodiment of a method 300 for recognizing a command speech input configured to cause a computing device to perform a specified action. Method 300 may be performed by a computing device configured to receive and process audio and depth input. At 302, method 300 includes receiving depth information from a depth camera, and at 304, receiving audio information from one or more microphones. At 306, method 300 comprises identifying one or more possible spoken from the audio information, and at 308, identifying contextual elements from the depth information. Contextual elements may include, but are not limited to, a gesture performed

Petitioner Exhibit 1004, Page 856 of 1518

by the user (e.g. movement of mouth, throat, tongue, body, etc.), as indicated at 310, a physical state of a user (e.g. whether a user is sitting, crouching or standing, whether a user's mouth is open or closed, how far a user is from a display, an orientation of the user's head, etc.), as indicated at 312, and/or an emotional state of the user, as indicated at 314. It will be understood that these contextual elements are described for the purpose of example, and are not intended to be limiting in any manner.

[0028[At 316, method 300 includes comparing the spoken words and the identified contextual elements. The spoken words and the contextual elements may be compared to determine, for example, whether the spoken words are intended as a speech input directing the computing device to perform a specified action based upon the one or more contextual elements identified from the depth information. For example, a particular gesture performed by the user and identified from the depth information may indicate that the spoken words are intended as user input. As more specific example, the user may direct a gesture at a speech recognition system device, such as pointing at the computing device/display/capture device/etc. while speaking, and/or the user may perform a gesture that matches a known gesture associated with a user input.

[0029[Further, an orientation of the user's head may be used to determine if the spoken words are intended as user input. For example, if the user is looking in a particular direction while speaking, such as at toward a speech recognition system device (e.g. a display, computing device, capture device, etc.), it may be determined that the words are intended as a user input to the computing device. Likewise, if the user is looking at another user in the physical space while speaking, it may be indicated that the words are not intended as a user input.

[0030[In a further example, one or more emotions of the user may be determined from the depth data and used to determine if the spoken words are intended as a user input. For example, if the user acting in a commanding and/or directive manner (e.g. deliberate, serious, not facially animated), it may be indicated that the words were intended as user input.

[0031[At 318, method 300 comprises determining from the comparison at 316 whether the spoken words are intended as user input based upon the contextual information. If the words are determined to be intended as speech input, then method 300 comprises, at 320, performing via the computing device the action associated with the speech input. Likewise, if the words are determined not to be intended as a speech input, then method 300 comprises, at 322, not performing an action via the computing device in response to the words.

Petitioner Exhibit 1004, Page 857 of 1518

[0032[FIG. 4 shows a flow diagram depicting an embodiment of a method 400 for identifying spoken words from a combination of audio and depth information. Method 400 may be performed by a computing device configured to receive audio and depth input, such as computing system 102.

[0033[At 402, method 400 comprises receiving depth information from a depth camera, and at 404, receiving audio information from one or more microphone(s). At 406, one or more of the user's mouth, tongue, and throat are located from the depth information. For example, feature extraction may be performed on the depth information to determine where each above-listed facial feature is located.

[0034[At 408, movements of the mouth, tongue, and/or throat may be identified. For example, a degree of opening of the user's mouth, position/shape of the tongue, shape/location of the user's lips, etc., as the user speaks may be tracked to identify the movements.

[0035[At 410, method 400 optionally includes triggering speech recognition to begin responsive to detecting identified movements of the mouth, tongue and/or throat that indicate the user is speaking. In this way, the operation of a resource-intensive speech recognition process may be avoided until identified movements indicate that the user is actually speaking.

[0036[At 412, method 400 comprises identifying a speech input of the user. As explained previously, the speech input may include a command for the computing device to perform an action, or may include input that is to be displayed (e.g. as text) on a display device and/or saved. Identifying the speech input may include for example, identifying one or more possible spoken words from the audio information at 414. The speech input may be identified from the audio data in any suitable manner. Further, as indicated at 416, identifying the speech input may include identifying one or more possible sounds, words, and/or word fragments from the depth information. For example, the mouth, tongue, and/or throat movements of the user may be used to identify sounds, words, etc.

[0037[Identifying the speech input also may include, at 418, comparing the one or more possible spoken words identified from the audio information to the one or more possible spoken words or sounds identified from the depth information. This may help to increase a confidence of possible spoken words identified via the audio data, to help disambiguate possibly ambiguous speech (for example, to identify boundaries between words via hand motion analysis), to identify additional words that were missed in the audio data, and/or may be used in any other suitable manner.

Petitioner Exhibit 1004, Page 858 of 1518

[0038[As a more specific example, movements of the user's mouth, tongue, and/or throat may be analyzed (e.g. by extracting movement data from the depth images and applying one or more classification functions to the movement data) to identify possible words/sounds spoken. Further, in some embodiments, confidence scores may be applied to the possible words/sounds spoken. Then, the determined possible spoken words/sounds determined from the depth information may be compared to the possible spoken words determined from the audio information, which likewise may include confidence score data in some embodiments. From this comparison, a most likely spoken word or words may be identified, e.g. from a highest combined confidence score, or other suitable metric. It will be understood that any suitable mechanism may be used for comparing the possible spoken sounds/words identified via the depth information and the possible spoken words identified via the audio information.

[0039[At 420, method 400 includes taking an action based on the speech input. As described above, any suitable action may be taken. For example, identified speech may be used as a command input to cause the computing device to take an action, may be displayed and/or saved as content, may be used to mark up content based upon a user's determined emotional state when speaking, and/or any other suitable action.

[0040[In some embodiments, the above described methods and processes may be tied to a computing system including one or more computers. In particular, the methods and processes described herein may be implemented as a computer application, computer service, computer API, computer library, and/or other computer program product.

[0041[FIG. 5 schematically shows a non-limiting embodiment of a computing system 500 that can enact one or more of the methods and processes described above. Computing system 500 is one non-limiting example of computing system 102. Computing system 500 is shown in simplified form. It will be understood that virtually any computer architecture may be used without departing from the scope of this disclosure. In different embodiments, computing system 500 may take the form of a mainframe computer, server computer, desktop computer, laptop computer, tablet computer, home-entertainment computer, network computing device, gaming device, mobile computing device, mobile communication device (e.g., smart phone), etc.

[0042[Computing system 500 includes a logic subsystem 502 and a storage subsystem 504. Computing system 500 may optionally include a display subsystem 506, input subsystem 508, communication subsystem 510, and/or other components not shown in FIG. 5.

Petitioner Exhibit 1004, Page 859 of 1518

[0043[Logic subsystem 502 includes one or more physical devices configured to execute instructions. For example, the logic subsystem may be configured to execute instructions that are part of one or more applications, services, programs, routines, libraries, objects, components, data structures, or other logical constructs. Such instructions may be implemented to perform a task, implement a data type, transform the state of one or more components, or otherwise arrive at a desired result.

[0044[The logic subsystem may include one or more processors configured to execute software instructions. Additionally or alternatively, the logic subsystem may include one or more hardware or firmware logic machines configured to execute hardware or firmware instructions. The processors of the logic subsystem may be single-core or multi-core, and the programs executed thereon may be configured for sequential, parallel or distributed processing. The logic subsystem may optionally include individual components that are distributed among two or more devices, which can be remotely located and/or configured for coordinated processing. Aspects of the logic subsystem may be virtualized and executed by remotely accessible, networked computing devices configured in a cloud-computing configuration.

[0045[Storage subsystem 504 includes one or more physical, non-transitory, devices configured to hold data and/or instructions executable by the logic subsystem to implement the methods and processes described herein. When such methods and processes are implemented, the state of storage subsystem 504 may be transformed—e.g., to hold different data.

[0046[Storage subsystem 504 may include removable media and/or built-in devices. Storage subsystem 504 may include optical memory devices (e.g., CD, DVD, HD-DVD, Blu-Ray Disc, etc.), semiconductor memory devices (e.g., RAM, EPROM, EEPROM, etc.) and/or magnetic memory devices (e.g., hard-disk drive, floppy-disk drive, tape drive, MRAM, etc.), among others. Storage subsystem 504 may include volatile, nonvolatile, dynamic, static, read/write, read-only, random-access, sequential-access, location-addressable, file-addressable, and/or content-addressable devices.

[0047[It will be appreciated that storage subsystem 504 includes one or more physical, non-transitory devices. However, in some embodiments, aspects of the instructions described herein may be propagated in a transitory fashion by a pure signal (e.g., an electromagnetic signal, an optical signal, etc.) that is not held by a physical device for a finite duration. Furthermore, data and/or other forms of information pertaining to the present disclosure may be propagated by a pure signal.

Petitioner Exhibit 1004, Page 860 of 1518

[0048[In some embodiments, aspects of logic subsystem 502 and of storage subsystem 504 may be integrated together into one or more hardware-logic components through which the functionally described herein may be enacted. Such hardware-logic components may include field-programmable gate arrays (FPGAs), program- and application-specific integrated circuits (PASIC / ASICs), program- and application-specific standard products (PSSP / ASSPs), system-on-a-chip (SOC) systems, and complex programmable logic devices (CPLDs), for example.

[0049[The term "module" may be used to describe an aspect of computing system 500 implemented to perform a particular function. In some cases, a module may be instantiated via logic subsystem 502 executing instructions held by storage subsystem 504. It will be understood that different modules may be instantiated from the same application, service, code block, object, library, routine, API, function, etc. Likewise, the same module may be instantiated by different applications, services, code blocks, objects, routines, APIs, functions, etc. The term "module" may encompass individual or groups of executable files, data files, libraries, drivers, scripts, database records, etc.

[0050[It will be appreciated that a "service", as used herein, is an application program executable across multiple user sessions. A service may be available to one or more system components, programs, and/or other services. In some implementations, a service may run on one or more server-computing devices.

[0051[When included, display subsystem 506 may be used to present a visual representation of data held by storage subsystem 504. This visual representation may take the form of a graphical user interface (GUI). As the herein described methods and processes change the data held by the storage subsystem, and thus transform the state of the storage subsystem, the state of display subsystem 506 may likewise be transformed to visually represent changes in the underlying data. Display subsystem 506 may include one or more display devices utilizing virtually any type of technology. Such display devices may be combined with logic subsystem 502 and/or storage subsystem 504 in a shared enclosure, or such display devices may be peripheral display devices.

[0052[When included, input subsystem 508 may comprise or interface with one or more user-input devices such as a keyboard, mouse, touch screen, or game controller. In some embodiments, the input subsystem may comprise or interface with selected natural user input (NUI) componentry. Such componentry may be integrated or peripheral, and the transduction and/or processing of input actions may be handled on- or off-board. Example NUI componentry may include one or more microphones for speech and/or voice

Petitioner Exhibit 1004, Page 861 of 1518

recognition; an infrared, color, steroscopic, and/or depth camera for machine vision and/or gesture recognition; a head tracker, eye tracker, accelerometer, and/or gyroscope for motion detection and/or intent recognition; as well as electric-field sensing componentry for assessing brain activity.

[0053[When included, communication subsystem 510 may be configured to communicatively couple computing system 500 with one or more other computing devices. Communication subsystem 510 may include wired and/or wireless communication devices compatible with one or more different communication protocols. As non-limiting examples, the communication subsystem may be configured for communication via a wireless telephone network, or a wired or wireless local- or wide-area network. In some embodiments, the communication subsystem may allow computing system 500 to send and/or receive messages to and/or from other devices via a network such as the Internet.

[0054[Further, computing system 500 may include a skeletal modeling module 512 configured to receive imaging information from a depth camera 520 (described below) and identify and/or interpret one or more postures and gestures performed by a user. Computing system 500 may also include a voice recognition module 514 to identify and/or interpret one or more voice commands or spoken words issued by the user detected via one or more microphones (coupled to computing system 500 or the depth camera). While skeletal modeling module 512 and voice recognition module 514 are depicted as being integrated within computing system 500, in some embodiments, one or both of the modules may instead be included in the depth camera 520.

[0055[Computing system 500 may be operatively coupled to the depth camera 520. Depth camera 520 may include an infrared light 522 and a depth camera 524 (also referred to as an infrared light camera) configured to acquire video of a scene including one or more human subjects. The video may comprise a time-resolved sequence of images of spatial resolution and frame rate suitable for the purposes set forth herein. As described above with reference to FIG. 1, the depth camera and/or a cooperating computing system (e.g., computing system 500) may be configured to process the acquired video to identify one or more postures and/or gestures of the user, determine a location of and track movements of a user's mouth, tongue, and/or throat, and to interpret such postures and/or gestures as device commands configured to control various aspects of computing system 500.

[0056[Depth camera 520 may include a communication module 526 configured to communicatively couple depth camera 520 with one or more other computing devices. Communication module 526 may include wired and/or wireless communication devices

Petitioner Exhibit 1004, Page 862 of 1518

compatible with one or more different communication protocols. In one embodiment, the communication module 526 may include an imaging interface 528 to send imaging information (such as the acquired video) to computing system 500. Additionally or alternatively, the communication module 526 may include a control interface 530 to receive instructions from computing system 500. The control and imaging interfaces may be provided as separate interfaces, or they may be the same interface. In one example, control interface 530 and imaging interface 528 may include a universal serial bus.

