(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 102543071 A
(43)申请公布日 2012.07.04

(21)申请号 201110424181.7

(22)申请日 2011.12.16

(71)申请人 安徽科大讯飞信息科技股份有限公司
地址 230088 安徽省合肥市高新开发区望江西路 666 号

(72)发明人 王海坤 何婷婷 王智国 胡国平 胡郁 刘庆峰

(74)专利代理机构 中科专利商标代理有限责任公司 11021
代理人 朱进桂

(51)Int.Cl.
G10L 15/00(2006.01)
G10L 15/28(2006.01)

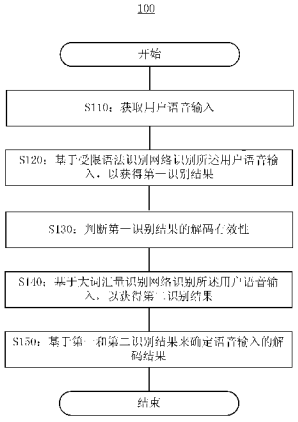
权利要求书 2 页 说明书 10 页 附图 5 页

(54)发明名称

用于移动设备的语音识别系统和方法

(57)摘要

本发明提供了一种应用于个人设备的语音识别系统和方法。该语音识别方法包括：获取用户语音输入，基于受限语法识别网络识别所述语音输入以获得第一识别结果；响应于第一识别结果不满足识别可接受条件，基于大词汇量连续语音识别网络识别所述语音输入以获得第二识别结果；以及选择所述第一和第二识别结果中的优选者作为所述语音输入的最终解码结果。本发明的实施例提供了一种新的语音识别方法和装置，其能够在统一的系统界面下支持对连续语音输入的智能响应及对简短语音命令的快速响应。



CN 102543071 A

1. 一种在移动设备或服务器上执行的语音识别方法，包括：

获取用户语音输入；

基于受限语法识别网络识别所述语音输入以获得第一识别结果；

响应于第一识别结果不满足识别可接受条件，基于大词汇量连续语音识别网络识别所述语音输入以获得第二识别结果；以及

选择所述第一和第二识别结果中的优选者作为所述语音输入的最终解码结果。

2. 根据权利要求 1 所述的语音识别方法，其中：

响应于第一识别结果满足识别可接受条件，直接以第一识别结果作为所述语音输入的最终解码结果。

3. 根据权利要求 2 所述的语音识别方法，其中：

所述识别可接受条件基于下述中的至少一种：所述语音输入的每帧语音的似然概率平均值、所述语音输入的各识别字符对应的概率得分、或置信度。

4. 根据权利要求 1 所述的语音识别方法，其中所述基于大词汇量连续语音识别网络识别所述语音输入以获得第二识别结果包括：

语音输入被提取为各个语音帧，以及通过在基于大词汇量连续语音识别网络定义的搜索空间中逐语音帧搜索最优路径来实现所述基于大词汇量连续语音识别网络的识别。

5. 根据权利要求 4 所述的语音识别方法，所述基于大词汇量连续语音识别网络的识别还根据实时解码状态提前终止搜索过程以提高解码效率，包括：

计算当前语音帧相应于所有活跃节点的最优历史路径并统计当前历史路径最大值 Si，

计算 Si 和当前语音帧在受限语法网络最优解码路径中的历史路径值 Si’的差值，以及

响应于上述差值小于预设的域值，终止所述搜索过程。

6. 一种用于移动设备的语音识别方法，包括：

获取用户语音输入；

基于受限语法识别网络识别所述语音输入以获得第一识别结果；

响应于第一识别结果不满足识别可接受条件，向服务器发送用户语音输入，以及从服务器接收基于大词汇量连续语音识别网络识别所述语音输入获得的第二识别结果；以及

选择所述第一和第二识别结果中的优选者作为所述语音输入的最终解码结果。

7. 一种语音识别系统，包括：

获取装置，用于获取用户语音输入，

第一识别装置，用于基于受限语法识别网络识别所述语音输入以获得第一识别结果，

第二识别装置，用于响应于第一识别结果不满足识别可接受条件，基于大词汇量连续语音识别网络识别所述语音输入以获得第二识别结果，以及

解码确定装置，用于选择所述第一和第二识别结果中的优选者作为所述语音输入的最终解码结果。

8. 根据权利要求 7 所述的语音识别系统，其中所述解码确定装置还响应于第一识别结果满足识别可接受条件，直接以第一识别结果作为所述语音输入的最终解码结果。

9. 根据权利要求 8 所述的语音识别系统，其中：

所述识别可接受条件基于下述中的至少一种：所述语音输入的每帧语音的似然概率平均值、所述语音输入的各识别字符对应的概率得分、或置信度。

10. 根据权利要求 7 所述的语音识别系统，其中所述获取装置还从将所述语音输入提取为各个语音帧，以及所述第二识别装置通过在基于大词汇量连续语音识别网络定义的搜索空间中逐语音帧搜索最优路径来实现所述基于大词汇量连续语音识别网络的识别。

11. 根据权利要求 10 所述的语音识别系统，其中，所述第二识别装置进一步包括：

第一计算装置，用于计算当前语音帧相应于所有活跃节点的最优历史路径并统计当前历史路径最大值 Si，

第二计算装置，用于计算 Si 和当前语音帧在受限语法网络最优解码路径中的历史路径值 Si’的差值，以及

判断装置，用于响应于上述差值小于预设的域值，终止所述搜索过程。

12. 一种移动设备或服务器，包括权利要求 6-9 中任意一项所述的语音识别系统。

13. 一种移动设备，包括：

获取装置，用于获取用户语音输入，

第一识别装置，用于基于受限语法识别网络识别所述语音输入以获得第一识别结果，

收发装置，用于响应于第一识别结果不满足识别可接受条件，向服务器发送用户语音输入，以及从服务器接收基于大词汇量连续语音识别网络识别所述语音输入获得的第二识别结果，以及

解码确定装置，用于选择所述第一和第二识别结果中的优选者作为所述语音输入的最终解码结果。

用于移动设备的语音识别系统和方法

技术领域

[0001] 本发明一般地涉及语音信号处理领域，特别地涉及一种用于移动设备的对用户语音输入执行识别的方法和装置。

背景技术

[0002] 实现人机之间人性化、智能化的有效交互，构建高效自然的人机交流环境，已经成为当前信息技术应用和发展的迫切需求。特别是近年来随着无线通讯网络的普及，各种智能化的便携式移动设备在人们生活中发挥了越来越重要的作用，日益增多的人机交互要求一种新的针对小屏幕设备的更为高效自然的交互手段。语音作为最自然人性的交互手段正发挥了越来越重要的作用。例如用户在开车等不方便拨打电话时希望可以通过语音输入如“打电话给王治国”来操控移动设备，又或者在短消息编辑等需要文字输入时希望直接通过语音输入和语音识别来实现。

[0003] 目前已经提出了多种语音识别技术。例如，在S.J.Young等人的“Token Passing：A Simple Conceptual Model for Connected Speech Recognition Systems”，Technical Report CUED/F-INFENG/TR38，Cambridge University Engineering Dept，1989，中公开了一种基于受限语法网络的语言识别系统。该系统对于简短的语音命令能够实现准确高效的识别，然而在随意说的普遍情况下，往往不能工作。