[0057[The nature and number of cameras may differ in various depth cameras consistent with the scope of this disclosure. In general, one or more cameras may be configured to provide video from which a time-resolved sequence of three-dimensional depth maps is obtained via downstream processing. As used herein, the term 'depth map' refers to an array of pixels registered to corresponding regions of an imaged scene, with a depth value of each pixel indicating the depth of the surface imaged by that pixel. 'Depth' is defined as a coordinate parallel to the optical axis of the depth camera, which increases with increasing distance from the depth camera.

[0058[In some embodiments, depth camera 520 may include right and left stereoscopic cameras. Time-resolved images from both cameras may be registered to each other and combined to yield depth-resolved video.

[0059[In some embodiments, a "structured light" depth camera may be configured to project a structured infrared illumination comprising numerous, discrete features (e.g., lines or dots). A camera may be configured to image the structured illumination reflected from the scene. Based on the spacings between adjacent features in the various regions of the imaged scene, a depth map of the scene may be constructed.

[0060[In some embodiments, a "time-of-flight" depth camera may include a light source configured to project a pulsed infrared illumination onto a scene. Two cameras may be configured to detect the pulsed illumination reflected from the scene. The cameras may include an electronic shutter synchronized to the pulsed illumination, but the integration times for the cameras may differ, such that a pixel-resolved time-of-flight of the pulsed illumination, from the light source to the scene and then to the cameras, is discernible from the relative amounts of light received in corresponding pixels of the two cameras.

[0061[Depth camera 520 may include a visible light camera 532 (e.g., RGB camera). Time-resolved images from color and depth cameras may be registered to each other and combined to yield depth-resolved color video. Depth camera 520 and/or computing system 500 may further include one or more microphones 534. One or more microphones may

Petitioner Exhibit 1004, Page 863 of 1518

determine directional and/or non-directional sounds coming from users in the physical space and/or other sources. Audio data may be recorded by the one or more microphones 534. Such audio data may be determined in any suitable manner without departing from the scope of this disclosure.

[0062[While depth camera 520 and computing system 500 are depicted in FIG. 5 as being separate devices, in some embodiments depth camera 520 and computing system 500 may be included in a single device. Thus, depth camera 520 may optionally include computing system 500.

[0063[It will be understood that the configurations and/or approaches described herein are exemplary in nature, and that these specific embodiments or examples are not to be considered in a limiting sense, because numerous variations are possible. The specific routines or methods described herein may represent one or more of any number of processing strategies. As such, various acts illustrated and/or described may be performed in the sequence illustrated and/or described, in other sequences, in parallel, or omitted. Likewise, the order of the above-described processes may be changed.

[0064[The subject matter of the present disclosure includes all novel and non-obvious combinations and sub-combinations of the various processes, systems and configurations, and other features, functions, acts, and/or properties disclosed herein, as well as any and all equivalents thereof.

Petitioner Exhibit 1004, Page 864 of 1518

CLAIMS

1. On a computing device, a method for recognizing speech of a user, comprising:

receiving depth information of a physical space from a depth camera;

receiving audio information from one or more microphones;

identifying a set of one or more possible spoken words from the audio information;

determining a speech input for the computing device based upon comparing the set of one or more possible spoken words from the audio information and the depth information; and

taking an action on the computing device based upon the speech input determined.

2. The method of claim 1, further comprising identifying contextual elements in one or more of the depth information from a depth camera, audio information from a directional microphone, and image information from a visible light camera, and comparing the set of one or more possible spoken words from the audio information to the contextual elements to determine the speech input.

3. The method of claim 2, wherein identifying the contextual elements comprises one or more of determining an identity of the user based on one or more of the depth information and information from a visible light camera, determining an emotional state of the user, determining a physical state of the user, determining a gesture performed by the user, and identifying one or more objects in a physical space of the user.

4. The method of claim 1, further comprising identifying a set of one or more possible spoken sounds and/or words from the depth information and comparing the set of one or more possible spoken words identified via the audio information to the set of one or more possible spoken sounds and/or words identified via the depth information to determine the speech input.

5. The method of claim 4, wherein identifying the set of one or more possible spoken sounds and/or words from the depth information further comprises identifying one or more mouth, tongue, and/or throat movements of the user, and identifying the set of one or more possible spoken sounds and/or words based on the movements.

Petitioner Exhibit 1004, Page 865 of 1518

6. The method of claim 1, wherein the speech input comprises a command, and wherein taking the action comprises performing the command.

7. The method of claim 1, further comprising identifying which user of a plurality of users is speaking based on one or more of mouth movements and gaze direction.

8. The method of claim 1, wherein the speech input is content to be stored, and wherein taking the action comprises storing the content.

9. The method of claim 1, wherein the speech input comprises content to be displayed on a display device, and wherein taking the action comprises sending the content to the display device.

10. The method of claim 1, wherein a boundary between possible spoken sounds and/or words is determined based on identified hand movements of the user.

Petitioner Exhibit 1004, Page 866 of 1518

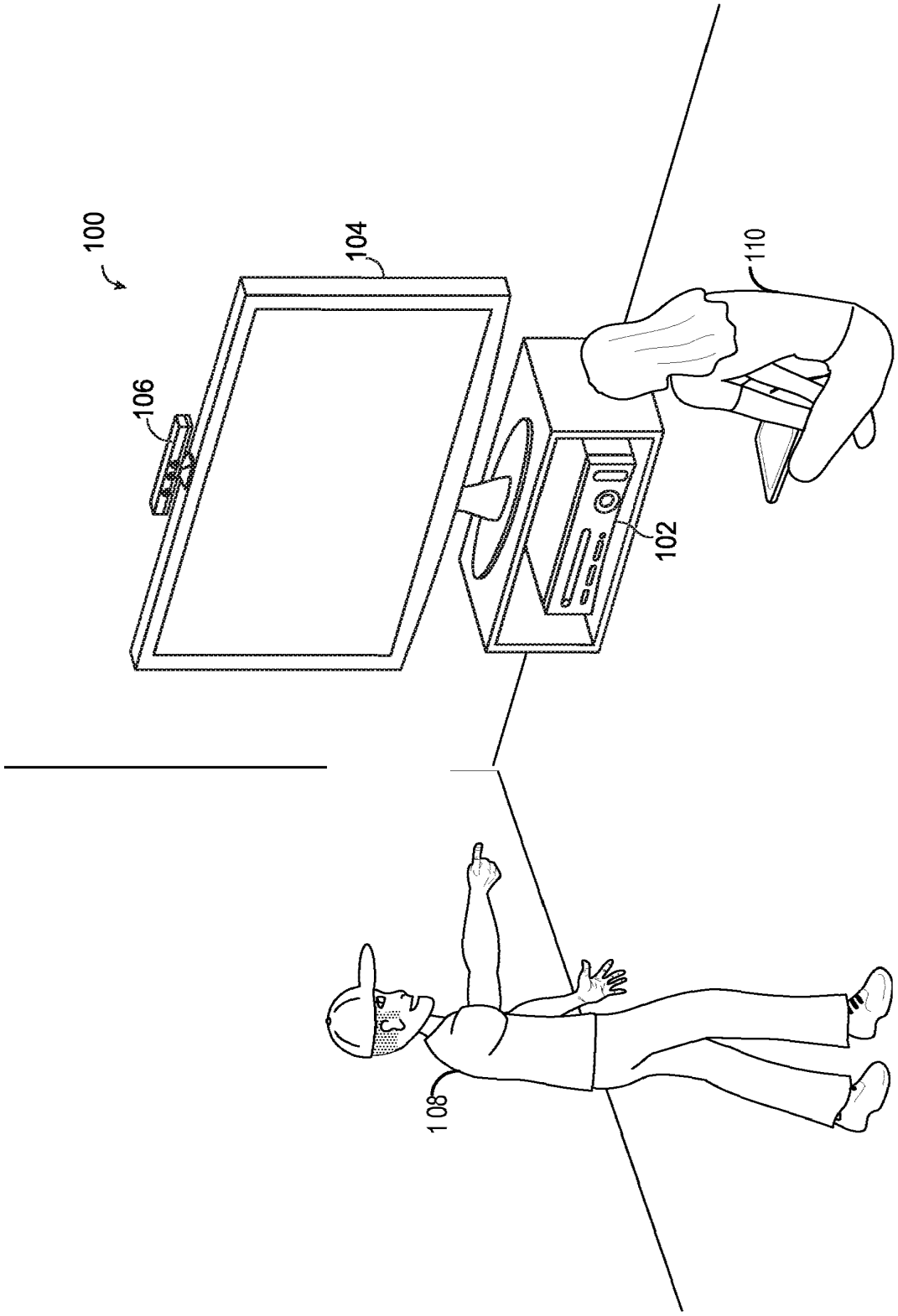


FIG. 1

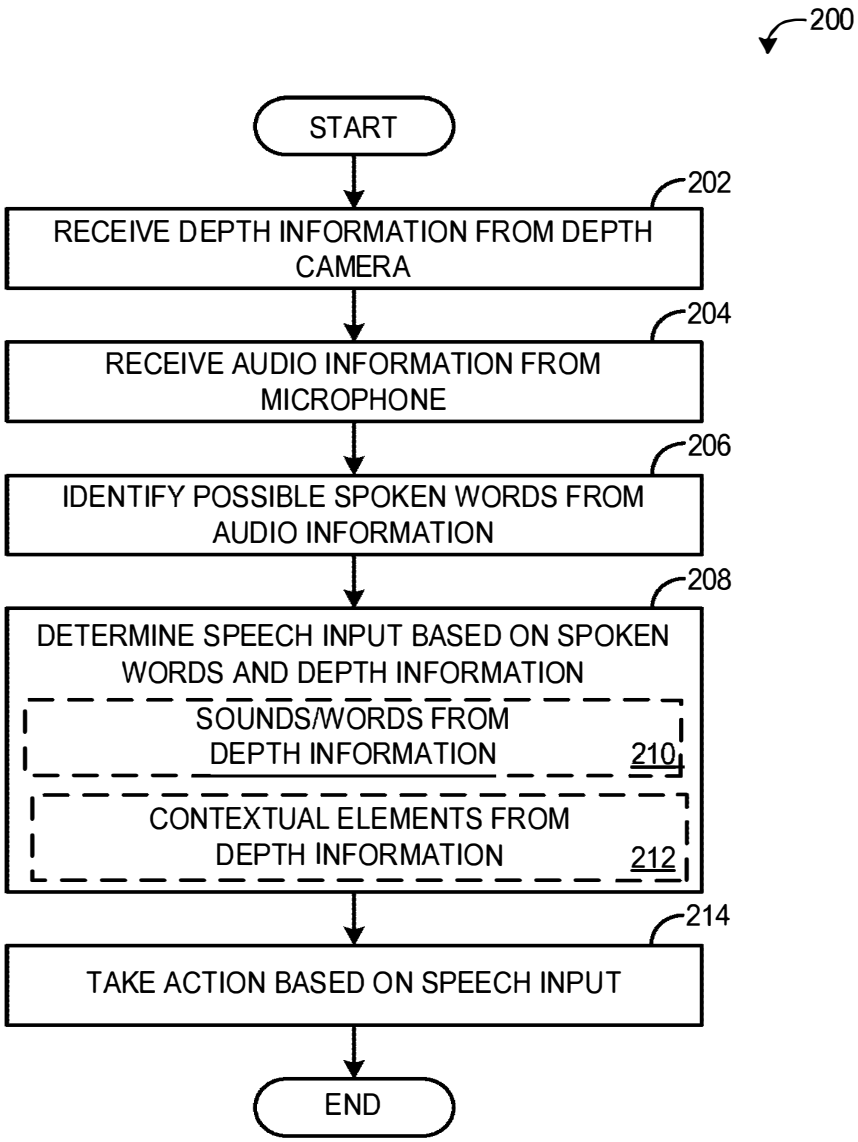

200
START
202
RECEIVE DEPTH INFORMATION FROM DEPTH CAMERA
204
RECEIVE AUDIO INFORMATION FROM MICROPHONE
206
IDENTIFY POSSIBLE SPOKEN WORDS FROM AUDIO INFORMATION
208
DETERMINE SPEECH INPUT BASED ON SPOKEN WORDS AND DEPTH INFORMATION
SOUNDS/WORDS FROM DEPTH INFORMATION 210
CONTEXTUAL ELEMENTS FROM DEPTH INFORMATION 212
214
TAKE ACTION BASED ON SPEECH INPUT
END