[0004] 例如，在Aubert X.等人的“Large Vocabulary Continuous Speech Recognition of Wall Street Journal Corpus.″，Proc.ICASSP′ 94，Adelaide，Australia，Vol.II，pp.129-132，1994，中公开了基于大词汇量连续语音识别网络的语言识别系统。然而，这种语音识别系统的一个缺点在于需要在由大规模声学模型和通用语言模型构成的巨大的搜索空间中搜索最优路径，简短语音命令需要的快速准确响应往往得不到保障。

[0005] 因此，需要一种新的用于移动设备的语音识别方法和系统，其能够实现在语音识别的准确度和效率之间平衡，提供对简短语音命令的快速准确响应，以及提供随意说的语音识别支持。

发明内容

[0006] 为了实现上述目的，本发明的实施例提出了一种新的语音识别方法和装置，其支持对连续语音输入的智能响应及对简短语音命令的快速响应。

[0007] 根据本发明的一个方面，提供了一种用于移动设备的语音识别方法，包括：获取用户语音输入；基于受限语法识别网络识别所述语音输入以获得第一识别结果；响应于第一识别结果不满足识别可接受条件，在本地端或通过向服务器端传输语音信号执行基于大词汇量连续语音识别网络识别所述语音输入以获得第二识别结果；以及选择所述第一和第二识别结果中的优选者作为所述语音输入的最终解码结果。

[0008] 根据本发明的另一个方面，提供了一种用于移动设备的语音识别系统，包括：获取装置，用于获取用户语音输入；第一识别装置，用于基于受限语法识别网络识别所述语音输

入以获得第一识别结果；第二识别装置，用于响应于第一识别结果不满足识别可接受条件，基于大词汇量连续语音识别网络识别所述语音输入以获得第二识别结果；以及解码确定装置，用于选择所述第一和第二识别结果中的优选者作为所述语音输入的最终解码结果。

[0009] 根据本发明的方案具有如下特点：

[0010] 用户可以在统一系统界面下实现对各类语音输入命令的识别，

[0011] 可以响应用户自由随意说的语音识别，

[0012] 可以快速准确响应简短语音命令识别，

[0013] 对本地移动设备关联的特定信息能实现准确识别。

附图说明

[0014] 通过结合附图参考下面对本发明的实施方式的详细描述，本发明的上述以及其他特征将更加明显。在附图中，

[0015] 图 1 示意性地示出了根据本发明一个实施例的用于移动设备的语音识别的方法的流程图；

[0016] 图 2 示出了根据本发明的一个实施例的示例受限语法识别网络；

[0017] 图 3 示出了根据本发明的一个实施例的判断语音输入的识别结果是否满足识别可接受条件的判断流程图；

[0018] 图 4 示出了根据本发明的一个优选实施例的用于基于大词汇量连续语音识别网络的连续语音识别的改进的 Viterbi 搜索方法的流程图；

[0019] 图 5 示意性地示出了根据本发明的一个实施例的综合评判识别结果确定语音输入的最终解码结果的流程图；

[0020] 图 6 示出了根据本发明一个实施例的用于移动设备的语音识别系统的框图；

[0021] 图 7 示出了在其中可以实现本发明的实施例的移动设备的示意框图。

[0022] 在附图中，相同或对应的标号表示相同或对应的部分。

具体实施方式

[0023] 在下文中，将参考附图通过实施方式对本发明的用于移动设备的语音识别方法和装置进行详细的描述。应当理解，给出这些实施例仅仅是为了使本领域技术人员能够更好地理解进而实现本发明，而并非以任何方式限制本发明的范围。

[0024] 下文中将主要以个人移动电话为例说明本发明，但是本发明可以用于各种可支持语音输入功能的设备，而不局限于移动电话。例如，本发明还可以用于个人数字助理 (PDA)、多媒体音乐播放器、平板计算机等等。

[0025] 在移动设备中，随着移动设备越来越多的承担起个人助理的职责，通常存在各种需要通过语音与设备交互的情形。在一些情况下，用户可能期望通过简短的语音命令来控制移动设备的操作。例如，可以通过语音命令来控制移动设备上的各种应用的启用或者结束。诸如，用户可能希望通过语音命令“打电话给张三”来启用对张三的电话呼叫，其中张三可以是该移动设备上的通讯录中的联系人之一。在另一些情况下，用户可能希望更自然地使用随意说的方式来与设备进行交互。例如希望通过语音输入“告诉张三今晚公司 7 点到 3 楼会议室开会”来让设备给通信录中的联系人张三发送具有相应内容“今晚公司 7 点

到 3 楼会议室开会”的短消息。显然为了实现对用户的各类语音输入命令的正确执行，其首要条件就是要正确的识别其语音内容。

[0026] 通常基于受限语法识别网络的语音识别系统往往仅能够处理简短语音命令，而对随意说的情况则不能很好处理。相反基于大词汇量连续语音识别网络的语音识别系统则又不适合于对简短语音命令的快速响应。目前的语音应用通常是针对具体应用程序的，用户首先选择进入指定的程序后系统再根据应用环境选择相应的识别系统。例如，用户在实施语音拨打电话的功能时，往往首先进入命令控制程序，然后系统利用基于受限语法识别网络的语音识别系统响应用户的简短拨号命令，如“打电话给张三”，“给张三打电话”等。再如在短消息编辑输入等需要实现随意语音转写应用时，用户在选择进入短消息应用程序后由系统根据应用环境相应地选择基于大词汇量连续语音识别网络的语音识别系统，响应用户的连续自由输入。这种通过预先选定具体应用程序，再启用语音功能的人机交互方式显得并不是很自然人性。

[0027] 针对上述情况，本发明的实施例提出了一种新的语音识别方法和装置，其采用混合识别网络，即基于受限语法的识别网络以及可支持随意说的大词汇量连续语音识别网络，实现了在统一系统界面下对简短语音命令的准确高效的识别以及对连续语音输入的转写。从而，本发明的实施例提高了用户使用基于移动设备的个人助理工具的语音识别的便利性。

[0028] 图 1 示意性地示出了根据本发明一个实施例的用于移动设备的语音识别的方法 100 的流程图。

[0029] 在步骤 S110 中，获取用户语音输入。用户可以在统一的系统界面下获取各种形式的用户语音输入，包括简短的语音命令或者随意说的任何语句。可以采用任何已知的或未来开发的语音信号跟踪技术来获取用户语音输入。可以对连续的语音信号进行数字采样，获得语音输入的数字化形式。

[0030] 可选地，可以对语音输入进行预处理。在优选的实施例中，为了提高系统的鲁棒性，可以对采集到的原始语音信号做前端降噪预处理。例如，首先通过对语音信号执行短时能量和短时过零率分析，将连续的语音信号分割成独立的语音片断和非语音片断。随后通过维纳滤波等技术对语音片断进行语音增强，进一步消除语音信号中的噪音，提高后续系统对该信号的处理能力。

[0031] 可选地，还可以对语音输入进行声学特征提取。考虑到降噪处理后的语音信号中依然存在大量语音识别无关的冗余信息，直接对其识别将导致运算量增加和识别准确率的下降，为此可以从语音能量信号中提取识别有效的语音特征，并存入特征缓存区内，以表征用户语音输入。在一个优选实施例中，提取语音的 MFCC 特征。例如，对窗长 25ms 帧移 10ms 的每帧语音数据做短时分析，得到 MFCC 参数及其一阶二阶差分，共计 39 维。一段语音输入可以量化为一 39 维的特征序列 O。在其他实施例中，还可以采用 PLP 特征 (Perceptual linear predictive) 或者 TANDEM特征等，来提取语音输入的特征以表征语音输入。为了避免模糊本发明的要点，在此对已知的语音信号跟踪技术、预处理技术和特征提取技术不再详述。