FIG. 2

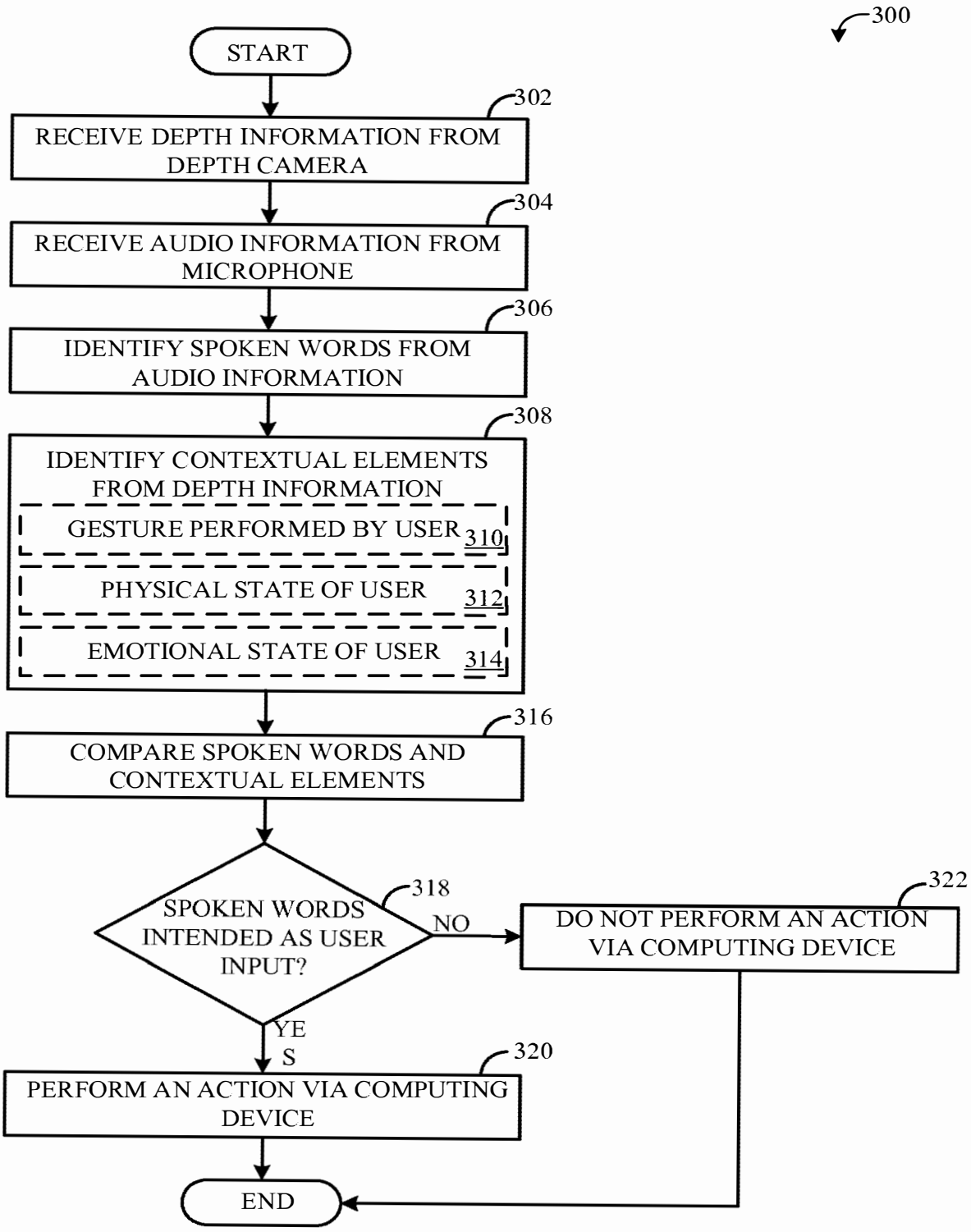

300
START
302
RECEIVE DEPTH INFORMATION FROM DEPTH CAMERA
304
RECEIVE AUDIO INFORMATION FROM MICROPHONE
306
IDENTIFY SPOKEN WORDS FROM AUDIO INFORMATION
308
IDENTIFY CONTEXTUAL ELEMENTS FROM DEPTH INFORMATION
GESTURE PERFORMED BY USER 310
PHYSICAL STATE OF USER 312
EMOTIONAL STATE OF USER 314
316
COMPARE SPOKEN WORDS AND CONTEXTUAL ELEMENTS
318
SPOKEN WORDS INTENDED AS USER INPUT?
NO
322
DO NOT PERFORM AN ACTION VIA COMPUTING DEVICE
YES
320
PERFORM AN ACTION VIA COMPUTING DEVICE
END

FIG. 3

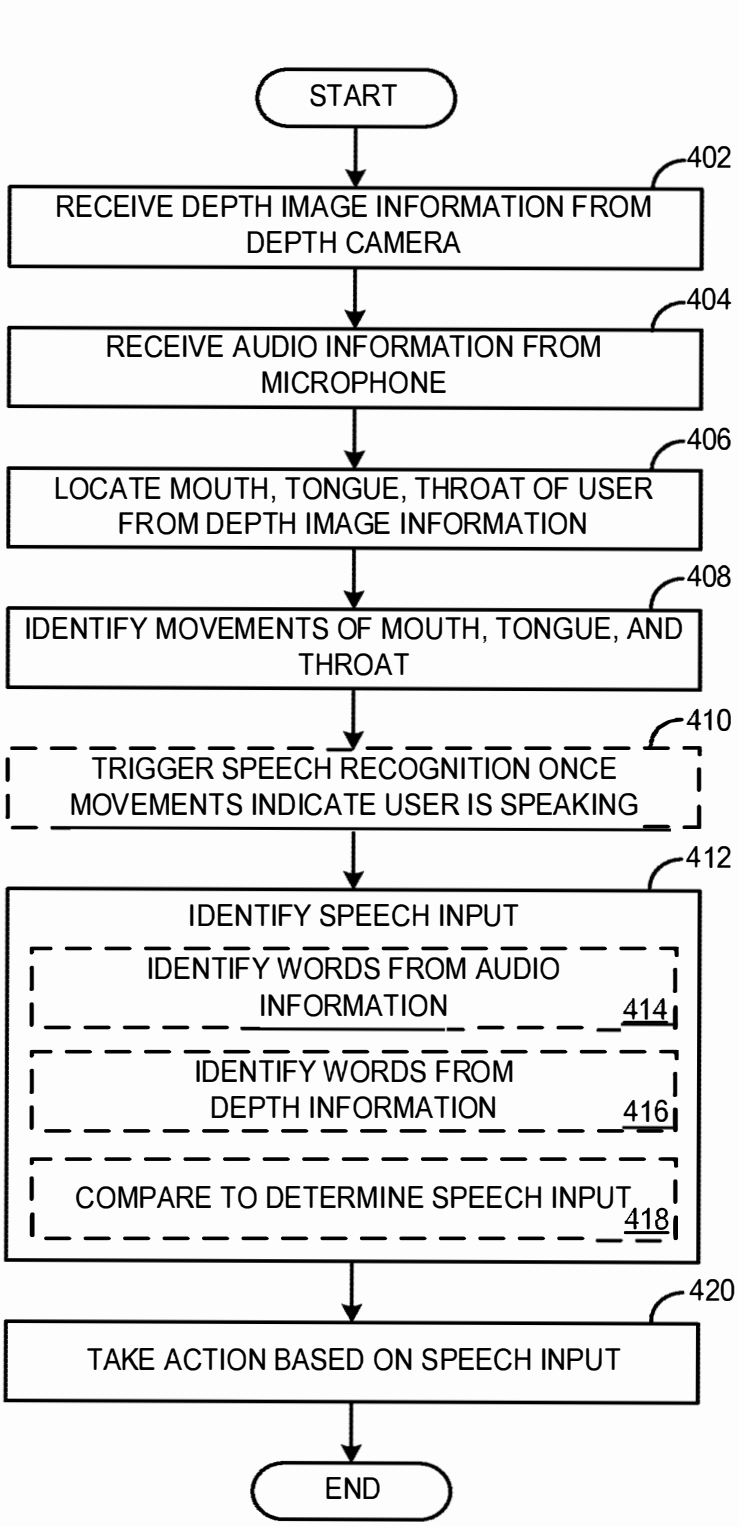

400
START
402
RECEIVE DEPTH IMAGE INFORMATION FROM DEPTH CAMERA
404
RECEIVE AUDIO INFORMATION FROM MICROPHONE
406
LOCATE MOUTH, TONGUE, THROAT OF USER FROM DEPTH IMAGE INFORMATION
408
IDENTIFY MOVEMENTS OF MOUTH, TONGUE, AND THROAT
410
TRIGGER SPEECH RECOGNITION ONCE MOVEMENTS INDICATE USER IS SPEAKING
412
IDENTIFY SPEECH INPUT
IDENTIFY WORDS FROM AUDIO INFORMATION
414
IDENTIFY WORDS FROM DEPTH INFORMATION
416
COMPARE TO DETERMINE SPEECH INPUT
418
420
TAKE ACTION BASED ON SPEECH INPUT
END

FIG. 4

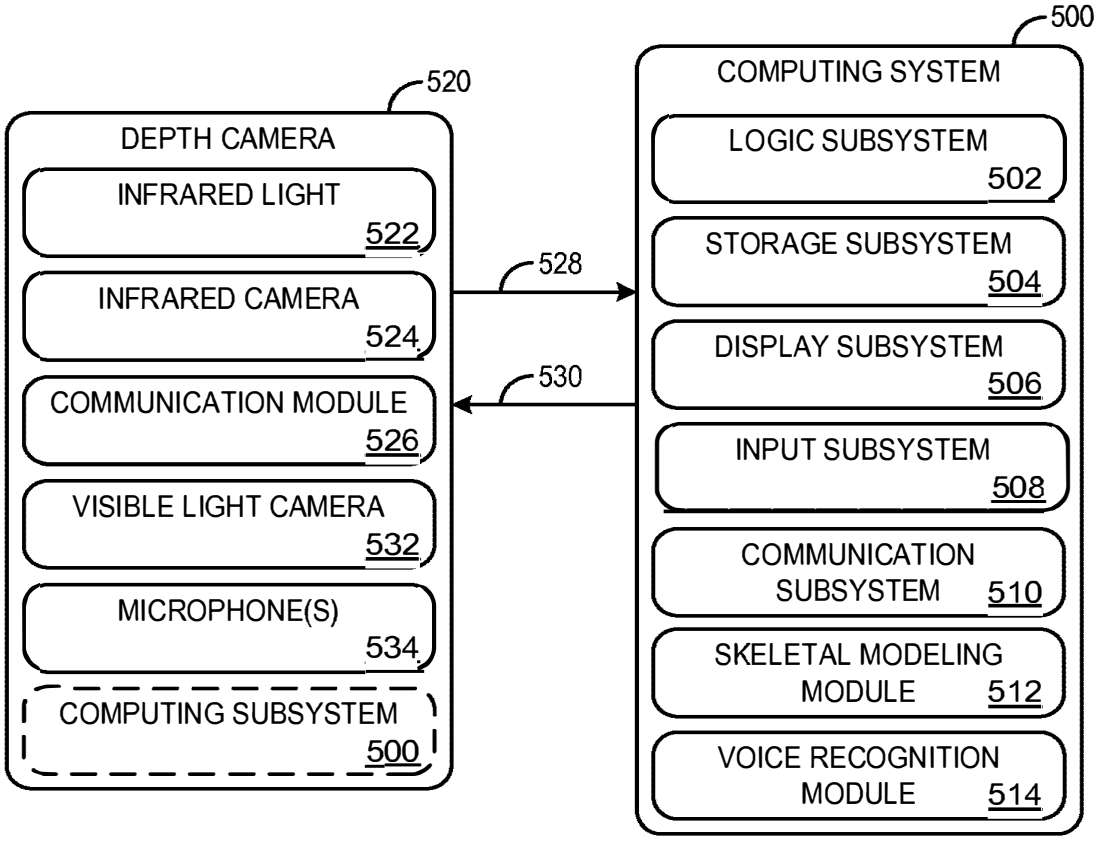

500
COMPUTING SYSTEM
LOGIC SUBSYSTEM 502
STORAGE SUBSYSTEM 504
DISPLAY SUBSYSTEM 506
INPUT SUBSYSTEM 508
COMMUNICATION SUBSYSTEM 510
SKELETAL MODELING MODULE 512
VOICE RECOGNITION MODULE 514
520
DEPTH CAMERA
INFRARED LIGHT 522
INFRARED CAMERA 524
COMMUNICATION MODULE 526
VISIBLE LIGHT CAMERA 532
MICROPHONE(S) 534
COMPUTING SUBSYSTEM 500
528
530

FIG. 5

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2013/065793

A. CLASSIFICATION OF SUBJECT MATTER
INV. G10L15/24 G06F3/01 G06K9/00 G06K9/62
ADD. G10L15/22 A63F13/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
A63F G10L G06K G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, WPI Data, COMPENDEX, INSPEC, IBM-TDB