[0032] 此外，应该理解，本发明的原始的或经预处理的用户语音输入或其特征标识可以存储在存储器中，并且不限于任何特定的存储格式。

[0033] 在步骤 S120 中，基于受限语法识别网络识别语音输入，以获得针对所述语音输入的第一识别结果。

[0034] 受限语法识别网络可以预先定义并且存储在设备中。受限语法识别网络主要用于实现对简短语音命令的支持，其支持的语法相对简单，包括诸如“发短信给 ×××”、“打电话给 ×××”等与有限的命令词相关的说辞。优选地，受限语法识别网络限定了与移动设备相关的个性化信息，例如与设备上支持的应用、通讯录中的信息等等相关。图 2 示出了根据本发明的一个实施例的示例受限语法识别网络。基于该受限语法识别网络，可以快速准确地识别类似“发短信给张菲”、“打电话给王智国”等简短的语音命令。

[0035] 可以通过下述步骤来实现基于受限语法识别网络搜索语音输入的识别结果。步一：载入声学模型及受限语法网络等系统参数。可选地，可以在方法 100 开始（例如初始化）时或者在执行步骤 120 中的实际识别之前的任何时间，载入声学模型及受限语法识别网络等系统参数。其中语法受限识别网络反映了本发明的语音识别系统支持的各类简单的语音命令，例如如图 2 所示。声学模型用于模拟字符的标准发音特征，在本实施例中采用语音识别领域常用的基于转移概率和传输概率的 HMM（隐马尔可夫）模型。应该理解，本发明还可以使用诸如神经网络 (Neural Network mode) 等其他声学模型。步二：根据受限语法识别网络生成基于声学模型的搜索网络。步三：在所述搜索网络定义的搜索空间中，搜索相应于步骤 S110 中获取的语音输入的最优路径。例如，可以根据语音输入提取各个语音帧。使用 Viterbi 搜索，对提取的每一语音帧，计算其相应于当前所有活跃节点的最优历史路径概率。利用动态规划思想依时间顺序搜索，在搜索到最后一帧语音矢量时，从终止状态回溯就得到最优解码状态序列及对应的历史路径概率。关于 Viterbi 算法例如可以详细参见 .J.Viterbi 的论文“Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm”，IEEE Transactions on Information Theory，Vol.IT-13，pp.260-269，April 1967，在此不再赘述。现在已知或者将来开发的其他搜索方法也是可行的，本发明的范围不局限于使用 Viterbi 算法的搜索方法。

[0036] 应该理解，在受限语法识别网络定义的搜索空间中，有可能搜索到用户语音输入的优选匹配路径（例如，用户语音输入是符合受限语法的简短语音命令），获得所述语音输入的第一识别结果，或者也可能其搜索到的优选匹配路径不合理（例如在用户随意说的情况下利用受限语法识别的解码结果路径得分往往很低），因此得不到所述语音输入的有效识别结果。

[0037] 在一个简化实施例中，如果当前的用户语音输入在语法受限网络中找到的匹配路径合理，也即获得了第一识别结果，则以所述识别结果作为用户语音输入的解码结果，方法 100 结束。否则，即没有找到合理的匹配路径，则方法 100 前进到步骤 S140，转入基于大词汇量连续语音识别网络的重新识别，以获得针对用户语音输入的第二识别结果。

[0038] 在优选的实施例中，方法 100 还包括步骤 S130，判断在步骤 S120 中基于受限语法识别网络搜索得到的第一识别结果是否满足识别可接受条件。如果第一识别结果满足可接受条件，则直接接受该第一识别结果作为用户语音输入的解码结果，从而方法 100 结束。这样可以节省识别时间，提高整体识别效率。如果第一识别结果不满足可接受条件，则方法 100 前进到步骤 S140，转入基于大词汇量语音识别网络的重新识别。

[0039] 图 3 示出了根据本发明的一个实施例的基于受限语法识别网络搜索得到的第一

识别结果是否满足识别可接受条件的判断流程。

[0040] 在步骤 S310 中：计算针对用户语音输入的识别结果中平均每帧语音的似然概率平均值。

[0041] 在步骤 S320 中：判断该帧平均值是否大于系统预先设置的域值，若不是则说明当前识别结果不可信，转入步骤 S360，否则转入步骤 S330。

[0042] 在步骤 S330 中：计算针对用户语音输入的各识别字符对应的概率得分。

[0043] 在步骤 S340 中：判断每个字符的概率得分是否大于其对应的域值。若是则说明当前识别结果可信，转入步骤 S350，否则转入步骤 S360。

[0044] 在步骤 S350 中：判定当前的识别结果满足可接受条件。

[0045] 在步骤 S360 中：判定当前的识别结果不满足可接受条件。

[0046] 即要求在两项概率得分均大于阈值时才可判断当前解码符合要求

[0047] 其中帧平均值对应的阈值和 / 或字符对应的域值可以由识别系统预先在海量训练数据上调试得到。

[0048] 应该理解，在图 3 示出的实施例中使用似然概率作为判断识别结果是否满足识别可接受条件仅是出于示例说明的目的，而非作为任何限制。本发明还可以使用置信度等其它参数来作为判断条件，参见 L.E.Baum, T.Petrie, G.Soules, 和 N.Weiss 等人的论文“A maximization technique occurring the statistical analysis of probabilistic functions of Markov chains,”Ann.Math.Stat., vol.41, no.1, pp.164-171, 1970。

[0049] 回到图 1，在步骤 S140 中，基于大词汇量连续语音识别网络重新识别语音输入，以获得针对所述语音输入的第二识别结果。

[0050] 基于大词汇量连续语音识别网络的语音识别采用大规模的声学模型和语言模型，不受语法限制可用于模拟任意自由语音输入。其解码流程具体如下所示。步一：载入预定的大规模声学模型及语言模型等系统参数。可选地，可以在方法 100 开始（例如初始化）时或者在执行步骤 140 中的实际识别语音之前的任何时间，执行所述载入。类似的，在本实施例中，声学模型采用了语音识别领域常用的基于转移概率和传输概率的 HMM（隐马尔可夫）模型，用于模拟字符标准发音特征。应该理解，本发明还可以使用诸如神经网络 (Neural Network mode) 等其他声学模型。步二：将带有词频概率的语言模型网络扩展成基于声学模型的搜索网络，以供后续路径搜索。步三：在所述搜索网络定义的搜索空间中，搜索相应于语音输入的最优路径。例如，可以使用 Viterbi 搜索，针对提取的语音帧序列，从搜索网络中找到其对应的最优单词序列，从而获得识别结果。

[0051] 优选地，在基于大词汇量连续语音识别网络的语音识别中，利用了在步骤 S120 中的基于受限语法网络的搜索中的最优解码路径的路径值，从而可以尽早地反馈识别结果。下面参考附图 4 进行详细说明实现步骤 S140 的一个优选实现。

[0052] 图 4 示出了根据本发明的一个优选实施例的用于基于大词汇量连续语音识别网络的连续语音识别的改进的 Viterbi 搜索方法 400 的流程图。

[0053] 在步骤 S410 中，初始化并设置当前语音帧 i = 1。

[0054] 在步骤 S420 中，计算当前语音帧相应于所有活跃节点的最优历史路径并统计当前历史路径最大值 Si。

[0055] 在步骤 S430 中，计算 Si 和当前语音帧在受限语法网络最优解码路径中的历史路

径值 Si’的差值。

[0056] 在步骤 S440 中 :判断上述差值是否大于系统预先设定的域值 S。若是则转入步骤 S450，否则转入步骤 S470。

[0057] 在步骤 S450 中，设置当前考察语音帧为下一语音帧 i++。

[0058] 在步骤 S460 中，判断当前考察语音帧是否大于语音帧总数 T，若是，则转入步骤 S470，否则转入步骤 S420，继续针对当前考察语音帧计算当前历史路径最大值 Si。其中，语音帧总数 T 是在受限语法网络解码时确定的当前语音输入总帧数。

[0059] 在步骤 S470 中，返回当前识别结果。优选地，可以返回历史路径得分，历史路径及已解码的总帧数等。

[0060] 在方法 400 中，利用了已经执行的基于受限语法识别网络的搜索结果，可以在不解码所有语音帧的情况下，提前结束基于大词汇量连续语音识别网络的识别过程。在该优选实施例中，对于当前语音帧，当其在基于大词汇量连续语音识别网络的搜索中的最优历史路径得分与其在基于受限语法网络搜索中的最优解码路径中的路径值之差小于预定的阈值时，可以提前结束基于大词汇量连续语音识别网络的搜索，直接返回基于受限语法识别网络的识别结果作为语音输入的识别结果。如果完成所有帧的基于大词汇量连续语音识别网络的识别，则在步骤 S470 中将返回基于大词汇量连续语音识别网络的第二识别结果。在方法 400 中，基于大词汇量连续语音识别网络的解码过程是否提前结束（即，没有完成），例如可以通过返回解码总帧数来指示。如果解码总帧数等于预定的语音帧总数 T，则说明基于大词汇量连续语音识别网络的解码已经完成，否则则是提前结束。备选地，也可以通过设置其他标志（如具有“真 / 假”值的二元比特）来指示是否识别过程是否提前结束。

[0061] 返回图 1，当步骤 S140 中结束基于大词汇量连续语音识别网络的识别时，方法 100 前进到步骤 S150。在步骤 S150 中，综合基于受限语法识别网络的识别结果和基于大词汇量连续语音识别网络的识别结果，确定所述语音输入的最终解码结果。如果基于大词汇量连续语音识别网络的识别没有完成（即提前结束），则确定基于受限语法识别网络的识别结果为用户语音输入的最终解码结果。如果基于大词汇量连续语音识别网络的识别已经完成，但是其识别结果的得分小于基于受限语法识别网络的识别结果的得分，则仍确定基于受限语法识别网络的识别结果为用户语音输入的最终解码结果，否则确定基于大词汇量连续语音识别网络的识别结果为用户语音输入的最终解码结果。

[0062] 在图 5 中示出了步骤 S140 的一个具体实现。

[0063] 在步骤 S510 中，判断基于大词汇量连续语音识别网络的解码过程是否完成，即解码到最后一帧。若是，则转入步骤 S520，否则转入步骤 S540。

[0064] 在步骤 S520 中，判断基于受限语法识别网络的识别中的系统最优路径得分是否小于基于大词汇量连续语音识别网络的识别中的系统最优路径得分。若是，则转入步骤 S530，否则转入步骤 S540。备选地，作为系统最优路径得分的替代或补充，也可以使用帧平均得分作为判断标准。

[0065] 在步骤 S530 中，输出基于大词汇量连续语音识别网络的连续语音识别结果作为最终的解码结果。

[0066] 在步骤 S540 中，输出基于语法受限识别网络的识别结果作为最终的解码结果。

[0067] 通常，在步骤 S150 中确定所述语音输入的最终解码结果之后，方法 100 结束。

[0068] 优选地，方法 100 获得语音输入的最终解码结果将用来触发移动设备中的相应应用，例如电话呼叫应用、短消息应用等。

[0069] 上面已经参考附图详细说明了在统一界面下对用户任意形式输入的语音识别的方法。应该注意，尽管在附图中以特定顺序描述了本发明方法的操作，但是，这并非要求或者暗示必须按照该特定顺序来执行这些操作，或是必须执行全部所示的操作才能实现期望的结果。相反，流程图中描绘的步骤可以改变执行顺序。附加地或备选地，可以省略某些步骤，将多个步骤合并为一个步骤执行，和 / 或将一个步骤分解为多个步骤执行，也可以增加其他步骤。

[0070] 此外，该方法可以基于各种具体实现，包括在移动设备本地端单独实现，以及移动设备本地端结合服务器端实现等。

[0071] 在一个实施例中，上述方法 100 可以完全在移动设备本地端实现。在该方案下，在移动设备处理的存储器中存储受限语法识别网络和基于大词汇量连续语音识别网络。

[0072] 备选地，在另一个实施例中，在移动设备的本地端获取用户语音输入（步骤 S110）。然后，移动设备向服务器发送获取的原始的或者经处理的用户语音输入。所述经处理的用户语音输入可以语音输入的数字形式，或者提取的特征序列。服务器接收该用户语音输入。然后服务器执行对语音输入的识别，包括 :基于受限语法识别网络的第一识别 (S120) ;判断第一识别结果的解码有效性 (S130) ;在第一识别结果无效时，执行基于大词汇量连续语音识别网络的第二识别 (S140) ;以及综合第一识别结果和第二识别结果，确定所述语音输入的最终解码结果 (S150)。然后，服务器向移动设备发送最终解码结果。

[0073] 在该实施例中，服务器端维护大词汇量连续语音识别网络。此外，服务器端还对每个移动设备或用户都维护一个个性化的信息库，例如受限语法识别网络，用于提高带有个性化信息的语音命令，如识别出“打电话给王智国”，而非大规模语言模型中的“王治国”。

[0074] 在又一个实施例中，在移动设备的本地端获取用户语音输入 (S110)，执行基于受限语法的第一识别 (S120)，以及判断第一识别结果的解码有效性 (S130)。在第一解码无效时，移动设备向服务器端发送获取的用户语音输入，其可以是语音信号或提取的特征序列。

[0075] 在服务器端利用其强大的解码运算能力和超大规模的模型库（例如，大词汇量连续语音识别网络）对用户语音输入进行连续语音解码，以获得第二识别结果 (S140)。优选地，为了提高服务器的解码效率，移动设备在向服务器传输语音特征序列时，可以同时传输本地端的解码结果（即第一识别结果），包括每帧的解码路径值。

[0076] 接着，移动设备可以从服务器接收第二识别结果。

[0077] 然后，移动设备可以综合基于受限语法识别网络的识别结果和基于大词汇量连续语音识别网络的识别结果，确定所述语音输入的最终解码结果 (S150)。

[0078] 在该实施例中，移动设备存储各自的受限语法识别网络。在服务器端存储超大规模的大词汇量连续语音识别网络。

[0079] 应该理解，本发明的方法不局限于所示出的具体示例和变形。在不脱离本发明的精神和范围的情况下，本领域技术人员可以想到其他修改、替代和变形。

[0080] 图 6 示出了根据本发明的一个优选实施例的用于移动设备的语音识别系统 600。系统 600 可以用于执行上述方法 100。例如，系统 600 可以是安装在移动设备上，或者分布在移动设备本地端和服务器上。