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2009/138805 A1 (HILDRETH EVAN [CA]) 28 May 2009 (2009-05-28)	1-3,6-8
Y	paragraphs [0057], [0151] - [0155] figure 2	10
X	US 2004/193413 A1 (WILSON ANDREW D [US] ET AL) 30 September 2004 (2004-09-30) paragraphs [0060], [0074], [0084], [0094], [0107], [0108], [0168], [0169] claims 1,5	1-3,9
X	US 2002/116197 A1 (ERTEN GAMZE [US]) 22 August 2002 (2002-08-22) paragraphs [0044], [0046], [0048], [0064] figure 2	1,4,5

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
4 December 2013	18/12/2013

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Geißler, Christian

1

Form PCT/ISA/210 (second sheet) (April 2005)

Petitioner Exhibit 1004, Page 872 of 1518

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2013/065793

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2004/046736 A1 (PRYOR TIMOTHY R [CA] ET AL) 11 March 2004 (2004-03-11) paragraphs [0324], [0327], [0519], [0578] - [0581]	1
X	US 2011/257971 A1 (MORRISON ANDREW R [US]) 20 October 2011 (2011-10-20) paragraphs [0018], [0026] - [0028]	1
Y	US 2010/070268 A1 (SUNG JUN HYUNG [KR]) 18 March 2010 (2010-03-18) paragraphs [0045], [0047], [0052], [0063] claim 8	10

1

Petitioner Exhibit 1004, Page 873 of 1518

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2013/065793

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2009138805 A1	28-05-2009	CN 101925916 A	22-12-2010
		CN 103369391 A	23-10-2013
		JP 2011504710 A	10-02-2011
		US 2009138805 A1	28-05-2009
		WO 2009067670 A1	28-05-2009
US 2004193413 A1	30-09-2004	US 2004193413 A1	30-09-2004
		US 2010138798 A1	03-06-2010
		US 2010151946 A1	17-06-2010
		US 2013190089 A1	25-07-2013
US 2002116197 A1	22-08-2002	AU 9645901 A	15-04-2002
		US 2002116197 A1	22-08-2002
		WO 0229784 A1	11-04-2002
US 2004046736 A1	11-03-2004	US 6720949 B1	13-04-2004
		US 2004046736 A1	11-03-2004
		US 2006202953 A1	14-09-2006
		US 2008122786 A1	29-05-2008
		US 2013267318 A1	10-10-2013
US 2011257971 A1	20-10-2011	US 2011257971 A1	20-10-2011
		WO 2011130083 A2	20-10-2011
US 2010070268 A1	18-03-2010	JP 2012502325 A	26-01-2012
		KR 20110095236 A	24-08-2011
		US 2010070268 A1	18-03-2010
		WO 2010030129 A2	18-03-2010

PATENT COOPERATION TREATY

PCT

INTERNATIONAL SEARCH REPORT

(PCT Article 18 and Rules 43 and 44)

Applicant's or agent's file reference	**FOR FURTHER ACTION**	see Form PCT/ISA/220 as well as, where applicable, item 5 below.
N0484.71178		
International application No.	International filing date *(day/month/year)*	(Earliest) Priority Date *(day/month/year)*
PCT/US2016/037495	15 June 2016 (15-06-2016)	
Applicant		
NUANCE COMMUNICATIONS, INC.		

This international search report has been prepared by this International Searching Authority and is transmitted to the applicant according to Article 18. A copy is being transmitted to the International Bureau.

This international search report consists of a total of 3 sheets.

[X] It is also accompanied by a copy of each prior art document cited in this report.

1. **Basis of the report**

a. With regard to the **language,** the international search was carried out on the basis of:

[X] the international application in the language in which it was filed

[ ] a translation of the international application into _____, which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b))

b. [ ] This international search report has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43.6*bis*(a)).

c. [ ] With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, see Box No. I.

2. [ ] **Certain claims were found unsearchable** (See Box No. II)

3. [ ] **Unity of invention is lacking** (see Box No III)

4. With regard to the **title,**

[X] the text is approved as submitted by the applicant

[ ] the text has been established by this Authority to read as follows:

5. With regard to the **abstract,**

[X] the text is approved as submitted by the applicant

[ ] the text has been established, according to Rule 38.2, by this Authority as it appears in Box No. IV. The applicant may, within one month from the date of mailing of this international search report, submit comments to this Authority

6. With regard to the **drawings,**

a. the figure of the **drawings** to be published with the abstract is Figure No. 1

[X] as suggested by the applicant

[ ] as selected by this Authority, because the applicant failed to suggest a figure

[ ] as selected by this Authority, because this figure better characterizes the invention

b. [ ] none of the figures is to be published with the abstract

Petitioner Exhibit 1004, Page 875 of 1518

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2016/037495

A. CLASSIFICATION OF SUBJECT MATTER
INV. G10L15/22
ADD. G10L15/08 G10L17/22

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 104 575 504 A (UNIV SHANGHAI) 29 April 2015 (2015-04-29) paragraph [0043] - paragraph [0049]; figure 3	1-53
A	EP 2 899 955 A1 (HUAWEI DEVICE CO LTD [CN]) 29 July 2015 (2015-07-29) paragraph [0020] - paragraph [0029]; figure 1	1-53
A	US 2014/249817 A1 (HART GREGORY MICHAEL [US] ET AL) 4 September 2014 (2014-09-04) paragraph [0025] - paragraph [0026]; figure 1	1-53

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
15 November 2016	05/12/2016

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Zimmermann, Elko

1

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2016/037495

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
CN 104575504 A	29-04-2015	NONE	
EP 2899955 A1	29-07-2015	CN 103595869 A	19-02-2014
		EP 2899955 A1	29-07-2015
		JP 2016502829 A	28-01-2016
		US 2015141079 A1	21-05-2015
		WO 2015070644 A1	21-05-2015
US 2014249817 A1	04-09-2014	EP 2965314 A1	13-01-2016
		US 2014249817 A1	04-09-2014
		WO 2014137751 A1	12-09-2014

PATENT COOPERATION TREATY

From the
INTERNATIONAL SEARCHING AUTHORITY

To:

see form PCT/ISA/220

PCT

WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43*bis*.1)

Date of mailing
(day/month/year) see form PCT/ISA/210 (second sheet)

Applicant's or agent's file reference
see form PCT/ISA/220

FOR FURTHER ACTION
See paragraph 2 below

International application No.	International filing date *(day/month/year)*	Priority date *(day/month/year)*
PCT/US2016/037495	15.06.2016	

International Patent Classification (IPC) or both national classification and IPC
INV. G10L15/22 ADD. G10L15/08 G10L17/22

Applicant
NUANCE COMMUNICATIONS, INC.

1. This opinion contains indications relating to the following items:

- ☒ Box No. I Basis of the opinion
- ☐ Box No. II Priority
- ☐ Box No. III Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- ☐ Box No. IV Lack of unity of invention
- ☒ Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step and industrial applicability; citations and explanations supporting such statement
- ☐ Box No. VI Certain documents cited
- ☐ Box No. VII Certain defects in the international application
- ☐ Box No. VIII Certain observations on the international application

2. **FURTHER ACTION**

If a demand for international preliminary examination is made, this opinion will usually be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notifed the International Bureau under Rule 66.1*bis*(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Form PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

Name and mailing address of the ISA:	Date of completion of this opinion	Authorized Officer
European Patent Office D-80298 Munich Tel. +49 89 2399 - 0 Fax: +49 89 2399 - 4465	see form PCT/ISA/210	Zimmermann, Elko Telephone No. +49 89 2399-0

Form PCT/ISA/237 (Cover Sheet) (January 2015)

Box No. I Basis of the opinion

1. With regard to the **language**, this opinion has been established on the basis of:

   ☒ the international application in the language in which it was filed.

   ☐ a translation of the international application into , which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1 (b)).

2. ☐ This opinion has been established taking into account the **rectification of an obvious mistake** authorized by or notified to this Authority under Rule 91 (Rule 43*bis*.1(a))

3. ☐ With regard to any **nucleotide and/or amino acid sequence** disclosed in the international application, this opinion has been established on the basis of a sequence listing:

   a. ☐ forming part of the international application as filed:

      ☐ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

4. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

5. Additional comments:

Box No. V Reasoned statement under Rule 43*bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

1. Statement

Novelty (N)	Yes:	Claims	1-53
	No:	Claims	
Inventive step (IS)	Yes:	Claims	1-53
	No:	Claims	
Industrial applicability (IA)	Yes:	Claims	1-53
	No:	Claims	

2. Citations and explanations

see separate sheet

Re Item V

Reasoned statement with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement

CN 104 575 504 A is regarded as being the prior art closest to the subject-matter of claim 1, and discloses a system wherein an acoustic signal is processed in parallel by a speech recognizer and a voiceprint identification unit in order to determine whether a wake-up command was received from a registered user, see figure 3 and paragraphs [0050] to [0056] of the computer-generated translation annexed to this opinion, which correspond to paragraphs [0043] to [0049] of CN 104 575 504 A.

The subject-matter of claim 1 differs from this known system in that the acoustic signal is interpreted using information indicative of the speaker's identity and is therefore new (Article 33(2) PCT).

The problem to be solved by the present invention may be regarded as improving the robustness of the wake-up word recognition.

The solution to this problem proposed in claim 1 of the present application is considered as involving an inventive step (Article 33(3) PCT) for the following reasons: By using information indicative of the speaker's identity components of the speech recognizer like vocabulary or acoustic model can be adapted in order to improve the speech recognizer's ability to accurately determine whether a speaker uttered any one of the wake-up words. In the prior art at hand there is no hint at interpreting the acoustic signal using information indicative of the speaker's identity, and it is not considered to be obvious in view of the parallel processing illustrated in figure 3 of CN 104 575 504 A.

Claims 20 and 37 correspond to claim 1, and claims 2-19, 21-36 and 38-53 are dependent claims and as such also meet the requirements of the PCT with respect to novelty and inventive step.

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 2 899 955 A1

(12) EUROPEAN PATENT APPLICATION

published in accordance with Art. 153(4) EPC

(43) Date of publication:
29.07.2015 Bulletin 2015/31

(21) Application number: **14814665.7**

(22) Date of filing: **01.08.2014**

(51) Int Cl.:
H04M 1/725 *(2006.01)* ***G10L 17/00*** *(2013.01)*

(86) International application number:
PCT/CN2014/083507

(87) International publication number:
WO 2015/070644 (21.05.2015 Gazette 2015/20)

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR
Designated Extension States:
BA ME

(30) Priority: **15.11.2013 CN 201310574762**

(71) Applicant: **Huawei Device Co., Ltd.**
Longgang District
Shenzhen
Guangdong 518129 (CN)

(72) Inventors:
- **WANG, Xiyong**
  Shenzhen
  Guangdong 518129 (CN)
- **JIANG, Hongrui**
  Shenzhen
  Guangdong 518129 (CN)
- **ZHENG, Weijun**
  Shenzhen
  Guangdong 518129 (CN)
- **WANG, Qing**
  Shenzhen
  Guangdong 518129 (CN)

(74) Representative: **Körber, Martin Hans**
Mitscherlich PartmbB
Patent- und Rechtsanwälte
Sonnenstrasse 33
80331 München (DE)

(54) TERMINAL VOICE CONTROL METHOD, DEVICE, AND TERMINAL

(57) Embodiments of the present invention disclose a terminal voice control method. The method includes: when a voice instruction of a user is received, identifying the voice instruction, and extracting wake-up indication information from the voice instruction; performing voice verification on the voice instruction if the wake-up indication information is obtained by extracting; and after the voice verification succeeds, extracting processing indication information from the voice instruction, and responding to the voice instruction of the user according to the processing indication information. According to the terminal voice control method in the embodiments of the present invention, a user may use one voice instruction to control a terminal to complete multiple tasks without requiring voice interaction for multiple times, which is simple to use, and has high security performance.

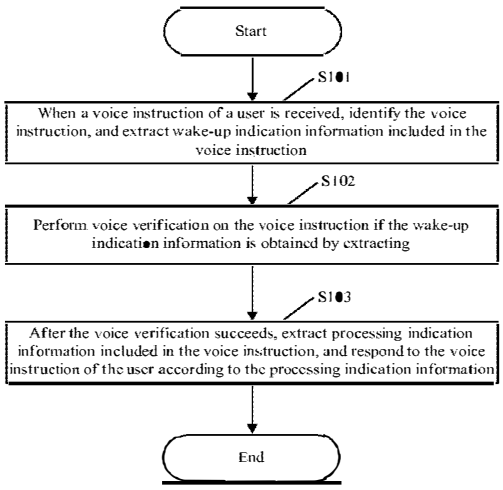


FIG. 1

EP 2 899 955 A1

Printed by Jouve, 75001 PARIS (FR)

Description

[0001] This application claims priority to Chinese Patent Application No. CN 201310574762.8, filed with the Chinese Patent Office on November 15, 2013 and entitled "TERMINAL VOICE CONTROL METHOD AND APPARETUS, AND TERMINAL", which is incorporated herein by reference in its entirety.

TECHNICAL FIELD

[0002] The present invention relates to the field of terminal technologies, and in particular, to a terminal voice control method and apparatus, and a terminal.

BACKGROUND

[0003] With the development of science and technology, smartphones have become necessities in people's lives. In order to facilitate a user to perform an operation on a smartphone, most smartphones currently have a voice control function. A user can issue a voice instruction to a mobile phone, and the mobile phone makes a response to the voice instruction. However, in the prior art, the foregoing solution is implemented by separately invoking different voice engines for multiple times. For example, a voice wake-up engine and a voiceprint recognition engine are used to identify whether a user is the owner first, and then the mobile phone is woken up, and the mobile phone then inquires about a requirement of the user and directs the user to answer questions step by step. After the user answers the questions, the mobile phone identifies a voice instruction that is issued by the user when the user answers the questions, and executes the voice instruction of the user according to an identification result.

[0004] A defect of the foregoing voice control manner is that: multiple voice instructions are inconsecutive, resulting in that a user needs to be inquired step by step and directed to answer questions; a next step can be started only after the user finishes answering the questions; a process of performing each step is cumbersome, which causes great inconvenience for the user to use a voice control function of a mobile phone. When the mobile phone is standby, no instruction can be directly issued to the mobile phone. The mobile phone needs to be woken up before a further operation can be performed on the mobile phone.

SUMMARY

[0005] According to a terminal voice control method provided in embodiments of the present invention, one voice instruction may be used to control a terminal to complete multiple tasks.

[0006] According to a first aspect, an embodiment of the present invention provides a terminal voice control method, including:

when a voice instruction of a user is received, identifying the voice instruction, and extracting wake-up indication information from the voice instruction;

performing voice verification on the voice instruction if the wake-up indication information is obtained by extracting; and

extracting, after the voice verification succeeds, processing indication information from the voice instruction, and responding to the voice instruction of the user according to the processing indication information.

[0007] With reference to the first aspect, in a first possible implementation manner of the first aspect, the extracting, after the voice verification succeeds, processing indication information from the voice instruction, and responding to the voice instruction of the user according to the processing indication information includes:

after the voice verification succeeds, if it is detected that a terminal is in a screen locking state, lighting up a terminal screen and unlocking the terminal screen, extracting the processing indication information from the voice instruction, and responding to the voice instruction of the user according to the processing indication information; and

after the voice verification succeeds, if it is detected that the terminal is not in a screen locking state, extracting the processing indication information from the voice instruction, and responding to the voice instruction of the user according to the processing indication information.

[0008] With reference to the first aspect, in a third possible implementation manner of the first aspect, the method further includes:

if no wake-up indication information is obtained by extracting, continuing to wait for receiving a voice instruction of the user.

[0009] With reference to the first aspect, in a fourth possible implementation manner of the first aspect, the method further includes:

prompting reverification if the voice verification fails, and stopping receiving a voice instruction of the user if the number of times of verification exceeds a threshold but the verification still fails.

[0010] With reference to the first aspect, in a fifth possible implementation manner of the first aspect, the extracting, after the voice verification succeeds, processing indication information from the voice instruction, and responding to the voice instruction of the user according to the processing indication information includes:

removing the wake-up indication information from

Petitioner Exhibit 1004, Page 883 of 1518

the voice instruction, extracting the processing prompt information in the voice instruction, and converting the processing indication information into a text; and

identifying a meaning of the processing prompt information according to the text, and making a response according to the obtained meaning.

[0011] According to a second aspect, an embodiment of the present invention provides a terminal voice control apparatus, including:

a receiving unit, configured to, when a voice instruction of a user is received, identify the voice instruction, and extract wake-up indication information from the voice instruction;

a first extracting unit, configured to perform voice verification on the voice instruction if the wake-up indication information is obtained by extracting; and

a first processing unit, configured to, after the voice verification succeeds, extract processing indication information included in the voice instruction, and respond to the voice instruction of the user according to the processing indication information.

[0012] With reference to the second aspect, in a first possible implementation manner of the second aspect, the first processing unit includes:

a first executing module, configured to, after the voice verification succeeds, if it is detected that a terminal is in a screen locking state, light up a terminal screen and unlock the terminal screen, extract the processing indication information from the voice instruction, and respond to the voice instruction of the user according to the processing indication information; and

a second executing module, configured to, after the voice verification succeeds, if it is detected that the terminal is not in a screen locking state, extract the processing indication information from the voice instruction, and respond to the voice instruction of the user according to the processing indication information.

[0013] With reference to the second aspect, in a third possible implementation manner of the second aspect, the apparatus further includes:

a second extracting unit, configured to, if no wake-up indication information is obtained by extracting, continue to wait for receiving a voice instruction of the user.

[0014] With reference to the second aspect, in a fourth possible implementation manner of the second aspect, the apparatus includes:

a second processing unit, configured to prompt reverification when the voice verification fails, and stop receiving a voice instruction of the user if the number of times of verification exceeds a threshold but the verification still fails.

[0015] With reference to the second aspect, in a fifth possible implementation manner of the second aspect, the first processing unit includes:

a converting module, configured to remove the wake-up indication information from the voice instruction, extract the processing prompt information in the voice instruction, and convert the processing indication information into a text; and

a responding module, configured to identify a meaning of the processing prompt information according to the text, and make a response according to the obtained meaning.

[0016] According to a third aspect, an embodiment of the present invention further provides a terminal, where the terminal includes an audio monitoring unit and a processor, where:

the audio monitoring unit is configured to pick up a voice instruction issued by a user; and

the processor is configured to, when the voice instruction of the user is received, identify the voice instruction, and extract wake-up indication information from the voice instruction; perform voice verification on the voice instruction if the wake-up indication information is obtained by extracting; and after the voice verification succeeds, extract processing indication information from the voice instruction, and respond to the voice instruction of the user according to the processing indication information.

[0017] According to the terminal voice control method in the embodiments of the present invention, a user may use one voice instruction to control a terminal to complete multiple tasks without requiring voice interaction for multiple times, which is simple to use, and has high security performance.

BRIEF DESCRIPTION OF DRAWINGS

[0018] To describe the technical solutions in the embodiments of the present invention or in the prior art more clearly, the following briefly introduces the accompanying drawings required for describing the embodiments or the prior art. Apparently, the accompanying drawings in the following description show some embodiments of the present invention, and a person of ordinary skill in the art may still derive other drawings from these accompanying drawings without creative efforts.

FIG. 1 is a flowchart of a terminal voice control meth-

Petitioner Exhibit 1004, Page 884 of 1518

od according to a first embodiment of the present invention;

FIG. 2 is a flowchart of a terminal voice control method according to a second embodiment of the present invention;

FIG. 3 is a flowchart of a terminal voice control method according to a third embodiment of the present invention;

FIG. 4 is a schematic structural diagram of a terminal voice control apparatus according to a fourth embodiment of the present invention;

FIG. 5 is a schematic structural diagram of a terminal voice control apparatus according to a fifth embodiment of the present invention;

FIG. 6 is a schematic structural diagram of a terminal voice control apparatus according to a sixth embodiment of the present invention; and

FIG. 7 is a schematic structural diagram of a terminal according to a seventh embodiment of the present invention.

DESCRIPTION OF EMBODIMENTS

[0019] The following clearly and completely describes the technical solutions in the embodiments of the present invention with reference to the accompanying drawings in the embodiments of the present invention. Apparently, the described embodiments are a part rather than all of the embodiments of the present invention. All other embodiments obtained by persons of ordinary skill in the art based on the embodiments of the present invention without creative efforts shall fall within the protection scope of the present invention.

[0020] Refer to FIG. 1, which is a flowchart of a terminal voice control method according to a first embodiment of the present invention. A terminal described in this embodiment of the present invention includes an electronic device such as a mobile phone, a PDA, or a tablet computer. The method specifically includes:

[0021] S101. When a voice instruction of a user is received, identify the voice instruction, and extract wake-up indication information from the voice instruction.

[0022] Specifically, the voice instruction of the user includes the wake-up indication information and processing indication information, where the wake-up indication information is used to wake up a terminal that is in a screen locking state, and the processing indication information is used to instruct the terminal to execute a specific action.

[0023] Identifying the voice instruction of the user is detecting whether the voice instruction of the user includes a preset wake-up word, for example, a word preset by the user, such as hi or hello, so as to extract the wake-up indication information included in the voice instruction of the user.

[0024] S102. Perform voice verification on the voice instruction if the wake-up indication information is obtained by extracting.

[0025] Specifically, performing the voice verification on the voice instruction is performing voiceprint recognition on a voice in the voice instruction, so as to determine whether the user who issues the voice instruction is a preset user, so that only a user who passes the verification can use the terminal and further perform a voice operation on the terminal, thereby improving security performance of the terminal.

[0026] It should be understood that voiceprint recognition (Voiceprint Recognition) is also referred to as speaker recognition, which refers to identifying, according to a voice feature of a person, who speaks a segment of voice. Generally, voiceprint recognition includes two aspects: speaker identifying and speaker determining, where the former is to determine which one of several persons speaks a segment of voice, and the latter is to determine whether a segment of voice is spoken by a specified person. The voiceprint recognition manner in this embodiment of the present invention is speaker determining, that is, whether a segment of voice is spoken by a preset specified person needs to be identified.

[0027] S103. After the voice verification succeeds, extract processing indication information from the voice instruction, and respond to the voice instruction of the user according to the processing indication information.

[0028] Specifically, after the voice verification succeeds, the wake-up indication information is removed from the voice instruction; the processing prompt information in the voice instruction is extracted, and the processing prompt information is converted into a text; and a meaning of the processing indication information is identified according to the text, and a response is made according to the obtained meaning.

[0029] According to the terminal voice control method provided in this embodiment of the present invention, a user may use one voice instruction to control a terminal to complete multiple tasks without requiring voice interaction for multiple times, which is simple to use.

[0030] Refer to FIG. 2, which is a flowchart of a terminal voice control method according to a second embodiment of the present invention. A terminal described in this embodiment of the present invention includes an electronic device such as a mobile phone, a PDA, or a tablet computer. The method specifically includes:

[0031] S201. When a voice instruction of a user is received, identify the voice instruction.

[0032] Specifically, identifying the voice instruction of the user is detecting whether the voice instruction of the user includes a preset wake-up word, for example, a preset word such as hi or hello, so as to extract wake-up indication information from the voice instruction of the user.

[0033] S202. Perform voice verification on the voice instruction if wake-up indication information is obtained by extracting.

[0034] Specifically, performing the voice verification on the voice instruction is performing voiceprint recognition on a voice in the voice instruction, so as to determine

Petitioner Exhibit 1004, Page 885 of 1518

whether the user who issues the voice instruction is a preset user, so that only a user who passes the verification can use the terminal, and further perform a voice operation on the terminal, thereby improving security performance of the terminal.

[0035] S203. After the voice verification succeeds, if it is detected that the terminal is in a screen locking state, light up a terminal screen and unlock the terminal screen, extract processing indication information from the voice instruction, and respond to the voice instruction of the user according to the processing indication information.

[0036] S204. After the voice verification succeeds, if it is detected that the terminal is not in a screen locking state, extract the processing indication information from the voice instruction, and respond to the voice instruction of the user according to the processing indication information.

[0037] Specifically, if it is detected that the terminal is in a screen locking state, the terminal that is in a screen locking state needs to be woken up first according to the wake-up indication information in the voice instruction, and then another step is performed. If the terminal is not in a screen locking state, the step of lighting up the terminal screen or unlocking the terminal screen does not need to be performed, and a next step is directly performed.

[0038] According to the terminal voice control method provided in this embodiment of the present invention, a user may use one voice instruction to control a terminal to complete multiple tasks without requiring voice interaction for multiple times, which is simple to use.

[0039] Refer to FIG. 3, which is a flowchart of a terminal voice control method according to a third embodiment of the present invention. A terminal described in this embodiment of the present invention includes an electronic device such as a mobile phone, a PDA, or a tablet computer. The method specifically includes:

[0040] S301. When a voice instruction of a user is received, identify the voice instruction, and extract wake-up indication information from the voice instruction.

[0041] Specifically, the voice instruction of the user includes the wake-up indication information and processing indication information, where the wake-up indication information is used to wake up a terminal that is in a screen locking state, and the processing indication information is used to instruct the terminal to execute a specific action.

[0042] Identifying the voice instruction of the user is detecting whether the voice instruction of the user includes a preset wake-up word, for example, a preset word such as hi or hello, so as to extract the wake-up indication information included in the voice instruction of the user.

[0043] S302. Perform voice verification on the voice instruction if the wake-up indication information is obtained by extracting.