[0081] 系统 600 包括获取装置 610、第一识别装置 620、第二识别装置 S640 和解码确定装置 650。

[0082] 根据本发明的一个实施例，获取装置 610 用于获取用户语音输入。优选地，获取装置 610 从用户语音输入中提取语音帧，将语音输入表示为一系列语音帧。获取装置 610 可以采用任何已知的或未来开发的语音信号跟踪技术来获取用户语音输入，可以对连续的语音信号进行数字采样，获得语音输入的数字化形式。优选地，获取装置 610 可以包括预处理装置，用于对语音输入进行预处理，以增强语音并且消除语音中的噪声。优选地，获取装置 610 还可以包括声学特征提取装置，用于从语音信号（特别是经预处理的语音信号）中提取声学特征以表征语音输入。

[0083] 第一识别装置 620 用于基于受限语法识别网络识别所述语音输入。第一识别装置 620 使用预先载入的声学模型及受限语法网络对语音输入进行识别，以获得第一识别结果。

[0084] 第二识别装置 640 用于响应于第一识别结果不满足识别可接受条件，基于大词汇量连续语音识别网络识别所述语音输入以获得第二识别结果。第二识别装置 640 使用预先载入的声学模型以及大词汇量连续语音识别网络对语音输入进行识别，以获得第二识别结果。

[0085] 解码确定装置 650 用于结合基于受限语法识别网络的识别结果和基于大词汇量连续语音识别网络的识别结果，确定所述语音输入的最终解码结果。如果第二识别装置 640 获得的基于大词汇量连续语音识别网络的识别结果的得分大于第一识别装置 620 获得的基于受限语法识别网络的识别结果的得分，则解码确定装置 650 确定第二识别装置 640 获得的识别结果为用户语音输入的最终解码结果，否则确定第一识别装置 620 获得的识别结果为用户语音输入的最终解码结果。

[0086] 根据本发明的一个优选实施例，系统 600 还包括解码有效性判定装置 630，用于判断基于受限语法识别网络识别的识别结果的解码有效性。在第一识别装置 620 获得第一识别结果后，如果解码有效性判定装置 630 判断该第一识别结果满足可接受条件，则其促使解码确定装置 650 确定该第一识别结果为用户语音输入的最终解码结果。

[0087] 根据本发明的一个优选实施例，第二识别装置 640 利用第一识别装置 620 已执行的基于受限语法网络的搜索中的最优解码路径的路径值判断是否提前结束其识别过程。当判断当前语音帧的基于大词汇量连续语音识别网络的搜索中的最优历史路径得分不大于其在基于受限语法网络搜索中的最优解码路径中的路径值，或者两者之差小于预定阈值时，可以提前结束基于大词汇量连续语音识别网络的识别。在提前结束基于大词汇量连续语音识别网络的识别时，第二识别装置 640 可以输出信号促使解码确定装置 650 确定第一识别装置 620 获得的第一识别结果为用户语音输入的最终解码结果。

[0088] 为清晰起见，在图 6 中并未示出各个装置所包含的子装置。然而，应当理解，系统 600 中记载的每个装置与参考图 1 描述的方法 100 中的各个步骤相对应。由此，上文针对图 1 描述的操作和特征同样适用于系统 600 及其中包含的装置和子装置，在此不再赘述。

[0089] 应该理解，尽管在上文详细描述中提及了系统的若干装置或子装置，但是这种划分仅仅并非强制性的。实际上，根据本发明的实施方式，上文描述的两个或更多装置的特征和功能可以在一个装置中具体化。反之，上文描述的一个装置的特征和功能可以进一步划分为由多个装置来具体化。

[0090] 此外，图 6 所示的系统仅是示例性的，而不是限制性的。系统 600 可以存在各种各样的变形。

[0091] 在一个实施例中，系统 600 安装在移动设备上。

[0092] 在另一个实施例中，系统 600 安装在服务器上。在该情况下，服务器还包括与移动设备的通信装置（未示出），用于在移动设备之间传输用户语音输入以及识别结果。

[0093] 在又一个实施例中，系统 600 分布在移动设备本地端和服务器二者上。在该实施例中，移动设备包括：获取装置，用于获取用户语音输入；第一识别装置，用于基于受限语法识别网络识别所述语音输入以获得第一识别结果；收发装置，用于响应于第一识别结果不满足识别可接受条件，向服务器发送用户语音输入，以及从服务器接收基于大词汇量连续语音识别网络识别所述语音输入获得的第二识别结果；以及解码确定装置，用于选择所述第一和第二识别结果中的优选者作为所述语音输入的最终解码结果。服务器包括：接收装置，用于从移动设备接收用户语音输入；第二识别装置，用于基于大词汇量连续语音识别网络识别所述语音输入以获得第二识别结果；发送装置，用于向移动设备发送第二识别结果。

[0094] 此外，系统 600 还可以包括其他装置，例如易失性或者非易失性存储装置，用于存储获取的语音输入和 / 或其识别结果。系统 600 还可以包括触发装置，用于根据语音输入的最终解码结果来触发设备中的相应应用，例如电话呼叫应用、短消息应用等。

[0095] 而且，系统 600 及其各个组成部分可以利用各种方式来实现。例如，在某些实施方式中，系统 600 可以利用软件和 / 或固件模块来实现。此外，系统 600 也可以利用硬件模块来实现。例如，系统 600 可以实现为集成电路 (IC) 芯片或专用集成电路 (ASIC)。系统 600 也可以实现为片上系统 (SOC)。现在已知或者将来开发的其他方式也是可行的，本发明的范围在此方面不受限制。

[0096] 图 7 示出了适于用来实现本发明实施方式的移动电话 700 的一个示例。然而应该理解，本发明的范围不限于所述的移动电话的具体类型。

[0097] 移动电话 700 可以是任何需要语音交互的移动终端。移动电话 700 可以包括用于容纳和保护其的外壳 30。移动电话 700 可以进一步包括液晶显示器形式的显示器 32。在本发明的其他实施方式中，显示器可以是适合于显示图像或文字的任何适当显示技术。移动电话 700 可以进一步包括小键盘 34。在本发明的其他实施方式中，可以采用任何适当的数据或用户接口机制。例如，可以将用户接口实现为虚拟键盘或数据录入系统以作为触敏显示器的一部分。该移动电话可以包括麦克风 36 或者可以是数字信号输入或模拟信号输入的任何适当音频输入。移动电话 700 可以进一步包括音频输出设备，其在本发明的实施方式中可以是以下任意一种：耳机 38、扬声器或者模拟音频或数字音频输出连接。移动电话 700 还可以包括电池 40（或者在本发明的其他实施方式中，该设备可以由任何适当的移动能量设备供电，诸如太阳能电池、燃料电池或发条发电机）。该移动电话可以进一步包括用于与其他设备进行短距离视线通信的红外端口 42。在其他实施方式中，移动电话 700 可以进一步包括任何适当的短距离通信方案，诸如蓝牙无线连接或 USB/ 火线有线连接。

[0098] 移动电话 700 可以包括用于对该移动电话 700 进行控制的控制器 56 或处理器。控制器 56 可以连接至存储器 58，该存储器 58 在本发明的实施方式中可以存储预设的声学模型、受限语法识别网络、大规模词汇量识别网络等，和 / 或还可以存储用于在控制器 56 上实

现的指令。控制器 56 可以进一步连接至编解码器电路 54，其适用于实施或辅助控制器 56 实施对音频和 / 或视频数据的编码和解码，包括根据本发明的实施例的语音识别。

[0099] 移动电话 700 可以进一步包括读卡器 48 和智能卡 46，例如 UICC 和 UICC 读卡器，其用于提供用户信息并且适合于提供认证信息以供在网络处对用户进行认证和授权。