[0044] Specifically, performing the voice verification on the voice instruction is performing voiceprint recognition on a voice in the voice instruction, so as to determine whether the user who issues the voice instruction is a preset user, so that only a user who passes the verification can use the terminal, and further perform a voice operation on the terminal, thereby improving security performance of the terminal.

[0045] S303. If no wake-up indication information is obtained by extracting, continue to wait for receiving a voice instruction of the user.

[0046] Specifically, if no wake-up indication information is obtained by extracting, the terminal does not perform any processing, and continues to wait for receiving a voice instruction of the user.

[0047] S304. After the voice verification succeeds, extract processing indication information from the voice instruction, and respond to the voice instruction of the user according to the processing indication information.

[0048] Specifically, after the voice verification succeeds, the wake-up indication information is removed from the voice instruction; the processing prompt information in the voice instruction is extracted, and the processing indication information is converted into a text; and a meaning of the processing prompt information is identified according to the text, and a response is made according to the obtained meaning.

[0049] S305. Prompt reverification if the voice verification fails, and stop receiving a voice instruction of the user if the number of times of verification exceeds a threshold but the verification still fails.

[0050] Specifically, if the number of times of verification exceeds the preset threshold but the verification still fails, a voice instruction of the user stops being received, and the user can only turn on a terminal screen manually, thereby ensuring security of using the terminal.

[0051] According to the terminal voice control method provided in this embodiment of the present invention, a user may use one voice instruction to control a terminal to complete multiple tasks without requiring voice interaction for multiple times, which is simple to use, and has high security performance.

[0052] Refer to FIG. 4, which is a schematic structural diagram of a terminal voice control apparatus according to a fourth embodiment of the present invention. A terminal described in this embodiment of the present invention includes an electronic device such as a mobile phone, a PDA, or a tablet computer. The apparatus 1 specifically includes a receiving unit 10, a first extracting unit 20, and a first processing unit 30.

[0053] The receiving unit 10 is configured to, when a voice instruction of a user is received, identify the voice instruction, and extract wake-up indication information from the voice instruction.

[0054] Specifically, the voice instruction of the user includes the wake-up indication information and processing indication information, where the wake-up indication information is used to wake up a terminal that is in a screen locking state, and the processing indication information is used to instruct the terminal to execute a specific action.

Petitioner Exhibit 1004, Page 886 of 1518

[0055] Identifying the voice instruction of the user is detecting whether the voice instruction of the user includes a preset wake-up word, for example, a preset wake-up word such as hi or hello, so as to extract the wake-up indication information included in the voice instruction of the user.

[0056] The first extracting unit 20 is configured to perform voice verification on the voice instruction if the wake-up indication information is obtained by extracting.

[0057] Specifically, performing the voice verification on the voice instruction is performing voiceprint recognition on a voice in the voice instruction, so as to determine whether the user who issues the voice instruction is a preset user, so that only a user who passes the verification can use the terminal, and further perform a voice operation on the terminal, thereby improving security performance of the terminal.

[0058] The first processing unit 30 is configured to, after the voice verification succeeds, extract processing indication information from the voice instruction, and respond to the voice instruction of the user according to the processing indication information.

[0059] The first processing unit 30 includes a converting module 31 and a responding module 32.

[0060] The converting module 31 is configured to remove the wake-up indication information from the voice instruction, extract the processing prompt information in the voice instruction, and convert the processing indication information into a text.

[0061] The responding module 32 is configured to identify a meaning of the processing prompt information according to the text, and make a response according to the obtained meaning.

[0062] Specifically, after the voice verification succeeds, the wake-up indication information is removed from the voice instruction; the processing prompt information in the voice instruction is extracted, and the processing indication information is converted into a text; and a meaning of the processing prompt information is identified according to the text, and a response is made according to the obtained meaning.

[0063] According to the terminal voice control method provided in this embodiment of the present invention, a user may use one voice instruction to control a terminal to complete multiple tasks without requiring voice interaction for multiple times, which is simple to use.

[0064] Refer to FIG. 5, which is a schematic structural diagram of a terminal voice control apparatus according to a fifth embodiment of the present invention. A terminal described in this embodiment of the present invention includes an electronic device such as a mobile phone, a PDA, or a tablet computer. The apparatus 2 is an optimization of the apparatus 1 in FIG. 4. A first processing unit 30 of the apparatus 2 includes a first executing module 33 and a second executing module 34.

[0065] The first executing module 33 is configured to, after voice verification succeeds, if it is detected that the terminal is in a screen locking state, light up a terminal screen and unlock the terminal screen, extract processing indication information from the voice instruction, and respond to the voice instruction of the user according to the processing indication information.

[0066] The second executing module 34 is configured to, after the voice verification succeeds, if it is detected that the terminal is not in a screen locking state, extract the processing indication information from the voice instruction, and respond to the voice instruction of the user according to the processing indication information.

[0067] Specifically, if it is detected that the terminal is in a screen locking state, the step of lighting up the terminal screen and unlocking the terminal screen needs to be performed; if the terminal is not in a screen locking state, the step of lighting up the terminal screen and unlocking the terminal screen does not need to be performed.

[0068] In addition, performing voice verification on the voice instruction is using a voiceprint recognition method to verify whether the user who issues the voice instruction is a preset user, thereby improving security of the terminal.

[0069] According to the terminal voice control method provided in this embodiment of the present invention, a user may use one voice instruction to control a terminal to complete multiple tasks without requiring voice interaction for multiple times, which is simple to use.

[0070] Refer to FIG. 6, which is a schematic structural diagram of a terminal voice control apparatus according to a sixth embodiment of the present invention. A terminal described in this embodiment of the present invention includes an electronic device such as a mobile phone, a PDA, or a tablet computer. The apparatus 3 is an optimization of the apparatus 1 in FIG. 4. The apparatus 3 further includes a second extracting unit 40 and a second processing unit 50.

[0071] The second extracting unit 40 is configured to, if no wake-up indication information is obtained by extracting, continue to wait for receiving a voice instruction of a user.

[0072] The second processing unit 50 is configured to prompt reverification when voice verification fails, and stop receiving a voice instruction of the user if the number of times of verification exceeds a threshold but the verification still fails.

[0073] Specifically, if the number of times of voice verification exceeds the preset threshold but the verification still fails, a voice instruction of the user stops being received, and the user can only turn on a terminal screen manually, thereby ensuring security of using the terminal.

[0074] According to the terminal voice control method provided in this embodiment of the present invention, a user may use one voice instruction to control a terminal to complete multiple tasks without requiring voice interaction for multiple times, which is simple to use, and has high security performance.

[0075] Refer to FIG. 7, which is a schematic structural diagram of a terminal according to a seventh embodiment

Petitioner Exhibit 1004, Page 887 of 1518

of the present invention. The terminal described in this embodiment of the present invention includes an electronic device such as a mobile phone, a PDA, or a tablet computer. The terminal 4 specifically includes an audio monitoring unit 60 and a processor 70.

[0076] The audio monitoring unit 60 is configured to pick up a voice instruction issued by a user.

[0077] The processor 70 is configured to, when the voice instruction of the user is received, identify the voice instruction and extract wake-up indication information from the voice instruction; perform voice verification on the voice instruction if the wake-up indication information is obtained by extracting; and after the voice verification succeeds, extract processing indication information from the voice instruction, and respond to the voice instruction of the user according to the processing indication information.

[0078] Specifically, identifying the voice instruction of the user is detecting whether the voice instruction of the user includes a preset wake-up word, for example, a word such as hi or hello preset by the user, so as to extract the wake-up indication information included in the voice instruction of the user.

[0079] Performing the voice verification on the voice instruction is performing voiceprint recognition on a voice in the voice instruction, so as to determine whether the user who issues the voice instruction is a preset user, so that only a user who passes the verification can use the terminal, and further perform a voice operation on the terminal, thereby improving security performance of the terminal.

[0080] After the voice verification succeeds, the wake-up indication information is removed from the voice instruction; the processing prompt information in the voice instruction is extracted, and the processing indication information is converted into a text; and a meaning of the processing prompt information is identified according to the text, and a response is made according to the obtained meaning.

[0081] Definitely, in an exemplary implementation manner of this embodiment of the present invention, the audio monitoring unit 60 may be a microphone.

[0082] According to the terminal voice control method provided in this embodiment of the present invention, a user may use one voice instruction to control a terminal to complete multiple tasks without requiring voice interaction for multiple times, which is simple to use, and has high security performance.

[0083] It should be understood that the first extracting unit and the second extracting unit, the first processing unit and the second processing unit, and the first executing module and the second executing module do not imply a sequence relationship or quantity unit, but are intended to distinguish different modules or units.

[0084] Persons of ordinary skill in the art may understand that, various aspects or possible implementations of various aspects of the present invention may be specifically implemented in a system, method, or computer program product. Moreover, various aspects or possible implementations of various aspects of the present invention may take a form of computer program product, where the computer program product refers to computer readable program code stored in a computer readable medium.

[0085] A computer readable medium may be a computer readable signal medium or computer readable storage medium. The computer readable storage medium includes but not limited to electronic, magnetic, optical, electromagnetic, infrared, or semi-conductive system, device, or apparatus, or any suitable combination thereof, for example, a random access memory (RAM), a read only memory (ROM), an erasable programmable read only memory (EPROM or FLASH), an optical fiber, a compact disc read only memory (CD-ROM).

[0086] A processor in a computer reads computer readable program code stored in the computer readable medium, so that a specific function and action in each step or in a combination of steps in the flowcharts may be executed by the processor; generates an apparatus for implementing a specific function and action in each block or in a combination of blocks in the block diagrams.

[0087] Computer readable program code may be all executed on a user computer alone, or a part thereof may be executed on a user computer as a standalone software package, or a part thereof may be executed on a local computer while the other part is executed on a remote computer, or may be all executed on a remote computer or server. It should also be noted that, in some alternative implementation solutions, each step in the flowcharts or specific functions in each block in the block diagrams may not occur in the illustrated order. For example, two consecutive steps in the illustration which are dependent on an involved function, or two blocks may in fact be executed substantially simultaneously, or these blocks may sometimes be executed in a reverse order.

[0088] It is apparent that persons skilled in the art can make various modifications and variations to the present invention without departing from the spirit and scope of the present invention. The present invention is intended to cover these modifications and variations provided that they fall within the scope of protection defined by the following claims or their equivalent technologies.

Claims

1. A terminal voice control method, comprising:

   when a voice instruction of a user is received, identifying the voice instruction, and extracting wake-up indication information from the voice instruction;

   performing voice verification on the voice instruction if the wake-up indication information is obtained by extracting; and

   extracting, after the voice verification succeeds,

Petitioner Exhibit 1004, Page 888 of 1518

processing indication information from the voice instruction, and responding to the voice instruction of the user according to the processing indication information.

2. The method according to claim 1, wherein the extracting, after the voice verification succeeds, processing indication information from the voice instruction, and responding to the voice instruction of the user according to the processing indication information comprises:

after the voice verification succeeds, if it is detected that a terminal is in a screen locking state, lighting up a terminal screen and unlocking the terminal screen, extracting the processing indication information from the voice instruction, and responding to the voice instruction of the user according to the processing indication information; and
after the voice verification succeeds, if it is detected that the terminal is not in a screen locking state, extracting the processing indication information from the voice instruction, and responding to the voice instruction of the user according to the processing indication information.

3. The method according to claim 1, further comprising:

if no wake-up indication information is obtained by extracting, continuing to wait for receiving a voice instruction of the user.

4. The method according to claim 1, further comprising:

prompting reverification if the voice verification fails, and stopping receiving a voice instruction of the user if the number of times of verification exceeds a threshold but the verification still fails.

5. The method according to claim 1, wherein the extracting, after the voice verification succeeds, processing indication information from the voice instruction, and responding to the voice instruction of the user according to the processing indication information comprises:

extracting the processing prompt information in the voice instruction after the wake-up indication information is removed from the voice instruction, and converting the processing indication information into a text; and
identifying a meaning of the processing prompt information according to the text, and making a response according to the obtained meaning.

6. A terminal voice control apparatus, comprising:

a receiving unit, configured to, when a voice instruction of a user is received, identify the voice instruction, and extract wake-up indication information from the voice instruction;
a first extracting unit, configured to perform voice verification on the voice instruction if the wake-up indication information is obtained by extracting; and
a first processing unit, configured to, after the voice verification succeeds, extract processing indication information from the voice instruction, and respond to the voice instruction of the user according to the processing indication information.

7. The apparatus according to claim 6, wherein the first processing unit comprises:

a first executing module, configured to, after the voice verification succeeds, if it is detected that a terminal is in a screen locking state, light up a terminal screen and unlock the terminal screen, extract the processing indication information from the voice instruction, and respond to the voice instruction of the user according to the processing indication information; and
a second executing module, configured to, after the voice verification succeeds, if it is detected that the terminal is not in a screen locking state, extract the processing indication information from the voice instruction, and respond to the voice instruction of the user according to the processing indication information.