[0100] 移动电话 700 可以包括无线电接口电路 52，其连接至控制器并且适合于生成无线通信信号，例如用于与蜂窝通信网络、无线通信系统或无线局域网通信。移动电话 700 可以进一步包括连接至无线电接口电路 52 的天线 44，用于传输和接收在无线电接口电路 52 处生成的射频信号。

[0101] 根据本发明的语音识别系统 600 可以作为硬件实现包括在移动电话 700 中。特别地，除硬件实施方式之外，根据本发明的设备 600 可以通过计算机程序产品的形式实现。例如，参考图 1 描述的方法 100 可以通过计算机程序产品来实现。该计算机程序产品可以存储在例如图 7 所示的存储器 58 中，或者通过网络从适当的位置下载到移动电话 700 上。计算机程序产品可以包括计算机代码部分，其包括可由适当的处理设备（例如，图 7 中示出的控制器 56 和 / 或编解码电路 54）执行的程序指令。所述程序指令至少可以包括 :用于获取用户语音输入的指令 ;用于基于受限语法识别网络识别所述语音输入的指令 ;用于基于大词汇量连续语音识别网络识别所述语音输入的指令 ;以及用于结合基于受限语法识别网络的识别结果和基于大词汇量连续语音识别网络的识别结果，确定所述语音输入的最终解码结果的指令。优选地，所述程序指令还包括用于判断基于受限语法识别网络识别的识别结果的解码有效性的指令。优先地，所述程序指令还包括利用所述基于受限语法网络的识别中的最优解码路径的路径值提前结束基于大词汇量连续语音识别网络的识别过程的指令。

[0102] 上文已经结合具体实施方式阐释了本发明的精神和原理。本发明的实施方式提供了一种新的语音识别系统和方法，可以向用户提供统一的系统界面简单高效完成与系统的交互，实现对移动设备的各类语音命令控制。通过采用结合基于受限语法识别网络以及可支持随意说的大词汇量连续语音识别网络的混合网络，实现了对简短语音命令的准确高效的识别以及对连续语音输入的转写。根据本发明的实施例不需要用户首先选择进入指定的程序，随后再根据当前应用环境选择对应的识别系统操作。例如，假设“王智国”是设备的通信录中的一个联系人。当用户输入语音输入“打电话给王智国”时，根据本发明的实施例将快速地输出基于受限语法识别网络的识别结果，并基于该识别结果可以调用通信录中王智国的信息给其打电话。当用户以随意说的方式提供语音输入“今晚公司 7 点到 3 楼会议室开会”时，根据本发明的实施例将快速输出基于大词汇量连续语音识别网络的识别结果“今晚公司 7 点到 3 楼会议室开会”以实现快速的语音文本转换。本发明的语音识别方法和系统更加准确高效，提供了更加智能便捷的人机交互方式。

[0103] 说明书中提及的术语“识别”、“解码”对于语音识别领域而言具有类似的含义，仅是出于不同语境下的选择，其均表示将音频的语音信号转换为对应的文字字符。

[0104] 虽然已经参考若干具体实施方式描述了本发明，但是应该理解，本发明并不限于所公开的具体实施方式。本发明旨在涵盖所附权利要求的精神和范围内所包括的各种修改和等同布置。所附权利要求的范围符合最宽泛的解释，从而包含所有这样的修改及等同结构和功能。

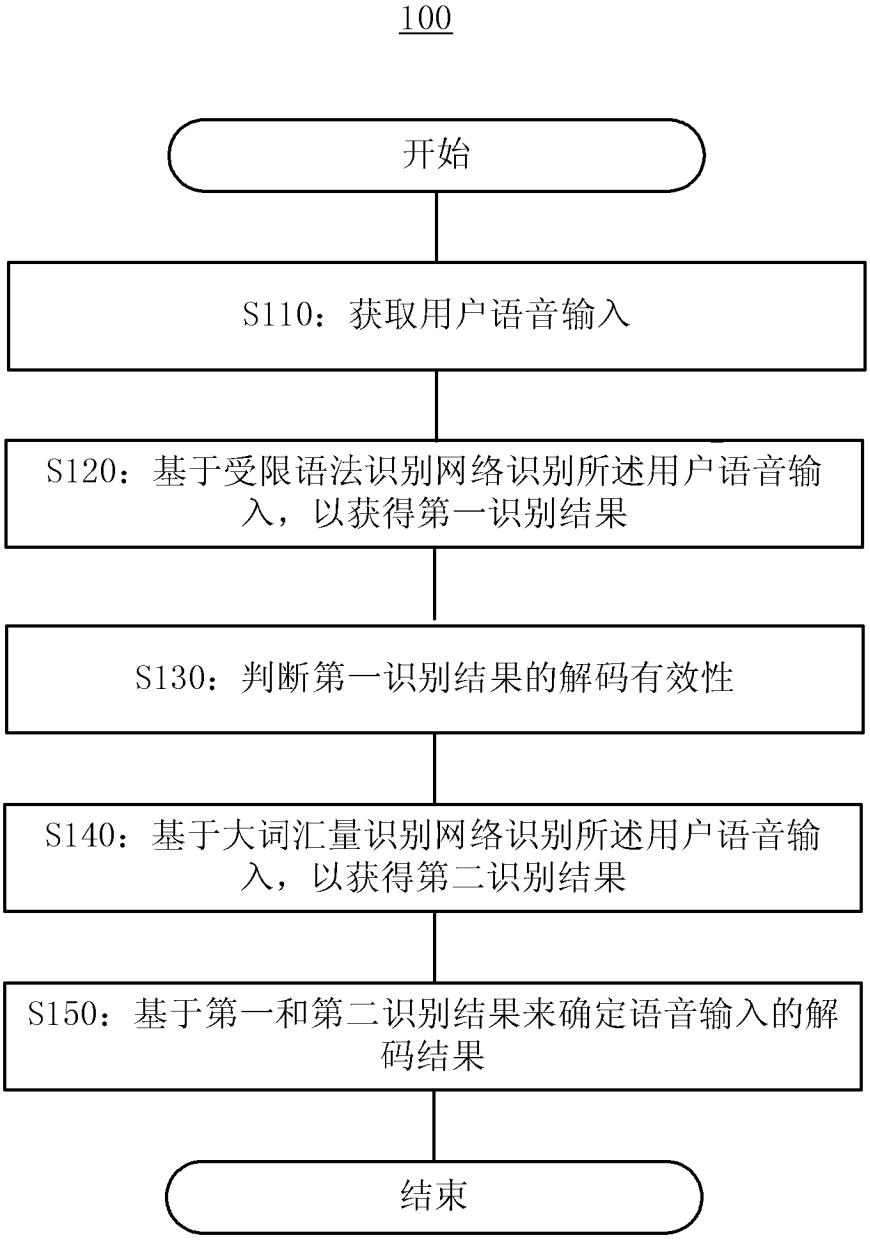

100
开始
S110：获取用户语音输入
S120：基于受限语法识别网络识别所述用户语音输入，以获得第一识别结果
S130：判断第一识别结果的解码有效性
S140：基于大词汇量识别网络识别所述用户语音输入，以获得第二识别结果
S150：基于第一和第二识别结果来确定语音输入的解码结果
结束