8. The apparatus according to claim 6, further comprising:

a second extracting unit, configured to, if no wake-up indication information is obtained by extracting, continue to wait for receiving a voice instruction of the user.

9. The apparatus according to claim 6, comprising:

a second processing unit, configured to prompt reverification when the voice verification fails, and stop receiving a voice instruction of the user if the number of times of verification exceeds a threshold but the verification still fails.

10. The apparatus according to claim 6, wherein the first processing unit comprises:

a converting module, configured to extract the processing prompt information in the voice instruction after the wake-up indication information is removed from the voice instruction, and convert the processing indication information in-

Petitioner Exhibit 1004, Page 889 of 1518

to a text; and
a responding module, configured to identify a meaning of the processing prompt information according to the text, and make a response according to the obtained meaning.

11. A terminal, comprising an audio monitoring unit and a processor, wherein:

the audio monitoring unit is configured to pick up a voice instruction issued by a user; and
the processor is configured to, when the voice instruction of the user is received, identify the voice instruction, and extract wake-up indication information from the voice instruction; perform voice verification on the voice instruction if the wake-up indication information is obtained by extracting; and after the voice verification succeeds, extract processing indication information from the voice instruction, and respond to the voice instruction of the user according to the processing indication information.

Petitioner Exhibit 1004, Page 890 of 1518

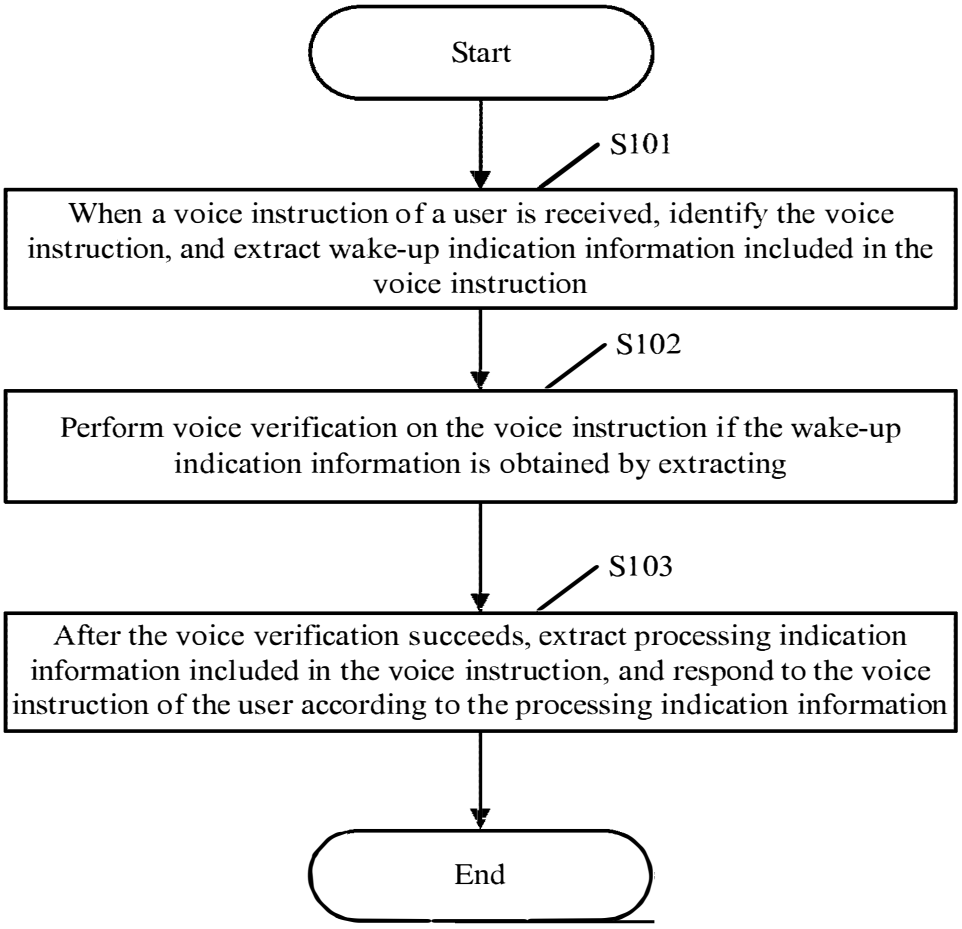


FIG. 1

Petitioner Exhibit 1004, Page 891 of 1518

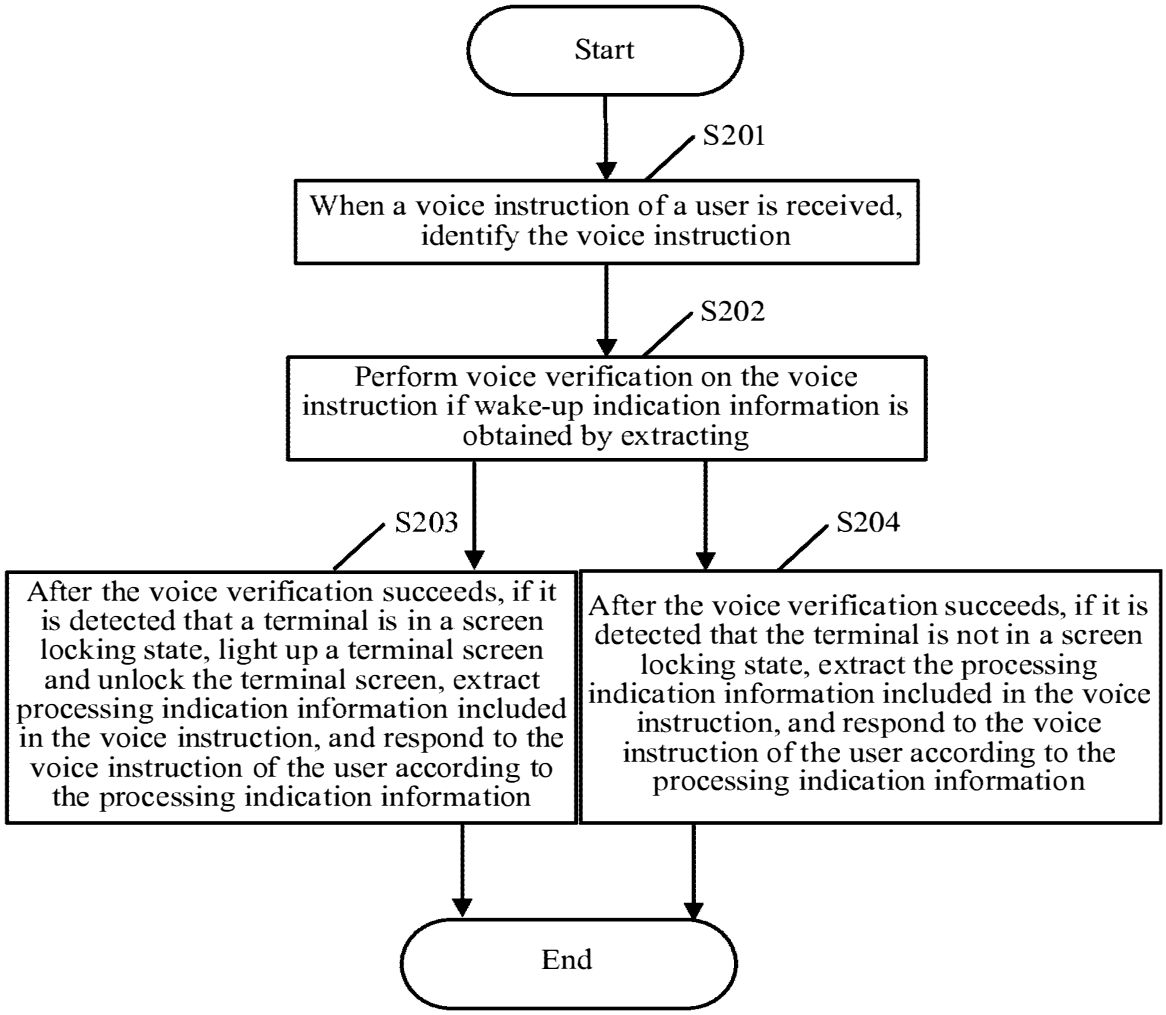


FIG. 2

Petitioner Exhibit 1004, Page 892 of 1518

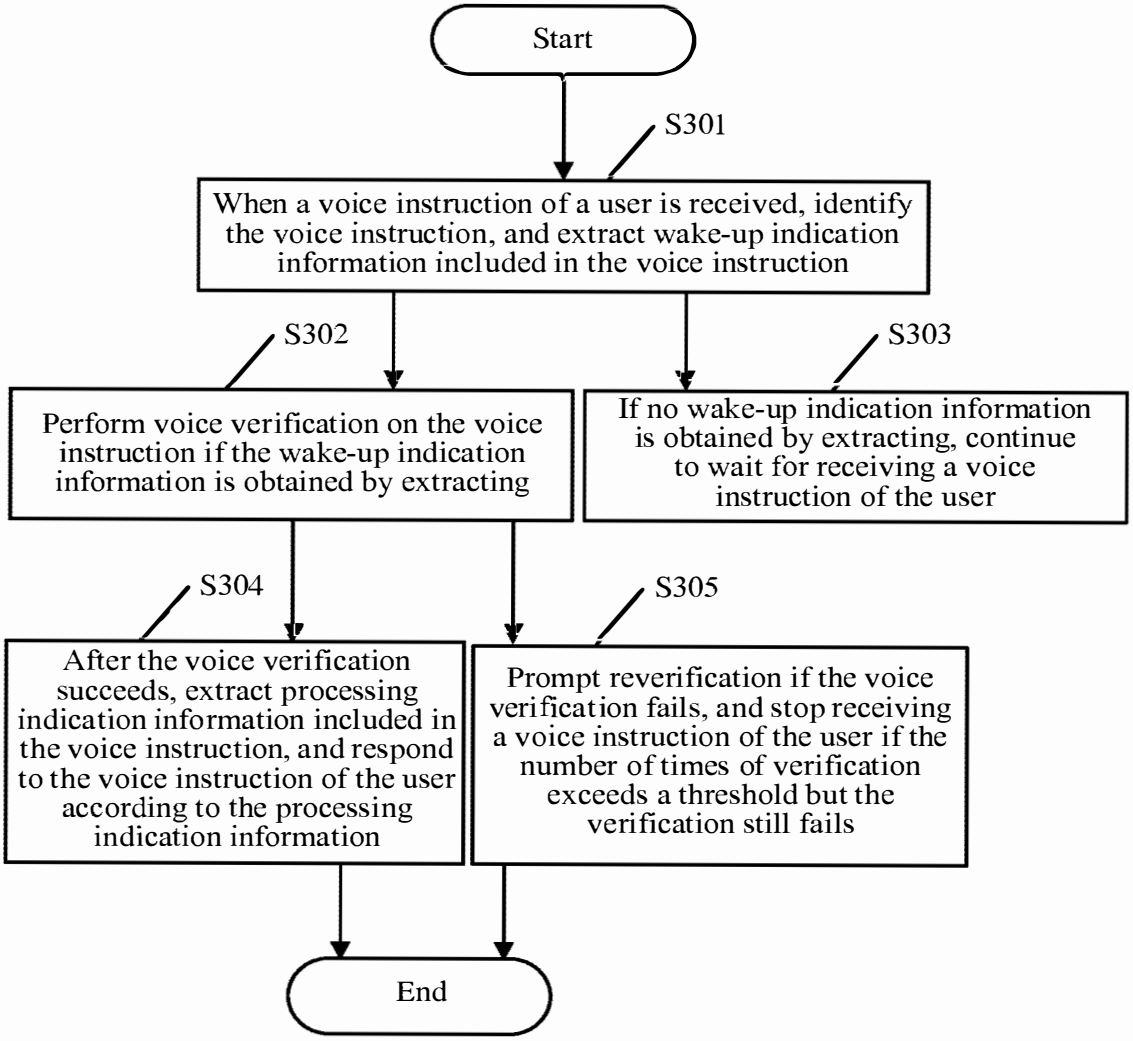


FIG. 3

Petitioner Exhibit 1004, Page 893 of 1518

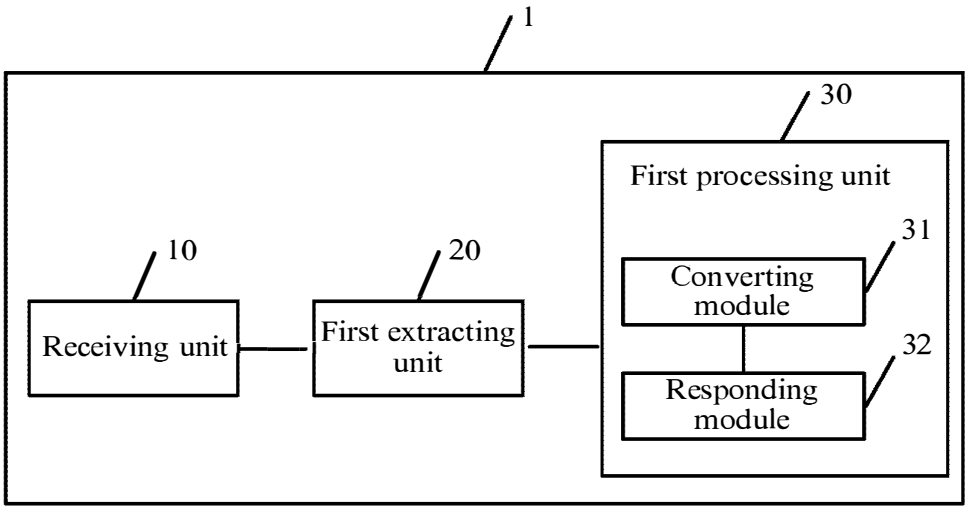


FIG. 4

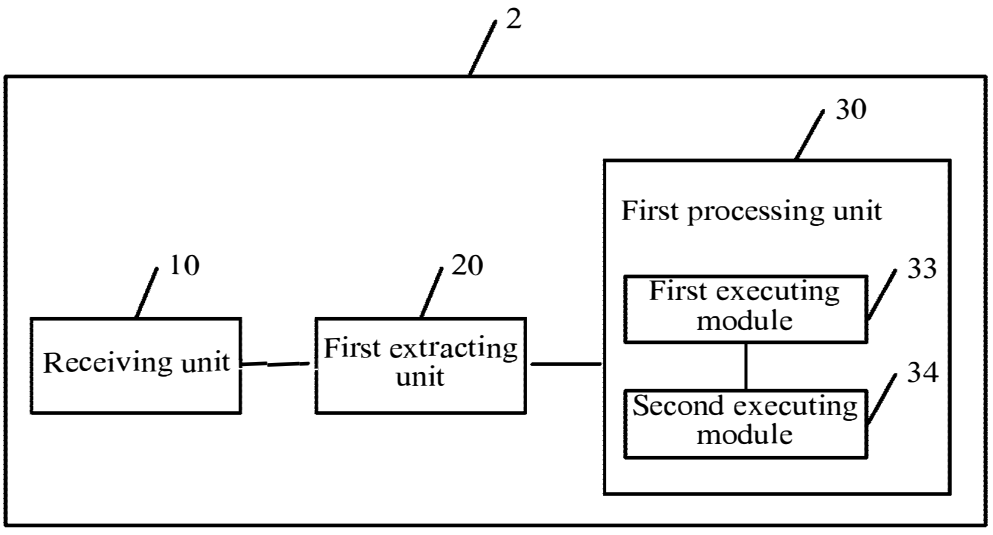


FIG. 5

Petitioner Exhibit 1004, Page 894 of 1518

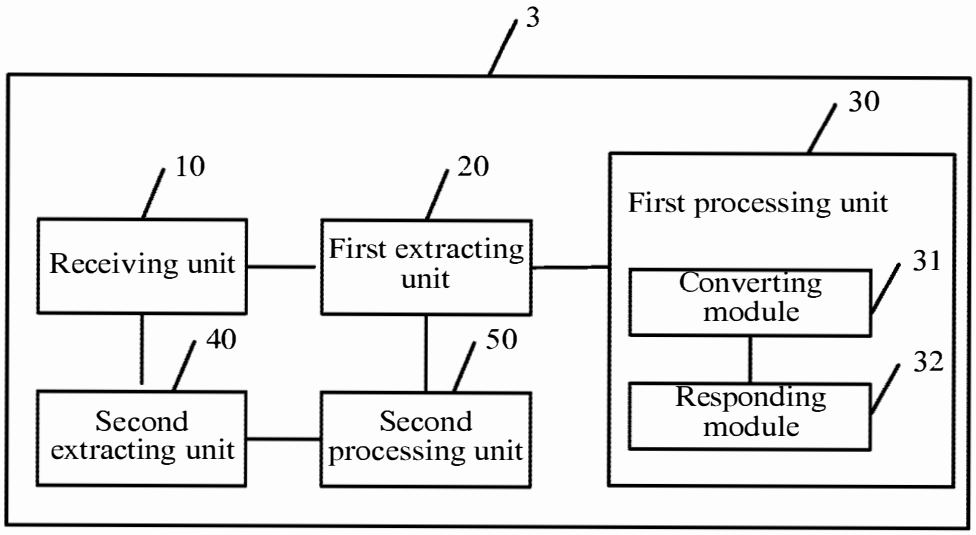


FIG. 6

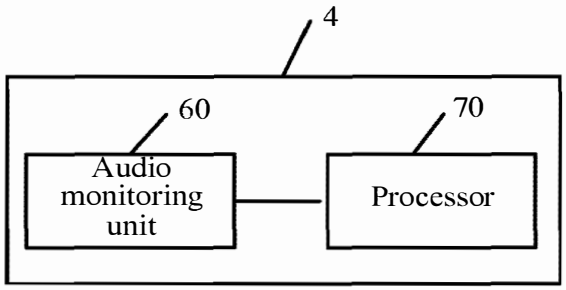


FIG. 7

Petitioner Exhibit 1004, Page 895 of 1518

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2014/083507

A. CLASSIFICATION OF SUBJECT MATTER

H04M 1/725 (2006.01) i; G10L 7/08 (2013.01) i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

H04M, H04W, G10L, G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CPRSABS, CNTXT, CNKI, VEN: handset, communication terminal, mobile telephone, mobile terminal, cellular phone, portable terminal, communication device, voice, sound, control, awak+, instruct+, command, unlock, recogniz+, extract, verify

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 103198831 A (VIA TECHNOLOGIES, INC.), 10 July 2013 (10.07.2013), description, paragraphs [0046]-[0076], claims 1-12, and figures 2-5	1, 3-4, 6, 8-9, 11
Y	CN 103198831 A (VIA TECHNOLOGIES, INC.), 10 July 2013 (10.07.2013), description, paragraphs [0046]-[0076], claims 1-12, and figures 2-5	2, 5, 7, 10
Y	CN 102932539 A (SHENZHEN ZTE MOBILE TELECOM CO., LTD.), 13 February 2013 (13.02.2013), description, paragraphs [0016]-[0038], claims 1-7, and figure 1	2, 5, 7, 10
A	CN 102546953 A (GIONEE COMMUNICATION EQUIPMENT CO., LTD.), 04 July 2012 (04.07.2012), the whole document	1-11
A	CN 102148899 A (GUANG DONG OPPO MOBILE TELECOMMUNICATIONS CO., LTD.), 10 August 2011 (10.08.2011), the whole document	1-11
PX	CN 103595869 A (HUAWEI DEVICE CO., LTD.), 19 February 2014 (19.02.2014), claims 1-11	1-11

☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
20 October 2014 (20.10.2014)	**04 November 2014 (04.11.2014)**
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer **LI, Zhenhua** Telephone No.: (86-10) **62087678**

Form PCT/ISA/210 (second sheet) (July 2009)

Petitioner Exhibit 1004, Page 896 of 1518

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2014/083507

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 103198831 A	10 July 2013	None	
CN 102932539 A	13 February 2013	None	
CN 102546953 A	04 July 2012	None	
CN 102148899 A	10 August 2011	None	
CN 103595869 A	19 February 2014	None	

Form PCT/ISA/210 (patent family annex) (July 2009)

Petitioner Exhibit 1004, Page 897 of 1518

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- CN 201310574762 **[0001]**

Petitioner Exhibit 1004, Page 898 of 1518

Publication number: CN104575504 A
Publication country: CHINA
Publication type: UNEXAMINED APPLIC. OPEN TO PUBLIC INSPECTION
Publication date: 20150429
Application number: CN201410840544
Application date: 20141224
Priority: CN201410840544 20141224 CN201410840544;
Assignee: SHANGHAI NORMAL UNIV ;
Assignee[std]: UNIV SHANGHAI ;
Inventor[std]: YE HONG ; WEI HAORAN ; LONG YANHUA ;
International class[8]: G10L17/04; G10L15/26; H04N21/439; H04N21/441;
Family members: CN104575504 A

Title: METHOD FOR PERSONALIZED TELEVISION VOICE WAKE-UP BY VOICEPRINT AND VOICE IDENTIFICATION

Title: 采用声纹和语音识别进行个性化电视语音唤醒的方法

Abstract:

The invention discloses a method for personalized television voice wake-up by voiceprint and voice identification, particularly a method for performing identity confirmation on a television user through voiceprint identification and controlling a television to perform personalized voice wake-up through confirmed identity and a voice identification result of user voice, and relates to voiceprint identification and voice identification technologies. A composition system comprises a voice control system (1), an information storage unit (2) and a television main controller (3) which are connected through electric signals. The method has the characteristics of short training time, very high voiceprint and voice identification speed and high identification rate. Voiceprint and voice identification can be finished by only offline training and testing, identification results do not need to be sent to a cloud server, use is convenient, and the safety of family information is guaranteed. The method also can be applied to user-personalized automatic voice channel change of the television, can be transplanted to a common high-speed DSP (digital signal processor) or chip for operation, and can be widely applied to the related fields of smart homes.

Abstract:

本发明采用声纹和语音识别进行个性化电视语音唤醒的方法，具体指一种通过声纹识别进行电视机使用用户的身份确认，并通过确认的身份和对用户语音的语音识别结果来控制电视机进行个性化的语音唤醒，涉及声纹识别和语音识别技术。本发明构成系统包括语音控制系统(1)，信息存储单元(2)和电视机主控制器(3)以电信号方式连接组成。本发明具

The present invention uses a voiceprint and voice recognition personalize Television Voice wake method, specifically refers to a voiceprint identification performed by the TV using the user's identification, and identification by the user's voice and speech recognition results to control the TV voice wake personalize involving voiceprint recognition and speech recognition technology.

有训练时间短，声纹和语音识别速度都很快，且识别率高的特点。声纹和语音识别均仅需离线训练和测试就可完成，无需发送到云端服务器，使用方便且保证了家庭信息的安全性。还可应用于电视机的用户个性化自动语音换台，同时可被移植到通用的DSP高速处理器或芯片上运行，能广泛应用于智能家居相关领域。

Constitute a system of the present invention includes a voice control system (1), an information storage unit (2) and TV host controller (3) connected to the electrical components. The present invention has the training time is short, voiceprint and voice recognition speed very quickly, and recognition rate characteristics. Voiceprint and voice recognition are just off-line training and testing can be completed without the need to send to the cloud server, easy to use and ensure the safety of family information. Users can also be used in televisions personalized automated voice for Taiwan, and can be ported to run on general-purpose high-speed processor or DSP chips, widely used in smart home related fields.

Claims:

1.一种采用声纹和语音识别进行个性化电视语音唤醒的方法，通 过采用声纹识别和语音识别系统，对用户的身份进行监测，实现个性 化电视语音唤醒，其特征在于，包括如下步骤：

A.个性化信息注册

采用声纹识别系统建立用户的声纹模型，采用语音识别系统识别 出用户最喜爱的电视频道信息；

B.信息存储

将用户声纹模型和最喜爱的电视频道信息进行关联并存储到电 视机信息存储单元；

C.语音唤醒

用户只需对电视机的语音采集单元说出预先设定的唤醒词，即可 根据声纹识别和语音识别系统对唤醒词的声纹和语音识别结果从电 视机信息存储单元中调出用户的个性化信息，自动完成打开电视并自 动切换到用户最喜爱的电视频道。

2.如权利要求1所述的采用声纹和语音识别进行个性化电视语 音唤醒的方法，其特征在于，所述步骤A、步骤B，在用户作第一次 操作时完成，以后只需操作步骤C，即可完成其个性化语音电视的唤醒；

当用户对电视频道的喜好发生改变时，通过重新执行步骤A，进 行用户个性化信息注册，对信息存储单元中的电视频道信息进行更新， 此时无需进行声纹建模，仅需进行声纹匹配和语音识别。

1. One kind of speech recognition using the voiceprint and personalize the TV voice wake method, by using the voiceprint recognition and speech recognition systems, to monitor the user's identity, personalized television voice awakened, characterized by comprising the steps of:

A. Personalized Information Register

Using the voiceprint identification system to establish the user's voiceprint model, using voice recognition system identifies the user's favorite television channel information;

B. Information storage

The user voiceprint model and favorite TV channels and stores information related to the TV information storage unit;

C. Voice Wake Up

Users only need to tell the television voice collecting unit pre-set wake-up word, you can wake up the words of the voiceprint and voice recognition results from the television tune information storage unit based on the voice recognition and speech recognition systems the user's personalized information, auto-complete turn on the TV and automatically switch to the user's favorite TV channels.

2. Voiceprint as claimed in using voice recognition and speech personalized TV Wake method according to claim 1, characterized in that said step A, step B, the user for the first When once the