图 1

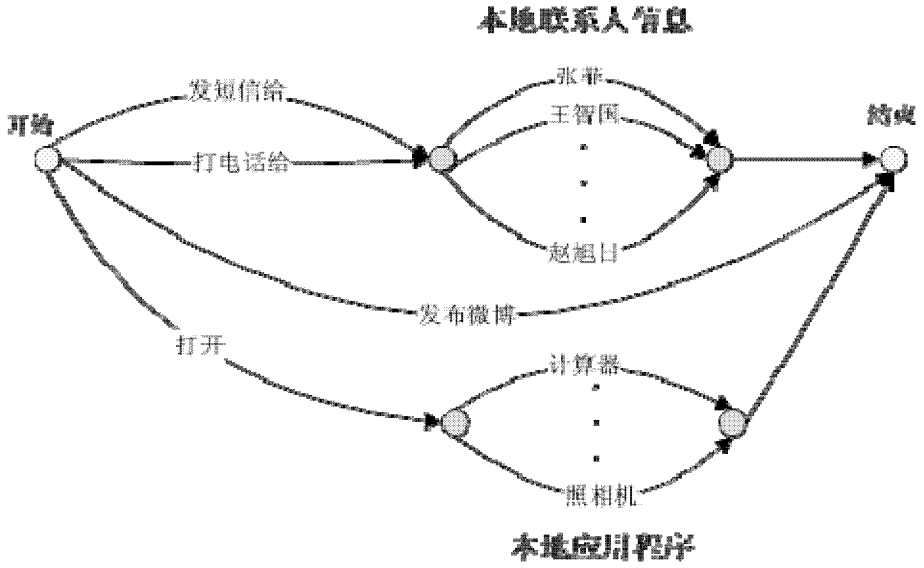

本地联系人信息
开始
发短信给
打电话给
张菲
王智国
赵旭日
结束
发布微博
打开
计算器
照相机
本地应用程序

图 2

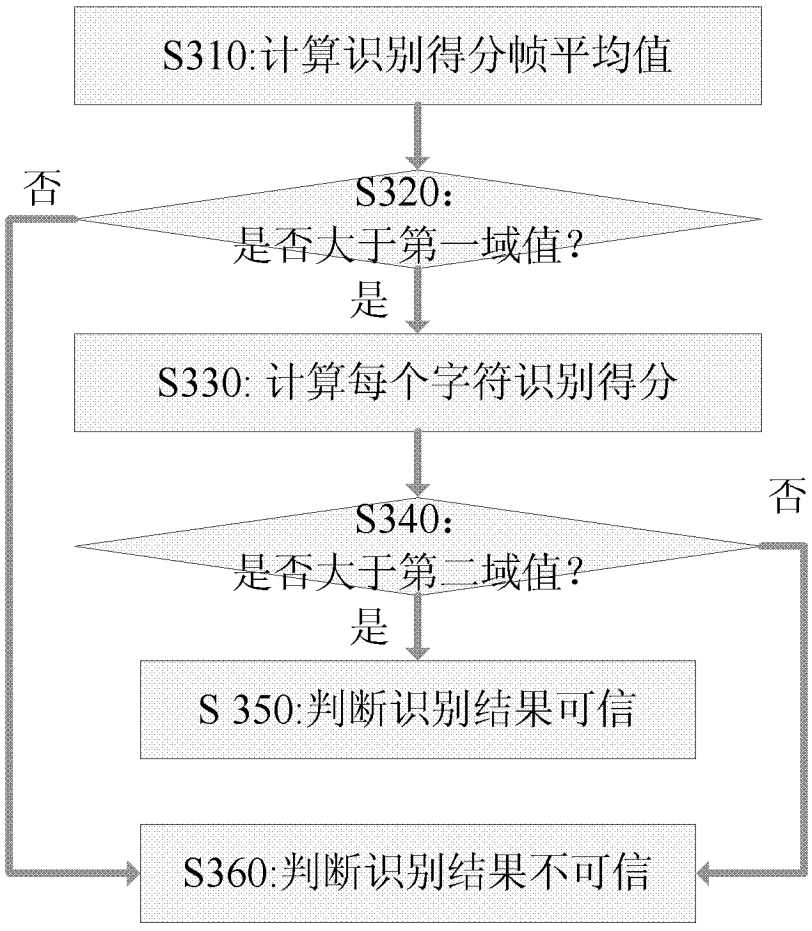

S310:计算识别得分帧平均值
否
S320:
是否大于第一域值?
是
S330: 计算每个字符识别得分
否
S340:
是否大于第二域值?
是
S 350:判断识别结果可信
S360:判断识别结果不可信

图 3

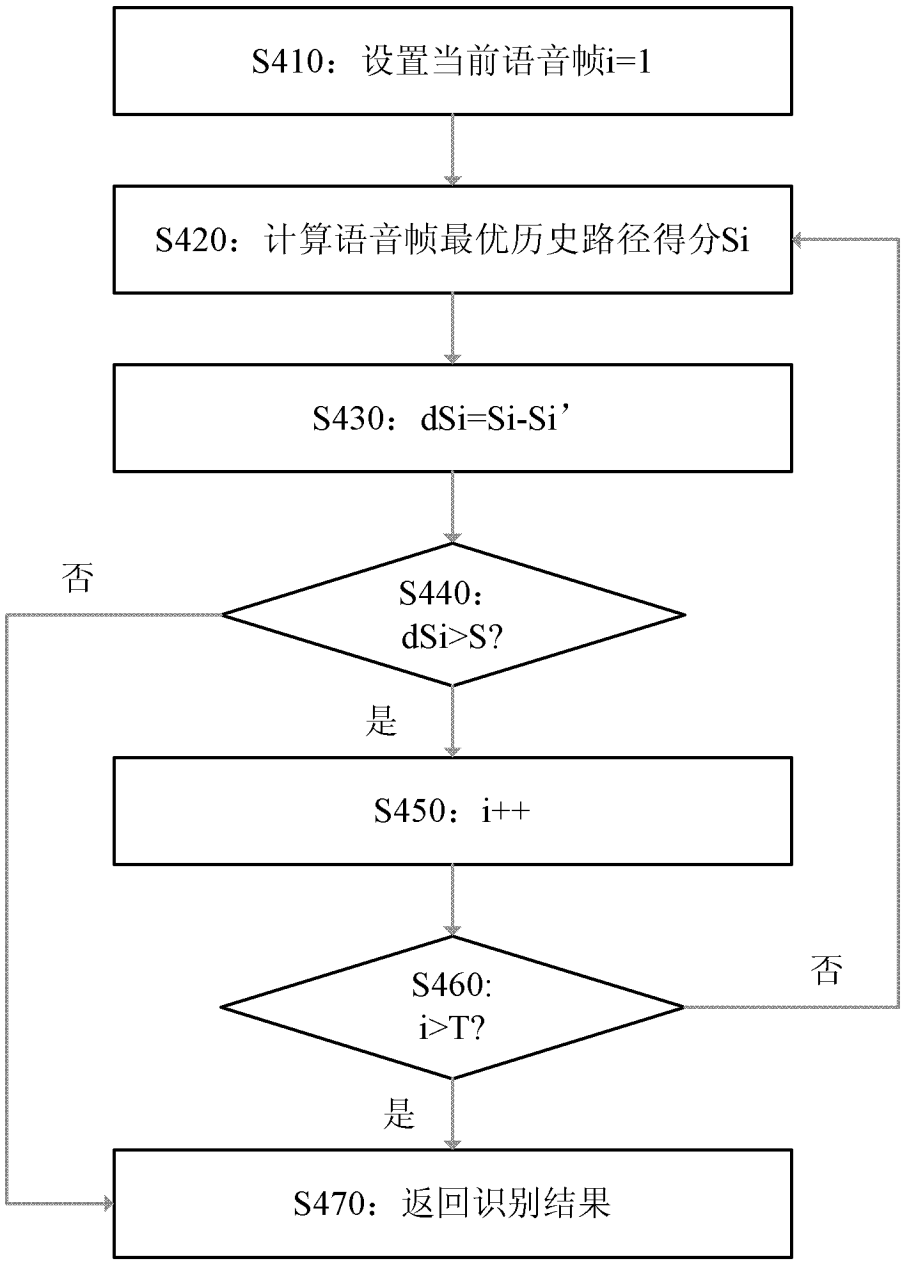

400
S410：设置当前语音帧i=1
S420：计算语音帧最优历史路径得分Si
S430：dSi=Si-Si’
S440：
dSi>S?
否
是
S450：i++
S460:
i>T?
否
是
S470：返回识别结果

图 4

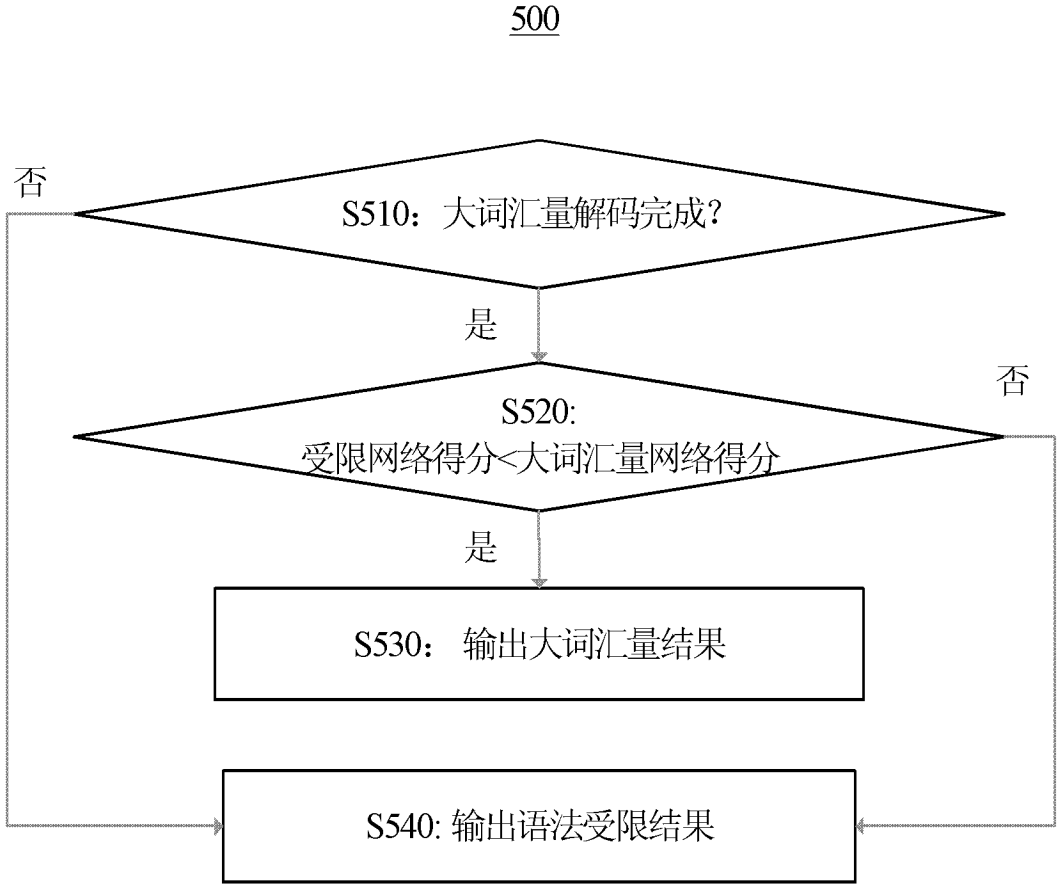

500
否
S510：大词汇量解码完成?
是
否
S520:
受限网络得分<大词汇量网络得分
是
S530： 输出大词汇量结果
S540: 输出语法受限结果

图 5

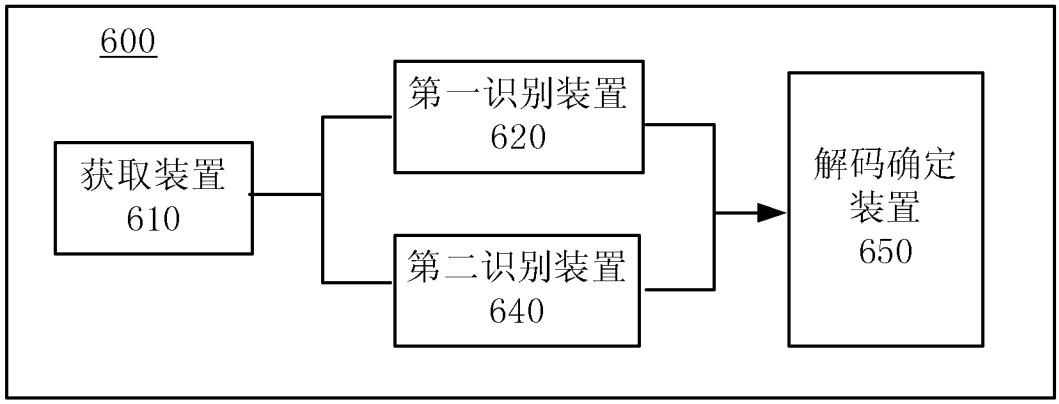

600
获取装置
610
第一识别装置
620
第二识别装置
640
解码确定
装置
650

图 6

700

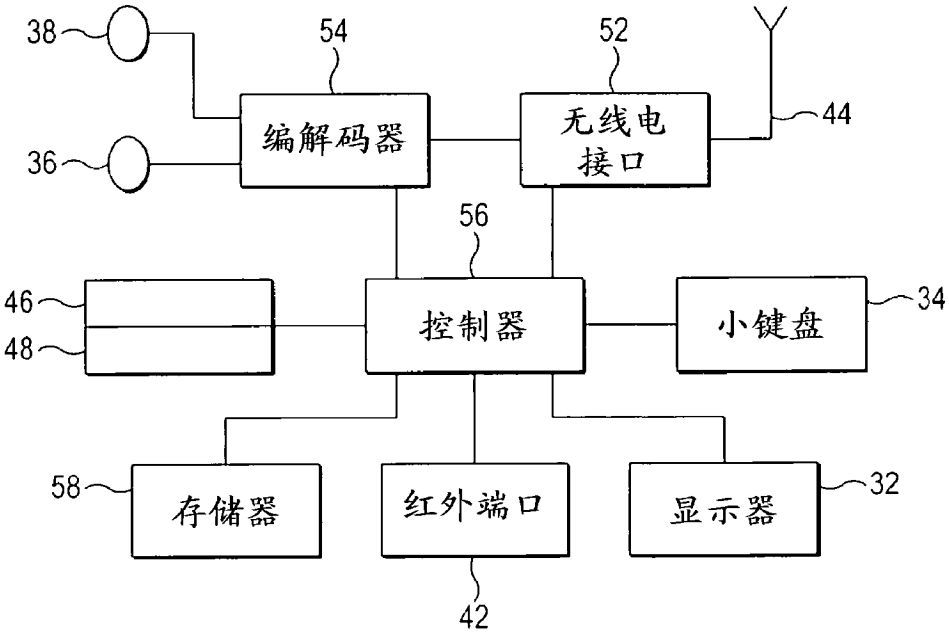


图 7